

Adaptive Resource Allocation for Scalable Video Transmission over OFDMA Systems

Zan Yang¹ · Yuping Zhao²

Published online: 27 June 2016
© Springer Science+Business Media New York 2016

Abstract We propose an adaptive resource allocation algorithm for scalable H.264 video transmission over orthogonal frequency-division multiple-access (OFDMA) systems. The algorithm explores the scalability of the video codec and the multiuser diversity of OFDMA systems to achieve enhanced efficiency and video quality. We use the peak signal-to-noise ratio (PSNR) as the measure metric of video quality, and jointly optimize the video layer extraction and radio resource allocation, considering the heterogeneity of video sources and channel conditions. Given the fact that different quality layers of the video have different importance, an adaptive modulation and coding scheme is proposed where a fixed coding rate is used for each quality layer, and unequal error protection is implemented for different layers. We also propose a subcarrier allocation algorithm to assign subcarriers for different layers of different users' videos to maximize the average PSNR of users while satisfying the base layer requirement. The proposed resource allocation algorithm has a low complexity and is suitable for practical implementations. We further develop an end-to-end simulation testbed to demonstrate the effectiveness of the proposed transmission scheme.

Keywords Scalable video · H.264 · OFDMA · Adaptive modulation and coding · Subcarrier allocation

✉ Zan Yang
0514yangzan@gmail.com
Yuping Zhao
yuping.zhao@pku.edu.cn

¹ Transport Planning and Research Institute, The Ministry of Transport, Beijing 100028, People's Republic of China

² State Key Laboratory of Advanced Optical Communication Systems and Networks, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, People's Republic of China

1 Introduction

With the development of both video compression technology and wireless networks, transmitting video programs over wireless channels becomes increasingly popular. On one hand, H.264/AVC enables more efficient and flexible video coding. On the other hand, the next generation (4G) wireless technologies achieves high data rate transmission, e.g., 1 Gbps for nomadic and 100 Mbps for mobile users [1]. However, wireless video transmission is still a challenging problem due to limited network bandwidth, the presence of channel errors, and the variability in consumer terminals [2]. In order to optimize the received video quality, it is important to map the video source to the radio resource in an efficient way by taking into account the video stream characteristics and the channel conditions.

The recently proposed scalable extension of H.264/AVC [3, 4], popularly known as SVC, inherits the superior compression efficiency and the error-resilient network adaptation layer (NAL) structure. SVC splits the information bitstreams into scalable layers to provide differentiated quality of service (QoS). The base layer guarantees the base-quality version of the content, whereas the refinement information encoded in the enhancement layers, allows video quality to progressively improve [5]. The scalability feature of SVC can be exploited to improve the video transmission efficiency over error-prone wireless networks. On the other hand, orthogonal frequency-division multiplexing (OFDM) is regarded as one of the promising techniques for future broadband wireless networks due to its ability to provide high data rates in the multi-path fading environment [6, 7]. OFDMA is a multiuser version of OFDM, where each user is assigned a subset of the subcarriers for exclusive data transmission such that multiuser diversity can be achieved. The resource allocation problems for OFDMA systems have been considered in a number of works, e.g., [8–10].

Resource allocation for scalable video transmission can efficiently improve the system performance, and has also been considered in literatures. Multiuser resource allocation for real-time video transmission in OFDM systems are treated in [11, 12] to maximize the total bit rates of users. In [13, 14], the received video distortion is used as the metric to select the application and physical layer parameters for single-user video transmission. In [15], scalable video transmission over MIMO systems is discussed. A joint transmit-precoding and scalable layer selection algorithm is proposed by taking into account the delay and buffer constraints. However, the design is based on ideal assumptions like Gaussian input alphabet and capacity-achieving codes, which is not appropriate for practical applications. Moreover, the manner of uniformly calculating the transmission rates of layers does not consider layers' unequal significance. A scalable multiuser framework for video transmission over OFDM networks is given in [16], where fairness and efficiency are considered in determining subcarrier allocation, power allocation, and modulation and channel coding rate. However, the analysis is based on the 3-D embedded wavelet video codec (EWW). Also, assigning different coding rates to each subcarrier has high signalling overhead and implementation complexity.

The goal of this paper is to design practical resource allocation algorithms for scalable H.264 streaming over OFDMA systems. By jointly considering video characteristics and channel conditions of different users, as well as the UEP requirement of layers, we develop a cross-layer design framework to optimize the video layer extraction, modulation and coding, and subcarrier allocation. The PSNR is used as the video quality metric, and an empirical model is used to relate the PSNR and the rate. Different from existing works

[15, 16], we consider the unequal importance of layers of users, and allocate video and radio resources not only between users, but also between scalable quality layers, such that the rate constraint and protection requirement of different layers can be satisfied separately. Based on the mutual information exponentially system mapping (MIESM) method, an adaptive modulation and coding scheme is first proposed where uniform channel coding rate is achieved for each scalable layer, and UEP is considered for different layers. We then formulate the subcarrier allocation problem to maximize the average PSNR of users while satisfying the base layer requirement of each user. The problem is found to be a typical 0–1 programming problem which is NP-hard. A fast suboptimal algorithm is further proposed with low complexity. The application-centric design makes the proposed algorithms suitable to practical implementation. Finally, we implement an end-to-end experimental testbed to show that the proposed scheme improves the delivered video quality.

The remainder of the paper is organized as follows. In Sect. 2, we provide some brief background on SVC and introduce the considered downlink OFDMA systems. The proposed adaptive modulation and coding scheme is described in Sect. 3. In Sect. 4, the proposed subcarrier allocation algorithm is developed. Section 5 discusses the simulation results. Finally Sect. 6 concludes the paper.

2 Background and System Descriptions

In this section, we first introduce the background of scalable video coding. Then we give the video rate-quality model which is used in the following algorithm. Finally, the considered downlink OFDMA system is described.

2.1 Scalable Video Coding

In general, a video bitstream is called scalable when parts of the stream can be removed in a way such that the resulting sub-stream forms another valid bitstream for some target decoder. In the SVC extension of the H.264/AVC standard, a stream has a base layer and several enhancement layers. The encoding and decoding of a higher enhancement layer are based on those of the base layer and all lower enhancement layers. The SVC provides graceful quality degradation in lossy transmission environments. As long as the base layer is received, the receiver can decode the video stream. As more enhancement layers are received, the decoded video quality is progressively improved.

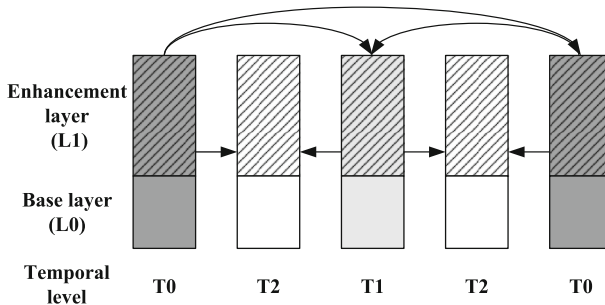


Fig. 1 The encoding structure of SVC

Specifically, this codec contains several forms of scalability, including the temporal scalability using a hierarchical prediction structure, the spatial scalability using inter-layer prediction mechanisms, and the quality (or SNR) scalability using quantized discrete cosine transform (DCT)-coefficients with different quantization parameters (QPs). In this paper, we employ the temporal scalability and the medium-grain scalability (MGS), which is a form of SNR scalability. However the proposed framework can be easily extended to other types of scalability. The encoding structure of the considered SVC is shown in Fig. 1, where 3 temporal levels [group of pictures (GOP = 4)], and 2 quality layers are included. The temporal levels are denoted by T0, T1, and T2; the quality layers are denoted by L0 (base layer) and L1 (enhancement layer). For a detailed overview of SVC, please refer to [4].

