MCast: High-Quality Linear Video Transmission With Time and Frequency Diversities

Chaofan He, Huiying Wang, Yang Hu, Yan Chen, *Senior Member, IEEE*, Xiaopeng Fan, *Senior Member, IEEE*, Houqiang Li, and Bing Zeng, *Fellow, IEEE*

***Abstract*— Uncoded linear video transmission has recently attracted people's attention due to its capacity to provide robust and scalable transmission. However, in reality, with the fluctuation of the wireless channels, the received quality may not be good enough. In such a case, the data may need to be transmitted multiple times to exploit both the time and frequency diversities to improve the received quality. Such a problem has never been investigated in the literature of uncoded video transmission. To resolve the problem, in this paper, we propose a framework, named MCast, to utilize the time and frequency diversities to achieve high-quality linear video transmission. We study how to optimally allocate the power and assign the channels at each time slot to the source data such that the overall performance is maximized. Specifically, we first derive a closed-form optimal power allocation solution for any given channel assignment. With the optimal power allocation, we then propose a suboptimal channel assignment scheme, where we sort the channels with their gains and assign the channels one-by-one to the corresponding block that can reduce the most reconstruction error. Finally, we compare the MCast system with four other systems that are based on Softcast and Parcast, and simulation results show that the MCast system can achieve better performance in terms of both the PSNR performance and visual quality.**

***Index Terms*— Uncoded linear video transmission, power allocation, channel assignment, multicast.**

I. Introduction

With the rapid development of mobile communication, wireless video is gradually to be the new generation of application for wireless networks. It is estimated that the volume of wireless videos will increase manyfold. How people can more easily enjoy high definition wireless video becomes an important issue, and as a result, substantial efforts have been made to improve the wireless video quality [1]–[4].

Traditionally, as a general framework, videos are compressed into a bit stream and transmitted over channels which go through quantization, entropy coding, channel coding and modulation, etc. [5], [6]. However, with the prevalence of wireless networks and mobile devices, the weakness of traditional framework becomes obvious. The main weakness is the cliff effect due to the mismatches between the estimated channel condition and the actual channel condition. If the actual channel condition is worse than the estimated channel condition, the packet error rate will be severe due to which the video quality will degrade significantly. On the other hand, if the actual channel condition is better than the estimated channel condition, the decoded video quality cannot be improved. For the multi-user scenario where different users encounter different channel conditions, the traditional framework will choose the bit rate conservatively to satisfy the user under the worst channel condition. In such a case, those users under good channel conditions cannot enjoy better video quality.

Furthermore, wireless medium unavoidably suffers from errors due to both interference and channel noise. The traditional framework utilizes the forward error correction (FEC) [7]–[9] to correct errors without retransmission. Besides, since different packets contribute differently in video reconstruction, it is beneficial to classify them based on their importance and provide different levels of protection through unequal error protection (UEP) [10]–[14]. Nevertheless, in spite of these efforts, errors can still happen, and the whole packet will be discarded once one single bit error happens after channel decoding.

To tackle the above challenges, cross-layer joint source and channel coding approaches have been proposed to improve the video quality received by different users in wireless networks. A pseudo-analog scalable video delivery scheme, known as SoftCast [15], skips quantization, entropy coding and channel coding, simply uses 3D-DCT to de-correlate the video source and transmits the uncoded DCT coefficients directly. Since only linear operations are used in the entire process, the received video quality varies with the channel quality smoothly. Therefore, there is no cliff effect and the scalability is improved. Also, with Softcast, the packet with errors will not be discarded. Instead, it is optimally reconstructed by the linear minimal mean square error estimator [16].

However, the reconstructed video quality by Softcast may not be good enough especially when the signal-to-noise-ratio (SNR) is low, i.e., the channel condition is not good

Manuscript received June 2, 2017; revised December 20, 2017 and March 11, 2018; accepted March 16, 2018. Date of publication March 21, 2018; date of current version April 20, 2018. This work was partly supported by the National Key Basic Research Program of China under Grant 2015CB351804, the National Natural Science Foundation of China under Grant 61672137 and 61602090, the 111 Project No. B17008. The work of Y. Chen was supported by the Thousand Youth Talents Program of China. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gene Cheung. *(Corresponding author: Yang Hu.)*

C. He, H. Wang, Y. Hu, Y. Chen, and B. Zeng are with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China, and also with the Center of Future Media, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: eecfhe@std.uestc.edu.cn; eehuiyingw@std.uestc.edu.cn; yanghu@uestc.edu.cn; eecyan@uestc.edu.cn).

X. Fan is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: fxp@hit.edu.cn).

H. Li is with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230000, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2018.2818019

enough. This is mainly due to the following two reasons. Firstly, Softcast fails to fully exploit the redundancy of the source, i.e., the 3D-DCT cannot completely de-correlate the source. Secondly, the channel information is not utilized in Softcast. To improve the reconstructed video quality, many efforts have been made in the literature. Fan *et al.* [17], [18] proposed a Dcast framework to exploit the inter-frame redundancy with the distributed video coding theory [19], [20]. The coset codes are transmitted in Dcast to reduce the transmission power cost, and the predicted frame is utilized as side information of the distributed video coding at the receiver to improve the reconstructed video quality. The Layercast in [21] further exploited the cross-layer redundancy through coset coding under the condition of excessive bandwidth. To satisfy users with different channel bandwidth, Layercast processed the DCT coefficients into multiple layers by coset coding. The users with narrow bandwidth are offered only the base layer while the users with wider bandwidth are offered both the base layer and the enhanced layers. In other words, when there are adequate channel bandwidth, instead of retransmitting packets as Softcast, Layercast transmits multiple layers which provides more information. The Linecast in [22] extended Softcast for broadcasting the satellite images in real time. Specifically, Linecast progressively encoded and transmitted the satellite image line by line. At the receiver, each line is reconstructed by the lines which have been restored, due to which the Linecast forms a low-delay flow line operation mode that cannot be realized by Softcast.

Efforts have also been made to exploit the channel information to improve the performance of Softcast. The Parcast in [23] considered the transmission over the multiple-input multiple-output (MIMO) wireless channels and proposed to transmit high-energy DCT components in high-gain channels. Specifically, the wireless channels are decomposed into subchannels through orthogonal frequency division multiplexing (OFDM) technique. Then, the authors matched DCT components to subchannels based on the respective sorted order of the energy levels and performed joint source-channel power allocation to optimize the total error performance. Later in [24], the authors proposed ParCast+ to further improve the performance through utilizing motion compensated temporal filtering (MCTF) to better de-correlate the source signals. It was shown that the MCTF together with 2D-DCT can better utilize the temporal redundancy than 3D-DCT. Hybrid digital and analog transmission approaches have also been proposed. In [25], Sharpcast divides a video sequence into structure part and content part. The structure part and the high-energy DCT blocks of the content part are transmitted in digital since they are more sensitive to visual attention. The rest of the content part is transmitted in analog as Softcast to save the energy consumption.

While the aforementioned approaches have achieved promising performance, most of them assume that the transmission is real-time, i.e., the transmitted data rate perfectly matches the bandwidth. However, in reality, due to the fluctuation of the wireless channels, the received quality may not be good enough. In such a case, we may need to transmit the data multiple times to exploit both the time and frequency diversities to improve the received quality. In other words, the source data will be transmitted across multiple time slots and multiple channels. By exploiting the time and frequency diversities, it is expected that better performance can be achieved. However, it is unclear how much gain can be achieved by optimally exploiting such diversities. The key problem here is how to optimally allocate the power and assign the channels at each time slot to the source data such that the overall performance is maximized. To resolve the problem, in this paper, we propose a framework, named MCast, to utilize the time and frequency diversities to achieve high-quality linear video transmission. Specifically, we first derive the end-to-end distortion measure, i.e., the mean square error between the original video at the transmitter and the reconstructed video at the receiver, as a function of the power allocation and channel assignment. Then, we formulate the optimal power allocation and channel assignment as an optimization, which is shown to be non-convex due to the coupling between the power allocation and channel assignment. We first derive a closed-form optimal power allocation solution for any given channel assignment. With the optimal power allocation, we then propose a suboptimal channel assignment scheme, where we sort the channels with their gains and assign the channels one-by-one to the corresponding block that can reduce the most reconstruction error. Finally, we compare MCast system with four other systems that are based on Softcast and Parcast, and simulation results show that MCast system can achieve better performance in terms of both PSNR performance and visual quality.

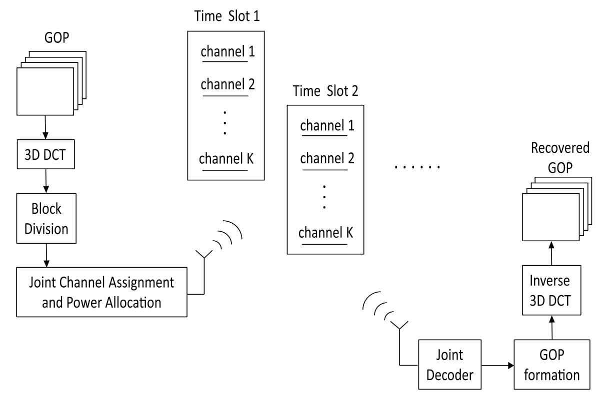


Fig. 1. System model of the MCast system.

