Joint Power and Channel Resource Optimization in Soft Multi-View Video Delivery

Ticao Zhang and Shiwen Mao, *Fellow, IEEE*

***Abstract*—Existing wireless multi-view video (MVV) transmission schemes use digital compression to achieve a better coding efficiency. However, the digital schemes suffer from the cliff effect, which refers to the phenomenon that the video quality is a step function of wireless channel quality. In this paper, we first consider a soft MVV transmission scheme where the correlations between the inter-view data and texture-depth data are exploited by a 5-dimensional discrete cosine transform (5D-DCT). The linearly transformed 5D-DCT signals are modulated in an analog manner so that the video quality gracefully improves when the channel quality becomes better. The cumbersome bit and rate controls in digital solutions are replaced by simple power controls. Second, as with the increase of the number of cameras and data depths, the data size of MVV increases linearly. To reduce the heavy data traffic in soft MVV transmission, we proposed efficient resource (bandwidth and power) allocation algorithms. Simulations results demonstrate that the proposed *distortion-resource* (DR) optimization algorithm can ensure a best viewing quality under a resource constraint and the proposed *resource-distortion* (RD) optimization algorithm can minimize the resource usage for a target video quality requirement. Third, the impact of power control across texture and depth frame and the impact of view positions on synthesized virtual view quality are investigated. The efficacy of the proposed algorithm on both the reference viewpoint as well as the virtual viewpoint is verified via simulations.**

***Index Terms*—Soft video delivery; SoftCast; Multi-view video; Resource allocation.**

I. Introduction

Multimedia has become the most popular application for emerging network paradigms [1]–[3]. Recently, Multi-view videos (MVV) are emerging in various application domains such as education, healthcare, and 3D-home entertainment. It is a fundamental technology in virtual-reality (VR), naked-eye 3D, and free-viewpoint video streaming [4], [5]. Fig. 1 shows an example of the MVV transmission system, where a number of cameras are deployed at different positions. Each camera captures both texture maps (images) and depth maps (distances from the objects). These texture and depth information is known as multi-view plus depth (MVD). The MVD information is encoded via texture and depth encoding and then transmitted to the receiver via, e.g., a wireless channel. After decoding, the receiver synthesizes intermediate virtual viewpoint using depth-image-based rendering (DIBR) from the received MVD frames [6]. As can be seen, virtual view 2 can be synthesized with the texture and depth data from its left view (i.e., view 1) and right view (i.e., view 3). The receiver can then select its favorite viewing angles and enjoy an enhanced viewing experience.

Compared with conventional single-view video streaming, an MVV streaming usually generates a large amount of video data. The video data grows linearly with the product of the number of cameras and the number of frame depths. When it comes to 4K (or 8K) videos, the heavy data traffic would consume a considerable amount of wireless resources and hence may become infeasible for realtime transmission. Currently, one solution is an independent coding on both texture and depth, which is a backward compatible extension of the H.264/AVC standard [7], [8]. Each viewpoint is encoded separately and only the view corresponding to the user's current selected viewpoint is transmitted [9], [10]. To further decrease the redundancy, multi-view video coding (MVC) is proposed as an extension of the H.264/MPEG-4 AVC standard. It introduces the concept of disparity-compensated prediction [11]. By exploiting the inter-view dependency, MVC enables a higher compression efficiency than separate view coding [12]. However, to arrive at a given level of video quality, the transmission rate still increases nearly linearly with the number of views. The large data overhead still poses a challenging problem.

Moreover, the digital video compression that current MVC adopts, relies on Shannon's *separate source and channel coding approach* [13]. The video is encoded at the transmitter first at a specific coding rate, which is called source coding. Then adaptive modulation and channel coding is adopted to facilitate reliable bit transmission. Over the past decades, coded transmission has dominated the existing wireless video transmissions. However, there still exists several weaknesses in this framework. First of all, the quantization process involved is a lossy process and the encoded video quality depends on the coding rate at the source. Once the video is encoded, its quality will not improve any more even if the channel condition allows the transmission of a better quality video. When the channel quality degrades, the received video may be garbled due to error spread and packet loss. This is called the *cliff effect* [14]. Moreover, the efficiency of channel coding greatly depends on the timeliness and precision of the channel feedback. In practice, the encoder adjusts the source coding rate according to the buffer size and the transmitter chooses the most appropriate modulation rate to transmit the packet. This usually requires a quite complicated bit and power allocation solution. In MVV,

Manuscript received Sept. 18, 2019; accepted Oct. 4, 2019. This work is supported in part by the US NSF under Grants IIP-1822055, ECCS-1923717, and CNS-1702957, and by the Wireless Engineering Research and Education Center (WEREC) at Auburn University, Auburn, AL, USA.

T. Zhang and S. Mao are with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201, USA. Email: tzz0031@tigermail.auburn.edu, smao@ieee.org.

DOI: 10.1109/ACCESS.2019.2946607

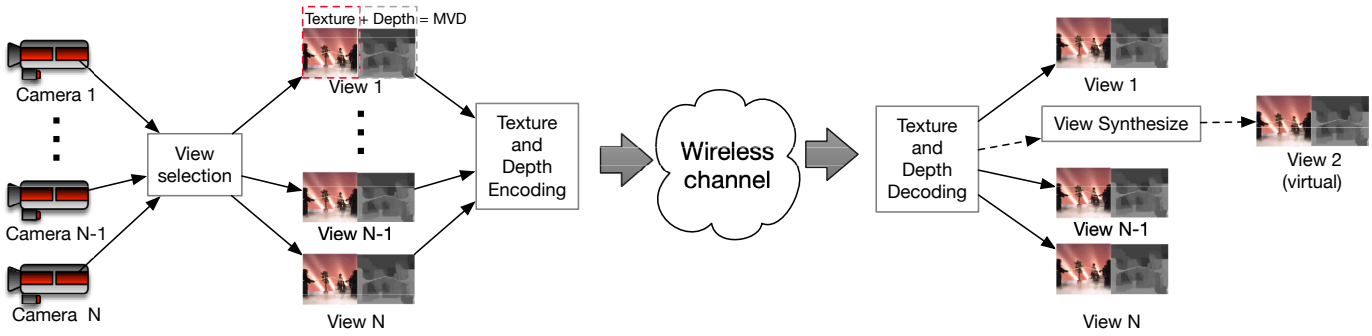

Fig. 1: Illustration of a multi-view video transmission system.

where a smooth navigation with 3D scenes with a minimum delay is required, timely and accurate feedback and a large buffer size are more indispensable to process and transmit the huge video data traffic.

To overcome the problems in conventional digital video transmission, soft video delivery is proposed in [15], where *joint channel and source coding* is exploited. By skipping quantization and entropy coding, the video frames are directly processed by a three-dimensional discrete cosine transformation (3D-DCT). Then the DCT coefficients are scaled like amplitude modulation (AM) to minimize the end-to-end distortion. Since all operations involved are linear, the pixel distortion is proportional to the noise power and there is no cliff effect. Users can gracefully improve the video quality commensurate with their wireless channel quality.

In this paper, we will incorporate the soft video delivery technique for wireless MVV transmissions. We use a five-dimensional discrete cosine transformation (5D-DCT) to jointly process video texture and depth frames from different cameras. The output is scaled and modulated in an analog manner. We investigate the complex resource control problem of soft MVV transmission in the form of two types of problems, i.e., *distortion-resource* (DR) optimization and *resource-distortion* (RD) optimization. Efficient algorithms are proposed to find an optimal solution to the formulated problems. The proposed schemes are evaluated using reference MVD videos as well as traditional single view monotone video sequences, while both the *objective performance metric* peak-signal-to-noise ratio (PSNR) and the *perceptual performance metric* structural similarity (SSIM) [16] are used for video quality assessment.

The main contributions made in this paper are summarized in the following:

- To the best of our knowledge, this is one of the first works that considers a resource allocation problem for practical soft MVV wireless transmissions. The resource allocation problem is NP-hard. We proposed efficient algorithms to find the optimal solutions to the DR problem and the RD problem. The resource usage can be greatly reduced without significant video quality degradation with the proposed schemes.
- We find that there exists an interesting tradeoff between channel and power usage in MVV transmission. The impact of power allocation across texture and depth frame and the impact of view positions are investigated to achieve a high video quality in each virtual viewpoint.
- Simulation results with both MVV videos and single-view videos demonstrate that the proposed algorithm works well not only in referenced views but also in synthesized virtual views, while considerable savings in channel usage can be achieved.

The remainder of this paper is organized as follows. In Section II, we introduce the related work on conventional digital based MVV transmission and soft video transmission. In Section III, the framework of soft MVV transmission is presented. We consider practical DR optimization in Section IV and and RD optimization in Section V. Then, extensive simulations are performed to demonstrate the advantages of the proposed resource allocation algorithms in MVV transmission in Section VI. Finally, Section VII concludes the paper.

II. Related work

We divide our discussion on related work into two parts. In this section, we first provide a brief introduction of the recent work on MVV and MVD transmissions. We then review related works on soft video transmission.