One way to evaluate the video transmission quality is to use the expected end-to-end distortion as a metric, given by

$$D_E = D_R + D_L \quad (1)$$

where D_R is the distortion due to source compression coding, and D_L is the distortion due to channel error [17]. D_R is determined by the quantization error and sub-stream extraction, whereas D_L is determined by the channel error and the error concealment technique employed at the decoder.

This paper focuses on the resource allocation for wireless video transmission over OFDMA systems. We mainly consider the source compression distortion D_R , while keeping the channel-induced distortion D_L at a controllable level. This is achieved by adapting the source rate to the variable channel condition, i.e., extracting a sub-stream of certain rate according to the channel state information (CSI), and selecting the appropriate MCS for transmission.

2.2 Video Rate-Quality Model

We consider the SVC bitstream is encoded with one base layer and one enhancement layer that support a rate interval. A sub-stream of certain rate within the rate interval can be extracted for transmission to provide differentiated video quality. The extraction takes into account the layers' priorities such that the base layer is extracted first, from low to high time level, then followed by the enhancement layer. We are interested in a downlink video transmission scenario where the basestation transmits K video streams, each to a designated user.

As in [14, 18], the PSNR is used as the metric of video quality. It is known that the PSNR is related to the rate of the extracted sub-stream through a piece-wise linear function [18]. The relationship between the PSNR Q_k of the k -th video stream and the video rate r can be expressed as

$$Q_k(r) = \begin{cases} Q_k^b + \beta_k^b(r - R_k^b), & r < R_k^b \\ Q_k^b + \beta_k^e(r - R_k^b), & R_k^b \leq r < R_k^b + R_k^e \\ Q_k^e, & r \geq R_k^b + R_k^e \end{cases} \quad (2)$$

where R_k^b and R_k^e are the rates of the base layer and the enhancement layer; Q_k^b and Q_k^e are the PSNR when only the base layer is extracted and that when the base layer plus the enhancement layer are extracted, respectively; β_k^b and β_k^e are the coefficients, which are determined by the QP values of the base and enhancement layers. Note that in general, we

have $\beta_k^b > \beta_k^e$. Moreover, β_k^b and β_k^e differ from video to video because of the content complexities and motion activities.

2.3 System Descriptions

In this paper, a single-cell downlink OFDMA system with N subcarriers and K users is considered. We assume a slow fading channel where the channel gains remain constant within each transmission interval. By performing resource allocation, each subcarrier is assigned to a particular layer of some user, with the associated MCS.

Denote γ_{kn} as the k -th user's signal to noise ratio (SNR) on the n -th subcarrier, which is given by

$$\gamma_{kn} = \frac{p_{kn}g_{kn}}{\sigma^2}, n = 0, 1, \dots, N - 1 \quad (3)$$

where p_{kn} is the transmission power for the k -th user on the n -th subcarrier; g_{kn} is the channel gain fed back from the receiver; σ^2 is the thermal noise power which is assumed to be the same for each subcarrier of different users. Because of limitation of power amplifier and consideration of co-channel interferences to other cells, the overall power is bounded by P_{\max} , i.e., $\sum_{k=0}^{K-1} \sum_{n=0}^{N-1} p_{kn} \leq P_{\max}$. To reduce computation complexity and signaling overhead, we use equal power allocation among subcarriers, i.e., $p_{kn} = P_{\max}/N$. It is already shown that a fixed power allocation leads to a negligible throughput penalty if power is poured on subcarriers with good channel gain, and adaptive rate scheme is implemented [19, 20].

The basestation extracts a sub-stream from each user's SVC bitstream. Then they are loaded onto the subcarriers after channel coding and modulation, based on the output of the resource allocation algorithm. In Sects. 3 and 4, we describe the proposed adaptive modulation and coding and subcarrier allocation schemes, respectively.

3 Adaptive Modulation and Coding

Adaptive modulation and coding provides high spectrum efficiency by adjusting each subcarrier's data transmission rate according to the channel condition. In this section, we first briefly introduce the MIESM method. Then an adaptive modulation and coding scheme is proposed based on the MIESM method, which assigns the channel coding rate and the modulation of subcarriers for different layers of users.

3.1 The MIESM Method

The MIESM method is based on the observation that when the normalized mutual information is the same, the corresponding block error rate (BLER) performance becomes identical regardless of the data modulation scheme, and is determined only by the coding rate [21]. This can be expressed as the function

$$P_{\text{err}}(x) = \phi(x, \rho) \quad (4)$$

where P_{err} is the BLER; x is the normalized mutual information, or mutual information per bit (MIB); ρ is the channel coding rate. Conversely, given the target BLER P_{err} , the

corresponding channel coding rate ρ needed to achieve such BLER can be obtained as $\rho = \phi^{-1}(x, P_{\text{err}})$.

Given a sequence $(\gamma_j, S_j), j = 0, 1, \dots, J - 1$, of the SNR and constellation pairs, the MIB can be calculated as $\sum_{j=0}^{J-1} I(\gamma_j, S_j) / \sum_{j=0}^{J-1} \log |S_j|$. We can achieve an acceptable BLER P_{err} with a practical code of rate [21]

$$\rho = \left(\frac{\sum_{j=0}^{J-1} I(\gamma_j, S_j)}{\sum_{j=0}^{J-1} \log |S_j|} - \Gamma \right)^+ \tag{5}$$

where $I(\gamma_j, S_j)$ denotes the coded-modulation mutual information; $(x)^+ = \max\{x, 0\}$. The value of $I(\gamma_j, S_j)$ can be computed numerically based on the channel capacity with discrete input [22]. The constant Γ captures the loss between practical codes and ideal codes, and can be obtained by checking the BLER versus MIB curve of practical channel codes with certain rates. Also, the value of Γ is affected by the choice of the acceptable BLER value. For instance, in 3GPP, the BLER performance as a function of MIB can be found in [21] where Γ is generally set as 0.1 with a target BLER of 0.1.

3.2 The Proposed Adaptive MCS Considering UEP

We next describe the proposed adaptive MCS for different layers of users. It is already known that UEP techniques can improve the received video quality of the video transmission [23, 24]. UEP recognizes the fact that bitstreams of different parts of the videos are not of equal importance, and correspondingly assigns unequal amounts of radio resource or degree of forward error correcting (FEC) protection according to the source significance information (SSI).

In SVC, considering the decoding dependencies and unequal importance to the reconstructed video, we protect different quality layers of the video stream with channel code of different rates. We first set the target BLERs $P_{\text{err}_k}^b$ and $P_{\text{err}_k}^e$ for the base layer and enhancement layer of the k -th user respectively, i.e., base layer targets at a lower BLER to better ensure their correct delivery. Then the corresponding Γ_k^b and Γ_k^e to achieve these BLERs can be obtained. Qualitatively, compared to the enhancement layer, the base layer targets at a smaller value of BLER, which results in a larger value of Γ_k^b , and correspondingly a lower coding rate.