The rest of this paper is organized as follows. The system model and problem formulation is described in detail in Section II. The joint power allocation and channel assignment for the proposed MCast system is discussed in Section III. The simulation results are shown in Section IV while the conclusions are drawn in Section V.

II. System Model

As shown in Fig. 1, in this paper, we consider a multiple time slots joint source and channel coding system, which we call "MCast" system. In the MCast system, a group of pictures (GOP) is first converted into frequency domain through 3D DCT transform and divided into blocks. Then, the blocks will be transmitted through multiple time slots and multiple channels to the receiver by exploiting the time and

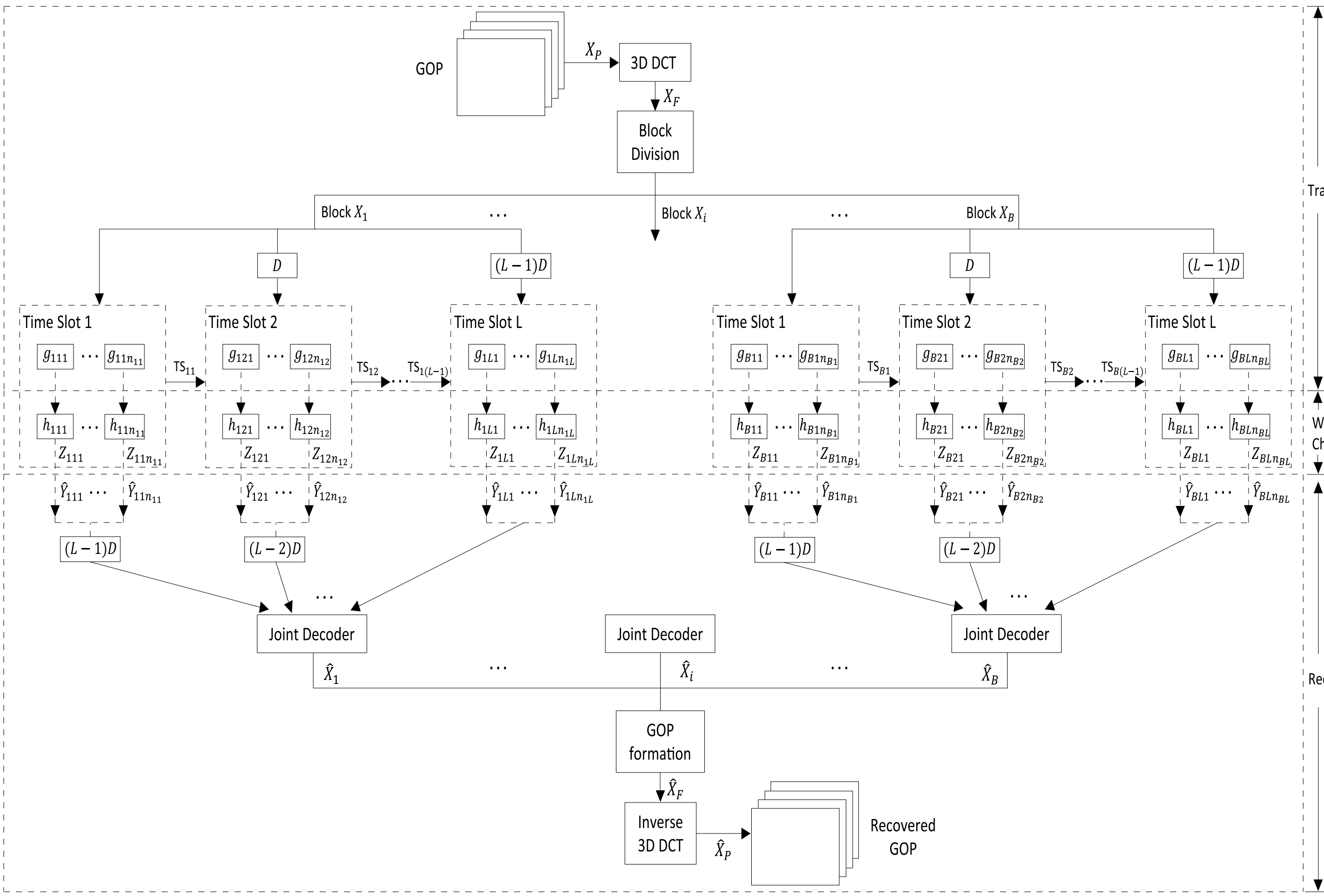


Fig. 2. Block diagram of the transmitter and receiver in the MCast system.

frequency diversity. The key problem here is how to assign channels and allocate power among different blocks, i.e., how to optimize the joint channel assignment and power allocation component in Fig. 1, to achieve the best receiving quality.

The detailed block diagram of the transmitter and receiver in the MCast system is shown in Fig. 2. Here, we assume that the GOP data is transmitted across L time slots, and in each time slot there are K channels. Let X_P denote the GOP data to be transmitted to the receiver. With the 3D DCT transform, X_P is converted to the frequency representation, denoted as X_F. Then, these DCT coefficients are divided into blocks with equal size, which are represented by $\{X_1, X_2, \ldots, X_B\}$. These blocks will be re-scaled and transmitted through different channels across different time slots. Specifically, let us assume that there are n_{ij} channels allocated to block X_i at the j^{th} time slot, with channel gains denoted as $H_i^j = [h_{ij1}, h_{ij2}, \ldots, h_{ijn_{ij}}]^T$. The corresponding power weights are denoted as $G_i^j = [g_{ij1}, g_{ij2}, \ldots, g_{ijn_{ij}}]^T$, where g_{ijk} is the re-scale power weight of $(ijk)^{th}$ channel.

At the receiver, for each block, multiple distorted copies will be received. For example, the received copies for i^{th} block are

$$Y_{ijk} = h_{ijk} g_{ijk} X_i + Z_{ijk}, \forall k = 1, \ldots, n_{ij}, \forall j = 1, \ldots, L, \quad (1)$$

where $X_i \in R^{N\times 1}$ is a vector of N DCT coefficients in the i^{th} block, the element of which is assumed to follow i.i.d. distribution with variance $\sigma_{x_i}^2$; $Z_{ijk} \in C^{N\times 1}$ is a vector of the Additive White Gaussian Noise (AWGN) with zero mean and variance σ_z^2. Here, without loss of generality, we assume the channel noise is i.i.d. for all channels and the channels are differentiated by the channel gains.

Let $n_i = \sum_{j=1}^{L} n_{ij}$, $H_i = [(H_i^1)^T, \ldots, (H_i^L)^T]^T \in C^{n_i L\times 1}$, $G_i = [(G_i^1)^T, \ldots, (G_i^L)^T]^T \in C^{n_i L\times 1}$, $Z_i = [Z_{i11}, \ldots, Z_{i1n_{i1}}, \ldots, Z_{iL1}, \ldots, Z_{iLn_{iL}}]^T \in C^{n_i L\times N}$ and $Y_i = [Y_{i11}, \ldots, Y_{i1n_{i1}}, \ldots, Y_{iL1}, \ldots, Y_{iLn_{iL}}]^T \in C^{n_i L\times N}$. Then, (1) can be re-written in a matrix form as

$$Y_i = (H_i \cdot G_i) X_i^T + Z_i, \quad (2)$$

where $\cdot$ denotes the Hadamard (i.e., entry-by-entry) product.

According to (2), we can see that the observations are linear in X_i with AWGN. Therefore, at the receiver, an optimal linear minimal mean square error (LMMSE) estimator can be performed to estimate X_i as follows [26]

$$\hat{X}_i^T = \frac{\sigma_{x_i}^2}{\sigma_{x_i}^2 \parallel H_i \cdot G_i \parallel^2 + \sigma_z^2} (H_i \cdot G_i)^T Y_i. \quad (3)$$

The corresponding expected estimation error can also be derived as follows

$$\epsilon(\hat{X}_i, X_i) = E(\hat{X}_i - X_i)^2 = \frac{\sigma_{x_i}^2 \sigma_z^2}{\sigma_{x_i}^2 \parallel H_i \cdot G_i \parallel^2 + \sigma_z^2}. \quad (4)$$

After obtaining the estimation of each block, we rearrange the estimated blocks back to the 3D DCT domain, represented as $\hat{X}_F$. Finally, we obtain the reconstruction of the original GOP $\hat{X}_P$ by taking the inverse 3D DCT of $\hat{X}_F$.