A. Multiview Video Transmission

MVD is a simple and effective extension of MVV by providing camera-depth information. It enables efficient depth-image based rendering so that virtual views can be generated from a limited number of source views. The texture data is captured by multiple cameras. Meanwhile, every texture is accompanied with depth information. The 3D HEVC standard [12] exploits the dependencies between texture and depth information to remove the redundancy. The encoding process is realized by spatial prediction within each frame, temporal motion-compensation between different frames, transform coding of the prediction residual, and entropy coding [17]. To code the depth data, new intra coding modes, modified motion compensation, and motion vector coding are used.

Delivering MVV/MVD content over existing MVV streaming networks faces many challenges including network bandwidth variation, packet loss, delay and client view selection uncertainty [18]. So far, interactive MVV streaming, 3D

video coding, and practical system implementation have been studied, e.g., see recent works [4], [5], [19]–[21]. However, very few works are focused on wireless MVV streaming. As introduced before, delivering 3D MVV video is a very challenging task over today's wireless networks. It requires to carry a potentially much larger data traffic generated from large number of cameras, with a strict requirement on latency on the complex wireless environments. In [22], the authors incorporated MVV with the multiple input multiple output (MIMO) technique that employed precoding and spatial multiplexing for simultaneous transmissions. A resource control algorithm was proposed to achieve unequal error protection against channel errors. Ref. [23] considered MVV transmission with multiple description coding. Multiple descriptions from texture and depth data of adjacent views were transmitted through separate wireless channels, so that the multi-path diversity could be exploited for improved reliability.

B. Soft Video Transmission

The interesting concept of soft video delivery (SoftCast) was first proposed in [24], [25]. Unlike traditional digital video transmission, SoftCast builds an analog code that achieves the compression-protection tradeoff with a suitable power allocation. Experiments demonstrate that the cliff effect in conventional digital video transmission can be avoided and users can enjoy a graceful video quality improvement according to the channel condition.

The SoftCast concept attracted considerable interest in the community. For example, Ref. [26] replaced the power allocation scheme in SoftCast with bit allocations, while Ref. [27], [28] combined the benefit of SoftCast and conventional digital video coding by considering an analog-digital hybrid coding scheme. In [29], the authors proposed an optimal channel and power allocation scheme under fast fading channels. The multiple antenna technique was exploited to improve the system performance. For example, Ref. [30] decomposed the MIMO channel into parallel sub-channels by MIMO precoding. By assigning high priority DCT coefficients to higher quality sub-channels, the reconstructed video quality could be optimized. Ref. [31] extended SoftCast to a wireless video multicast scenario with receiver antenna heterogeneity. The proposed scalable video multicast system allowed receivers to have a reconstructed video quality that was commensurable with the number of equipped antennas. In [32], a curve-fitting based source control algorithm was developed to find the cost distortion relationship, where cost consisted of bandwidth and transmit power, for soft video delivery.

To the best of our knowledge, Ref. [33] was the first work that investigated the soft video transmission for MVD. In this work, the metadata overhead could be greatly reduced with the proposed Gaussian Markov random filed (GMRF) model, and thus a better video quality could be achieved. However, compared with the overhead incurred by metadata, there exist a huge redundancy in the coded video data, especially in MVD transmission. How to jointly optimize the resources used, while achieving a satisfactory video quality, is still an open problem. In [34], we proposed a blind data detection method that recovered received video from the squared amplitude of received signals, which was almost metadata free. This work [34] was designed for a generic video and the AWGN channel.

III. Soft Video Transmission for MVV Delivery

In MVV delivery, the transmitter adopts multiple cameras to record the multi-color texture and depth frames. When the transmitter is notified via the feedback channel of the receiver preferred virtual viewpoint, it captures the data at several adjacent cameras near the requested virtual viewpoint. These captured data are then encoded and transmitted to the receiver. At the decoder side, the requested viewpoint is synthesized from the decoded texture and depth frames via DIBR.

In this section, we consider the case when there is plenty of bandwidth in the transmission channel and all the DCT coefficients will be transmitted to the receiver. Therefore we will focus on *power allocation problem*. At the encoder, a 5D-DCT is used for the entire texture and depth frames in one group of picture (GOP), which is a sequence of successive MVD video frames. After power allocation for each DCT coefficient, the DCT coefficients are then mapped to I (in-phase) and Q (quadrature-phase) components for analog wireless transmission.

Specifically, the DCT coefficients are divided into N rectangular chunks with size $h \times w$. Let $x_i[j]$ denote the jth DCT coefficient in the ith chunk. We scale all the DCT coefficients in the ith chunk by a common scaling factor g_i for noise reduction. The scaled coefficient $s_i[j]$ is denoted as follows

$$s_i[j] = g_i \cdot x_i[j]. \tag{1}$$

This analog-like scaling is also called power allocation. The optimal power scaling factor is obtained by minimizing the end-to-end distortion under a constrained power budget P as follows [29].

$$(\mathbf{P0}) \quad \min_{\rho_i} \ \mathrm{MSE} = \mathbb{E}\left[(x_i[j] - \hat{x}_i[j])^2\right] = \frac{1}{N}\sum_{i=1}^{N} \frac{\lambda_i}{\rho_i + 1} \tag{2}$$

$$\text{s.t.} \ \ \frac{1}{N}\sum_{i=1}^{N} \rho_i \leq \frac{P}{Nhw\sigma_n^2} := \widetilde{P}, \tag{3}$$

where $\mathbb{E}$ denotes the expectation, $\hat{x}_i[j]$ is the estimated DCT coefficient at the receiver, N is the number of DCT chunks, $\lambda_i = \mathbb{E}[|x_i[j]|^2]$ is the average power of all the DCT coefficients in the ith chunk, σ_n^2 is the noise variance, $\rho_i = \frac{g_i^2 \lambda_i}{\sigma_n^2}$ is the signal noise ratio for chunk i after power allocation, and $\widetilde{P}$ is the signal-to-noise ratio (SNR) for each GOP.

To obtain the optimal power allocation, we solve ($\mathbf{P0}$) with the Lagrange multiplier method [24]. That is, we first define a Lagrange multiplier $\gamma > 0$ and the corresponding Lagrange function $\mathcal{L}$ as

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \frac{\lambda_i}{\rho_i + 1} - \gamma\left(\frac{1}{N}\sum_{i=1}^{N} \rho_i - \widetilde{P}\right). \tag{4}$$

By setting $\frac{\partial \mathcal{L}}{\partial \rho_i} = 0$, $i = 1, 2, ..., N$, and $\frac{\partial \mathcal{L}}{\partial \gamma} = 0$, we obtain the optimal solution as

$$\rho_i^* = \frac{N\sqrt{\lambda_i}}{\sum_{i=1}^{N}\sqrt{\lambda_i}}(\widetilde{P}+1) - 1 \approx \frac{N\sqrt{\lambda_i}}{\sum_{i=1}^{N}\sqrt{\lambda_i}}\widetilde{P} \tag{5}$$

$$g_i^* = \sqrt{\frac{\sigma_n^2 \cdot \rho_i}{\lambda_i}} \tag{6}$$

After demodulation, the receiver receives $y_i[j] = s_i[j] + n_i[j]$, where $n_i[j]$ is the additive white Gaussian noise (AWGN) with a variance σ_n^2. The DCT coefficients are then extracted from the I and Q components using a *linear least square estimator* (LLSE) filter as

$$\hat{x}_i[j] = \frac{g_i \cdot \lambda_i}{g_i^2 \cdot \lambda_i + \sigma_n^2} \cdot y_i[j]. \tag{7}$$

The decoder then takes an inverse 5D-DCT on the DCT coefficients $\hat{x}_i[j]$ to recover the video sequence. Finally, the decoder synthesizes the virtual viewpoint from the received texture and depth frames with DIBR.

IV. Distortion Resource (DR) Optimization

We next consider the more realistic case with limited channel resource (in the form of time slots or frequency bands). In soft video delivery, each scaled DCT chunk is transmitted in different time slots or frequency bands. To transmit a video of a large size, e.g., MVDs, a considerable amount of channel resources is required. This makes it hard for real time delivery. Fortunately, due to the compacting nature of DCT, most of the DCT components in high spatial frequency domain tend to have very small values. Therefore, we can discard a certain amount of high-frequency DCT chunks to satisfy the channel resource constraints, while not degrading the video quality too much. It is also worth noting that even if there are sufficient channel resources, it may still be helpful to drop some DCT chunks, since the saved power (by dropping some chunks) can be utilized more efficiently by re-allocating it to other more important chunks, especially when the power constraint is stringent. Then we have a *joint chunk selection and power allocation problem*.