We derive the transmission rate in the considered OFDMA system. First we define the binary indicator variables $\{a_{kn}^b\}$ such that $a_{kn}^b = 1$ if the n -th subcarrier is allocated to the base layer of the k -th user, and $a_{kn}^b = 0$ otherwise. Similarly, define the binary indicator variables $\{a_{kn}^e\}$ such that $a_{kn}^e = 1$ if the n -th subcarrier is allocated to the enhancement layer of the k -th user, and $a_{kn}^e = 0$ otherwise. Denote $\mathcal{A}_k^b = \{n | a_{kn}^b = 1\}$ and $\mathcal{A}_k^e = \{n | a_{kn}^e = 1\}$ as the sets of subcarriers allocated to the base layer and the enhancement layer of the k -th user, respectively. From (5), it follows that the effective transmission rate of each symbol is $I(\gamma_j, S_j) - \Gamma \log |S_j|$. Then the transmission rate of the n -th subcarrier, if it is allocated to the k -th user, is given by

$$r_{kn}(\gamma_{kn}, S_{kn}) = \begin{cases} I(\gamma_{kn}, S_{kn}) - \Gamma_k^b \log |S_{kn}|, & n \in \mathcal{A}_k^b \\ I(\gamma_{kn}, S_{kn}) - \Gamma_k^e \log |S_{kn}|, & n \in \mathcal{A}_k^e \end{cases} \tag{6}$$

where S_{kn} is a modulation symbol of unit energy from a constellation set \mathcal{S} . Given γ_{kn} , the maximum transmission rate is obtained by the SNR-dependent constellation selection, which is expressed as

$$\hat{r}_{kn} = \max_{S_{kn} \in \mathcal{S}} r_{kn}(\gamma_{kn}, S_{kn}). \tag{7}$$

The corresponding optimal constellation is then given by

$$\hat{S}_{kn} = \operatorname{argmax}_{S_{kn} \in \mathcal{S}} r_{kn}(\gamma_{kn}, S_{kn}). \tag{8}$$

It is seen from (6) and (7) that the same subcarrier may have different transmission rates if it is allocated to different layers of users. The subcarrier allocation will be discussed in Sect. 4.

We use uniform channel coding rate for each layer of users. From (5)–(8), the channel coding rate for the base layer of the k -th user is given by

$$\rho_k^b = \left(\frac{\sum_{n \in \mathcal{A}_k^b} I(\gamma_{kn}, \hat{S}_{kn})}{\sum_{n \in \mathcal{A}_k^b} \log |\hat{S}_{kn}|} - \Gamma_k^b \right)^+. \tag{9}$$

Similarly, the channel coding rate for the enhancement layer of the k -th user is given by

$$\rho_k^e = \left(\frac{\sum_{n \in \mathcal{A}_k^e} I(\gamma_{kn}, \hat{S}_{kn})}{\sum_{n \in \mathcal{A}_k^e} \log |\hat{S}_{kn}|} - \Gamma_k^e \right)^+. \tag{10}$$

In practice, ρ_k^b and ρ_k^e can be quantized to the closest rate when a finite set of channel codes with different rates is employed. Compared to the schemes that assign different coding rates to subcarriers [16], the method proposed here uses fixed coding rate for the same layer across different subcarriers, therefore has less signalling overhead and lower implementation complexity. Compared to the schemes that set coding rates regardless of the channel condition, the proposed method has greater flexibility, and better exploits the channel resources.

4 The Proposed Subcarrier Allocation Algorithm

In this section, we first formulate the subcarrier allocation problem. Then the proposed subcarrier allocation algorithm is given. Finally, the convergence and complexity of the proposed algorithm are discussed.

4.1 Problem Formulation

We now formulate the subcarrier allocation problem to assign subcarriers to different layers of the users. Note that $\sum_{n=0}^{N-1} a_{kn}^b \hat{r}_{kn}^b$ and $\sum_{n=0}^{N-1} a_{kn}^e \hat{r}_{kn}^e$ are the allocated transmission rates of the k -th user’s base and enhancement layers, respectively. Using the rate-quality model given in (2), each layer of the k -th user is considered *completed* when the allocated transmission rate is larger than its rate bound, e.g., for the base layer, it is expressed as

$$\sum_{n=0}^{N-1} a_{kn}^b \hat{r}_{kn}^b \geq R_k^b. \tag{11}$$

Note that in (11), the source rate R_k^b is normalized by $1/T_s$, where T_s is the OFDM symbol interval. Because of the video decoding dependency, for any user, the transmission rate of the enhancement layer is meaningful only when the base layer is completed.

We then formulate the resource allocation problem to maximize the average PSNR of all users, while meeting their base layer requirement. Based on the rate-quality model (2), we have the following formulation:

$$\max_{\{a_{kn}^b, a_{kn}^e\}} \sum_{k=0}^{K-1} \beta_k^e \sum_{n=0}^{N-1} a_{kn}^e \hat{r}_{kn}^e \tag{12a}$$

$$\text{s.t. } \sum_{k=0}^{K-1} (a_{kn}^b + a_{kn}^e) \leq 1, \quad a_{kn}^b, \quad a_{kn}^e \in \{0, 1\}, \quad \forall n \tag{12b}$$

$$\sum_{n=0}^{N-1} a_{kn}^b \hat{r}_{kn}^b \geq R_k^b, \quad \forall k \tag{12c}$$

$$\sum_{n=0}^{N-1} a_{kn}^e \hat{r}_{kn}^e \leq R_k^e, \quad \forall k \tag{12d}$$

The constraint of (12b) imposes that one subcarrier can be assigned to at most one layer of one user. The fairness constraint of (12c) imposes that the base layer rate requirement of all users must be satisfied. The constraint of (12d) imposes that any extra allocation to the enhancement layer exceeding its rate bound R_k^e is a waste. The objective function (12a) is the sum PSNR achieved by the enhancement layers of all users. Since the base layer requirement is satisfied by (12c), from (2) it is easy to find that (12a) has the same meaning as maximizing the average PSNR of all users. Note that (12c) assumes that by the optimal allocation, the base layer rates of all users can be completed using the given channel resources.

Clearly (12) is a standard linear 0–1 programming problem, which is NP-hard. In order to find the optimal solution, an exhaustive search involves searching over $(2K)^N$ possible allocations of $\{a_{kn}^b\}$ and $\{a_{kn}^e\}$. Although standard methods such as branch-and-bound can reduce the complexity, they are still prohibitively complex when the numbers of the users and subcarriers are large [25]. We next propose a low-complexity suboptimal solution to (12).

4.2 The Proposed Subcarrier Allocation Algorithm

We first allocate subcarriers to maximize the objective function of (12a) without considering the rate constraints in (12c) and (12d). This yields an optimum solution corresponding to the unconstrained problem, which can be obtained using a greedy approach. However, the solution is not in the feasible set. We then use this solution as a starting point for the search process, and reallocate the subcarriers step by step, until all the rate constraints have been satisfied.

