EURASIP Journal on
Wireless Communications and Networking
a SpringerOpen Journal

RESEARCH   **Open Access**

A resource allocation algorithm for OFDM-based cellular system serving unicast and multicast services

Daeyeon Kim[1*], Takeo Fujii[1] and Kyesan Lee[2]

Abstract

In this article, we propose a resource allocation algorithm for orthogonal frequency division multiplexing-based cellular system that serves users each of which requests a regular service or a multimedia broadcast multicast service (MBMS). We assume that the MBMSs are considered for video services such as mobile TV. In order to efficiently serve the users with heterogeneous channel conditions, we employed channel-adaptive video streams by using scalable video coding (SVC), where a realistic multicast configuration is considered for multicasting the SVC streams. Therefore, we designed the cellular system to serve constant and flexible bit rates for the regular services and the MBMSs, respectively. Thus, we consider three objectives when allocating the bandwidth resources, where total transmit power is regulated. First, the bit rate of each regular service has to be guaranteed. Second, the minimum bit rate of each MBMS for achieving the minimum video quality has to be guaranteed. Third, the average video quality of the users requesting MBMSs has to be maximized. In order to achieve the objectives, we propose to quantify the service quality of each service, and then find the resource allocation that maximizes the total service quality. By using the computer simulation results, we could verify that the three objectives can efficiently be achieved. Also, we could verify that the performance of the proposed algorithm is superior to that of the previous algorithm designed to maximize the total bit rate while guaranteeing the minimum bit rates.

Keywords: OFDM, Resource allocation, Multicast, SVC

1. Introduction

In order to serve the high-speed wireless services, orthogonal frequency division multiplexing (OFDM) has been considered, which can be used to mitigate the inter-symbol interference and dynamically allocate the given bandwidth, where an OFDM channel is configured with multiple subchannels. By allocating the subchannels according to the channel conditions of each user to serve, the service quality of the users with the heterogeneous channel conditions can be more efficiently achieved [1-7]. Such utilizations of the OFDM also can be considered for multicast services such as multimedia broadcast multicast service (MBMS) which is defined in long-term evolution (LTE) standards [8]. However, in order to robustly multicast a common service to multiple users, transmission

parameters such as transmit power or modulation and coding scheme (MCS) have to be determined according to the user with the worst channel condition, and therefore the heterogeneous behavior of the channels degrades the multicast performance. Therefore, in order to improve the multicast performance by using the OFDM, adaptive configurations of users to multicast can be considered for each subchannel as described in [9-11].

In this article, we consider a cellular system for serving regular services and MBMSs. In order to efficiently serve the users with heterogeneous channel conditions, we can consider bit rate-flexible bit streams which can be decoded only with a part of the whole bit stream. As an example of such bit streams, scalable video coding (SVC) stream can be considered [12]. As the SVC stream is configured with multiple video layers, we can serve the users more adaptively to the channel conditions by adjusting the transmission of the video layers. Thus, we consider the bit rate-flexible SVC streams for MBMSs,

* Correspondence: daeyeon@awcc.uec.ac.jp
[1]Advanced Wireless Communication Research Center, The University of Electro-communications, Tokyo, Japan
Full list of author information is available at the end of the article

where we assume that the bit rate for serving each regular service is constant.

Therefore, we propose a resource allocation algorithm for achieving three objectives: First, for each user that requires a regular service, the system must guarantee the constant bit rate required for the regular service. Second, for each user that requests MBMSs, the system must guarantee the minimum bit rate required for achieving the minimum video quality. Third, the system must maximize the average video quality of the users requesting the MBMSs. Although the resource allocation algorithms in [13,14] are also designed to serve the multicast services over the OFDM channels by using the SVC streams, we found that there are defects in three aspects: the multicast configuration, the system objective, and the optimality of the resource allocations. Therefore, we design the proposed algorithm to overcome the defects in the previous algorithms.

This article is configured as follows. In Section 2, we review the SVC and the related resource allocation algorithms, and then address the defects in the previous algorithms. In Section 3, we present the configuration of the cellular system assumed in this article. After that, we introduce the proposed resource allocation algorithm in Section 4. In Section 5, we provide some simulation results to verify the effectiveness of the proposed algorithm, and then we conclude this article in Section 6.

2. Related works
2.1. SVC
The SVC is designed to provide adaptation in required bit rate and decoded video quality (in this article, we use peak-signal-to-noise ratio (PSNR) for measuring the video quality, which is commonly used in the studies such as [12,15]). For example, even if a part of the compressed video stream is discarded, the remaining part can be uncompressed and reconstructed to the video pictures (also, referred to as frames in this article). Each of the truncatable parts is referred to as video layers, where the layers that represent basic information and details are referred to as base layer (BL) and enhancement layer (EL), respectively. Accordingly, by adjusting the transmission of the video layers, the quality of the video content can be adapted to the channel conditions.

There are three types of video scalabilities referred to as spatial, temporal, and quality scalabilities. By considering the spatial and the temporal scalabilities, the SVC bit stream can be decoded into multiple resolutions (sizes of frames) and frame rates (the number of frames per second). In addition, for each combination of the available resolutions and the frame rates, quality scalability can be considered to provide multiple bit rate options. In this article, we focus on the quality scalability. Figure 1 shows an example of hierarchical-B coding structure [12] of a group

of pictures (GOP) configured with nine frames and three quality layers (QLs). Each box in Figure 1 denotes each SVC packet, where the two indicators in each box denote the frame and the QL indices, respectively. The frames each of which is configured with the packets denoted by the same frame index can be coded by using the temporal prediction of the hierarchical-B structure in order to improve the coding efficiency and to enable the temporal scalability [12]. Each QL is configured with the packets denoted by the same QL index, where the QL with QL index 1 and each of the other QLs are BL and EL, respectively.

By using the three QLs, we can obtain three PSNR options, i.e., we can obtain the lowest PSNR by decoding the BL and improve the PSNR by additively decoding the ELs in ascending order of the QL index. This quality scalability scheme is referred to as coarse grain scalability (CGS). Therefore, the QLs also referred to as CGS layer in the rest of this article. In addition, the standard SVC [12] allows decoding each packet of each EL independently from other packets in the same EL in order to achieve more bit rate options. This quality scalability scheme is referred to as medium grain scalability (MGS). In the rest of this article, the packets in the ELs are referred to as MGS packets. In order to achieve more efficient bit rate adaptabilities in terms of bit rate–PSNR performance, priorities in transmitting order can be considered for the MGS packets according to the bit length and the contribution to the PSNR of each MGS packet [15]. For example, 19 bit rate options can be achieved by the coding structure in Figure 1. Figure 2 shows an example of the bit rate options coded by joint scalable video model (JSVM) [16], where first 9 frames of test videos "Football" and "Crew", common intermediate format (CIF), and 30 frames per second (fps), are used [17]. The QLs are coded with the quantization parameters 40, 34, and 28 [12], respectively. For each test video, 3 performance points of CGS and additional 16 points by considering MGS can be achieved. As shown in Figure 2, the bit rate–PSNR behavior of MGS is almost linear within a single CGS layer as analyzed in [18]. When allocating the resources, this behavior can be used for improving the received PSNR performance as discussed in Section 4.1.