III. Joint Channel Assignment and Power Allocation for the Proposed MCast System

In this section, we will discuss how to optimize the channel assignment and power allocation component in the proposed MCast system. Specifically, our objective is to minimize the end-to-end distortion, i.e., the mean square error (MSE) between the original video X_P and the reconstructed video $\hat{X}_P$. Since X_F and $\hat{X}_F$ are the 3D DCT versions of X_P and $\hat{X}_P$, respectively, and the 3D DCT is an orthogonal linear operator, the MSE between X_F and $\hat{X}_F$ is equal to the MSE between X_P and $\hat{X}_P$.

While the blocks can be transmitted through multiple time slots to exploit the time-frequency diversities, the actual transmission is slot-by-slot since the channels may vary slot-by-slot. In such a case, the power allocation and channel assignment should be performed slot-by-slot. For each time slot, there is a total power constraint for all channels at that time slot as follows

$$\sum_{i=1}^{B}\sum_{k=1}^{n_{ij}} g_{ijk}^2\sigma_{x_i}^2 \leq P_t, \quad \forall j = 1,\ldots,L, \tag{5}$$

where P_t is the total power constraint at each time slot.

Moreover, since there are K channels at each time slot, the channel assigned to all blocks at each time slot should not exceed K, i.e., there is a channel constraint as follows

$$\sum_{i=1}^{B} n_{ij} \leq K, \quad \forall j = 1,\ldots,L. \tag{6}$$

With power constraint in (5) and the channel constraint in (6), the joint channel assignment and power allocation problem can be formally formulated as follows

$$\begin{aligned} \min_{H,G} \quad & \sum_{i=1}^{B}\frac{\sigma_{x_i}^2\sigma_z^2}{\sigma_{x_i}^2 \parallel H_i\cdot G_i \parallel^2 + \sigma_z^2}, \\ \text{subject to} \quad & \sum_{i=1}^{B} n_{ij} \leq K, \quad \forall j = 1,\ldots,L, \\ & \sum_{i=1}^{B}\sum_{k=1}^{n_{ij}} g_{ijk}^2\sigma_{x_i}^2 \leq P_t, \quad \forall j = 1,\ldots,L, \end{aligned} \tag{7}$$

where $H = [(H_1)^T,\ldots,(H_B)^T]^T \in C^{n_i LB\times 1}$ and $G = [(G_1)^T,\ldots,(G_B)^T]^T \in C^{n_i LB\times 1}$ are the possible channel assignment and power allocation, respectively.

The above equation (7) tries to optimize the channel assignment and power allocation across all time slots. However, in practice, the channel information is obtained slot-by-slot, and thus it is impossible for us to obtain the optimal channel assignment and power allocation across multiple time slots at one time. Instead, the problem needs to be solved slot-by-slot, and while solving the j^{th} time slot, all the channel assignment and power allocation at previous time slots are known. Therefore, the joint channel assignment and power allocation at the j^{th} time slot can be written as

$$\begin{aligned} \min_{H^j,G^j} \quad & \sum_{i=1}^{B}\frac{\sigma_{x_i}^2\sigma_z^2}{\sigma_{x_i}^2 \parallel H_i^j\cdot G_i^j \parallel^2 + \sigma_z^2}, \\ \text{subject to} \quad & \sum_{i=1}^{B} n_{ij} \leq K, \\ & \sum_{i=1}^{B}\sum_{k=1}^{n_{ij}} g_{ijk}^2\sigma_{x_i}^2 \leq P_t, \end{aligned} \tag{8}$$

where $H^j = [(H_1^j)^T,\ldots,(H_B^j)^T]^T \in C^{n_i B\times 1}$ and $G^j = [(G_1^j)^T,\ldots,(G_B^j)^T]^T \in C^{n_i B\times 1}$.

The optimization problem in (8) is obviously not convex due to the coupling between H^j and G^j, i.e., the Hadamard product between H^j and G^j. In the following subsections, we first discuss the power allocation problem by treating H^j as parameters. Then, we discuss how to perform the channel assignment.

A. Optimal Power Allocation

When given channel assignment, we can expand $||H_i^j\cdot G_i^j||^2$ as follows

$$\begin{aligned} \parallel H_i^j\cdot G_i^j \parallel^2 &= \sum_{m=1}^{j-1}\sum_{k=1}^{n_{im}} |h_{imk}|^2 g_{imk}^2 + \sum_{k=1}^{n_{ij}} |h_{ijk}|^2 g_{ijk}^2, \\ &= b_{ij} + \sum_{k=1}^{n_{ij}} |h_{ijk}|^2 g_{ijk}^2, \end{aligned} \tag{9}$$

where $b_{ij} = \sum_{m=1}^{j-1}\sum_{k=1}^{n_{im}} |h_{imk}|^2 g_{imk}^2$, if $j > 1$, and $b_{ij} = 0$ if $j = 1$, which is fixed at time slot j.

Let $u_{ijk} = g_{ijk}^2$, then by substituting (9) into (8), the optimization problem in (8) can be re-written as follows

$$\begin{aligned} \min_{U_j} J(U_j) &= \sum_{i=1}^{B}\frac{\sigma_z^2}{\sum_{k=1}^{n_{ij}} |h_{ijk}|^2 u_{ijk} + b_{ij} + \sigma_z^2/\sigma_{x_i}^2}, \\ \text{subject to} & \sum_{i=1}^{B}\sum_{k=1}^{n_{ij}} u_{ijk}\sigma_{x_i}^2 \leq P_t, \quad u_{ijk} \geq 0, \end{aligned} \tag{10}$$

where $U_j = \{u_{1j1},\ldots,u_{1jn_{1j}},\ldots,u_{Bj1},\ldots,u_{Bjn_{Bj}}\}$.

Since the objective function in (10) is monotonically decreasing in term of u_{ijk}, the equality in the constraint should be hold. Therefore, the power allocation optimization becomes

$$\begin{aligned} \min_{U_j} J(U_j) &= \sum_{i=1}^{B}\frac{\sigma_z^2}{\sum_{k=1}^{n_{ij}} |h_{ijk}|^2 u_{ijk} + b_{ij} + \sigma_z^2/\sigma_{x_i}^2}, \\ \text{subject to} & \sum_{i=1}^{B}\sum_{k=1}^{n_{ij}} u_{ijk}\sigma_{x_i}^2 = P_t, \quad u_{ijk} \geq 0. \end{aligned} \tag{11}$$

Theorem 1: The optimization problem on power allocation defined in (11) is convex.

Proof: The second order derivative of the objective function in (11) in the term u_{ijk} can be derived as

$$\frac{\partial^2 J(U_j)}{\partial u_{ijk}^2} = \frac{2\sigma_z^2|h_{ijk}|^4}{(\sum_{k=1}^{n_{ij}} |h_{ijk}|^2 u_{ijk} + b_{ij} + \sigma_z^2/\sigma_{x_i}^2)^3} > 0. \tag{12}$$

Therefore, the objective function in (11) is a strict convex function in term of u_{ijk}. Since the constraints are linear, the optimization problem in (11) is a convex optimization problem. ■

To solve the convex optimization (11), we first derive the following theorem.

Theorem 2: Under the optimal channel assignment, the number of channels assigned to each block cannot exceed one at each time slot.

Proof: Let us assume that under optimal channel assignment, $(ij1)^{th}$ and $(ij2)^{th}$ channels are assigned to the i^{th} block at j^{th} time slot, where the corresponding optimal power weights are u_{ij1} and u_{ij2}. Without loss of generality, we assume that their channel gains satisfy $|h_{ij1}|^2 > |h_{ij2}|^2$. Then, let us consider another channel assignment, where only $(ij1)^{th}$ channel is assigned to the i^{th} block at j^{th} time slot with the allocated power $u_{ij1}+u_{ij2}$. Obviously, both these two assignment schemes share the same constraint in (11), and we have

$$\frac{\sigma_{z_i}^2}{|h_{ij1}|^2 \sum_{k=1}^{2} u_{ijk} + b_{ij} + \sigma_z^2/\sigma_{x_i}^2} < \frac{\sigma_{z_i}^2}{\sum_{k=1}^{2} |h_{ijk}|^2 u_{ijk} + b_{ij} + \sigma_z^2/\sigma_{x_i}^2}. \quad (13)$$

From the above inequality, we can see that the new assignment scheme achieves better performance, which contradicts with the assumption. Therefore, under the optimal channel assignment, the number of channels assigned to each block cannot exceed one at each time slot. ■

Remark 1: Notice that similar result of Theorem 2 has been shown in [27], [28], and [29] although the system model is slightly different. Here, we provide the proof to Theorem 2 from a different perspective.