4.2.1 Unconstrained optimization

Without considering (12c) and (12d), the optimization problem of (12) becomes

$$\max_{\{a_{kn}^b, a_{kn}^e\}} \sum_{k=0}^{K-1} \beta_k^e \sum_{n=0}^{N-1} a_{kn}^e \hat{r}_{kn}^e \tag{13a}$$

$$\text{s.t. } \sum_{k=0}^{K-1} (a_{kn}^b + a_{kn}^e) \leq 1, \quad a_{kn}^b, \quad a_{kn}^e \in \{0, 1\}, \quad \forall n \tag{13b}$$

It is obvious to find that (13) can be solved by finding for the n -th subcarrier,

$$k_n^1 = \operatorname{argmax}_{k=0, \dots, K-1} \beta_k^e \hat{r}_{kn}^e \tag{14}$$

and then let

$$a_{kn}^e = \begin{cases} 1, & k = k_n^1 \\ 0, & \forall k \neq k_n^1 \end{cases} \tag{15}$$

and

$$a_{kn}^b = 0, \quad \forall k. \tag{16}$$

That is, the maximum objective function is achieved by assigning each subcarrier to the enhancement layer of the user who has the largest value of $\beta_k^e \hat{r}_{kn}^e$. Then the subcarrier allocation sets are obtained as $\mathcal{A}_k^e = \{n | a_{kn}^e = 1\}$, $\mathcal{A}_k^b = \phi, \forall k$.

4.2.2 Subcarrier Reallocation

The solution of (15) and (16) does not satisfy the rate constraints of (12c) and (12d), thus a reallocation is needed. Note that any reallocation from (15) and (16) will inevitably cause a decrease in the objective function (12a). Therefore, optimal reallocation should cause the least reduction of the objective function. Moreover, the number of reallocation operations should be kept as low as possible considering the implementation complexity.

In this paper a three-step reallocation procedure is proposed. We first reallocate the subcarriers between users to satisfy the rate constraints of (12c) and (12d), respectively. Then we reallocate subcarriers between layers of each user.

4.2.2.1 Step 1: Reallocation for the Lower Bound Requirement In order to satisfy the constraint of (12c), we first consider the lower bound constraint of each user’s total transmission rate. The key idea is that we do not discriminate base and enhancement layers’ allocation at first, provided the subcarriers are reallocated to the same user. Then the lower bound of the k -th user’s rate can be obtained by substituting a_{kn}^b with a_{kn}^e in (12c), yielding

$$\sum_{n=0}^{N-1} a_{kn}^e \hat{r}_{kn}^b \geq R_k^b, \quad \forall k. \tag{12c'}$$

That is, at least, the total subcarriers allocated to each user should be enough to complete the base layer.

Type-I cost function: Define the cost function of reallocating the n -th subcarrier to the base layer of the k -th user instead of the originally assigned user k_n^1 , which is given by

$$c_{kn}^1 = \frac{\beta_{k_n^1}^e \hat{r}_{k_n^1 n}^e}{\hat{r}_{kn}^b}, \quad \forall n. \tag{17}$$

Note that c_{kn}^1 is proportional to the decrease in the PSNR of the k_n^1 -th user's enhancement layer, and inversely proportional to the increase in the rate of the k -th user's base layer.

Then the reallocation algorithm is summarized as follows. For each user who has not met its lower bound requirement of (12c'), each time we select the subcarrier with the lowest value of the Type-I cost function. If the subcarrier's reallocation won't violate any constraint that is already satisfied, the selected subcarrier is reallocated to this user. Otherwise, the reallocation is not allowed. We skip the subcarrier from this step, and find another subcarrier to reallocate, in the rest of the subcarriers. The above procedure is summarized with Algorithm 1.

Algorithm 1: The reallocation algorithm for the lower bound requirement

Initialization: set subcarrier index set $\mathcal{W} = \{0, 1, \dots, N - 1\}$, and let user index $k = 0$.

Repeat

1. Calculate the k -th user's cost function $c_{kn}^1, \forall n \in \mathcal{W}$.
2. **Repeat**
 - 1) Find the subcarrier with the least reallocation cost, i.e., $n' = \arg \min_{n \in \mathcal{W}} c_{kn}^1$. The user to whom this subcarrier is originally assigned is denoted as $k_{n'}^1$.
 - 2) **If** $\sum_{n=0}^{N-1} a_{k_n^1, n}^e \hat{r}_{k_n^1 n}^b - \hat{r}_{k_{n'}, n'}^b \geq R_{k_{n'}}^b$, i.e., the reallocation won't violate the $k_{n'}^1$ -th user's lower bound constraint, then
 - i. update $a_{kn'} = 1$, and add subcarrier n' to set \mathcal{A}_k^e .
 - ii. update $a_{k_{n'}, n'}^e = 0$, and remove subcarrier n' from set $\mathcal{A}_{k_{n'}^1}^e$.
 - iii. remove subcarrier n' from set \mathcal{W} .
 - 3) **Else** remove subcarrier n' from set \mathcal{W} , then **Goto** Step 1).
 - 4) **End If**

Until The lower bound constraint of the k -th user is satisfied, i.e., $\sum_{n=0}^{N-1} a_{kn}^e \hat{r}_{kn}^b \geq R_k^b$ **or** all subcarriers have been used up, i.e., $\mathcal{W} = \phi$.
3. $k = k + 1$.

Until All users have been treated, i.e., $i \geq K$.

After the reallocation in Step 1, each user has been allocated with a certain number of subcarriers, which can satisfy the base layer rate constraint of (12c), after the reallocation between layers to be introduced in Step 3.

4.2.2.2 Step 2: Reallocation for the Upper Bound Requirement We next reallocate subcarriers to satisfy the constraint of (12d). Similarly, we allocate subcarriers for the base and enhancement layers in a united way, and consider the upper bound constraint of the each user’s total transmission rate, which can be obtained by rewriting (12d) as

$$\sum_{n=0}^{N-1} a_{kn}^e \hat{r}_{kn}^e \leq R_k^b + R_k^e, \quad \forall k. \tag{12d'}$$

Proposition 1 The upper bound constraint of the k -th user’s rate can be approximated by (12d’).

Proof In (12d’), $R_k^b + R_k^e$ is the total source rate, $\sum_{n=0}^{N-1} a_{kn}^e \hat{r}_{kn}^e$ is used to approximate the total transmission rate of the k -th user. Note that since we have $\hat{r}_{kn}^b < \hat{r}_{kn}^e$ considering the UEP requirement, (12d’) actually has a more stringent sum rate constraint than that of (12c) and (12d), and reduces the feasible solution space. However, for good practical codes, the performance is close to the Shannon limit, and the values of Γ_k^b and Γ_k^e are small. Moreover, it is known that with SNR/SINR-dependent constellation selection, in the normal operating SNR range, the MIESM method works in the high coding rate range. Thus from (6) we find that although the UEP for layers results in difference of \hat{r}_{kn}^b and \hat{r}_{kn}^e , the difference is not large. Therefore, (12d’) is a good approximation of the upper bound of the total transmission rate. □

Type-II cost function: Define the cost function of reallocating the n -th subcarrier to the enhancement layer of the k -th user instead of the originally assigned user k_n^1 , which is given by

$$c_{kn}^2 = \frac{\beta_{k_n^1}^e \hat{r}_{k_n^1 n}^e - \beta_k^e \hat{r}_{kn}^e}{\hat{r}_{k_n^1 n}^e}, \quad \forall n. \tag{18}$$

Similarly, c_{kn}^2 is proportional to the decrease in the total PSNR, and inversely proportional to the decrease in the rate of the k_n^1 -th user’s enhancement layer.