2.2. Resource allocation algorithms for OFDM-based multicast service
In [9-11,13,14], resource allocation algorithms for OFDM-based multicast services are considered. Especially for [13,14], the video services using the SVC are considered, where the minimum bit rates for transmitting the BLs, i.e., the essential parts of the SVC streams, has to be guaranteed to all the users. More specifically, the algorithms are designed to find resource allocation that maximizes the total bit rate while guaranteeing the minimum bit rates, where the total power is constrained below a predefined
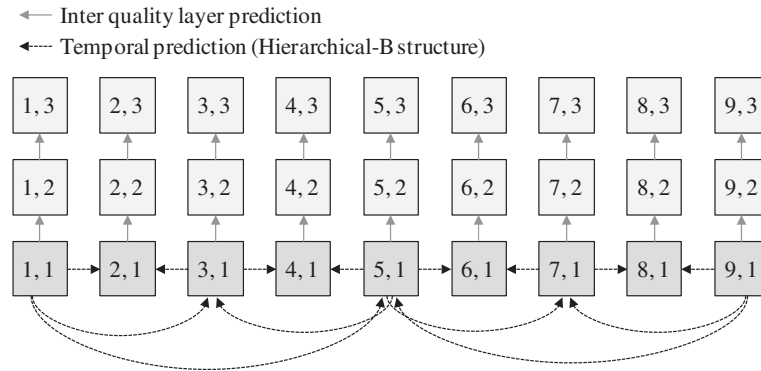
Figure 1 Coding structure of a GOP configured with nine frames and three QLs, where the two indicators in each box denote the frame and the QL indices, respectively.

power level P_{MAX}, i.e., allowed interference to other surrounding wireless systems.

The objective in [13,14] can be denoted as

$$\max\left(\sum_{u\in U}R_u\right)\text{ subject to }R_u\geq R_{\vec{s}(u)}^{\mathrm{MIN}}\text{and}\sum_{c\in C}p_c\leq P_{\mathrm{MAX}},\tag{1}$$

where the terms R_u and R_s^{MIN}, and $\vec{s}(u)$ denote the bit rate allocated to user u, the minimum bit rate of service s, and the index of service that is requested by user u, respectively. The terms $C = \{c|1 \leq c \leq C\}$ and p_c in (1) denote the transmit power allocated to subchannel c and the subchannel index set, where C is the number of the subchannels.

The previous algorithms first split the subchannels into two sets, a set to guarantee the minimum bit rates and the other set to maximize the total bit rate (in the rest of this article, we denote the former and the latter subchannel sets by subchannel sets 1 and 2, respectively). Next, the previous algorithms equally distribute the P_{MAX} over the C subchannels, i.e., the p_c is initialized to P_{MAX}/C. Then, the algorithms quantify how much each subchannel is suitable to be used as a subchannel of subchannel set 1 than that of subchannel set 2. In the articles, the quantified value is referred to as "suitability". After that, the subchannels are allocated to subchannel set 1 in descending order of the suitability until the minimum bit rates are fulfilled, where the rest subchannels are allocated to subchannel set 2. Then, the total power minimization while guaranteeing the minimum bit rates is accomplished for subchannel set 1. After that, the total bit rate maximization is accomplished for subchannel set 2 while suppressing the sum of the total power allocated to both subchannel sets below the P_{MAX}.

2.3. Limitations in related works

Although the previous algorithms improve the service qualities of the OFDM-based cellular systems designed to multicast the SVC streams, there are defects in three aspects which can be summarized as follows.

1. System objective
2. Multicast configuration
3. Optimality in resource allocation

First, the algorithms are designed to maximize the total bit rate which is not directly related to users' satisfaction, i.e., PSNR. Although the algorithms are designed to avoid the worst cases that any user is not available of the BL as far as possible, the given resources still can be wasted. For example, a large part of the resources can be used to allocate the bit rates exceeding the maximum bit rates (required for transmitting all the video packets) to
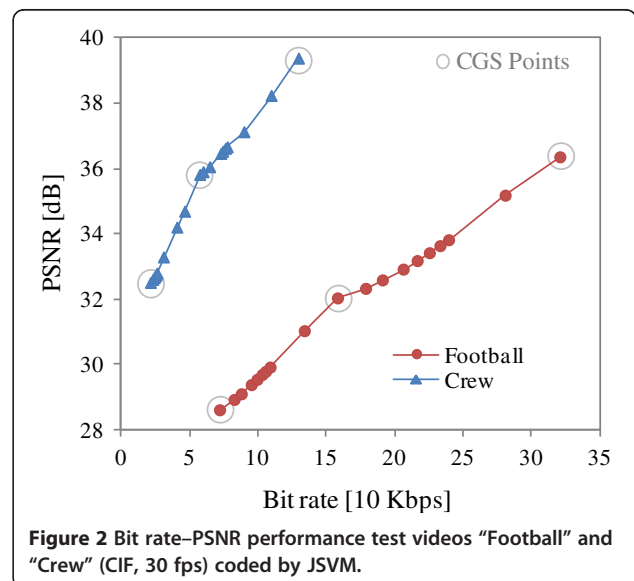

Figure 2 Bit rate–PSNR performance test videos "Football" and "Crew" (CIF, 30 fps) coded by JSVM.
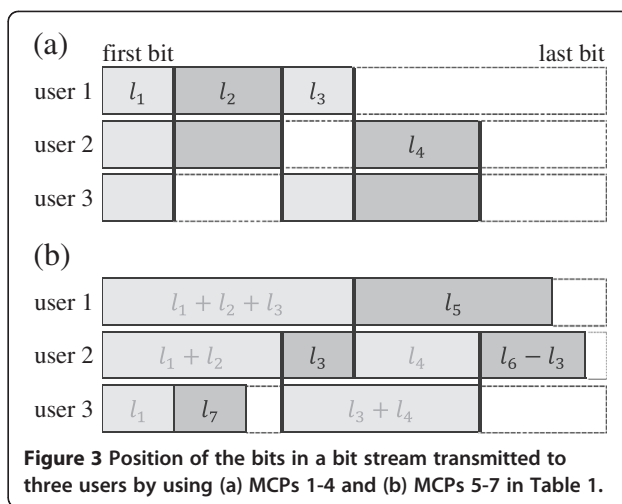
some of the users, while allocating only the minimum bit rates to others. Other than these algorithms, algorithms in [4-7] are designed to maximize average received PSNR for unicast services, and therefore such inefficient resource allocation can be avoided. Hence, such objectives need to be considered also for the multicast services in order to improve the PSNR performance.