A. Problem Statement

Given a set of N chunks of DCT coefficients with average energy denoted by $\lambda_1, \lambda_2, ..., \lambda_N$. Without loss of generality, we assume $\lambda_i \geq \lambda_j$, for all $i > j$. Let M be the amount of available channel resources (e.g., time or frequency slots) and P be the total power constraint for each chunk (and $\widetilde{P}$ be the SNR budget for each chunk). We use a binary channel allocation vector $\mathbf{k} = [k_1, k_2, ..., k_N]$ to denote the chunk selection of each GOP: $k_i = 0$ indicates that chunk i is discarded, and $k_i = 1$ means that chunk i is transmitted via a channel resource slot. We aim to find the optimal channel allocation $\mathbf{k}^*$ and the optimal power allocation $\rho^* = [\rho_1^*, \rho_2^*, ..., \rho_N^*]$, so that the total video distortion is minimized. Mathematically, the problem can be formulated as follows.

$$(\mathbf{P1}) \quad \min_{\mathbf{k},\rho} \ \text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\frac{\lambda_i}{k_i\rho_i + 1} \tag{8}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} k_i \leq M \tag{9}$$

$$\frac{1}{N}\sum_{i=1}^{N} k_i\rho_i \leq \widetilde{P} \tag{10}$$

$$k_i \in \{0, 1\}, \quad 1 \leq i \leq N \tag{11}$$

$$\rho_i \geq 0, \quad 1 \leq i \leq N. \tag{12}$$

Intuitively, since k_i only takes binary values, M should be in the range of $[0, N]$. If $M = N$, Problem (**P1**) will be exactly the same as Problem (**P0**). If $M < N$, to minimize distortion, we will retain the largest M chunks, which we refer to as *high-priority* (HP) data, and discard the remaining smaller chunks, which we call *low-priority* (LP) data, i.e.,

$$k_i^* = \begin{cases} 1, & \text{if } i \leq M \\ 0, & \text{if } i > M. \end{cases} \tag{13}$$

The problem then becomes finding the optimal value M^* and the optimal power allocation $\{\rho_i^*\}$. Similarly, by the Lagrange multiplier method, we can derive the optimal solution of ρ_i as

$$\rho_i^* \approx \frac{N\sqrt{\lambda_i}}{\sum_{i=1}^{M}\sqrt{\lambda_i}} \cdot \widetilde{P} \cdot k_i. \tag{14}$$

Then the MSE can be expressed as

$$\text{MSE} = \frac{1}{N}\left(\sum_{i=1}^{M}\frac{\lambda_i}{\rho_i + 1} + \sum_{i=M+1}^{N}\lambda_i\right). \tag{15}$$

It can be seen that the total MSE in (15) can be expressed as a function of the power budget P and the channel resource constraint M. Increasing the power budget will lead to an increase of ρ_i, which will help decrease the distortion. In other words, MSE is a monotone function in terms of P (or $\widetilde{P}$). However, it is still unclear how to find the optimal value of M because of its discrete value and that it appears in the superscription of the summation term in (14). Such discrete nature makes it hard to obtain the optimal value M^* in closed-form as what we did in the case of ρ_i^*.

B. A Greedy Search Approach

To find the optimal value M^*, we propose an exhaustive search based algorithm, as presented in Algorithm 1. The main idea is that, the transmitter has full knowledge of λ_i's and the total power budget P, it can find the number of chunks that minimize the MSE by searching all the possible discrete channel resources in an exhaustive manner. With the optimal chunk selection, the video is actually compressed but without too much performance degradation, and users can enjoy a better experience since the transmission time is saved and the amount of video traffic is reduced. Meanwhile, the saved channel resources can be utilized by other users in the network.

In Algorithm 1, we first compute the initial energy distribution of the chunks in each GOP. Based on the information,

Algorithm 1 Distortion Resource Optimization Algorithm

1: Initialize $\lambda_i = \mathbb{E}[|x_i[j]|^2]$, for all i ;
2: **for** $n = 1, 2, ..., N$ **do**
3: $M \leftarrow n$;
4: Calculate $\rho_{i,n}$ according to (13) and (14) ;
5: Calculate MSE_n according to (15) ;
6: Calculate PSNR_n as in (30) in Section VI-A ;
7: **end for**
8: $\mathrm{PSNR}_{\max} = \max_{n\in\{1,2,...,N\}} \mathrm{PSNR}_n$;
9: **for** $n = 1, 2, ..., N$ **do**
10: **if** $\mathrm{PSNR}_n > \alpha \cdot \mathrm{PSNR}_{\max}$ **then**
11: $M^* \leftarrow n$;
12: break ;
13: **end if**
14: **end for**
15: Calculate $\{k_i^*\}$ and $\{\rho_i^*\}$ using M^* as in (13)-(14) ;
16: Output M^*, $\{k_i^*\}$, and $\{\rho_i^*\}$;

we find the optimal channel allocation and power allocation in an exhaustive manner. By Line 8, the algorithm can actually terminate and output the optimal channel number M^*, the optimal chunk selection $\{k_i^*\}$, and the optimal power control $\{\rho_i^*\}$. However, in Section VI, we observe from simulations that the Distortion-Resource (DR) curve tends to have a flat tail (e.g., see Figs. 3 and 4). This means that we are using a much larger number of channel resources to achieve only a slight improvement in PSNR. This is obviously inefficient. Hence, we introduce a control parameter $0 < \alpha \leq 1$ in Line 10 to search for a sub-optimal solution. By slightly sacrificing the PSNR performance, we can significantly reduce the video traffic and the channel resource usage.

C. Complexity Analysis

In Algorithm 1, we traverse all the possible n values in a greedy manner and for each n, the involved operations are all linear. Hence the complexity of Algorithm 1 is $\mathcal{O}(N)$, which is negligible. For one GOP, the main complexity comes from the energy sorting process of the N chunks, with complexity $\mathcal{O}(N \log N)$. However, in practice, we do not necessarily need to sort all these chunks strictly according to their energy distribution. Instead, a more feasible way would be to sort these chunks in a *zigzag* scanning manner, which is used in the JPEG image compression. In this way, the sorting process can be avoided and the complexity can be greatly reduced. Moreover, in simulations, we find that for consecutive video frames, the optimal value M^* does not change too much. This means that we may only need to find the corresponding M^* for the first GOP, and then applies this M^* to the remaining GOPs, which further reduces the complexity.

V. Resource Distortion (RD) Optimization

A. Problem Statement

Note that for a specific video sequence, different GOPs may have different levels of compressibility. In DR optimization, under a fixed power and channel budget, the distortion of consecutive GOPs may have large variations, which become quite annoying for viewers (although the overall PSNR could be maximized). In addition, human eyes are less sensitive to the differences of videos when the PSNR is very high. Based on these two observations, keeping the distortion relatively more stable may be a better choice. The saved power and channel resources this way can also be utilized by other users. Therefore, our problem becomes to find a good combination of chunk selection and power allocation for a target distortion. This problem is called resource distortion (RD) optimization.

Since RD optimization involves distortion, channel, and power usage, it is a three-dimensional optimization problem, which is hard to solve. However, by fixing one factor, we can decompose the difficult problem into two sub-problems. For example, we can formulate a power distortion optimization problem that aims to minimize the power resources usage under a target distortion constraint $\overline{\mathrm{MSE}}$ and a channel usage constraint M. The problem can be stated as follows.

$$(\mathbf{P2}) \quad \min_{\rho} \ \widetilde{P} = \frac{1}{N}\sum_{i=1}^{m} \rho_i \tag{16}$$

$$\text{s.t.} \ \ \frac{1}{N}\left(\sum_{i=1}^{m} \frac{\lambda_i}{\rho_i + 1} + \sum_{i=m+1}^{N} \lambda_i\right) \leq \overline{\mathrm{MSE}} \tag{17}$$

$$1 \leq m \leq M,\ m \in \mathbb{Z}. \tag{18}$$

Alternatively, we can formulate a channel distortion optimization problem that aims to minimize the channel resource usage under a distortion constraint $\overline{\mathrm{MSE}}$ and a power budget $\widetilde{P}$ as follows.

$$(\mathbf{P3}) \quad \min_{\rho} \ M \tag{19}$$

$$\text{s.t.} \ \ \frac{1}{N}\left(\sum_{i=1}^{M} \frac{\lambda_i}{\rho_i + 1} + \sum_{i=M+1}^{N} \lambda_i\right) \leq \overline{\mathrm{MSE}} \tag{20}$$

$$\frac{1}{N}\sum_{i=1}^{M} \rho_i \leq \widetilde{P}.s \tag{21}$$

Since the channel usage variable M is discrete while the power scaling factor g_i is continuous, these two problems belong to the class of mixed integer non-linear programming problems, which is generally NP-hard.

B. Power Distortion Optimization

To find the minimal power use in Problem ($\mathbf{P2}$), we search all the feasible $m \in [1, M]$ in an exhaustive manner. For each fixed channel resource usage m, we solve the following sub-problem

$$(\mathbf{P2a}) \quad \min_{\rho} \ \widetilde{P} = \frac{1}{N}\sum_{i=1}^{m} \rho_i \tag{22}$$

$$\text{s.t.} \ \ \frac{1}{N}\left(\sum_{i=1}^{m} \frac{\lambda_i}{\rho_i + 1} + \sum_{i=m+1}^{N} \lambda_i\right) \leq \overline{\mathrm{MSE}} \tag{23}$$

$$\rho_i \geq 0. \tag{24}$$

Define Lagrange multiplier $\mu > 0$, then the Lagrange function can be written as

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{m}\rho_i + \mu\left(\frac{1}{N}\left(\sum_{i=1}^{m}\frac{\lambda_i}{\rho_i+1} + \sum_{i=m+1}^{N}\lambda_i\right) - \overline{\mathrm{MSE}}\right). \tag{25}$$

Now we set $\frac{\partial \mathcal{L}}{\partial \rho_i} = 0$, $i = 1, 2, ..., m$, and $\frac{\partial \mathcal{L}}{\partial \mu} = 0$, to have

$$\rho_i^* = \begin{cases} \frac{\frac{1}{N}\cdot\sqrt{\lambda_i}\cdot\sum_{j=1}^{m}\sqrt{\lambda_j}}{\overline{\mathrm{MSE}} - \frac{1}{N}\sum_{j=m+1}^{N}\lambda_j} - 1, & i = 1, 2, ..., m \\ 0, & i = m+1, ..., N. \end{cases} \tag{26}$$

The corresponding power distortion optimization algorithm is presented in Algorithm 2. Note that constraint (23) in Problem (**P2a**) implies that $\overline{\mathrm{MSE}} - \frac{1}{N}\sum_{i=m+1}^{N}\lambda_i > 0$ for a certain m. If this condition is violated, there will be no feasible solution and we will search for the next value of m. As shown in Algorithm 2, Line 5, the objective function of Problem (**P2a**) will have a closed form. By comparing all such $\widetilde{P}$s, we choose the one that has the smallest value and adopt the corresponding power allocation $\{\rho_i^*\}$.