Then the following reallocation algorithm is proposed. For each user who has exceeded its upper bound, each time we find a subcarrier originally assigned to the user, as well as another user to whom to reallocate this selected subcarrier. In other words, we find a subcarrier-user reallocation pair, denoted as (n', k') . Again, the reallocation with the lowest value of the Type-II cost function is considered with the highest priority. If the reallocation won’t violate the already satisfied constraints, it is allowed to proceed. Otherwise, the selected user is skipped from this step, and we repeat the above procedure by finding another set of subcarrier-user pair, in the rest of the users. Different from Algorithm 1, each time a subcarrier-user reallocation pair is jointly selected. The procedure is summarized with Algorithm 2.

Algorithm 2: The reallocation algorithm for the upper bound requirement

Initialization: set user index set $\mathcal{K} = \{0, 1, \dots, K - 1\}$, and let user index $k = 0$.

Repeat

1. Among the subcarriers allocated to the k -th user, calculate the reallocation cost function of each subcarrier-user pair, i.e., evaluate $c_{in}^2, \forall n \in \mathcal{A}_k^e, \forall i \in \mathcal{K}, i \neq k_n^1$.
2. **Repeat**
 - 1) Find the subcarrier-user reallocation pair with the least reallocation cost, i.e., $(n', k') = \arg \min_{n \in \mathcal{A}_k^e, i \in \mathcal{K}, i \neq k_n^1} c_{in}^2$.
 - 2) **If** $\sum_{n=0}^{N-1} a_{k'n}^e \hat{r}_{k'n}^e + \hat{r}_{k'n'}^e \leq R_{k'}^b + R_{k'}^e$, i.e., the reallocation won't violate the k' -th user's rate constraint, then
 - i. update $a_{k'n'} = 1$, and add subcarrier n' to set \mathcal{A}_k^e .
 - ii. update $a_{kn'} = 0$, and remove subcarrier n' from set \mathcal{A}_k^e .
 - 3) **Else** remove user k' from set \mathcal{K} , then **Goto** Step 1).
 - 4) **End If**
- Until** The upper bound constraint of the k -th user is satisfied, i.e., $\sum_{n=0}^{N-1} a_{kn}^e \hat{r}_{kn}^e \leq R_k^b + R_k^e$ **or** all users have reached their upper bound constraint, i.e., $\mathcal{K} = \phi$.
3. **If** $\mathcal{K} = \phi$, then
 - 1) find the surplus subcarriers $\bar{n} = \arg \min_{n \in \mathcal{A}_k^e} \sum_{n=0}^{\bar{n}} a_{kn}^e \hat{r}_{kn}^e > R_k^b + R_k^e$.
 - 2) let the subcarriers carry no bits, i.e., set $a_{kn}^e = 0, \forall n \geq \bar{n}$.
4. **End If**
5. $k = k + 1$.

Until All users have been treated, i.e., $i \geq K$.

After the reallocation in Step 2, the subcarrier allocation of each user can satisfy the enhancement layer rate constraint of (12d) after the reallocation between layers to be introduced in Step 3.

4.2.2.3 Step 3: Reallocation Between Layers After subcarrier allocation between users, we allocation subcarriers within layers of each user. For each user, we simply reallocate the subcarriers originally allocated to the enhancement layer to the base layer, until the base layer is completed. The adjustment procedure is summarized as follows.

Algorithm 3: The procedure of reallocation between layers

Initialization: set user index $k = 0$.

Repeat

1. **Repeat**

1) Find subcarrier $\hat{n} = \min(\mathcal{A}_k^e)$, set $a_{k\hat{n}}^e = 0$, $a_{k\hat{n}}^b = 1$.

2) Add subcarrier \hat{n} to \mathcal{A}_k^b , remove subcarrier \hat{n} from \mathcal{A}_k^e .

Until The base layer requirement of the k -th user is satisfied, i.e., $\sum_{n=0}^{N-1} a_{kn}^b \hat{r}_{kn}^b \geq R_k^b$, **or** all subcarriers have been used up, i.e., $\mathcal{A}_k^e = \phi$.

2. $k = k + 1$

Until All the users have been treated, i.e., $k \geq K$.

4.3 Convergence and Complexity

We discuss the convergence of the proposed subcarrier allocation algorithm.

Proposition 2 *The proposed subcarrier allocation algorithm converges to a feasible solution.*

Proof Since the unconstrained problem of (13) always has an optimum solution using the greedy approach given by (14)–(16), it suffices to show that the iterative reallocation procedure given by Algorithm 1 and Algorithm 2 converges to a feasible solution. This is guaranteed by noting that the reallocation only take place when the already satisfied constraints of (12c) and (12d) are not violated, which ensures that the number of satisfied constraints increases monotonically in the subcarrier reallocation process. Moreover, we assume that the base layer rates of all users can be completed using the given channel resources, and allow surplus subcarriers that carry no bits. Therefore, the number of reallocation operations is finite. The proposed algorithm converges to a feasible solution, and divergence is avoided. \square

We next analyze the computation complexity of the proposed subcarrier allocation algorithm part-by-part. First, the unconstrained subcarrier allocation in 4.2.1 has the complexity of $O(N)$. That is because each subcarrier is allocated to the user with the largest value of the objective function. Next the Algorithm 1 in 4.2.2.1 has the complexity of $O(N)$ since at most it needs N times evaluation of c_{kn}^1 and $\sum_{n=0}^{N-1} a_{k,n}^e \hat{r}_{k,n}^b - \hat{r}_{k,n}^b$. The Algorithm 2 in 4.2.2.2 has the complexity of $O(NK)$. The reason is that each time the reallocation algorithm chooses a subcarrier-user pair. At most, there are NK times calculation of c_{kn}^2 . Finally, the complexity of the Algorithm 3 in 4.2.2.3 is $O(N)$, since the maximum number of the iterations is N . Overall, the complexity of the proposed algorithm is in polynomial time. Compared to the exponential-complexity exhaustive search, the proposed algorithm is suitable for practical implementations.

5 Simulation Results

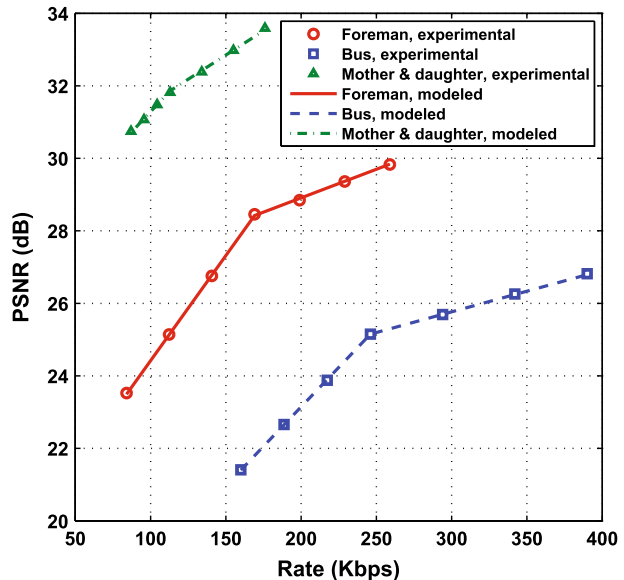
In this section, we provide extensive simulation results to demonstrate the effectiveness of the proposed resource allocation algorithm for scalable video transmission over OFDMA systems.