Second, multicast configuration in [13,14] can have the users difficult to decode the received bits properly due to following reasons: In order to robustly multicast a common service to multiple users, transmission parameters, e.g., transmit power and MCS, have to be decided according to the worst channel quality of the users. Thus, in [13,14], adaptive multicast patterns (MCPs) for each subchannel according to the subchannel qualities are considered in order to improve the worst subchannel quality. Table 1 shows an example of seven MCPs for three users, i.e., all the possible MCPs of the users. According to the seven MCPs of Table 1, the bits of the common service can be transmitted to each of the three users as shown in Figure 3, where l_k is referred to as the total bit rate transmitted by using MCP k. In Figure 3a, the first l_1 bits are transmitted to the three users of Table 1. Then, the next l_2 bits are transmitted to users 2 and 3 as described in Figure 3a. In this way, the subsequent l_3 and l_4 bits can be transmitted as shown in Figure 3a. Then, by using the subchannels allocated to MCP 6, the l_3 bits after the first $l_1 + l_2$ bits can be transmitted to user 2 as shown in Figure 3b. After that, the rest $l_6 - l_3$ can be transmitted after the first $l_1 + l_2 + l_3 + l_4$ bits. In this case, user 2 is difficult to understand the position of the received bits unless the user is informed that the bits of MCP 6 starts from $(l_1 + l_2 + 1)$th bit and restarts from $(l_1 + l_2 + 1)$th bit. In addition, in case of user 3, because the l_7 is smaller than the l_2 user 3 cannot receive the bits properly. Therefore, the configuration of the bits has to be considered when allocating the subchannels to the MCPs. These problems become more complex as we consider more users, which are not discussed in [13,14].

Third, the resource allocation determined by the previous algorithms can be inefficient in maximizing the average PSNR. That is, the algorithms configure the two subchannel sets by assuming that the total power is equally distributed. Accordingly, it is not obvious that the subchannel allocation is also the best allocation scheme for the resulting power allocation. Furthermore,

Figure 3 Position of the bits in a bit stream transmitted to three users by using (a) MCPs 1-4 and (b) MCPs 5-7 in Table 1.

in [13,14], only the MCP including all the users, e.g., MCP 1 in Table 1, are allowed to be used for subchannel set 1. That is, the algorithms exclude the cases that the remaining MCPs are used to guarantee the minimum bit rates so improve the performance. For example, MCP 4 together with MCP 5 in Table 1 can also be used to guarantee the minimum bit rates.

Therefore, we considered the aforementioned defects when designing the resource allocation algorithm which is introduced in Section 4.

3. Proposed system

In this article, we consider an OFDM-based cellular system that serves users each of who subscribes one of the regular services and the MBMSs, where downlink channel is focused. We assume that a constant bit rate is required for each regular service requested by only one user (i.e., unicast). On the other hand, we employ bit rate-flexible SVC streams for MBMS, and therefore multiple bit rates can be considered for each MBMS which can be requested by multiple users (i.e., multicast). We assume that the BS transmits services $S = \{s | 1 \leq s \leq S\}$ (each of which is either of the regular service or the MBMS) to users $U = \{u | 1 \leq u \leq U\}$, where S and U are the numbers of the services and the users, respectively. Therefore, the U users can be divided into S groups according to the service requests, each of which is referred to as service group in the rest of this article.

We define U_s^{srv} the indices of the users of service group s, i.e., the users requesting service s. As each user requests one of the services, the service groups are configured exclusively. That is, we can denote that

$$\bigcap_{s \in S} U_s^{srv} = \phi, \quad \text{where } U = \bigcup_{s \in S} U_s^{srv}. \tag{2}$$

The BS transmits the requested services over OFDM subchannels C, where we assume that subchannel quality

Table 1 MCPs which can be considered for three users

User	MCPs						
	1	2	3	4	5	6	7
1	O	O	O	–	O	–	–
2	O	O	–	O	–	O	–
3	O	–	O	O	–	–	O

of user u over subchannel c, denoted as $h_{c,u}$, is exponentially distributed. We assume that the average subchannel quality (ACQ) of user u is determined by path loss which results from distance between BS and each user [19]. Also, we assume that perfect channel state information (i.e., the $h_{c,u}$'s) is transmitted from the users to the BS via the uplink channel. As discussed in Section 2, in case of the MBMSs, the transmission parameters, e.g., transmit power and MCS, have to be determined according to the lowest channel quality. Therefore, we consider adaptive transmit power, MCS, and MCP for each OFDM subchannel so each subchannel is efficiently utilized. In order to improve the service quality, we consider three objectives as follows.

1. Guarantee the constant bit rates required for serving the regular services.
2. Guarantee the minimum bit rates required for the lowest PSNRs of the MBMSs.
3. Maximize the average PSNR of the MBMS users.

In the following section, we introduce a resource allocation algorithm for realizing the three objectives above.
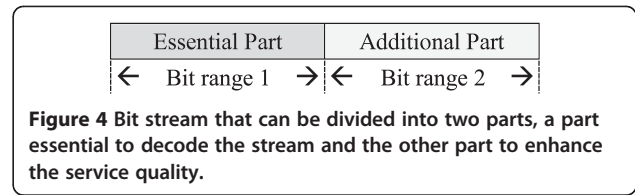
4. Proposed algorithm

In order to achieve the objectives stated in Section 3, we propose to weight the bits of the services according to the bit rate requirements of the services, and then maximize the weighted sum of the bits. The weight is referred to as bit value in the rest of this article. That is, we consider significantly higher bit values for the bits that have to be preferentially transmitted (i.e., the bits in the regular services or the BLs of the MBMSs) than bit values for the bits in the ELs of the MBMSs so the former bits are guaranteed to be transmitted. Also, we propose to set the bit values for the latter bits according to the bit rate–PSNR behaviors of the MGS packets (see Figure 2) so the average PSNR can be maximized. Then, the proposed resource allocation algorithm maximizes the total service quality, where each service quality is quantified as the total bit value (TBV) of each service. The resource allocator operates on the basis of convex optimization which requires the objective function to be concave.

In this section, we first define the bit values and the service qualities. Then, we introduce a realistic multicast configuration that is considered for avoiding the problems described in Figure 3. After that, we present the resource allocation algorithm for maximizing the total service quality.

4.1. Concave approximation of service quality

In this section, we discuss how to quantify the bit value and the service quality in order to achieve the multiple objectives introduced in Section 3. Figure 4 shows an example of a bit stream that can be divided into two parts,

Figure 4 Bit stream that can be divided into two parts, a part essential to decode the stream and the other part to enhance the service quality.

the part that is essential to decode the stream (e.g., the BL of an SVC stream) and the additional part that is not essential but can be used to enhance the service quality (e.g., the EL of the SVC stream). If we assume that the bit value of the first part is sufficiently larger than that of the second part, and we desire to maximize the sum of the TBVs received by all the users requesting the bit stream, we need to preferentially transmit the first part of the bit stream to all the users. In the rest of this article, each part configured with the bits of the same bit value is referred to as bit range. For service s, we denote the last bit of bit range i and the bit value of the range as $X_{s,i}$ and $\beta_{s,i}$, respectively.