Algorithm 2 Power Distortion Optimization Algorithm

1: Initialize $\lambda_i = \mathbb{E}[|x_i[j]|^2]$, for all i, and $\overline{\mathrm{MSE}}$;
2: **for** $m = 1, 2, ..., M$ **do**
3: $M \leftarrow m$;
4: Calculate $\{\rho_{i,m}\}$ according to (26) ;
5: Calculate $\widetilde{P}_m$ according to (22) ;
6: **end for**
7: $\widetilde{P}^* = \min_{m\in\{1,2,...,M\}} \widetilde{P}_m$;
8: Output $\widetilde{P}^*$ and the corresponding $\{\rho_i^*\}$;

C. Channel Distortion Optimization

For channel distortion optimization, we can search the value M in a descending order. For each M, we solve the following subproblem.

$$(\mathbf{P3a}) \quad \min_{\rho} \ \mathrm{MSE} = \frac{1}{N}\left(\sum_{i=1}^{M}\frac{\lambda_i}{\rho_i+1} + \sum_{i=M+1}^{N}\lambda_i\right) \tag{27}$$

$$\text{s.t.} \ \ \frac{1}{N}\sum_{i=1}^{M}\rho_i \le \widetilde{P}. \tag{28}$$

If the objective value of Problem (**P3a**) is less than $\overline{\mathrm{MSE}}$, then corresponding M is feasible. Then we decrease the value of M by 1 and solve Problem (**P3a**) again, until we find an M that is infeasible. For each fixed M, Problem (**P3a**) is a simple convex optimization problem. By the Lagrange multiplier method, we obtain the optimal solution as

$$\rho_i^* \approx \frac{N\cdot\sqrt{\lambda_i}}{\sum_{j=1}^{M}\sqrt{\lambda_j}}\cdot\widetilde{P}. \tag{29}$$

We present the procedure in Algorithm 3. Note that Algorithms 1–3 are all greedy search algorithms and they share similar procedures. Thus the complexity analysis for Algorithm 1 also applies to Algorithm 2 and Algorithm 3.

Algorithm 3 Channel distortion optimization algorithm

1: Initialize $\lambda_i = \mathbb{E}[|x_i[j]|^2]$, for all i, $\overline{\mathrm{MSE}}$, $m = N+1$;
2: **repeat**
3: $m = m - 1$;
4: Calculate $\rho_{i,m}$ according to (29) ;
5: Calculate $\mathrm{MSE_m}$ as in (27) ;
6: **until** $\mathrm{MSE_m} < \overline{\mathrm{MSE}}$
7: $M^* \leftarrow m$, $\rho_i^* \leftarrow \rho_{i,m}$, for all i ;
8: Output M^* and $\{\rho_i^*\}$;

D. Power and Channel Usage Tradeoff

For a target video quality, we can either optimize the power consumption under a given channel usage constraint or optimize the channel usage under a constrained power budget. Therefore, we can obtain a power and channel usage tradeoff curve. On each point of the curve, the combination of the corresponding channel and power usage achieves the same target video quality. This tradeoff curve provides us a useful guideline for choosing a suitable power and channel usage pair based on practical resource constraints. In multi-user system, different viewers may have diverse power and channel resource budgets, where a joint optimization can be applied to save resource consumption.

E. View Synthesis

In MVV transmissions, the texture video data contains detailed video content information, while the depth data plays important roles in view synthesis. The quality of both the texture and depth frames determines the virtual view quality. In digital MVV transmissions, bit allocation and power assignment are performed to ensure a good virtual view synthesis performance, which is usually quite complicated [12], [35]. In [35], the view synthesis optimization algorithm is integrated into the encoding process to enable rate-distortion optimization. To achieve a bit rate distribution balance between the texture data and the depth data, a complex combinatorial optimization problem has to be solved. For a two-view scenario, the video/depth rate distribution can be 86%/14%. This way, the depth data is encoded at a low cost. Similar to the bit rate distribution in digital transmissions, in soft video transmissions, we will investigate the power allocation across the texture and depth video data.

Suppose there are N views in the system, as shown in Fig. 2. For simplicity, in this work, we consider the case where equal power control among different reference views is assumed. Power allocation between texture data and depth data is investigated. Specifically, before the 5D-DCT operation, each texture view is scaled by a common factor of β/N and each depth view is scaled by a factor of $(1-\beta)/N$. After 5D-DCT, all the video data are linearly transformed and modulated in an analog manner. At the decoder, an inverse process is performed. Followed by a digital renderer to generate the user's preferred virtual views. Compared with the complicated bit allocation in digital video transmissions, the proposed soft video framework simplifies the process into a power allocation

problem. Our target becomes to investigate the impact of the scaling factor β on the quality of the synthesized virtual view. We will provide our study and discussion on parameter β in Section VI-D1.

VI. Simulation Study

A. Parameter Setting

1) Performance Metric: In our simulation study, we use both the *objective performance metric* PSNR and the *perceptual metric* SSIM [16] for video quality assessment. PSNR is defined (in dB) as

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{(2^B-1)^2}{\mathrm{MSE}}\right), \tag{30}$$

where B is the number of bits used to encode pixel luminance (usually 8 bits) and MSE is the mean squared error between all the pixels between the decoded and the original video. In soft video delivery, since DCT is a linear transformation, the MSE stays the same after the transformation. Hence, we substitute (15) into (30) and we get the corresponding distortion. Generally, improvements of PSNR of magnitude larger than 0.5dB are visually noticeable. A PSNR below 20dB is considered not acceptable.

We also use SSIM to measure the similarity of the original and reconstructed images to test the performance of the proposed method [16]. For two $N \times N$ images x and y, the SSIM index is computed as [16]

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \tag{31}$$

where μ_x and μ_y are the means of x and y, respectively; σ_x^2 and σ_y^2 are the variances of x and y, respectively; σ_{xy} is the covariance of x and y; $c_1 = (k_1 \cdot L)^2$ and $c_2 = (k_2 \cdot L)^2$; L is the dynamic range of the pixel values; $k_1 = 0.01$, and $k_2 = 0.03$. An SSIM value closer to 1 suggests higher perceptual similarity between the original and the decoded image.

2) Test Video: We use two standard reference MVD videos, *balloons* and *kendo*, at 30 fps. We choose view points 1, 3, 5. Three cameras are used with a distance 10cm away. Their resolution is 1024×768 pixels for texture and depth frames at a frame rate of 20 fps. The video sequences are selected from the video database [36]. In addition, some standard single view monotone CIF video sequences from video database [37] are used in our simulations.

3) Parameter Setting: For soft video delivery, we set the GOP size to 4. In existing chunk-based schemes of soft video delivery, we divide each frame into $8 \times 8 = 64$ chunks. For MVD video, we read both the texture and depth data from the three cameras. Thus one GOP consists of $3 \times 2 \times 4 \times 64 = 1536$ chunks. The camera configuration is summarized in Table I. For Algorithm 1, we choose $\alpha = 0.98$. We use the 3D HEVC test model (HTM) v15.0 software [38] renderer to synthesize a virtual viewpoint from the received texture and depth frames. We assume the AWGN channel in the simulations.

TABLE I: Camera Configuration for Video Sequences *kendo.yuv* and *balloons.yuv*

View	FocalLength	Position	CShift	ZNear	ZFar
1	2241.25607	5.0	701.5	448.251214	11206.280350
3	2241.25607	15.0	701.5	448.251214	11206.280350
5	2241.25607	25.0	701.5	448.251214	11206.280350

B. DR Optimization Performance

We first evaluate the DR optimization performance. We investigate the maximum PSNR that can be achieved under a given resource (i.e., channel and power) constraint. As mentioned before, there will be 1536 chunks in each GOP. Suppose in each channel slot (e.g., time or frequency), only one chunk can be transmitted and let the maximum number of available channel slots M for each GOP be 1536. We fix the noise variance to be 1 and vary the total transmit power budget P for each GOP [1]. The DR optimization results for video sequences *kendo.yuv* and *balloons.yuv* are shown in Fig. 3 and Fig. 4, respectively.