5.1 Simulation Setup

We have implemented a software testbed that simulates the end-to-end SVC bitstream transmission over the OFDMA system. The video sequences used in the experiments are downloaded from [26], and encoded according to the H.264 extended SVC standard using the JSVM software [27]. The sequences are coded at 30 frames per second. The spatial resolution is CIF (352×288). The GOP size is 8, and $QP = 47$. One base layer, and one enhancement layers are generated in the video encoder. Each frame of certain layer is further divided into 18 independently coded NAL units. A 32-bit cyclic redundancy check (CRC) is added to each NAL for error detection.

The basestation provides services for $K = 3$ video subscribers simultaneously, each of which has a video stream request. The NAL units from the same layer of each user are then concatenated, coded and modulated, mapped to subcarriers, and transmitted over the OFDM channel. The OFDMA system has 64 subcarriers, and the OFDM symbol interval is 0.4 ms. A nine-path Rayleigh fading channel with an exponentially decayed power profile is assumed between each user and the base station. The channel code used is the turbo code specified in 3GPP with the Max-Log-Map decoder [28]. The simulated coding rates set is $\mathcal{R} = \{1/3, 1/2, 3/5, 2/3, 3/4, 5/6\}$, and the length of the channel coding block is 640. The quadrature amplitude modulation (QAM) is used with constellations chosen from $\mathcal{S} = \{QPSK, 16 QAM, 64 QAM\}$. The UEP is considered by setting BLERs $P_{err}^b = 0.001$, $P_{err}^e = 0.01$ for the users' base layer and enhancement layer, respectively. At the receiver,

Fig. 2 The rate-PSNR measurements and models for different videos



we use perfect channel knowledge to demodulate the received signal, and then decode the video bitstreams. Whenever an uncorrectable bit is detected by the CRC, the corresponding NAL is dropped. If it is a base-layer NAL, the corresponding NAL of the nearest previously available frame is simply copied for error concealment. The associated additional information (header information) is assumed to be transmitted without errors. The JSVM software [27] is modified to support the above implementations. The reconstructed videos are compared with the original ones for PSNR calculation.

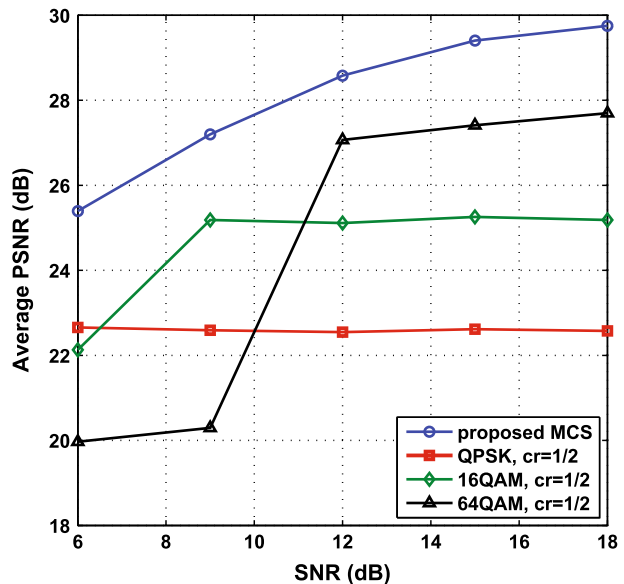
5.2 Rate-Quality Model Simulation

We first use the video sequences to verify the rate-quality model given in Sect. 2.2. The video sequences in our simulation are Foreman, Bus, and Mother and daughter. Both the PSNR-rate sample pairs and the model that we obtained are plotted in Fig. 2. It is seen that the model given by (2) matches the experimental results. It is also observed that bits from different quality layers contribute differently to the reconstructed video, which is indicated by the different slopes of the segments. Even with the same QP, the rate-quality relationships of videos differ from each other because of their content complexities and motion activities.

5.3 MCS Simulation

We compare the performance of the proposed adaptive modulation and coding scheme with the fixed modulation and coding schemes. In Fig. 3, the average PSNR of users is plotted with different SNRs. In fixed modulation and coding schemes, we simulate fixed modulation of QPSK, 16QAM, 64QAM with coding rate of 1/2 ($cr = 1/2$). It is seen that higher modulation types achieve better PSNR performance only in high SNR regions.

Fig. 3 The average PSNR performance comparisons



The proposed adaptive MCS scheme outperforms the fixed modulation and coding schemes in all the simulated SNR values because of better exploitation of the channel.

5.4 Subcarrier Allocation Simulation

We next show the performance of the proposed subcarrier allocation algorithm, and compare it with the conventional greedy method, the round-robin method, and the solution to the linear programming (LP) relaxation of (12).

In the greedy method, each subcarrier is allocated to the user who has the best channel condition, i.e., $\hat{k}_n = \operatorname{argmax}_k \hat{r}_{kn}, \forall n$. Such a greedy approach can maximize the sum data rate, but it does not consider fairness among users. On the other hand, the round-robin method assigns each subcarrier to each user in a round-robin fashion so that fairness among users is ensured, however the multiuser diversity is not exploited. In both the greedy and round-robin approaches, after allocation to users, the subcarriers are further allocated between quality layers of each user. The base layer is first allocated, followed by the enhancement layer, until all allocated subcarriers have been used up. As an upper bound on the performance of the subcarrier allocation, we also consider the LP relaxation of (12), where the indicator variables are allowed to take decimal numbers, i.e., $0 \leq a_{kn}^b \leq 1, 0 \leq a_{kn}^e \leq 1, \forall k, \forall n$. Since it allows fractional occupation of a single subcarrier by several users, the solution of the LP relaxation actually serves as the upper bound of the solution of (12).

The performances of the four algorithms are shown in Fig. 4, where the average PSNR of the users are plotted under different SNRs. It is seen that the proposed allocation algorithm outperforms the conventional greedy and round-robin methods by 1–5 dB in SNR. That's because the conventional methods do not consider the characteristics of the

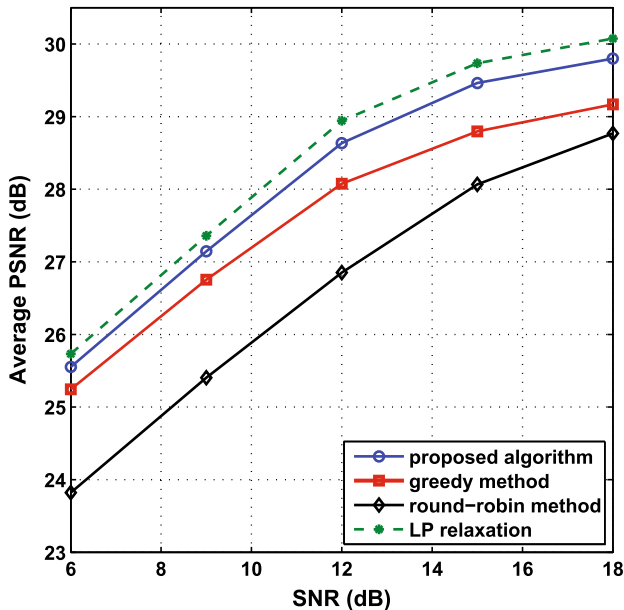


Fig. 4 The average PSNR performance comparisons

video sources, and improperly allocate subcarriers. It is also seen that the difference the proposed allocation algorithm and the LP relaxation is less than 0.5 dB in SNR.