As the resource allocation is designed on the basis of convex optimization, the service quality has to be concave function which has zero or negative second derivation. Therefore, we define continuous bit value $y(s,x)$ for the continuous bit length x, i.e., continuous version of the number of bits, where $y(s,x)$ is $\beta_{s,i}$ when the x is bounded by bit range i of service s. As the service quality is quantified as the TBV, the continuous service quality, which is denoted by $Y(s,x)$, can be obtained by

$$Y(s,x) = \int_0^x y(s,w)dw \qquad (3)$$
$$= \sum_{i=1}^{N_s} \int_{X_{s,i-1}}^{X_{s,i}} \beta_{s,i} \zeta(w,x)dw,$$

where N_s is the number of bit ranges and $X_{s,0}$ is 0 for any s, i.e., any service, respectively. The $\zeta(w,x)$ in (3) is decreasing step function, which is defined as

$$\zeta(x_1,x_2) = \begin{cases} 1, & 0 \leq x_1 \leq x_2. \\ 0, & x_2 < x_1. \end{cases} \qquad (4)$$

By defining cumulative step function

$$Z(x_1,x_2) = \int_0^{x_1} \zeta(w,x_2)dw, \qquad (5)$$

the continuous service quality in (3) can be written as

$$Y(s,x) = \sum_{i=1}^{N_s-1} \left(\beta_{s,i} - \beta_{s,i+1}\right) Z\left(x, X_{s,i}\right)$$
$$+ \beta_{s,N_s} Z\left(x, X_{s,N_s}\right). \qquad (6)$$

From (6), we can see that the continuous service quality is weighted sum of the cumulative step functions. Therefore, if the function $Z(x_1, x_2)$ is concave, the service quality is concave. However, as the derivative of $Z(x_1, x_2)$, i.e., the step function $\zeta(x_1, x_2)$, is not continuous when x_1 is x_2, the function $Z(x_1, x_2)$ is not concave when the x_1 is x_2.

Therefore, we define an alternative function $\hat{\zeta}(x_1, x_2)$ as

$$
\hat{\zeta}(x_1, x_2) = \begin{cases} 1 - \dfrac{1}{1+\tau}\left(\dfrac{x_1}{x_2}\right)^\xi, & 0 < x_1 \le x_2 \\[2ex] \dfrac{\tau}{1+\tau}\left(1 - \dfrac{x_1 - x_2}{\tau x_2}\right)^\xi, & x_2 < x_1 \end{cases}
$$
(7)

which is continuous approximation of the function $\zeta(x_1, x_2)$. The shape of the curve can be controlled by adjusting the terms ξ and τ in (7) as described in Figure 5. In the rest of this article, we denote the curve with $\xi = \dot{\xi}$ and $\tau = \dot{\tau}$ as "Curve $(\dot{\xi}, \dot{\tau})$". As shown in Figure 5, the approximation curve become closer to the step function as we consider larger ξ or smaller τ. By plugging $\hat{Z}(x_1, x_2)$ which is the cumulative $\hat{\zeta}(x_1, x_2)$ into (6), the concave approximation of the service quality can be obtained by

$$
\hat{Y}(s, x) = \sum_{i=1}^{N_s - 1}\left(\beta_{s,i} - \beta_{s,i+1}\right)\hat{Z}\left(x, X_{s,i}\right) \\ + \beta_{s,N_s}\hat{Z}\left(x, X_{s,N_s}\right)
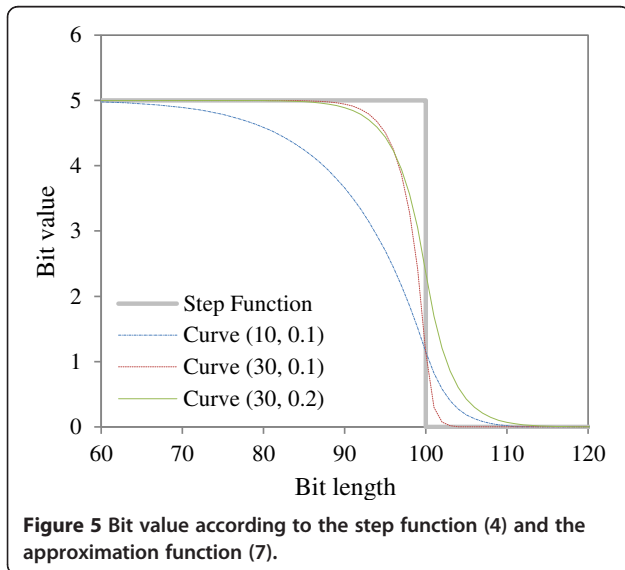$$
(8)

Figure 5 Bit value according to the step function (4) and the approximation function (7).

Because the bit value is derivative of the service quality, we can obtain the continuous approximation of the bit value as

$$
\hat{y}(s, x) = \frac{\partial \hat{Y}(s, x)}{\partial x} = \sum_{i=1}^{N_s - 1}\left(\beta_{s,i} - \beta_{s,i+1}\right)\hat{\zeta}\left(x, X_{s,i}\right) \\ + \beta_{s,N_s}\hat{\zeta}\left(x, X_{s,N_s}\right).
$$
(9)

By using the curves of (8) and (9), we can define the concave service quality and the continuous bit value for the regular services and MBMSs. As only one bit rate requirement is considered for the regular services, the bit value is defined as
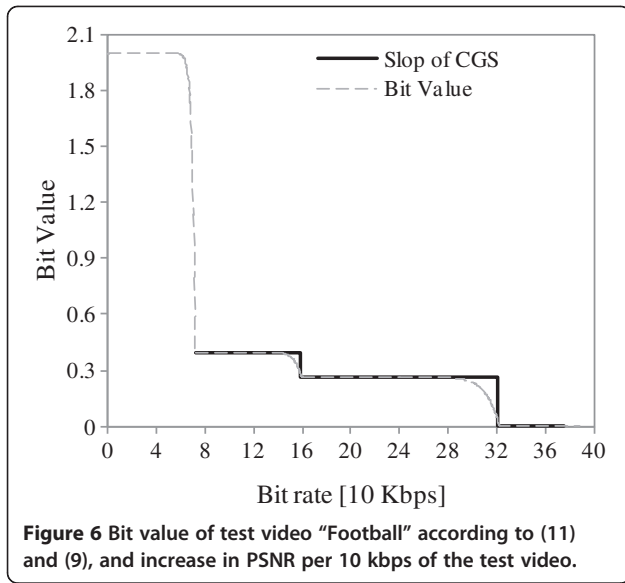
$$
\beta_{s,1} = \beta_0
$$
(10)

where the β_0 is a predefined bit value for guaranteeing the essential bits, i.e., the bit length required for serving each regular service. On the other hand, for the MBMSs, the bit value is defined as

$$
\beta_{s,i} = \begin{cases} \beta_0, & i = 1, \\ v_{s,i}, & 1 < i \le N_{s'} \end{cases}
$$
(11)

where the N_s in this case is the same as the number of the CGS layers (see Figure 2). Accordingly, the $\beta_{s,i}$ is the bit value of ith CGS layer. The $v_{s,i}$ in (11) is the slop between $(i-1)$th and ith CGS points, which can be obtained by

$$
v_{s,i} = \frac{V_{s,i} - V_{s,i-1}}{X_{s,i} - X_{s,i-1}},
$$
(12)

where the $V_{s,i}$ denotes the PSNR of ith CGS point of service s. We want to remind that the bit rate–PSNR behavior of MGS is almost linear within a single CGS layer as discussed in Section 2.1, and therefore the slop of the PSNR by additively decoding each MGS packets within ith CGS layer is almost the same as the $v_{s,i}$. By setting β_0 as significantly larger than any possible value of the $v_{s,i}$'s and by maximizing the total service quality, we can preferentially guarantee the required bit rate of the regular services and the minimum bit rate of the MBMSs, where the resource allocation algorithm that maximizes the total service quality is described in Section 4.3. Figure 6 shows an example of the slops of the CGS points of "Football" in Figure 2 (increase of the PSNR per 10 kbps) and the bit values according to (9) and (11), where the β_0, the ξ, and τ are set to 2, 30, and 0.1, respectively. The horizontal axis of Figure 6 is the bit rate $R = x/T$, where the T is deadline which is considered for the real-time service. In order to support 30 fps, the nine frames in the GOP as shown in Figure 1 have to be transmitted in 0.3 s, and therefore the T is 0.3. In Figure 6, as the PSNR is not defined[a] when the bit rate is lower than the

Figure 6 Bit value of test video "Football" according to (11) and (9), and increase in PSNR per 10 kbps of the test video.

minimum bit rate (approximately 72 kbps), the slope is denoted only when the bit rates higher than the minimum.