With Theorem 2, the power allocation optimization in (11) can be simplified as follows

$$\min_{U_j} J(U_j) = \sum_{i=1}^{B} \frac{\sigma_z^2}{|h_{ij}|^2 u_{ij} + b_{ij} + \sigma_z^2/\sigma_{x_i}^2},$$

$$\text{subject to } \sum_{i=1}^{B} u_{ij}\sigma_{x_i}^2 = P_t, \quad u_{ij} \geq 0. \quad (14)$$

By using the Lagrange multipliers, the optimal power allocation can be derived as follows

$$u_{ij}^* = \left(\frac{\sigma_z}{\sqrt{\lambda}\sigma_{x_i}|h_{ij}|} - \frac{b_{ij}\sigma_{x_i}^2 + \sigma_z^2}{\sigma_{x_i}^2 |h_{ij}|^2} \right)^+, \quad (15)$$

where $(x)^+$ is equal to x, if $x > 0$, and equal to 0 if $x \leq 0$, and λ is a parameter satisfies $\sum_{i=1}^{B} u_{ij}^* \sigma_{x_i}^2 = P_t$.

B. Channel Assignment

With the optimal power allocation in the previous subsection, we discuss in this subsection about the channel assignment. Finding the optimal channel assignment in (8) is computationally expensive as all the possible permutations of available channels at j^{th} time slot need to be evaluated. To resolve the problem, we propose in this subsection a suboptimal channel assignment scheme, where we first sort the channels with their powers and then assign the channels one-by-one to the corresponding block that can reduce the most MSE.

Without loss of generality, let us assume $|h_{1j}|^2 > |h_{2j}|^2 > \ldots > |h_{Kj}|^2$ and suppose that we are now assigning $(ij)^{th}$ channel. According to Theorem 2, each block will be assigned at most one channel under optimal channel assignment. Thus, we assume that $\forall k < i$, the $(kj)^{th}$ channel has been assigned to block x_k. Denote $S_i = \{x_1, \ldots, x_{i-1}\}$ be the block set containing those blocks which have been assigned channels, and $\bar{S}_i$, the complementary of S_i, be the block set containing the blocks which can be assigned to $(ij)^{th}$ channel. To choose a proper block for $(ij)^{th}$ channel, we define a measure γ_{il} for each block $l \in \bar{S}_i$ as follows

$$\gamma_{il} = \frac{\sigma_z^2 \sigma_{x_l}^2}{b_{lj}\sigma_{x_l}^2 + \sigma_z^2} - \frac{\left(\sum_{d=1}^{i-1} \frac{\sigma_z \sigma_{x_d}}{|h_{dj}|} + \frac{\sigma_z \sigma_{x_l}}{|h_{ij}|} \right)^2}{P_t + \sum_{d=1}^{i-1} \frac{b_{dj}\sigma_{x_d}^2 + \sigma_z^2}{|h_{dj}|^2} + \frac{b_{lj}\sigma_{x_l}^2 + \sigma_z^2}{|h_{ij}|^2}}, \quad (16)$$

and the block that maximizes γ_{il} will be chosen for $(ij)^{th}$ channel, i.e.,

$$x_i = \arg\max_{l \in \bar{S}_i} \gamma_{il}. \quad (17)$$

In the following, we discuss some properties of the channel assignment scheme in (17).

Lemma 1: Let x_i be the block selected for $(ij)^{th}$ channel by the channel assignment scheme in (17) and P_{ij} be the corresponding optimal power allocation. Then, for any other $x_n \in \bar{S}_i$, the following equation holds

$$\frac{\sigma_z^2 \sigma_{x_i}^2}{b_{ij}\sigma_{x_i}^2 + \sigma_z^2} - \frac{\sigma_z^2 \sigma_{x_i}^2}{|h_{ij}|^2 P_{ij} + b_{ij}\sigma_{x_i}^2 + \sigma_z^2} \geq \frac{\sigma_z^2 \sigma_{x_n}^2}{b_{nj}\sigma_{x_n}^2 + \sigma_z^2} - \frac{\sigma_z^2 \sigma_{x_n}^2}{|h_{ij}|^2 P_{ij} + b_{nj}\sigma_{x_n}^2 + \sigma_z^2}. \quad (18)$$

Proof: We first re-write (16) as follows

$$\gamma_{il} = \frac{\sigma_z^2 \sigma_{x_l}^2}{b_{lj}\sigma_{x_l}^2 + \sigma_z^2} - \left(\sum_{d=1}^{i-1} \frac{\sigma_z^2 \sigma_{x_d}^2}{|h_{dj}|^2 P_{dj} + b_{dj}\sigma_{x_d}^2 + \sigma_z^2} + \frac{\sigma_z^2 \sigma_{x_l}^2}{|h_{ij}|^2 P_{lj} + b_{lj}\sigma_{x_l}^2 + \sigma_z^2} \right) \quad (19)$$

where P_{dj} is power allocated to the d^{th} block at j^{th} time slot through the optimal power allocation algorithm discussed in the previous subsection.

Since $(ij)^{th}$ channel is assigned to block x_i, we have

$$\frac{\sigma_z^2 \sigma_{xi}^2}{b_{ij}\sigma_{x_i}^2 + \sigma_z^2} - \left(\sum_{d=1}^{i-1} \frac{\sigma_z^2 \sigma_{x_d}^2}{|h_{dj}|^2 P_{dj} + b_{dj}\sigma_{x_d}^2 + \sigma_z^2} + \frac{\sigma_z^2 \sigma_{x_i}^2}{|h_{ij}|^2 P_{ij} + b_{ij}\sigma_{x_i}^2 + \sigma_z^2} \right)$$

$$> \frac{\sigma_z^2\sigma_{x_n}^2}{b_{nj}\sigma_{x_n}^2+\sigma_z^2} - \left(\sum_{d=1}^{i-1}\frac{\sigma_z^2\sigma_{x_d}^2}{|h_{dj}|^2P'_{dj}+b_{dj}\sigma_{x_d}^2+\sigma_z^2} + \frac{\sigma_z^2\sigma_{x_n}^2}{|h_{ij}|^2P'_{nj}+b_{nj}\sigma_{x_n}^2+\sigma_z^2}\right) \quad (20)$$

where $\left\{P'_{dj}, P'_{nj}\right\}, d = 1, \ldots, (i-1)$ is the optimal power allocation if $(ij)^{th}$ channel is assigned to x_n, which can lead to the following

$$\frac{\sigma_z^2\sigma_{x_n}^2}{b_{nj}\sigma_{x_n}^2+\sigma_z^2} - \left(\sum_{d=1}^{i-1}\frac{\sigma_z^2\sigma_{x_d}^2}{|h_{dj}|^2P'_{dj}+b_{dj}\sigma_{x_d}^2+\sigma_z^2} + \frac{\sigma_z^2\sigma_{x_n}^2}{|h_{ij}|^2P'_{nj}+b_{nj}\sigma_{x_n}^2+\sigma_z^2}\right)$$
$$> \frac{\sigma_z^2\sigma_{x_n}^2}{b_{nj}\sigma_{x_n}^2+\sigma_z^2} - \left(\sum_{d=1}^{i-1}\frac{\sigma_z^2\sigma_{x_d}^2}{|h_{dj}|^2P_{dj}+b_{dj}\sigma_{x_d}^2+\sigma_z^2} + \frac{\sigma_z^2\sigma_{x_n}^2}{|h_{ij}|^2P_{ij}+b_{nj}\sigma_{x_n}^2+\sigma_z^2}\right) \quad (21)$$

By combining (20) and (21), we derive the following

$$\frac{\sigma_z^2\sigma_{x_i}^2}{b_{ij}\sigma_{x_i}^2+\sigma_z^2} - \frac{\sigma_z^2\sigma_{x_i}^2}{|h_{ij}|^2P_{ij}+b_{ij}\sigma_{x_i}^2+\sigma_z^2} > \frac{\sigma_z^2\sigma_{x_n}^2}{b_{nj}\sigma_{x_n}^2+\sigma_z^2} - \frac{\sigma_z^2\sigma_{x_n}^2}{|h_{ij}|^2P_{ij}+b_{nj}\sigma_{x_n}^2+\sigma_z^2} \quad (22)$$

■

Based on *Lemma 1*, we can show in the following theorem that if $(ij)^{th}$ channel is assigned to block x_i, the optimal power allocated to $(kj)^{th}$ channel is positive, for $\forall k < i$.

Theorem 3: If $(ij)^{th}$ channel is assigned to block x_i, the optimal power allocated to $(kj)^{th}$ channel is positive, for $\forall k < i$, i.e., $P_{kj} > 0, \forall k < i$.