It can be seen that for a given channel resource use N, the PSNR generally increases with the power budget $\widetilde{P}$. This is also confirmed by our discussion of (15). However, for a given power budget $\widetilde{P}$, the PSNR does not necessarily increase with the channel usage, especially when the power budget is low. For example, when the power budget $\widetilde{P}$ is 5dB, the maximum PSNR point for the video sequences *kendo.yuv* and *balloons.yuv* are attained when $N = 250$ and $N = 196$, respectively. A maximum channel usage of 1536 does not always lead to the highest PSNR. The reason is in soft video delivery, different chunks are not of equal importance although each of them consumes one channel use. Under a fixed power budget, allocating more power to HP chunks and allocating less power to LP chunks (or even discarding them) helps improve the PSNR performance. Finally, we note that these DR curves generally have a *flat* tail, which means when the PSNR is above a certain level, improving the channel use is not efficient as improving the power budget. For example, for video sequence *kendo.yuv*, when the power budget $\widetilde{P} = 10$dB and the channel use $N = 668$, the PSNR is 40.74dB. Now improving channel usages doesn't improve PSNR value any more. However, a power increase from 10dB to 15 dB brings a PSNR improvement of 4.52dB and a power increase from 10dB to 15dB brings a PSNR improvement of 8.28dB.

To further clarify the channel usage saving, we plot the corresponding performance for video sequence *kendo.yuv* in Fig. 5. As can be seen, conventional Softcast uses all the channel and power resources to achieve a good PSNR. However, with the proposed algorithm, we can achieve a slightly higher PSNR value with a reduced channel usage. For example, in Fig. 5 (d) when the power budget $\widetilde{P} = 5$dB, in conventional SoftCast, the channel usage is 1536 and the achieved PSNR is 35.903dB. However, with the proposed algorithm, we achieve

[1]Equivalently, we are changing the value of the SINR $\widetilde{P}$ for each GOP. Since we assume a fixed noise level, in the following, we do not distinguish P and $\widetilde{P}$ anymore.

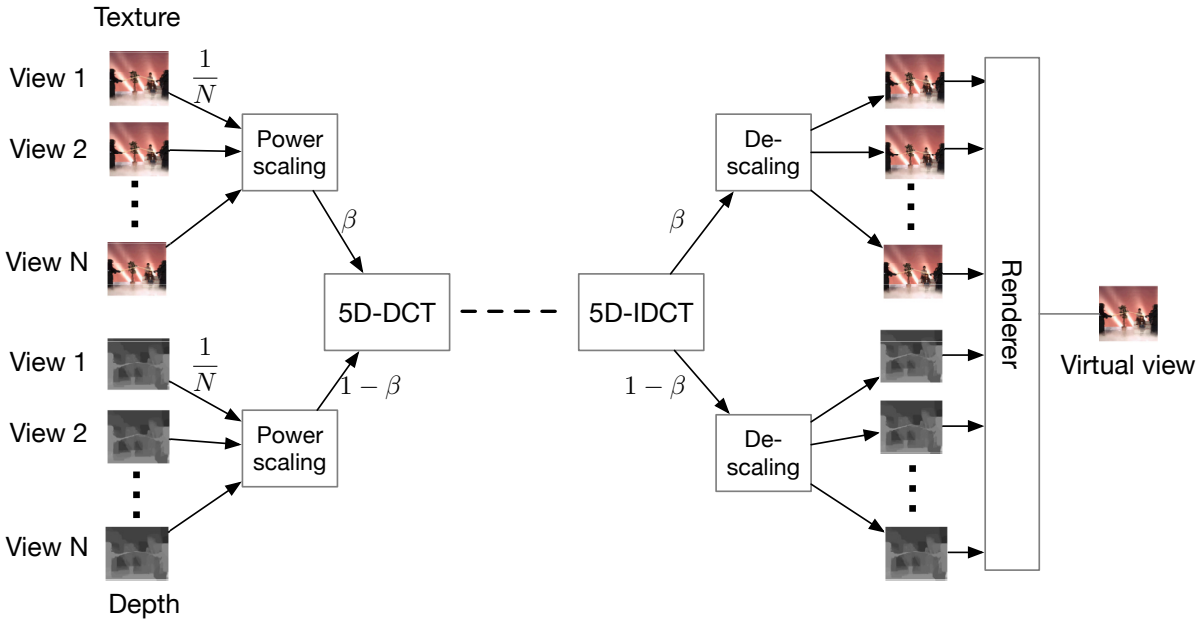

Fig. 2: Diagram of the proposed soft multi-view video delivery system.

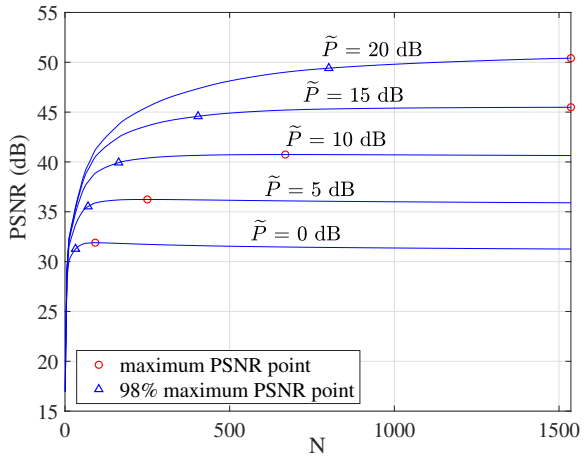

Fig. 3: DR optimization performance for MVD video sequence *kendo.yuv*.

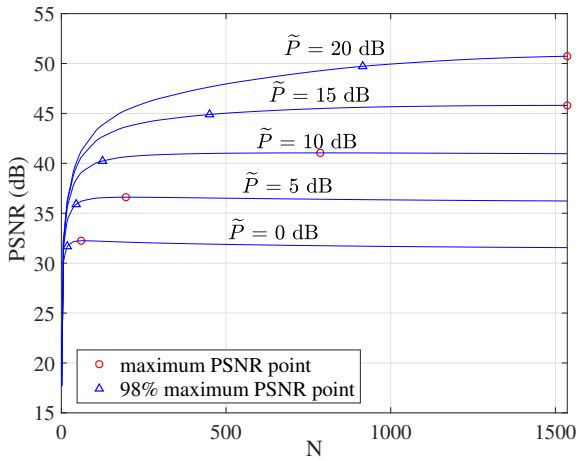

Fig. 4: DR optimization performance for MVD video sequence *balloons.yuv*.

the maximum PSNR of 36.236dB with only $250/1536 = 16.2\%$ of the original channel use. Moreover, due to the flat tail of the distortion resource curve, by slightly lowering the PSNR requirement (e.g., $\alpha = 98\%$ of the maximum PSNR), the channel usage can be further greatly reduced from 250 to 70, as shown in Fig. 5 (d). Comparing these figures, we also note that the proposed algorithm saves more channel usage when $\widetilde{P}$ is low. Hence the proposed method is more suitable for the case when channel condition is not good or the total power budget is limited.

Table II lists the channel usage comparison for different video sequences. The video file names in bold are MVD videos while others are standard monotone CIF video sequences used in SoftCast test, with a resolution of 352×288 at a frame rate of 20 fps. The proposed method significantly reduces the channel usage while still maintaining a satisfactory performance.

Note that the proposed method not only applies to MVV videos but also conventional standard videos. Considering that the channel usage grows linearly with the number of depth and camera data for MVV video transmissions, our proposed method is naturally more suitable to deal with the heavy data burden challenge caused by MVV videos.

C. RD Optimization Performance

1) Power Distortion Optimization Performance: Suppose the user requires a PSNR of 35dB, under a constant channel usage of 1536 chunks per GOP. We plot the power allocation

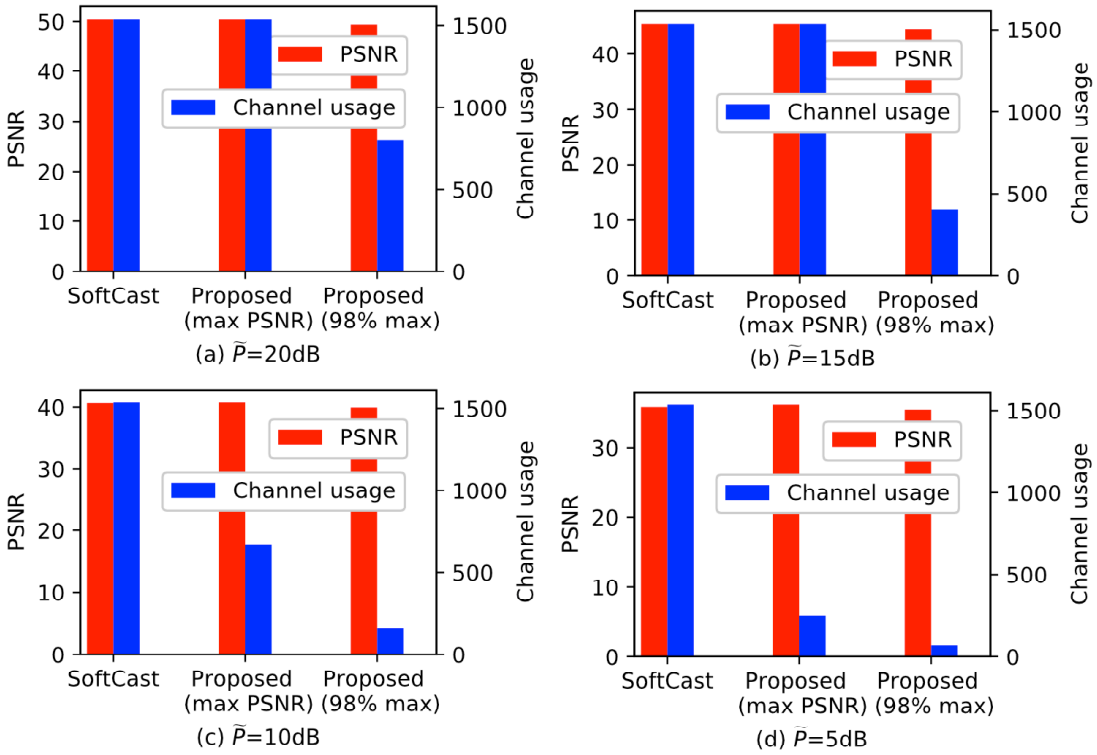

Fig. 5: Channel usage savings for MVD sequence *kendo.yuv*.