4.2. Multicast configuration

In order to improve the total service quality by using the various MCPs for a large number of users, we propose to divide users U into Q exclusive user groups according to the channel quality of each user, i.e., the ACQ, each of which is referred to as channel quality group. As the channel quality groups are configured exclusively, it can be denoted that

$$\bigcap_{q\in Q} U_q^{qlt} = \phi, \text{where} U = \bigcap_{q\in Q} U_q^{qlt}. \qquad (13)$$

where the $Q = \{q|1 \le q \le Q\}$ is channel quality index set and the U_q^{qlt} is the index set of the users with quality index q. That is, one of the channel quality indices is given to each user. In this article, we configure the channel quality groups so the number of users in each group is similar to other groups. For example, if we have five users to be divided into two channel quality groups, we group the three users with the worst ACQs to the U_1^{qlt} and the remaining users to the U_2^{qlt}. Accordingly, the highest ACQ of users U_q^{qlt} is lower than the lowest ACQ of users U_p^{qlt} if q is smaller than p. Accordingly, we can define $S \times Q$ groups $U_{s,q} = U_s^{srv} \cap U_q^{qlt}$, each of which is referred to as component group (CG) in the rest of this study.

We consider that each MCP is configured with a number of CGs. Table 2 shows an example of MCPs for four CGs. We want to note that if each CG is configured with only one user, MCPs 4, 5, 6, and 7 are exactly the same as unicasts. In order to avoid the improper cases described in Figure 3b, we propose to hierarchically configure the MCPs when the MCP is to be configured with multiple

Table 2 Proposed MCPs for four CGs

Users	MCPs						
	1	2	3	4	5	6	7
$U_{3,1}$	O	–	–	O	–	–	–
$U_{3,2}$	O	O	–	–	O	–	–
$U_{3,3}$	O	O	O	–	–	O	–
$U_{3,4}$	O	O	O	–	–	–	O

CGs, e.g., MCPs 1, 2, and 3 of Table 2. That is, if a CG is included in an MCP with multiple CGs, then all the CGs with higher quality indices have to be included in the MCP. Figure 7 shows an example that how the bits can be allocated to the four CGs by considering the MCPs in Table 2. Figure 7a describes the bits that can be transmitted by MCPs 1, 2, and 3. First, by using MCP 1, the first l_1 bits can be transmitted to all the CGs. Then, by using MCP 2, the next l_2 bits can be transmitted to the CGs $U_{s,2}$, $U_{s,3}$, and $U_{s,4}$. After that, by using MCP 3, the next l_3 bits can be transmitted to the CGs $U_{s,3}$ and $U_{s,4}$. Then, as described in Figure 7b, the subsequent bits can be transmitted to each CG by using the remaining MCPs. Therefore, by informing the users the MCP configurations as Table 2 and the index of MCP allocated to each subchannel, each user can decode the received bits properly, because the improper cases as users 2 and 3 in Figure 3b do not occur. This configuration of MCPs can also be considered for the multiple services. Table 3 shows an example of the MCP configuration of the two MBMSs (services 1 and 2) and two regular services (services 3 and 4). The proposed algorithm first configures the CGs according to the subchannel qualities, and then configures the MCPs according to the user request as shown in
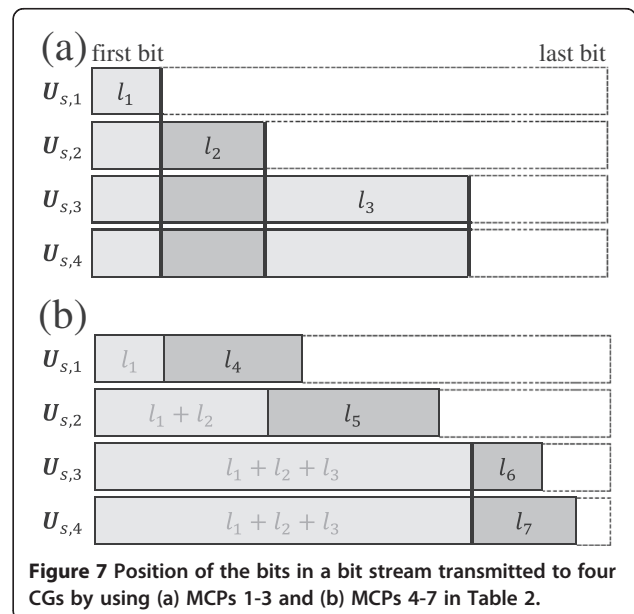

Figure 7 Position of the bits in a bit stream transmitted to four CGs by using (a) MCPs 1-3 and (b) MCPs 4-7 in Table 2.

Table 3 Proposed MCPs for seven CGs requesting two regular services and two MBMSs

Users	MCPs									
	1	2	3	4	5	6	7	8	9	10
U_{11}	O	–	O	–	–	–	–	–	–	–
U_{12}	O	O	–	O	–	–	–	–	–	–
U_{13}	O	O	–	–	O	–	–	–	–	–
$U_{2,1}$	–	–	–	–	–	O	O	–	–	–
$U_{2,2}$	–	–	–	–	–	O	–	O	–	–
$U_{3,1}$	–	–	–	–	–	–	–	–	O	–
$U_{4,1}$	–	–	–	–	–	–	–	–	–	O

Table 3, where only one MCP with one CG configured with only one user is considered for the regular services, e.g., CGs $U_{3,1}$ and $U_{4,1}$. After that, the subchannels are allocated to one of the MCPs so the system objectives described in Section 3 are achieved. The transmit power and the MCS are determined for each subchannel so the bits are robustly transmitted to all the users of each MCP allocated to each subchannel.