Proof: Suppose $(kj)^{th}$ channel $(k < i)$ is allocated with zero power, i.e., $P_{kj} = 0$, and the corresponding total MSE can be derived as follows

$$\epsilon_1 = \sum_{d\neq k,i}\frac{\sigma_z^2\sigma_{x_d}^2}{|h_{dj}|^2P_{dj}+b_{dj}\sigma_{x_d}^2+\sigma_z^2} + \frac{\sigma_z^2\sigma_{x_k}^2}{b_{kj}\sigma_{x_k}^2+\sigma_z^2} + \frac{\sigma_z^2\sigma_{x_i}^2}{|h_{ij}|^2P_{ij}+b_{ij}\sigma_{x_i}^2+\sigma_z^2} \quad (23)$$

Then, we consider the case that the power allocated to $(kj)^{th}$ and $(ij)^{th}$ channels are $P'_{kj} > 0$ and $P'_{ij} > 0$, respectively, with $P'_{kj} + P'_{ij} = P_{ij}$. Then, we have

$$\epsilon_2 = \sum_{d\neq k,i}\frac{\sigma_z^2\sigma_{x_d}^2}{|h_{dj}|^2P_{dj}+b_{dj}\sigma_{x_d}^2+\sigma_z^2} + \frac{\sigma_z^2\sigma_{x_k}^2}{|h_{kj}|^2P'_{kj}+b_{kj}\sigma_{x_k}^2+\sigma_z^2} + \frac{\sigma_z^2\sigma_{x_i}^2}{|h_{ij}|^2P'_{ij}+b_{ij}\sigma_{x_i}^2+\sigma_z^2} \quad (24)$$

With (23) and (24), we calculate difference between ϵ_1 and ϵ_2 as follows

$$\epsilon_1-\epsilon_2 = \frac{\sigma_z^2\sigma_{x_k}^2}{b_{kj}\sigma_{x_k}^2+\sigma_z^2} - \frac{\sigma_z^2\sigma_{x_k}^2}{|h_{kj}|^2P'_{kj}+b_{kj}\sigma_{x_k}^2+\sigma_z^2} + \frac{\sigma_z^2\sigma_{x_i}^2}{|h_{ij}|^2P_{ij}+b_{ij}\sigma_{x_i}^2+\sigma_z^2} - \frac{\sigma_z^2\sigma_{x_i}^2}{|h_{ij}|^2P'_{ij}+b_{ij}\sigma_{x_i}^2+\sigma_z^2} \quad (25)$$

According to *Lemma 1* and the fact that $|h_{kj}| > |h_{ij}|$, we have

$$\frac{\sigma_z^2\sigma_{x_k}^2}{b_{kj}\sigma_{x_k}^2+\sigma_z^2} - \frac{\sigma_z^2\sigma_{x_k}^2}{|h_{kj}|^2P'_{kj}+b_{kj}\sigma_{x_k}^2+\sigma_z^2} > \frac{\sigma_z^2\sigma_{x_i}^2}{b_{ij}\sigma_{x_i}^2+\sigma_z^2} - \frac{\sigma_z^2\sigma_{x_i}^2}{|h_{ij}|^2P'_{kj}+b_{ij}\sigma_{x_i}^2+\sigma_z^2} \quad (26)$$

With (26), we can have a low bound for (25) as follows

$$\begin{aligned}\epsilon_1-\epsilon_2 >& \frac{\sigma_z^2\sigma_{x_i}^2}{b_{ij}\sigma_{x_i}^2+\sigma_z^2} - \frac{\sigma_z^2\sigma_{x_i}^2}{b_{ij}\sigma_{x_i}^2+\sigma_z^2+|h_{ij}|^2P'_{kj}} \\ &- \left(\frac{\sigma_z^2\sigma_{x_i}^2}{b_{ij}\sigma_{x_i}^2+\sigma_z^2+|h_{ij}|^2P'_{ij}} - \frac{\sigma_z^2\sigma_{x_i}^2}{b_{ij}\sigma_{x_i}^2+\sigma_z^2+|h_{ij}|^2P'_{ij}+|h_{ij}|^2P'_{kj}}\right) \\ &= (\sigma_z^2\sigma_{x_i}^2)[f(b_{ij}\sigma_{x_i}^2+\sigma_z^2) \\ &-f(b_{ij}\sigma_{x_i}^2+\sigma_z^2+|h_{ij}|^2P'_{ij})]\end{aligned} \quad (27)$$

where $f(x)$ is defined as

$$f(x) = \frac{1}{x} - \frac{1}{x+a} \quad (28)$$

with $a > 0$ being a positive constant and $x > 0$.

By taking the first-order derivative of $f(x)$ over x, we have

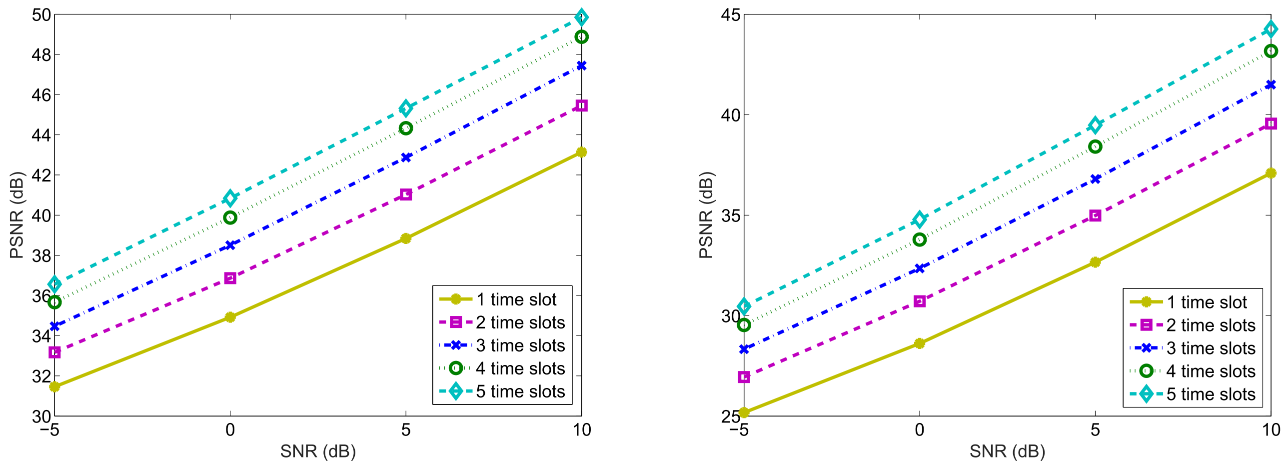
$$\frac{df(x)}{dx} = -\frac{1}{x^2} + \frac{1}{(x+a)^2} < 0. \quad (29)$$

Therefore, $f(x)$ is a decreasing function in term of x for $x > 0$. According to (27), we have $\epsilon_2 < \epsilon_1$, which contradicts with our assumption that the ϵ_1 is the minimal MSE. This completes the proof. ■

The channel assignment procedure can be terminated as long as we find a channel with zero power allocation, according to the following Theorem.

Theorem 4: If $(ij)^{th}$ channel is allocated with zero power through the optimal power allocation algorithm, i.e., $P_{ij} = 0$, then, $P_{kj} = 0, \forall k > i$.

Proof: Suppose that all the $(lj)^{th}$ channels $(l < i)$ are allocated with positive power through the optimal power allocation algorithm, i.e., $P_{lj} > 0, \forall l < i$. When assigning $(ij)^{th}$ channel, we evaluate all possible blocks, $x_m \in \bar{S}_i$. With (17), block x_i is chosen, and the $\sqrt{\lambda}$ in (15) can be obtained



(a) (b)

Fig. 3. The PSNR performance of the proposed MCast system: (a) *foreman* sequence; (b) *bus* sequence.

as

$$\sqrt{\lambda} = \frac{\sum_{d=1}^{i-1}\frac{\sigma_z\sigma_{x_d}}{|h_{dj}|} + \frac{\sigma_z\sigma_{x_i}}{|h_{ij}|}}{P_t + \sum_{d=1}^{i-1}\frac{b_{dj}\sigma_{x_d}^2+\sigma_z^2}{|h_{dj}|^2} + \frac{b_{ij}\sigma_{x_i}^2+\sigma_z^2}{|h_{ij}|^2}} = \frac{\alpha + \frac{r_i}{|h_{ij}|}}{\beta + \frac{s_i}{|h_{ij}|^2}} \tag{30}$$

where $\alpha = \sum_{d=1}^{i-1}\frac{\sigma_z\sigma_{x_d}}{|h_{dj}|} > 0$, $\beta = P_t + \sum_{d=1}^{i-1}\frac{b_{dj}\sigma_{x_d}^2+\sigma_z^2}{|h_{dj}|^2} > 0$, $r_i = \sigma_z\sigma_{x_i}$, and $s_i = b_{ij}\sigma_{x_i}^2 + \sigma_z^2$.

With the optimal power allocation, the power allocated to x_i is

$$P_{ij} = \left(\frac{\alpha s_i|h_{ij}| - \beta r_i}{\beta|h_{ij}|^2 + |h_{ij}|s_i}\right)^+ \tag{31}$$

The $P_{ij} = 0$ iff the following condition holds

$$|h_{ij}| \le \frac{\beta r_m}{\alpha s_m} \tag{32}$$

Since $|h_{kj}| \le |h_{ij}|, \forall k > i$, (32) holds for all $(kj)^{th}$ channels ($k > i$), which means that $P_{kj} = 0, \forall k > i$. This completes the proof. ■

C. Managing Metadata

The variance of each block is needed to be transmitted to the receiver. With such information, the receiver could calculate the power weights and assigned channels for each block slot-by-slot through the proposed joint channel assignment and power allocation algorithm. Then, the receiver could use the joint decoding to reconstruct the transmitted signal. In our simulations, each frame is divided into 12 blocks, thus the required metadata of the MCast system is no more than 0.002bbp (bits per pixel), which is similar to that of the Softcast system with the same block size.