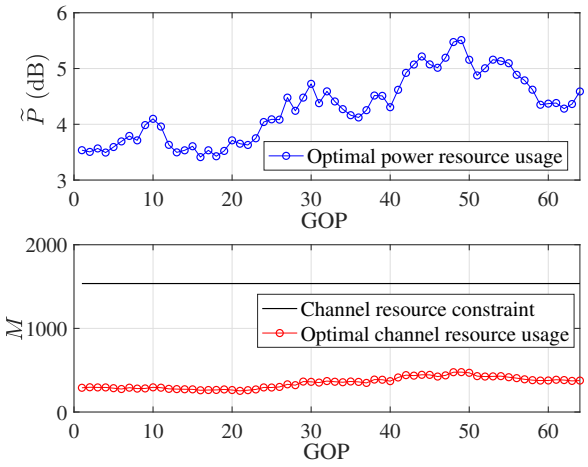

Fig. 6: Power allocation for MVV video sequence *kendo.yuv*.

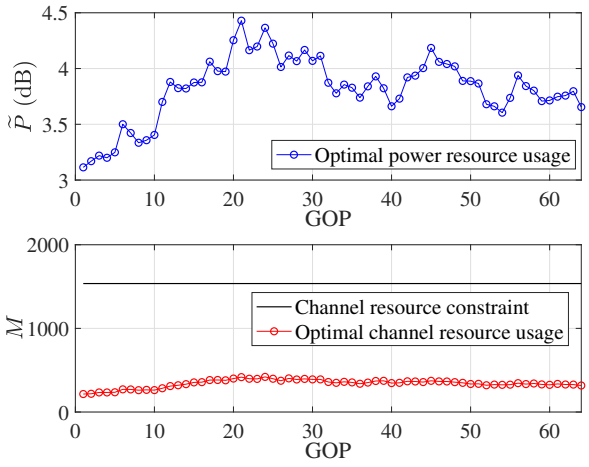

Fig. 7: Power allocation for MVV video sequence *balloons.yuv*.

curve for consecutive MVV video sequences *kendo.yuv* in Fig. 6 and that for *balloons.yuv* in Fig. 7. Specifically, we plot the resource usage curve for the first 64 GOPs in the figures. It can be seen that the power allocation of consecutive GOPs have very small variations and the channel usage is maintained at a relatively low level. This intricate resource control helps the MVV video quality remain at the prescribed PSNR value of 35dB. Hence the viewer can enjoy a favorable viewing experience and the saved wireless resources can be utilized by other users.

2) Channel Distortion Optimization Performance: Similarly, still suppose the user requires a PSNR of 35dB under a constant transmit power budget $\widetilde{P}$ = 10dB. We plot the channel allocation curve for different videos in Fig. 8 and Fig. 9. We note that the channel usage fluctuates at a relatively low and stable level compared with the large chunk number (1536) in one GOP. Hence, in practice, we can actually progressively allocate slightly more channel resources (say 100) without running algorithm 3 many times. The computational cost, therefore, can be further reduced.

TABLE II: Channel Usage Comparison for Different Video Sequences

(a) $\widetilde{P} = 20$dB

video	SoftCast	Proposed (max PSNR)	Proposed (98% max)
akiyo.yuv	256	256	53
bridge-close.yuv	256	256	203
forman.yuv	256	256	164
highway	256	256	205
stefan.yuv	256	256	150
mother-daughter.yuv	256	256	113
kendo.yuv	1536	1536	801
balloons.yuv	1536	1536	915

(b) $\widetilde{P} = 15$dB

video	SoftCast	Proposed (max PSNR)	Proposed (98% max)
akiyo.yuv	256	163	35
bridge-close.yuv	256	256	128
forman.yuv	256	224	116
highway	256	256	135
stefan.yuv	256	210	112
mother-daughter.yuv	256	190	63
kendo.yuv	1536	1536	404
balloons.yuv	1536	1536	450

(c) $\widetilde{P} = 10$dB

video	SoftCast	Proposed (max PSNR)	Proposed (98% max)
akiyo.yuv	256	52	27
bridge-close.yuv	256	171	49
forman.yuv	256	158	57
highway	256	185	49
stefan.yuv	256	152	93
mother-daughter.yuv	256	105	26
kendo.yuv	1536	669	163
balloons.yuv	1536	786	125

(d) $\widetilde{P} = 5$dB

video	SoftCast	Proposed (max PSNR)	Proposed (98% max)
akiyo.yuv	256	37	18
bridge-close.yuv	256	56	22
forman.yuv	256	65	21
highway	256	56	9
stefan.yuv	256	106	59
mother-daughter.yuv	256	38	12
kendo.yuv	1536	250	70
balloons.yuv	1536	196	45

(e) $\widetilde{P} = 0$dB

video	SoftCast	Proposed (max PSNR)	Proposed (98% max)
akiyo.yuv	256	21	8
bridge-close.yuv	256	22	7
forman.yuv	256	20	6
highway	256	10	2
stefan.yuv	256	54	23
mother-daughter.yuv	256	16	5
kendo.yuv	1536	92	32
balloons.yuv	1536	60	18

3) Channel Power Tradeoff: For a target video quality, there exists a tradeoff between power usage and channel usage.

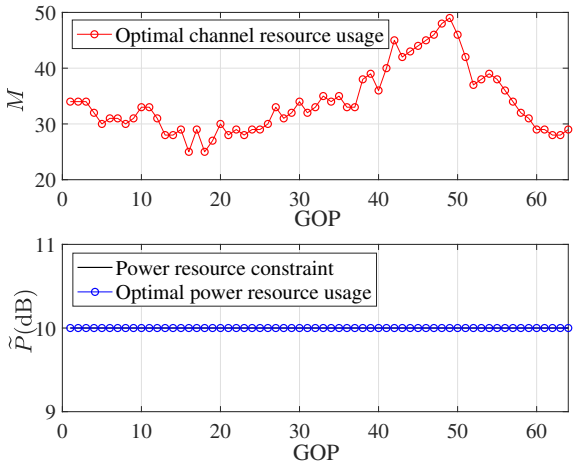

Fig. 8: Channel control for video sequence *kendo.yuv*.

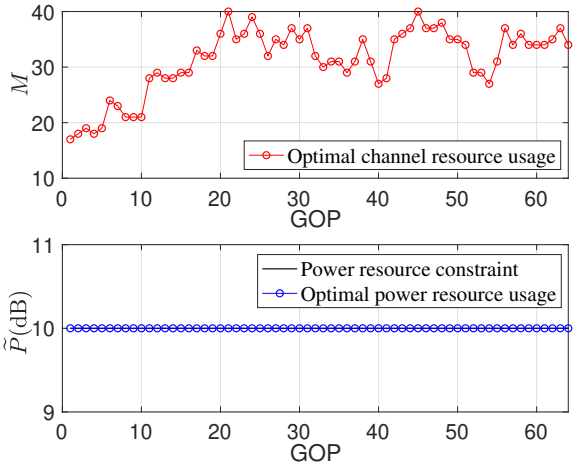

Fig. 9: Channel control for video sequence *balloons.yuv*.

We plot the tradeoff curve in Fig. 10 and Fig. 11. It can be observed that this kind of tradeoff varies with different target PSNR values. Under a lower target PSNR, both the power and channel usage required are quite low. Moreover, these curves tend to have a very sharp turning point when M is relatively low. Hence, maintaining the power and channel usage pair near the turning point would be an efficient strategy. This observation has also been confirmed by the previous simulations.

D. View Synthesis

1) Impact of Parameter β : As mentioned before, the power scaling factor β determines the power allocation ratio between the texture data and the depth data, which has a joint impact on the quality of synthesized virtual view. For each view, the power ratio between the texture data and the depth is $\beta/(1-\beta)$. $\beta = 0.5$ means an equal power control between texture data and depth data and $\beta > 0.5$ means more power is allocated to the texture data. We plot the impact of parameter β on the quality of the synthesized virtual view in Fig. 12.