4.3. Resource allocation

The resource allocation algorithm allocates each subchannel to one of the MCPs and determines the transmit power and the MCS for each subchannel so the total service quality is maximized. As a result, the requested bit rate for each regular service and the minimum bit rate for each MBMS can be guaranteed and the average PSNR of the MBMS users can be maximized. As introduced in Section 4.1, we propose to quantify the service quality in order to achieve the objectives, where the service quality can be written in function of bit rate as

$$F(s,R) = \hat{Y}(s,RT) \tag{14}$$

and

$$\frac{\partial}{\partial R}F(s,R) = f(s,R) = T\hat{y}(s,RT). \tag{15}$$

The total service which we want to maximize can be written as

$$F_{total} = \sum_{s \in M} \sum_{u \in U_s^{srv}} F(s,R_u). \tag{16}$$

The proposed algorithm is designed to find the resource allocation that maximizes the total service quality subject to total power constraint. Therefore, we can denote the objective as

$$\max(F_{total})s.t.P \leq P_{\max}, \tag{17}$$

where the F_{total} is total power allocated to subchannels. As discussed in Section 4.2, each user is included in only

one of the CGs. Therefore, for user that included in CG $U_{s,q}$, the bit rate can be written as

$$R_u = R_{s,q}, \forall u \in U_{s,q}. \tag{18}$$

Accordingly, total service quality of CG $U_{s,q}$ can be written as

$$\sum_{u \in U_{s,q}} F(s,R_u) = U_{s,q}F(s,R_{s,q}), \tag{19}$$

where the $U_{s,q}$ is the number of users included in CG $U_{s,q}$. Therefore, the total service quality (16) can be written as

$$F_{\text{total}} = \sum_{s \in S} \sum_{q \in Q} U_{s,q}F(s,R_{s,q}). \tag{20}$$

The bit rate of each user of CG, $R_{s,q}$, in (20) can be determined by

$$R_{s,q} = \sum_{k \in K} A_{k,s,q}r_k, \tag{21}$$

where the K is index set of the MCPs and the indicator $A_{k,s,q}$ denotes that whether MCP k transmit service s to users $U_{s,q}$ ($A_{k,s,q} = 1$) or not ($A_{k,s,q} = 0$). For example, in case of Table 3, $A_{1,1,1}$ is 1 and $A_{2,1,1}$ is 0. The r_k in (21) is total bit rate of all the subchannels allocated to MCP k, which can be written as

$$r_k = \sum_{c \in C} a_{c,k}r_{c,k}. \tag{22}$$

The indicator $a_{c,k}$ in (22) denotes whether subchannel c is allocated to MCP k ($a_{c,k} = 1$) or not ($a_{c,k} = 0$), where each subchannel can be allocated up to one MCP, i.e., $\sum_{k \in K} a_{c,k} \leq 1$, $\forall c \in C$. Bit rate $r_{c,k}$ in (22) is the bit rate of subchannel provided that the subchannel is allocated to the MCP k. According to [20], capacity of LTE system can be approximated by Shannon capacity by using bandwidth loss factor ε_{bw} and SNR loss factor ε_{SNR}. That is, if we assume an LTE system, the maximum bit rate that can be robustly transmitted to users of MCP k with transmit power $p_{c,k}$ over subchannel c can be approximated by

$$r_{c,k} = \varepsilon_{bw}W\log_2\left(1 + \varepsilon_{\text{SNR}}P_{c,k}^*H_{c,k}\right), \tag{23}$$

where the W is the bandwidth of each subchannel. The $H_{c,k}$ is the worst channel quality of users U_k, i.e., $H_{c,k} = \min_{u \in U_k}(h_{c,u})$, where U_k is the index set of users included in MCP k.

In addition, the total power distributed over all the subchannels can be written as

$$P = \sum_{c \in C} \sum_{k \in K} a_{c,k} p_{c,k}. \qquad (24)$$

In this article, we consider the two resource allocation functions for maximizing the total service quality. That is, we introduce power allocation function (PAF) designed to find the power allocation $p_{out} = \{p_{c,k}^{out}|c \in C, k \in K\}$ that maximizes the total service quality for a given subchannel allocation $a_{in} = \{a_{c,k}^{in}|c \in C, k \in K\}$. Also, we present the subchannel allocation function (CAF) designed to find the subchannel allocation $a_{out} = \{a_{c,k}^{out}|c \in C, k \in K\}$ for a given power allocation $p_{in} = \{p_{c,k}^{in}|c \in C, k \in K\}$. That is, by cooperatively operating the two functions, we can determine the optimal power allocation a^* and subchannel allocation r^*. More specifically, if the subchannel allocation determined by the CAF with the r^* is identical to a^*, and the power allocation determined by the PAF with the a^* is also identical to r^* (as shown in Figure 8), then a^* and r^* can be considered as the optimal resource allocation that maximizes the total service quality. Therefore, we consider cooperatively (and iteratively) operating the two functions as shown in Figure 8b, where the power or subchannel allocation determined by one of the functions is fed back to the other function. If the a_{out} and the p_{out} (or the a_{in} and the p_{in}) become stable over the iterative operations in Figure 8b, the resource allocation can be considered as the optimal resource allocation as described in Figure 8a. In the rest of this section, we present the two functions, respectively. Then, we explain the iterative operation of the two functions in detail.

4.3.1. PAF

As shown in Figure 8b, the PAF is designed to find the p_{out} that maximizes the total service quality, given the a_{in} (determined by the CAF). According to (17), the Lagrange function can be written as

$$\Lambda = F_{total} + \lambda(P_{MAX} - P), \qquad (25)$$

where λ denotes the Lagrange multiplier. As the optimal value of power $p_{c,k}$ that Λ maximizes the makes the derivative of the Λ zero, it can be written as

$$\frac{\partial \Lambda}{\partial p_{c,k}} = 0. \qquad (26)$$

According to (20) and (24), Equation (26) can be written as

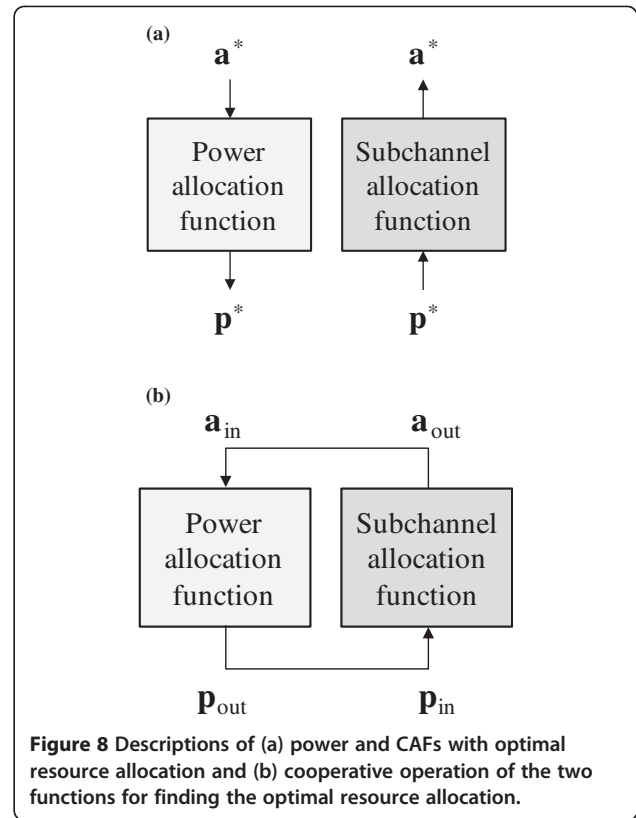$$\frac{\partial F_{total}}{\partial p_{c,k}} = \lambda a_{c,k}. \qquad (27)$$

Figure 8 Descriptions of (a) power and CAFs with optimal resource allocation and (b) cooperative operation of the two functions for finding the optimal resource allocation.