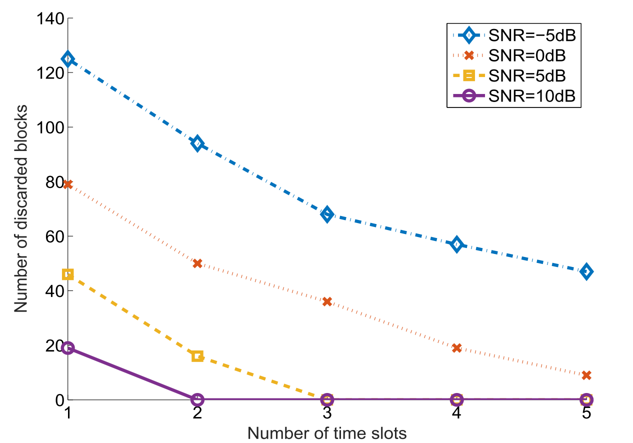


Fig. 4. The number of discarded blocks with different SNRs and time slots.

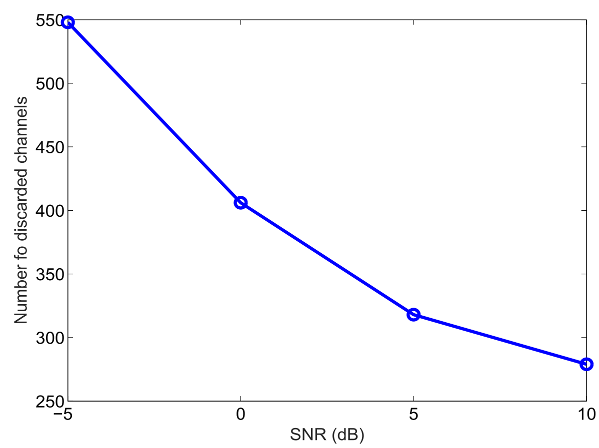


Fig. 5. The number of unused channels versus the SNR when the number of time slots is five.

IV. Simulation Results

To evaluate the performance of the proposed MCast system, we conduct multiple simulations under different settings in this section. We also compare with four other approaches

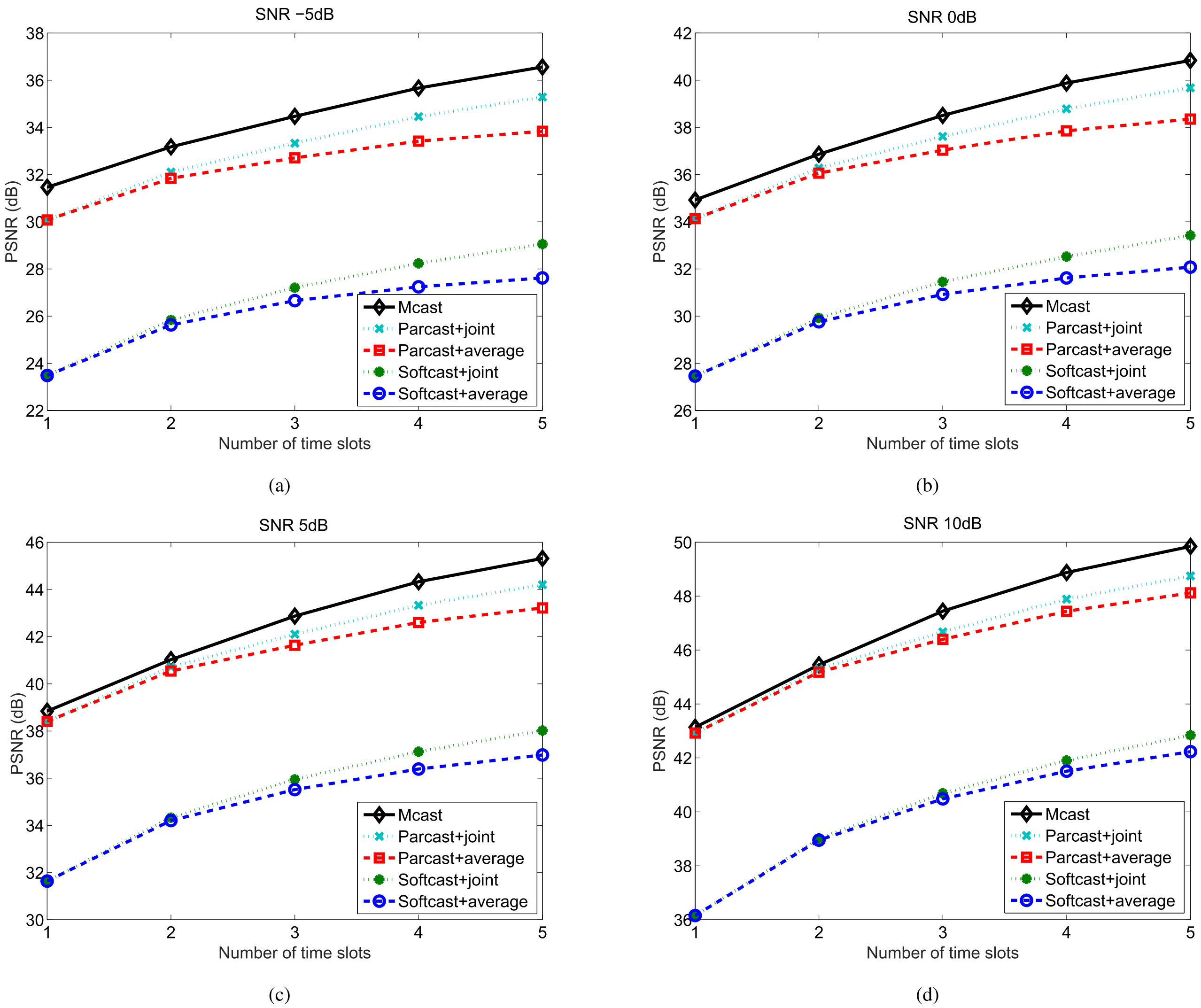


Fig. 6. The PSNR performance comparison among different approaches with different SNRs: (a) SNR=−5dB; (b) SNR=0dB; (c) SNR=5dB; (d) SNR=10dB.

to demonstrate the superiority of the proposed MCast system. These four approaches, Softcast+average, Softcast+joint, Parcast+average, and Parcast+joint, are direct extensions of Softcast [15] and Parcast [23], where "average" means that the receiver decodes the received signal of each time slot independently and average the results while "joint" means that the receiver jointly decodes the received signals of all time slots.

A. Simulation Settings

Without loss of generality, we only use the luminance component of videos chosen from Xiph [30]. Most of test vidoes are in standard CIF format (352×288 pixels, 60 fps), including $akiyo$, $bream$, bus, $coastguard$, $football$, $foreman$, $mobile$, $news$, $setfan$ and $templete$. Therefore, the source bandwidth is 6.08 MHz. The size of GOP is set to be 16 frames, which is processed as a whole through 3D-DCT. After 3D-DCT, the DCT coefficients are equally divided into 192 blocks, each of which contains 88×72 pixels. Similar to most of the traditional wireless communication systems [31], the Rayleigh fading channels are assumed in our MCast system. We further assume that there are 192 channels, however, some deep fading channels may not be used with our channel assignment algorithm. In addition, we also test some high definition videos (1920×1080 pixels, 60 fps), including $Kimono$, $BasketballDrive$, $BQTerrace$, $Cactus$ and $ParkScene$.

B. Performances of the Proposed MCast System

In this subsection, we conduct simulations to study the performance of the proposed MCast system. We assume current time slot is the j^{th} time slot and the channel assignment and power allocation in all previous $(j-1)$ time slots are known.

We first evaluate the quality of the reconstructed sequences versus the transmitted SNR when the video sequences are transmitted across different number of time slots, and the results are shown in Fig. 3. We can see that the PSNR increases as the channel SNR increases when the number of used time slots is fixed. This is reasonable since the reconstruction will be better with better channel condition. From Fig. 3, we can

Fig. 7. The visual quality comparison among different approaches with five time slots and SNR −5dB: (a) Original frame; (b) Reconstructed by Softcast+average; (c) Reconstructed by Softcast+joint; (d) Reconstructed by Parcast+average; (e) Reconstructed by Parcast+joint; (f) Reconstructed by the proposed MCast system.

also observe that with any fixed channel SNR, the PSNR increases as the number of time slots increases, but the increment becomes smaller when the number of time slots is larger. Such a phenomenon is because that with more time slots, we can exploit better time and frequency diversities to improve the reconstruction, but the benefit will saturate when there are too many diversities.