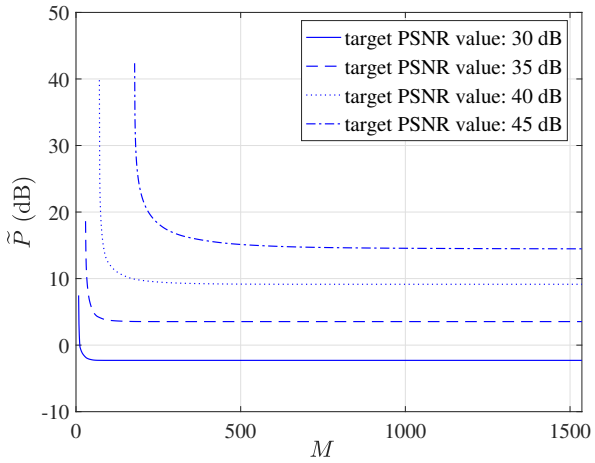

Fig. 10: Power and channel usage tradeoff curve for video sequence *kendo.yuv*.

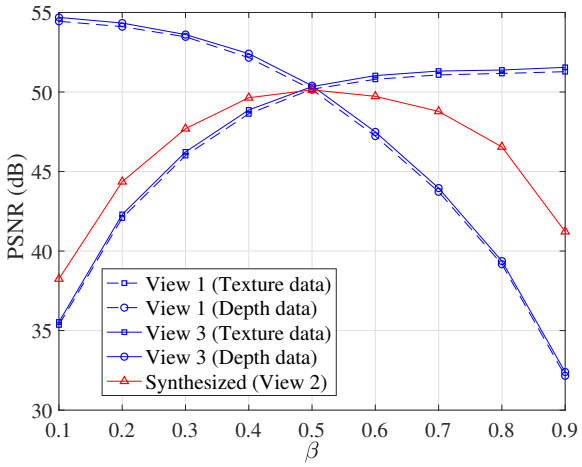

Fig. 12: Impact of parameter β on view 1, synthesized virtual view 2, and view 3.

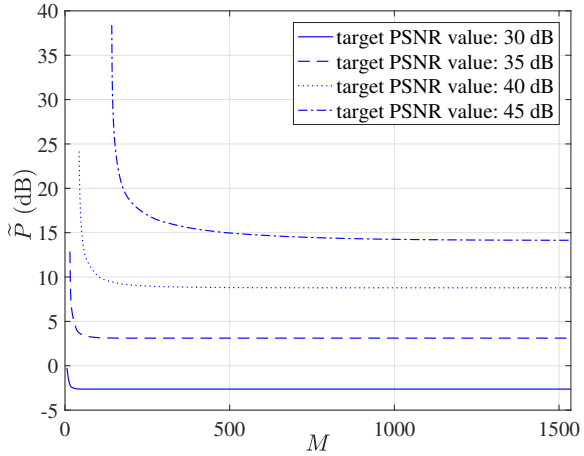

Fig. 11: Power and channel usage tradeoff curve for video sequence *balloons.yuv*.

The test video sequence is *kendo.yuv*. We investigate the PSNR performance for the first frame in the first GOP of view 1, view 3, and the corresponding virtual views. The virtual view is synthesized based on the texture and depth data from view 1 and view 3. Recall that 5D-DCT is a linear transformation and it is performed on all the texture and depth data from each relevant views. Hence the MSE distortion of each component (texture data and depth data from each view) can be actually approximated by the average MSE of its corresponding GOP. This explains why the texture and depth data quality curve from view 1 crossed the curve from view 3 when $\beta = 0.5$. If we increase the value of β, we intend to allocate more power to texture data, hence the PSNR of the texture data increases and the PSNR of the depth data decreases. Remember that in this paper we assume power is equally allocated between different view points, hence view 1 and view 3 is of equal importance in viewing synthesis. That is why the PSNR quality curve of view 1 fits well with that of view 3. In real systems, there may be 50 or even more cameras, while a viewer may only be interested in one specific view. An equal power allocation for each view is obviously not optimal. Instead of performing an equal power control among different cameras, it may be a better strategy to allocate more power to the adjacent views of the user's chosen one. This problem will be addressed in our future work. Finally, we note that the synthesized virtual view 2 reaches its best quality around $\beta = 0.5$. Moreover, we note that when β is between 0.4 to 0.6, the virtual view quality remains at a relatively high level. As β gets closer to either end of the interval [0,1], the synthesized virtual view quality drops dramatically. This is because when β is small (or large), there is a huge distortion in the texture (or the depth) data. This kind of imbalance between the texture data and depth data degrades the quality of the synthesized view. From this figure, we can see a good choice of β is between 0.5 and 0.6 where the texture view qualities at view 1 and view 3 are slightly increased and the view quality at virtual view 2 almost remains at a constant high level.

2) Impact of Virtual View Positions: In this experiment, view 3 and view 5 are reference views and the views between them are synthesized virtual views. Fig. 13 shows the impact of the virtual view position on the video perceptual quality. We consider the first frame of video sequence *kendo.yuv*. Video power is equally distributed between the texture and depth data ($\beta = 0.5$). We change the SNR value from 0dB to 20dB. Both the PSNR performance and the SSIM performance are presented. As can be seen, with the increase of SNR, both PSNR and SSIM increase. When SNR is 0dB and 10dB, the view performance at the mid-point (virtual view 4) tends to have a higher perceptual quality than the view quality at other positions. This is due to the equal power allocation between adjacent views. Hence, if a user is more interested in the virtual view that is closer to view 3 (e.g., virtual view at position 3.2), we may want to allocated more power to the view at position 3. When SNR is 20dB, there is generally no big difference between the video quality data at different positions.

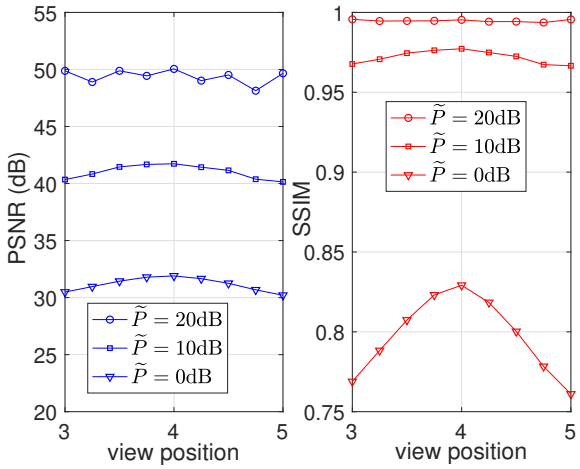

Fig. 13: Impact of virtual view positions.

3) Performance the Proposed Algorithm: In this subsection, we test the performance of our proposed DR algorithm on the quality of synthesized virtual views. Each simulation is performed 10 times and we present the averaged results.

Fig. 14 presents the synthesis quality at view point 2 for different algorithms, along with the synthesized frames. Under a channel constraint 1536, the conventional SoftCast scheme utilizes all the channel resources and allocate power on each chunk. In contrast, our proposed DR algorithm discard LP chunks and only retain the HP chunks. The channel usage is greatly saved and the PSNR performance is improved. For example, when power usage is 10dB, for *kendo.yuv*, both the PSNR and SSIM slightly increases and the channel usage is only $163/1536 = 10.6\%$ of the conventional SoftCast scheme. The proposed algorithm achieves an even better performance when SNR is low. Considering the huge data traffic in MVV, the saved wireless resources would be considerable. Finally, we have to mention that although the proposed algorithm is designed based on the PSNR metric, it still works well in terms of the SSIM metric. For video sequence *balloons.yuv*, the proposed algorithm improves the SSIM from 0.84184 to 0.91023 when SNR is 0dB. We also perform similar experiments on the RD algorithms, where similar observations are made. For space limitation, we omit the RD results here.

VII. Conclusions

In this paper, we integrated applying soft video delivery for MVV transmissions. Compared with the conventional digital based solutions, the proposed scheme improves video quality gracefully and the cliff effect can be avoided. Furthermore, complex bit allocation and rate control in digital systems can be replaced by a simple power allocation scheme. To handle the heavy data traffic caused by MVV, we proposed a resource control algorithm. The proposed DR optimization algorithm achieves the best viewing quality under a resource constraint. The proposed RD optimization algorithm minimizes the resource usage for a target video quality requirement. Hence the viewer can enjoy a stable viewing quality, which is favorable for MVV video streaming. With the proposed scheme, we also investigated the impact of the power control across the texture and depth data and view positions on synthesized virtual view quality. Simulation results demonstrated that the proposed algorithm worked well not only for referenced views but also for synthesized virtual views. Despite all these merits brought by soft video delivery, we find that soft video transmission may not be resilient to packet loss and is prone to channel variations. An extension toward hybrid digital and analog video coding would be a promising direction to fully exploit the benefit of both the digital and analog video coding.

References

[1] S. Mao, *Video Streaming over Cognitive Radio Networks: When Compression Meets the Spectrum*, 1st ed. New York, NY: Springer Science+Business Media, Jan. 2014.

[2] Y. Xu and S. Mao, "A survey of mobile cloud computing for rich media applications," *IEEE Wireless Commun.*, vol. 20, no. 3, pp. 46–53, June 2013.

[3] M. Amjad, M. Rehmani, and S. Mao, "Wireless multimedia cognitive radio networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 1056–1103, Second Quarter 2018.