According to (15) and (20), it can be written as

$$\frac{\partial F_{total}}{\partial p_{c,k}} = \sum_{s \in S} \sum_{q \in Q} U_{s,q} \frac{\partial r_{c,k}}{\partial p_{c,k}} \frac{\partial r_k}{\partial r_{c,k}} \frac{\partial R_{s,k}}{\partial r_k} f(s, R_{s,q}). \qquad (28)$$

Therefore, (27) can be written as

$$\frac{a_{c,k} W \varepsilon_{bw} \varepsilon_{SNR} H_{c,k} B_k}{(1 + \varepsilon_{bw} p_{c,k} H_{c,k}) \ln(2)} = \lambda a_{c,k} \qquad (29)$$

according to (21)–(23), where the B_k is

$$B_k = \sum_{s \in S} \sum_{q \in Q} U_{s,q} A_{k,s,q} f(s, R_{s,q}). \qquad (30)$$

According to (15), we can see that the B_k is sum of the bit values of the users U_k, multiplied by T. Therefore, the B_k is referred to as TBV of MCP k in the rest of this article. If the $a_{c,k}$ is 0, the equality of (29) is satisfied regardless of the $p_{c,k}$. If not, the $p_{c,k}$ that satisfies the equality can be obtained by

$$p_{c,k} = \frac{B_k W \varepsilon_{bw}}{\lambda \ln(2)} - \frac{1}{\varepsilon_{SNR} H_{c,k}}. \qquad (31)$$

As the power according to (31) can be negative, the subchannel allocation is determined by

$$a_{c,k} = \begin{cases} a_{c,k'}^{in}, & p_{c,k} > 0. \\ 0, & p_{c,k} \leq 0. \end{cases} \quad (32)$$

According to (24) and (31), the total power can be written as

$$P = \sum_{c \in C} \sum_{k \in K} a_{c,k} \left(\frac{B_k W \varepsilon_{bw}}{\lambda \ln(2)} - \frac{1}{\varepsilon_{SNR} H_{c,k}} \right). \quad (33)$$

In order to fully utilize the given total power P_{MAX}, we need to find λ^* which is defined as

$$\lambda^* = \frac{arg}{\lambda}(P = P_{MAX}). \quad (34)$$

According to the λ^*, the power can be obtained by

$$p_{c,k}^* = \frac{B_k W \varepsilon_{bw}}{\lambda^* \ln(2)} - \frac{1}{\varepsilon_{SNR} H_{c,k}}. \quad (35)$$

Finally, p_{out} is determined by

$$p_{c,k}^{out} = \max\left(p_{c,k'}^*, 0\right) \quad (36)$$

Because the given subchannel allocation may not be the optimal allocation, the PAF also finds the power for the MCPs that are not allocated to each subchannel (according to the) so the CAF can modify the subchannel allocation by using the determined power allocation as described in Figure 8b.

4.3.2. CAF

As shown in Figure 8b, given the p_{in} (determined by the PAF), the CAF is designed to find the a_{out} that maximizes the total utility of the subchannels, where the utility of subchannel c by allocating it to MCP k can be determined by

$$\Delta_{c,k} = \Psi_{c,k} - \Omega_{c,k'} \quad (37)$$

The $\Psi_{c,k}$ (the increase in the total service quality by using subchannel c for MCP k) and the $\Omega_{c,k}$ (the cost of using subchannel c for MCP k) above can be obtained by

$$\Psi_{c,k} = \sum_{s \in S} \sum_{q \in Q} A_{k,s,q} U_{s,q}$$
$$\left(F\left(s, R_{s,q}^{-c} + r_{c,k}^{in}\right) - F\left(s, R_{s,q}^{-c}\right) \right) \quad (38)$$

and

$$\Omega_{c,k} = \lambda^* p_{c,k'}^{in} \quad (39)$$

respectively. The $R_{s,q}^{-c}$ in (38) is the bit rate of each user in CG $U_{s,q}$, which is allocated by the subchannels excluding subchannel c and the $r_{c,k}^{-c}$ is the bit rate determined by the $p_{c,k}^{in}$ according to (23). The λ^* in (39) is determined by PAF according to (34). In order to maximize the total utility, each subchannel is allocated to the MCP that results in the largest utility. By defining $\kappa(c)$ the index of the MCP allocated to subchannel c, we can denote that

$$a_{c,k} = \begin{cases} 1, & k = \kappa(c). \\ 0, & k = \kappa(c). \end{cases} \quad (40)$$

That is, if the $\kappa(c)$ is identical to $\kappa^*(c)$ for all the subchannels, the subchannel allocation maximizes the total utility, where the $\kappa^*(c)$ is defined as

$$\kappa^*(c) = \arg\max_{k \in K}\left(\Delta_{c,k}\right), \quad (41)$$

However, the $\Delta_{c,k}$ is dependent on the $R_{s,q}^{-c}$ which is dependent on the MCPs allocated to the other subchannels. Therefore, we propose to implement the CAF as described in Algorithm 1, where the C_{count} in Algorithm 1 is the number of subchannels that $\kappa(c)$ is identical to $\kappa^*(c)$.

Algorithm 1. CAF

for $c = 1$ to C

Determine $\kappa^*(c)$, and then $\kappa(c)$ to $\kappa^*(c)$.

end for

Initiate C_{count} to 0.
Initiate c to 1.

while C_{count} is identical to c, iterate:

Determine $\kappa^*(c)$.
if $\kappa^*(c)$ is identical to $\kappa(c)$, increase C_{count} by 1.
else set C_{count} to 0 and set $\kappa(c)$ to $\kappa^*(c)$.
increase c by 1 (if $c = C$, set c to 1).

end while

First, the CAF allocates the MCP with the largest utility to each subchannel. Then, initiate the C_{count} to 0. After that, from subchannel 1, the CAF checks if the allocated MCP is the MCP with the largest utility. If the allocated MCP is identical to the MCP with the largest utility, the C_{count} is increased by 1. If not, the C_{count} is set to 0, and then the subchannel is allocated to the MCP with the largest utility. This operation is iterated over the subchannels until all the subchannels are allocated to the MCP with the largest utility, i.e., C_{count}

is increased to C. After the algorithm is terminated, the subchannel allocation is determined by

$$a_{c,k}^{\text{out}} = \begin{cases} 1, & k = \kappa(c) = \kappa^*(c). \\ 0, & k \neq \kappa(c) = \kappa^*(c). \end{cases} \quad (42)$$

The determined subchannel allocation is fed back to the PAF as shown in Figure 8b in order to find the optimal resource allocation.

4.3.3. Cooperative operation of power and CAFs

The cooperative operation of the two functions described in Figure 8b is implemented as described in Algorithm 2, where TBVs $B = \{B_k | k \in K\}$ in (31) also have to be given in addition to a_{in} in order to operate the PAF. In Algorithm 2, the terms $a_{(g)}$, $p_{(g)}$, and $B_{(g)}$ denote the gth a, p, and B determined by the algorithm, respectively. Also, the $B_{k,g}$ in Algorithm 2 denotes gth TBV of MCP k determined by the algorithm.

Algorithm 2 Proposed resource allocation algorithm described in Figure 8b

Initiate the g to 1.