In Fig. 4, we illustrate the number of discarded blocks, i.e., the blocks with zero power allocation, versus the number of time slots with different SNRs. We can see that the number of discarded blocks decreases as the number of time slots increases and/or the SNR increases. When the SNR is 10dB, all blocks are transmitted if more than one time slot is utilized for transmission. However, when the SNR is −5dB, there are about 50 blocks discarded even if five time slots are utilized for transmission. That is because when the channel condition is not good, assigning good channels to a few important blocks with large power and discarding some less important blocks can lead to smaller total MSE. We also illustrate the number of unused channels versus the SNR when the number of time slots that are utilized for transmission is five, and the results are shown in Fig. 5. We can see that as the SNR increases, the number of unused channels reduces. This is reasonable since when the channel condition is better, more channels should be utilized for transmission to improve the reconstruction quality.

C. Comparison With Other Systems

In this subsection, we compare the performance of the MCast system with those of Parcast+joint, Parcast+average, Softcast+joint and Softcast+average. In the Parcast+joint and Parcast+average systems, the channel assignment and power allocation at each time slot are the same as those employed by Parcast. Specifically, at each time slot, both the channels and blocks are sorted in a descend order according to the channel powers and block energy, respectively. Then, each channel is assigned to the corresponding block with the same sorted order, i.e., higher gain channel is assigned to more important block. With such a channel assignment scheme, the power allocated to each channel can be written as below [23],

$$g_i = \sqrt{\frac{P_t}{\sigma_{x_i}|h_i|\sum_j(\sigma_{x_i}/|h_i|)}} \tag{33}$$

where g_i is the power weight allocated to the i^{th} block corresponding to the i^{th} largest channel whose gain is h_i, σ_{x_i} is the standard deviation of the i^{th} block's energy and P_t is total power at each time slot.

In Softcast+joint and Softcast+average systems, there is no channel assignment procedure and the power allocation at each time slot is the same as that in Softcast [15], where the allocated power only depends on the energy of each block as below

$$g_i = \sqrt{\frac{P_t}{\sigma_{x_i}\sum_j \sigma_{x_i}}} \tag{34}$$

where g_i is the power weight allocated to i^{th} block.

The PSNR performance comparisons versus the number of time slots that are utilized for transmission at different SNRs are shown in Fig. 6. We can see the Softcast+average performs worst, and Softcast+joint performs slightly better with joint decoding, especially when the number of time slots utilized increases. By utilizing the channel information in both the channel assignment and power allocation, Parcast+average and Parcast+joint can greatly improve the PSNR performance, and again the joint decoding can further improve the performance. By exploiting the time and frequency diversities across multiple time slots, the proposed MCast system can achieve the best PSNR performance. The performance improvement

Fig. 8. The visual quality comparison among different approaches with 1080p video 'Kimono' when the number of utilized time slots is five and SNR is −5dB: (a) Original frame; (b) Reconstructed by Softcast+average; (c) Reconstructed by Softcast+joint; (d) Reconstructed by Parcast+average; (e) Reconstructed by Parcast+joint; (f) Reconstructed by the proposed MCast system.

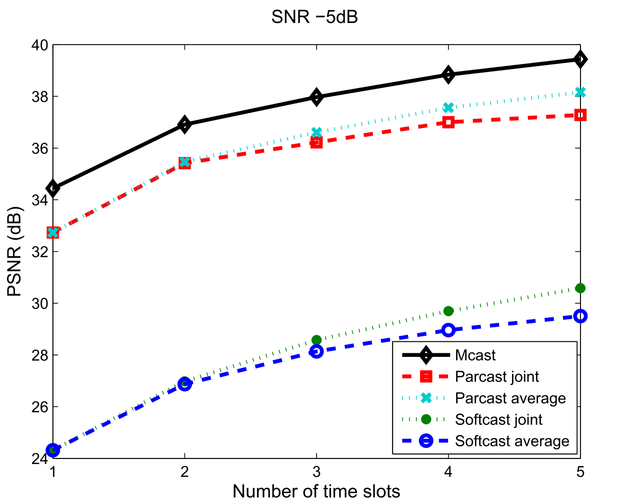


Fig. 9. The PSNR performance comparison among different approaches with 1080p video 'Kimono' and SNR −5dB;

is larger when the number of time slots utilized increases and/or the SNR decreases. This is because when the channel condition is not good, it is more important to select the right channel and allocate a proper power. Also, when the number of time slots utilized is large, there are more diversities to exploit and thus the performance gain is larger.

The visual quality of the reconstructed video with different approaches are also evaluated, and the results are shown in Fig. 7. Due to page limitation, we only show the results of $foreman$, bus and $stefan$ sequences while the number of time slots utilized is five and the SNR is $-5dB$. The results on other video sequences under different settings are similar. We can clearly see that the visual quality of the reconstruction generated by Softcast+avearge and Softcast+joint are unacceptable. This is mainly because the channel information becomes critical with low SNR. Without utilizing the channel information, the performance of Softcast+avearge and Softcast+joint is too bad. By utilized the channel information, the performance of Parcast+avearge and Parcast+joint improves. However, the background noise is still very obvious. With the proposed MCast system, we can see that the visual quality of the reconstruction is greatly improved by exploiting the time and frequency diversities.

We have also evaluated the MCast system on the high definition videos. In Fig. 9, we show the PSNR performance versus the number of utilized time slots on the 1080p video $Kimono$ with −5dB SNR. We can see that similar to the previous results, the MCast system achieves the best PSNR performance. The visual quality of the reconstructed video frame generated by different approaches are shown in Fig. 8 when five time slots are used and the SNR is −5dB. We can see that the MCast system achieves the best visual quality of reconstruction. Note that due to the page limitation, we only discuss the results on the 1080p $Kimono$ sequence with −5dB SNR. The results on all sequences with different SNRs (including the $Kimono$ sequence with different SNRs) are similar.

V. Conclusion

In this paper, we propose a MCast system to exploit the time and frequency diversities to achieve high-quality linear video transmission. The MCast system solves the problem of how to optimally allocate the power and assign the channels at each time slot to the source data to maximize the overall performance. Specifically, a closed-form optimal power allocation solution is derived when the channel assignment is given. With the optimal power allocation, a suboptimal channel assignment scheme is proposed, where the channels are first sorted according to their powers and then assigned one-by-one to the source data that can reduce the most distortion. Some properties of the channel assignment scheme are also discussed. Simulation results show that compared with existing schemes, the proposed MCast system can better exploit the time and frequency diversities to achieve better performance in terms of both PSNR performance and visual quality.

References

[1] Y. Chen, Y. Hu, O. C. Au, H. Li, and C. W. Chen, "Video error concealment using spatio-temporal boundary matching and partial differential equation," *IEEE Trans. Multimedia*, vol. 10, no. 1, pp. 2–15, Jan. 2008.

[2] Y. Chen, B. Wang, and K. J. R. Liu, "Multiuser rate allocation games for multimedia communications," *IEEE Trans. Multimedia*, vol. 11, no. 6, pp. 1170–1181, Oct. 2009.

[3] Y. Chen, Y. Wu, B. Wang, and K. J. R. Liu, "Spectrum auction games for multimedia streaming over cognitive radio networks," *IEEE Trans. Commun.*, vol. 58, no. 8, pp. 2381–2390, Aug. 2010.

[4] C. Y. Wang, Y. Chen, H. Y. Wei, and K. J. R. Liu, "Scalable video multicasting: A stochastic game approach with optimal pricing," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2353–2367, May 2015.

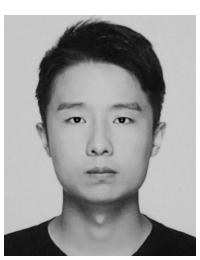
[5] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.

[6] A. Kolmogorov, "On the Shannon theory of information transmission in the case of continuous signals," *IRE Trans. Inf. Theory*, vol. 2, no. 4, pp. 102–108, Dec. 1956.

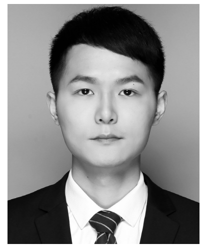
[7] D. Taubman and J. Thie, "Optimal erasure protection for scalably compressed video streams with limited retransmission," *IEEE Trans. Image Process.*, vol. 14, no. 8, pp. 1006–1019, Aug. 2005.

[8] R. Xiong, D. Taubman, and V. Sivaraman, "Optimal PET protection for streaming scalably compressed video streams with limited retransmission based on incomplete feedback," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2382–2395, Sep. 2010.