In Fig. 5, we show a typical frame-by-frame PSNR performance of the proposed algorithm and the greedy method, which is measured at the transmitter, i.e., the distortion is only due to the allocation-dependent sub-stream extraction. The SNR is 15 dB. It is seen that under the simulated channel realization, the two algorithms behave similarly with Foreman, but quite differently with Bus and Mother and daughter. In Mother and daughter, the greedy method outperforms the proposed algorithm for the second half of the video sequence. This is because both methods fulfill the base layers' rates, and the greedy method assigns more subcarriers to the enhancement layer of the user. However, in Bus, the greedy method suffers from some violent fluctuation in PSNR, simply because the base layer's bits of the user are not completed, and some temporal frames are lost. Although the frame-copy error concealment is used in the decoder, there is severe degradation in the quality of the reconstructed video. Figure 6 shows sample frames of Bus in Fig. 5. It is seen that compared that of (b), the frame of (a) suffers only "decent" degradation in quality. Comparing (c) and (d), we note that the allocation of the greedy method loses the original Frame 49, and replaces it using the previous Frame 48 in the video decoder for error concealment. This severely degrades the PSNR performance and causes annoying "pause and rush" in the decoded video. Overall, the proposed subcarrier allocation algorithm achieves decent video quality for all users, and a higher average PSNR.

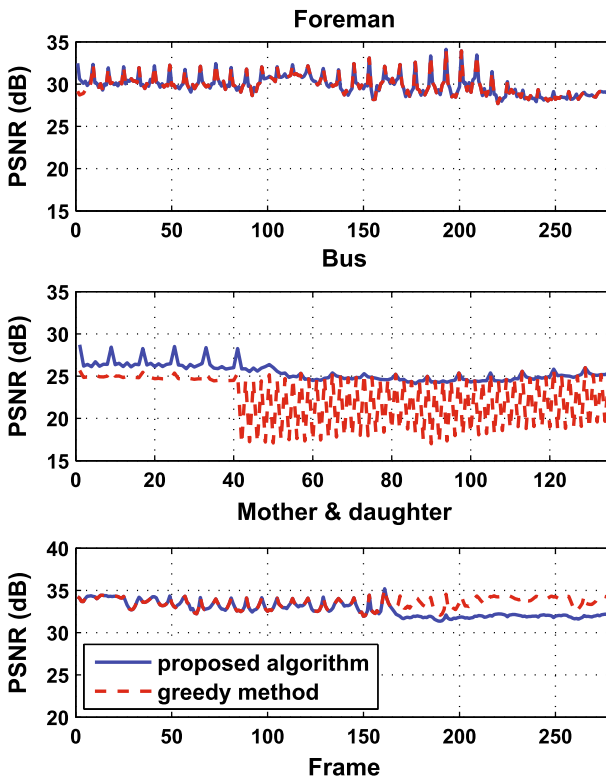


Fig. 5 Frame-by-frame PSNR performance comparison of the proposed algorithm and the greedy method



Fig. 6 Sample frames in the Bus sequence in Fig. 5: **a** Frame 32, proposed algorithm **b** Frame 32, greedy method **c** Frame 49, proposed algorithm **d** Frame 49, greedy method

5.5 UEP Simulation

Finally, we show the PSNR performance with different error protection strategies. In Fig. 7, we plot the frame-by-frame PSNR performance of the decoded videos at the receiver, i.e., the distortion is caused by the sub-stream extraction as well as the channel-induced error. The PSNR performance of proposed UEP scheme is compared with that of equal error protection (EEP) scheme. In the latter, BLERs $P_{\text{err}}^b = P_{\text{err}}^e = 0.01$ are set for both the base and enhancement layers. For fair comparison, the same amount of radio resource is used for both schemes in Fig. 7. Therefore, the EEP scheme can transmit a larger amount of the enhancement layer, but with a higher possibility of losing the base layer. It is seen that both schemes suffer from some occasional PSNR degradation caused by the random channel errors. However, overall, the UEP scheme outperforms the EEP scheme. A sample frame is also shown in Fig. 8. It is observed that the reconstructed frame with UEP has better quality than the one with EEP.

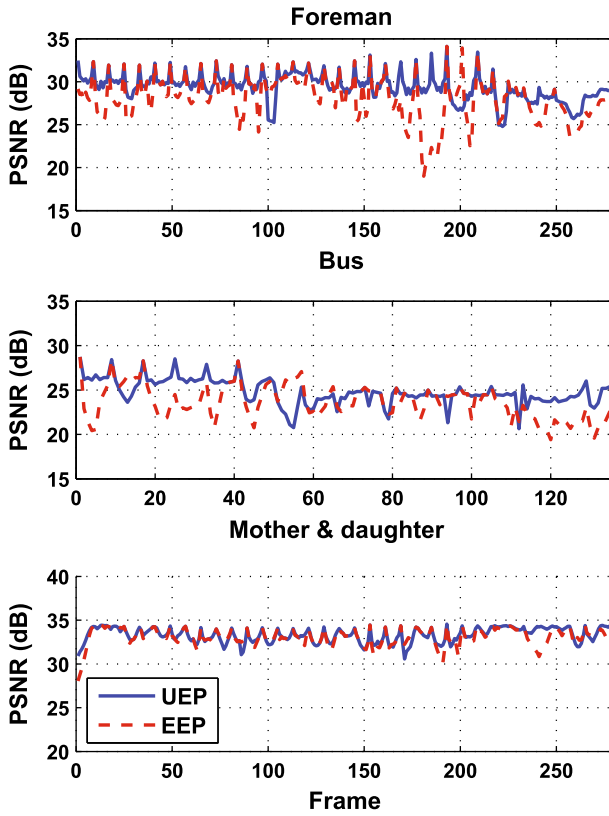


Fig. 7 Frame-by-frame PSNR performance comparison of UEP and EEP

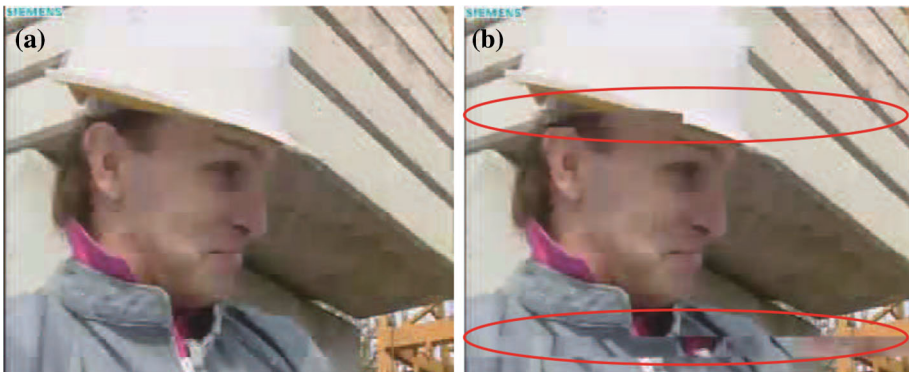


Fig. 8 Frame 167 in the Foreman sequence: **a** reconstructed with UEP **b** reconstructed with EEP

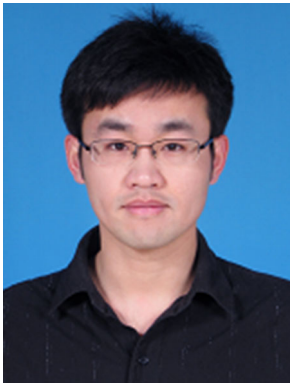
6 Conclusions

We have developed a resource allocation strategy for scalable H.264 videos over downlink OFDMA systems, taking into account the video characteristic, the channel state information, and the UEP requirement. Using a piece-wise linear PSNR-rate model for the videos, we have proposed an adaptive coding and modulation scheme for unequal error protection of different quality layers of the videos, and a subcarrier allocation algorithm to map different layers of different users' videos to the OFDMA subcarriers. Moreover, we have implemented an end-to-end simulation platform to show the effectiveness of the proposed scheme.