[4] Y. Chen, M. M. Hannuksela, T. Suzuki, and S. Hattori, "Overview of the MVC+D 3D video coding standard," *J. Vis. Commun. Image Represent.*, vol. 25, no. 4, pp. 679–688, Apr. 2013.

[5] G. Tech, Y. Chen, K. Müller, J.-R. Ohm, A. Vetro, and Y.-K. Wang, "Overview of the multiview and 3D extensions of high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 35–49, Jan. 2016.

[6] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," in *Proc. SPIE Conf. Stereoscopic Displays and Virtual Reality Systems XI*, CA, U.S.A., Jan. 2004, pp. 93–105.

[7] T. Zhang and S. Mao, "An overview of emerging video coding standards," *ACM GetMobile*, vol. 22, no. 4, pp. 13–20, Dec. 2018.

[8] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Multi-view video plus depth representation and coding," in *IEEE Int. Conf. on Image Process. (ICIP'07)*, San Antonio, TX, Sep. 2007.

[9] Z. Liu, G. Cheung, and Y. Ji, "Optimizing distributed source coding for interactive multiview video streaming over lossy networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1781–1794, Oct. 2013.

[10] L. Toni, G. Cheung, and P. Frossard, "In-network view synthesis for interactive multiview video systems," *IEEE Trans. Multimedia*, vol. 18, no. 5, pp. 852–864, May. 2016.

[11] Y. Chen, Y.-K. Wang, K. Ugur, M. M. Hannuksela, J. Lainema, and M. Gabbouj, "The emerging MVC standard for 3D video services," *EURASIP J. Adv. Signal Process.*, vol. 2009, p. 8, Jan. 2009.

[12] K. Müller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F. H. Rhee *et al.*, "3D high-efficiency video coding for multi-view video and depth data," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3366–3378, Sept. 2013.

[13] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding: Turbo-codes. 1," in *Proc. 1993 IEEE Int. Conf. on Communications*, vol. 2. IEEE, 1993, pp. 1064–1070.

[14] T. Kratochvíl and R. Štukavec, "DVB-T digital terrestrial television transmission over fading channels," *Radioengineering*, vol. 17, no. 3, pp. 96–102, Sep. 2008.

[15] S. Jakubczak and D. Katabi, "SoftCast: One-size-fits-all wireless video," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 4, pp. 449–450, Oct. 2011.

[16] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[17] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[18] J. Chakareski, "Adaptive multiview video streaming: Challenges and opportunities," *IEEE Commun. Mag.*, vol. 51, no. 5, pp. 94–100, May. 2013.

[19] A. De Abreu, P. Frossard, and F. Pereira, "Optimizing multiview video plus depth prediction structures for interactive multiview video streaming," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 3, pp. 487–500, Apr. 2015.

Fig. 14: Synthesize quality at virtual view point 2 for different algorithms.

[20] E. Ekmekcioglu, C. G. Gurler, A. Kondoz, and A. M. Tekalp, "Adaptive multiview video delivery using hybrid networking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 6, pp. 1313–1325, Jun. 2017.

[21] R. Suenaga, K. Suzuki, T. Tezuka, M. P. Tehrani, K. Takahashi, and T. Fujii, "A practical implementation of free viewpoint video system for soccer games," in *Proc. SPIE Three-Dimensional Image Process., Meas., Appl.*, vol. 9393. International Society for Optics and Photonics, May. 2015, p. 93930G.

[22] Z. Chen, X. Zhang, Y. Xu, J. Xiong, Y. Zhu, and X. Wang, "MuVi: Multiview video aware transmission over MIMO wireless systems," *IEEE Trans. Multimedia*, vol. 19, no. 12, pp. 2788–2803, Jun. 2017.

[23] Z. Liu, G. Cheung, J. Chakareski, and Y. Ji, "Multiple description coding and recovery of free viewpoint video for wireless multi-path streaming," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 1, pp. 151–164, Feb. 2015.

[24] S. Jakubczak and D. Katabi, "SoftCast: One-size-fits-all wireless video," *ACM SIGCOMM Comput. Commu. Rev.*, vol. 40, no. 4, pp. 449–450, Oct. 2010.

[25] ——, "A cross-layer design for scalable mobile video," in *Proc. ACM MobiCom'11*, Las Vegas, NV, Sept. 2011, pp. 289–300.

[26] S. Aditya and S. Katti, "FlexCast: Graceful wireless video streaming," in *Proc. 17th Annu. Int. Conf. Mobile Comput. Netw*, Sept., Las Vegas, Nevada 2011, pp. 277–288.

[27] Y. Li, Z. Li, Y. Liu, and Y. Wang, "SCAST: Wireless video multicast scheme based on segmentation and SoftCast," in *Proc. IEEE WCNC'17*, San Francisco, CA, Mar. 2017, pp. 1–6.

[28] B. Tan, J. Wu, Y. Li, H. Cui, W. Yu, and C. W. Chen, "Analog coded SoftCast: A network slice design for multimedia broadcast/multicast," *IEEE Trans. Multimedia*, vol. 19, no. 10, pp. 2293–2306, Oct. 2017.

[29] H. Cui, C. Luo, C. W. Chen, and F. Wu, "Robust uncoded video transmission over wireless fast fading channel," in *Proc. IEEE INFOCOM'14*, Toronto, Canada, Apr./May 2014, pp. 73–81.

[30] X. L. Liu, W. Hu, C. Luo, Q. Pu, F. Wu, and Y. Zhang, "ParCast+: Parallel video unicast in MIMO-OFDM WLANs," *IEEE Trans. Multimedia*, vol. 16, no. 7, pp. 2038–2051, Nov. 2014.

[31] H. Cui, C. Luo, C. W. Chen, and F. Wu, "Scalable video multicast for MU-MIMO systems with antenna heterogeneity," *IEEE Trans. Circ. Syst. Video Technol.*, vol. 26, no. 5, pp. 992–1003, May. 2016.

[32] D. Liu, J. Wu, H. Cui, D. Zhang, C. Luo, and F. Wu, "Cost-distortion optimization and resource control in pseudo-analog visual communications," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3097–3110, Nov. 2018.

[33] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "FreeCast: Graceful free-viewpoint video delivery," *IEEE Trans. Multimedia*, Apr. 2018.
[34] T. Zhang and S. Mao, "Metadata-reduction for soft video delivery," *IEEE Networking Letters*, vol. 1, no. 2, pp. 84–88, June 2019.
[35] W.-S. Kim, A. Ortega, P. Lai, and D. Tian, "Depth map coding optimization using rendered view distortion for 3D video coding," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3534–3545, Nov. 2015.
[36] "Mobile 3DTV content delivery optimization over DVB-H system." [Online]. Available: http://sp.cs.tut.fi/mobile3dtv/stereo-video/
[37] "Xiph.org video test media [derf's collection]." [Online]. Available: https://media.xiph.org/video/derf/
[38] Fraunhofer Heinrich Hertz Institute, "3D high efficiency video coding (3D-HEVC) — JCT-VC," https://hevc.hhi.fraunhofer.de/3dhevc.

Ticao Zhang received the B.E. degree in 2014 and the M.S. degree in 2017 from School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China. He is currently pursuing a Ph.D. degree in Electrical and Computer Engineering at Auburn university. His research interests include video coding and communications, machine learning, and optimization and design of wireless multimedia networks.

Shiwen Mao [S'99-M'04-SM'09-F'19] received his Ph.D. in electrical and computer engineering from Polytechnic University, Brooklyn, NY (now New York University Tandon School of Engineering). He joined Auburn University, Auburn, AL in 2006 as an Assistant Profess in Electrical and Computer Engineering and held the McWane Associate Professorship from 2012 to 2015. Currently, he is the Samuel Ginn Distinguished Professor and Director of the Wireless Engineering Research and Education Center at Auburn University.

His research interests include wireless networks, multimedia communications, and smart grid. He is a Distinguished Speaker (2018-2021) and was a Distinguished Lecturer (2014-2018) of the IEEE Vehicular Technology Society. He was the chair of IEEE ComSoc Multimedia Communications Technical Committee (MMTC) for 2016-2018. He is an Area Editor of IEEE Open Journal of the Communications Society and IEEE Internet of Things Journal, and an Associate Editor of IEEE Transactions on Network Science and Engineering, IEEE Transactions on Mobile Computing, IEEE Transactions on Multimedia, IEEE Multimedia, IEEE Networking Letters, and ACM GetMobile, among others.

He received the IEEE ComSoc TC-CSR Distinguished Technical Achievement Award in 2019, the IEEE ComSoc MMTC Distinguished Service Award in 2019, Auburn University Creative Research & Scholarship Award in 2018, the 2017 IEEE ComSoc ITC Outstanding Service Award, the 2015 IEEE ComSoc TC-CSR Distinguished Service Award, the 2013 IEEE ComSoc MMTC Outstanding Leadership Award, and NSF CAREER Award in 2010. He is a co-recipient of the IEEE ComSoc MMTC Best Journal Paper Award in 2019, the IEEE ComSoc MMTC Best Conference Paper Award in 2018, the Best Demo Award from IEEE SECON 2017, the Best Paper Awards from IEEE GLOBECOM 2016 & 2015, IEEE WCNC 2015, and IEEE ICC 2013, and the 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems. He is a Fellow of the IEEE.