[9] R. Xiong, D. S. Taubman, and V. Sivaraman, "PET protection optimization for streaming scalable videos with multiple transmissions," *IEEE Trans. Image Process.*, vol. 22, no. 11, pp. 4364–4379, Nov. 2013.
[10] W.-T. Tan and A. Zakhor, "Video multicast using layered FEC and scalable compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 373–386, Mar. 2001.
[11] M. Gallant and F. Kossentini, "Rate-distortion optimized layered coding with unequal error protection for robust Internet video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 357–372, Mar. 2001.
[12] V. M. Stankovic, R. Hamzaoui, Y. Charfi, and Z. Xiong, "Real-time unequal error protection algorithms for progressive image transmission," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1526–1535, Dec. 2003.
[13] J. Kim, R. M. Mersereau, and Y. Altunbasak, "Error-resilient image and video transmission over the Internet using unequal error protection," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 121–131, Feb. 2003.
[14] N. Thomos, N. V. Boulgouris, and M. G. Strintzis, "Wireless image transmission using turbo codes and optimal unequal error protection," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1890–1901, Nov. 2005.
[15] S. Jakubczak and D. Katabi, "A cross-layer design for scalable mobile video," in *Proc. 17th Annu. Int. Conf. Mobile Comput. Netw. (MOBICOM)*, Las Vegas, NV, USA, Sep. 2011, pp. 289–300.
[16] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[17] X. Fan, F. Wu, D. Zhao, and O. C. Au, "Distributed wireless visual communication with power distortion optimization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 6, pp. 1040–1053, Jun. 2013.
[18] X. Fan, F. Wu, D. Zhao, O. C. Au, and W. Gao, "Distributed soft video broadcast (DCAST) with explicit motion," in *Proc. Data Compress. Conf.*, Snowbird, UT, USA, Apr. 2012, pp. 199–208.
[19] A. Aaron, R. Zhang, and B. Girod, "Wyner–Ziv coding of motion video," in *Proc. Conf. Rec. 36th Asilomar Conf. Signals, Syst. Comput.*, vol. 1. Nov. 2002, pp. 240–244.
[20] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2005.
[21] X. Fan, R. Xiong, D. Zhao, and F. Wu, "Layered soft video broadcast for heterogeneous receivers," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 11, pp. 1801–1814, Nov. 2015.
[22] F. Wu, X. Peng, and J. Xu, "LineCast: Line-based distributed coding and transmission for broadcasting satellite images," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1015–1027, Mar. 2014.
[23] X. L. Liu, W. Hu, Q. Pu, F. Wu, and Y. Zhang, "ParCast: Soft video delivery in MIMO-OFDM WLANs," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw. (Mobicom)*, Istanbul, Turkey, 2012, pp. 233–244.
[24] 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.
[25] D. He, C. Luo, C. Lan, F. Wu, and W. Zeng, "Structure-preserving hybrid digital-analog video delivery in wireless networks," *IEEE Trans. Multimedia*, vol. 17, no. 9, pp. 1658–1670, Sep. 2015.
[26] H. Cui, C. Luo, C. W. Chen, and F. Wu, "Robust linear video transmission over Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2790–2801, Aug. 2014.
[27] K.-H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. Commun.*, vol. COM-24, no. 12, pp. 1283–1290, Dec. 1976.
[28] S. Zheng *et al.*, "Softcast with per-carrier power-constrained channels," in *Proc. IEEE Int. Conf. Image Process., (ICIP)*, Phoenix, AZ, USA, Sep. 2016, pp. 2122–2126.
[29] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1179–1197, May 2004.
[30] *Xiph.Org Media*. Accessed: Mar. 1, 2018. [Online]. Available: https://media.xiph.org/video/derf
[31] T. S. Rappaport, *Wireless Communications: Principles and Practice*. vol. 2. Englewood Cliffs, NJ, USA: Prentice Hall, 1996,

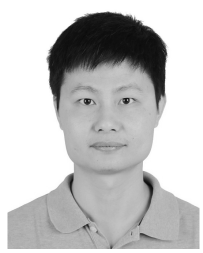


Chaofan He received the B.S. degree from the University of Electronic Science and Technology of China, Chengdu, in 2016, where he is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering. His research interests include multimedia signal processing, wireless multimedia communication and networking, and multimedia social network.



Huiying Wang received the B.S. degree from the University of Electronic Science and Technology of China, Chengdu, in 2016, where he is currently pursuing the master's degree with the School of Information and Communication Engineering. His research interests include multimedia signal processing and wireless multimedia communication.

Yang Hu received the B.S. and Ph.D. degrees in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2004 and 2009, respectively. She was with the Institute for Advanced Computer Studies, University of Maryland, as a Research Associate from 2010 to 2015. She is currently an Associate Researcher with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, China. Her current research interests include computer vision, machine learning, and multimedia signal processing.



Yan Chen (SM'14) received the bachelor's degree from the University of Science and Technology of China in 2004, the M.Phil. degree from The Hong Kong University of Science and Technology in 2007, and the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 2011. He was with Origin Wireless Inc., as a Founding Principal Technologist. Since 2015, he has been a Full Professor with the School of Information and Communication Engineering, University of Electronic Science and Technology of China. His research interests include multimedia, signal processing, game theory, and wireless communications.

Dr. Chen was a recipient of multiple honors and awards, including the Chinese Government Award for Outstanding Students Abroad in 2010, the Future Faculty Fellowship and the Distinguished Dissertation Fellowship Honorable Mention from the Department of Electrical and Computer Engineering in 2010 and 2011, respectively, the Finalist of the Dean's Doctoral Research Award from the A. James Clark School of Engineering, University of Maryland, in 2011, the Best Paper Award at the IEEE GLOBECOM in 2013, the Best Student Paper Award at the IEEE ICASSP in 2016, and the Best Student Paper Award at the PCM in 2017.

Xiaopeng Fan (S'07–M'09–SM'17) received the B.S. and M.S. degrees from the Harbin Institute of Technology (HIT), Harbin, China, in 2001 and 2003, respectively, and the Ph.D. degree from The Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2009. He joined HIT in 2009, where he is currently a Professor. From 2003 to 2005, he was with Corporation (China) as a Software Engineer. From 2011 to 2012, he was with Microsoft Research Asia as a Visiting Researcher. From 2015 to 2016, he was with HKUST as a Visiting Assistant Professor. He has authored one book and over 100 articles in refereed journals and conference proceedings. His current research interests include video coding and transmission, image processing, and computer vision. He has received the Outstanding Contributions Award for the Development of IEEE Standard 1857 from the IEEE in 2013. He served as the Program Chair of PCM2017, the Chair of the IEEE SGC2015, and a Co-Chair of MCSN2015. He has been an Associate Editor of the IEEE 1857 standard since 2012.

Houqiang Li received the B.S., M.Eng., and Ph.D. degrees from the University of Science and Technology of China (USTC), Hefei, China, in 1992, 1997, and 2000, respectively, all in electronic engineering. He is currently a Professor with the Department of Electronic Engineering and Information Science, USTC. He has authored or co-authored over 100 papers in journals and at conferences. His research interests include video coding and communication, multimedia search, and image/video analysis. He was a Senior Author of the Best Student Paper of the 5th International Mobile Multimedia Communications Conference in 2009. He was a recipient of the Best Paper Award for the International Conference on Mobile and Ubiquitous Multimedia from the ACM in 2011, the Best Paper Award for Visual Communications and Image Processing in 2012, and the Best Paper Award for the International Conference on Internet Multimedia Computing and Service in 2012. He has served on technical/program committees, organizing committees, and as a program co-chair or track/session chair for over 10 international conferences. He served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUIT AND SYSTEMS FOR VIDEO TECHNOLOGY from 2010 to 2013.

Bing Zeng (M'91–SM'13–F'16) received the B.Eng. and M.Eng. degrees in electronic engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1983 and 1986, respectively, and the Ph.D. degree in electrical engineering from the Tampere University of Technology, Tampere, Finland, in 1991.

He was a Post-Doctoral Fellow with the University of Toronto from 1991 to 1992 and a Researcher with Concordia University from 1992 to 1993. He then joined The Hong Kong University of Science and Technology (HKUST). After 20 years of service at HKUST, he returned to UESTC through the Chinas 1000-Talent-Scheme in 2013. At UESTC, he leads the Institute of Image Processing to work on image and video processing, 3D and multi-view video technology, and visual big data.

During his tenure at HKUST and UESTC, he graduated over 30 master's and Ph.D. students, received about 20 research grants, filed eight international patents, and authored over 260 papers. He hold significant three works—one paper on fast block motion estimation, published in the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (TCSVT) in 1994, has so far been SCI-cited over 1000 times (Google cited over 2200 times) and currently stands at the eighth position among all papers published in this Transactions, one paper on smart padding for arbitrarily-shaped image blocks, published in the IEEE TCSVT in 2001, led to a patent that has been successfully licensed to companies, and one paper on directional discrete cosine transform, published in the IEEE TCSVT in 2008, received the 2011 IEEE CSVT Transactions Best Paper Award. He also received the Best Paper Award at ChinaCom three times (2009 Xian, 2010 Beijing, and 2012 Kunming).

Dr. Zeng was elected as an IEEE Fellow in 2016 for contributions to image and video coding. He served as an Associate Editor for the IEEE TCSVT for eight years. He received the Best Associate Editor Award in 2011 and the 2nd Class Natural Science Award (the first recipient) from the Chinese Ministry of Education in 2014. He was the General Co-Chair of VCIP-2016 and PCM-2017.