References

1. Costa, J. M. (2007). *More frequencies needed for mobiles*. No. 8: ITU News.
2. Sun, H., Vetro, A., & Xin, J. (2007). An overview of scalable video streaming. *Wireless Communications and Mobile Computing*, 7(2), 159–172.
3. Reichel, J., Schwarz, H., & Wien, M. (2005). Scalable working draft - working draft 1. *Joint video team (JVT), Doc. JVT-N020*, Hong Kong, China, 2005.
4. Schwarz, H., Marpe, D., & Wiegand, T. (2007). Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(9), 1103–1119.
5. Tillo, T., Baccaglioni, E., & Olmo, E. (2011). Unequal Protection of video data according to slice relevance. *IEEE Transactions on Image Processing*, 20(6), 1572–1582.
6. Nee, R. V. & Prasad, R. (2000). OFDM for wireless multimedia communications. *Artech House, Inc.*
7. Prasad, R. (2004). OFDM for wireless communications systems. *Artech House, Inc.*
8. Fraimidis, I. G., & Kotsopoulos, S. A. (2013). Queue-aware resource allocation for multi-cell OFDMA systems with QoS provisioning. *Wireless Personal Communications*, 71(4), 3033–3044.
9. Ashourian, Z., Salimian, R., & Nasab, H. M. (2013). A low complexity resource allocation method for OFDMA system based on channel gain. *Wireless Personal Communications*, 71(1), 519–529.
10. Tu, S. Y., Chen, K. C., & Prasad, R. (2009). Spectrum sensing of OFDMA systems for cognitive radio networks. *IEEE Transactions on Vehicular Technology*, 58(7), 3410–3425.
11. Xu, J., Shen, X., Mark, J. W., & Cai, J. (2008). Quasi-optimal channel assignment for real-time video in OFDM wireless systems. *IEEE Transactions on Wireless Communications*, 7(4), 1417–1427.
12. Yang, Z., & Zhao, Y. (2013). Scalable video multicast over multi-antenna OFDM systems. *Wireless Personal Communications*, 70(4), 1487–1504.
13. Jubran, M. K., Bansal, M., Kondi, L. P., & Grover, R. (2009). Accurate distortion estimation and optimal bandwidth allocation for scalable H.264 video transmission over MIMO systems. *IEEE Transactions on Image Processing*, 18(1), 106–116.
14. Kim, I.-M., & Kim, H.-M. (2002). A new resource allocation scheme based on a PSNR criterion for wireless video transmission to stationary receivers over gaussian channels. *IEEE Transactions on Wireless Communications*, 1(3), 393–401.
15. Xu, J., Hormis, R., & Wang, X. (2010). MIMO video broadcast via transmit-precoding and SNR-scalable video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(3), 456–466.
16. Su, G.-M., Han, Z., Wu, M., & Liu, K. J. R. (2006). A scalable multiuser framework for video over OFDM networks: Fairness and efficiency. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(10), 1217–1231.
17. Katsaggelos, A. K., Eisenberg, Y., Zhai, F., Berry, R., & Pappas, T. N. (2005). Advances in efficient resource allocation for packet-based real-time video transmission. *Proceedings of IEEE*, 93(1), 135–147.
18. Zhang, H., Zheng, Y., Khojastepour, M. A., & Rangarajan, S. (2010). Cross-layer optimization for streaming scalable video over fading wireless networks. *IEEE Journal on Selected Areas in Communications*, 28(3), 344–353.
19. Zhang, Y., & Letaief, K. B. (2004). Multiuser adaptive subcarrier-and-bit allocation with adaptive cell selection for OFDM systems. *IEEE Transactions on Wireless Communications*, 3(5), 1566–1575.

20. Chung, S. T., & Goldsmith, A. J. (2001). Degrees of freedom in adaptive modulation: A unified view. *IEEE Transactions on Communications*, 49(9), 1561–1571.
21. Yang, K., Prasad, N., & Wang, X. (2009). An auction approach to resource allocation in uplink OFDMA systems. *IEEE Transactions on Signal Processing*, 57(11), 4482–4496.
22. Ungerboeck, G. (1982). Channel coding with multilevel/phase signals. *IEEE Transactions on Information Theory*, 28(1), 55–67.
23. Baccaglioni, E., Tillo, T., & Olmo, G. (2008). Slice sorting for unequal loss protection of video streams. *IEEE Signal Processing Letters*, 15, 581–584.
24. Tsai, Y.-R., & Chen, Y.-C. (2011). Multilevel coupling modulation for multi-resolution multimedia broadcast/ multicast service in OFDM systems. *IEEE Transactions on Communications*, 59(1), 141–150.
25. Reklaitis, G. V., Ravindran, A., & Ragsdell, K. M. (1983). *Engineering optimization, methods, and applications*. New York: Wiley.
26. *Video test sequences*. <http://trace.eas.asu.edu/yuv/>.
27. http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software/.
28. Multiplexing and channel coding (FDD). (2006). *3GPP TS 25.212*.



Zan Yang received the B.S. degree in Electronics Engineering from Nankai University, Tianjin, P. R. China, in 2007. He received the Ph.D. degree in communications and information systems from Peking University, Beijing, P. R. China, in 2013. During 2011, he was a visiting Ph.D. student at Columbia University, New York, USA. Currently, he is an engineer at Transport Planning and Research Institute, The Ministry of Transport, Beijing, P. R. China. His current research interests include vehicular communications, scalable video coding, resource allocation, and wireless sensor networks.



Yuping Zhao received the B.S. and M.S. degrees in Electrical Engineering from Northern Jiaotong University, Beijing, P. R. China, in 1983 and 1986, respectively. She received the Ph.D. and Doctor of Science degrees in wireless communications from Helsinki University of Technology, Helsinki, Finland, in 1997 and 1999, respectively. She was a System Engineer for telecommunication companies in China and Japan. She worked as a research engineer at the Helsinki University of Technology, Helsinki, Finland, and at the Nokia Research Center in the field of radio resource management for wireless mobile communication networks. Currently, she is a Professor in the State Key Laboratory of Advanced Optical Communication Systems and Networks, School of Electronics Engineering and Computer Science, Peking University, Beijing, P. R. China. Her research interests include the areas of wireless communications and corresponding signal processing, especially for OFDM, UWB and MIMO systems, cooperative networks, cognitive radio, and wireless sensor networks.