Initiate each element of $p_{(g)}$ to P_{MAX}/C.

while every element in a, p and B converges, iterate:

Determine $a_{(g)}$ by using the CAF with $p_{(g)}$.
According to $a_{(g)}$ and $p_{(g)}$, determine $B_{(g)}$.
if g is larger than 1, **then**
    Modify $B_{k,g}$ to $(B_{k,1} + B_{k,2} + B_{k,g})/g$ for all $k \in \mathbf{K}$.

end if

Determine $P_{(g+1)}$ by using the PAF, $a_{(g)}$ and $B_{(g)}$.
Increase the g by 1.

end while

First, the algorithm equality distribute the P_{MAX} over the subchannels. Then, determine the subchannel allocation according to the CAF. After that, the TBVs are determined according to (30), and then the PAF is operated to determine the power allocation. Then, CAF is operated to determine the subchannel allocation, and therefore the TBVs can be determined. However, from the second iteration of the PAF, each TBV is averaged with the previously determined TBVs in order to achieve the convergence over the iterations. This cooperative operation of the two functions is iterated until the resource allocation becomes stable and finds the optimal resource allocation scheme, where each averaged TBV converges

to the currently determined TBV as the resource allocation become stable.

5. Simulation results

In this section, we provide computer simulation results to verify the performance of the proposed algorithm, where each OFDM symbol is configured with 500 subchannels. We assume that the bandwidth of each subchannel is 10 kHz, and therefore the total bandwidth of the cellular system is 5 MHz. Also, we assume that the subchannel qualities are exponentially distributed, where the ACQ of each user is determined by the distance from the BS. By using the path loss model in [19], we determined the ACQ as $\overline{\text{h}}_u = 20 - 35 \times \log_{10}(D_u)$, where the terms $\overline{\text{h}}_u$ and D_u denote the ACQ and the distance of user u from the BS, respectively. As we assume that the subchannel qualities are exponentially distributed, the $h_{c,u}$'s are determined according to the probability density function $f_{PDH}(h, \overline{\text{h}}_u) = \exp(-h/\overline{\text{h}}_u)/\overline{\text{h}}_u$, where h is the random variable for setting $h_{c,u}$'s. As discussed in [20], we set the bandwidth and the SNR losses to 0.75 and 0.8, respectively. In addition, we assumed the ten MCSs considered [20]. More specifically, according to (35), we find one of the MCS modes, i.e., MCS modes $M = \{m | 1 \leq m \leq M\}$, for each subchannel, where the M is the number of available MCS modes, i.e., 10. That is, the MCS index can be found as

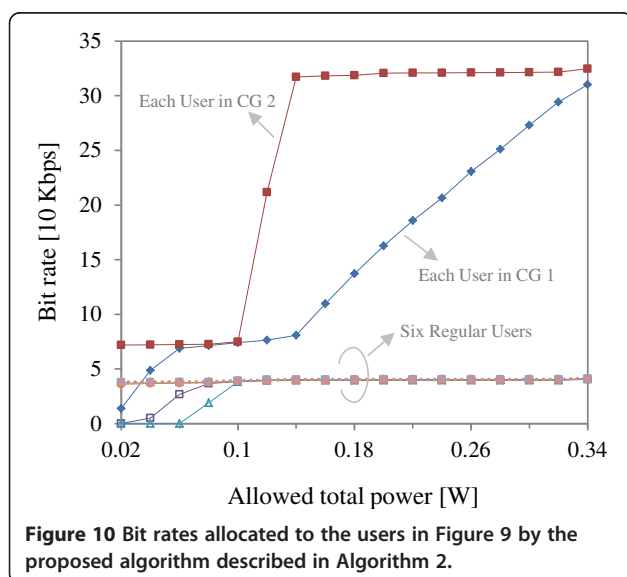$$m^* = \underset{\{m | \Gamma_m \leq r_{c,k}^*\}}{\arg \max}(\Gamma_m), \quad (43)$$

where the $r_{c,k}^*$ is the bit rate determined by the resource allocation algorithm described in Algorithm 2 and the Γ_m is the bit rate of mth transmission mode. As a result, the realistic bit rate that is provided by the system is Γ_{m^*}.

5.1. Bit rate allocation of proposed algorithm

As introduced in Section 3, we assume the system serves the regular services and the MBMSs. In order to verify the performance of the proposed algorithm for the two types of the services, we can consider a configuration of a BS and users in a 4-km^2 square as shown in Figure 9, where the BS on the center serves six users requesting regular services and four users requesting an MBMS. We assume that the MBMS users request "Football" in Figure 2, and therefore the bit rate of 72 kbps has to be guaranteed, where the bit rate exceeding 320 kbps does not improve the PSNR. Figure 10 shows the bit rates allocated by the proposed algorithm, where we assume that each of the bit rates required for the regular services is 40 kbps. Also, we set the number of the quality indices, Q, to 2, and therefore the MBMS users are divided into two CGs each of which is configured with two of the MBMS users. As discussed in Section 4.2, the same bit

Figure 9 Configuration of a BS and users in a 4-km² square.
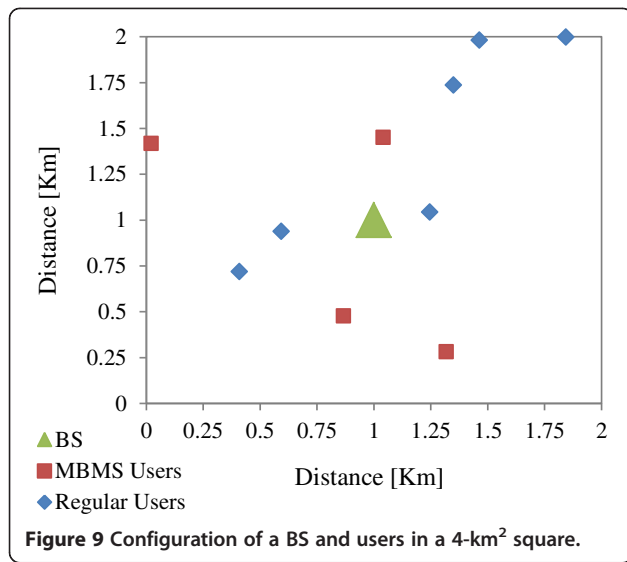
rate is allocated to all the users in each CG, and therefore the bit rate curves of the two CGs and the six regular-service users are denoted in Figure 10. According to (10) and (11), the bit values of the bits that can be transmitted with 40 kbps (for the regular services) and 72 kbps (for the MBMS) are significantly higher than the bit values of the remaining bits of the MBMS, i.e., the bits in ELs. Therefore, when the total power is lower than a certain power level (i.e., 0.1 W in Figure 10), the bit values of each of the regular-service users and the MBMS users do not exceed 40 and 72 kbps, respectively. If the power is higher than the power level of 0.1 W, only the bit rates of the MBMS users are increased, because the bit values of the regular services approach to zero (e.g., as shown in Figure 5). Also, as shown in Figure 6, the bit value of the MBMS approaches to zero when the allocated bit rate

Figure 10 Bit rates allocated to the users in Figure 9 by the proposed algorithm described in Algorithm 2.
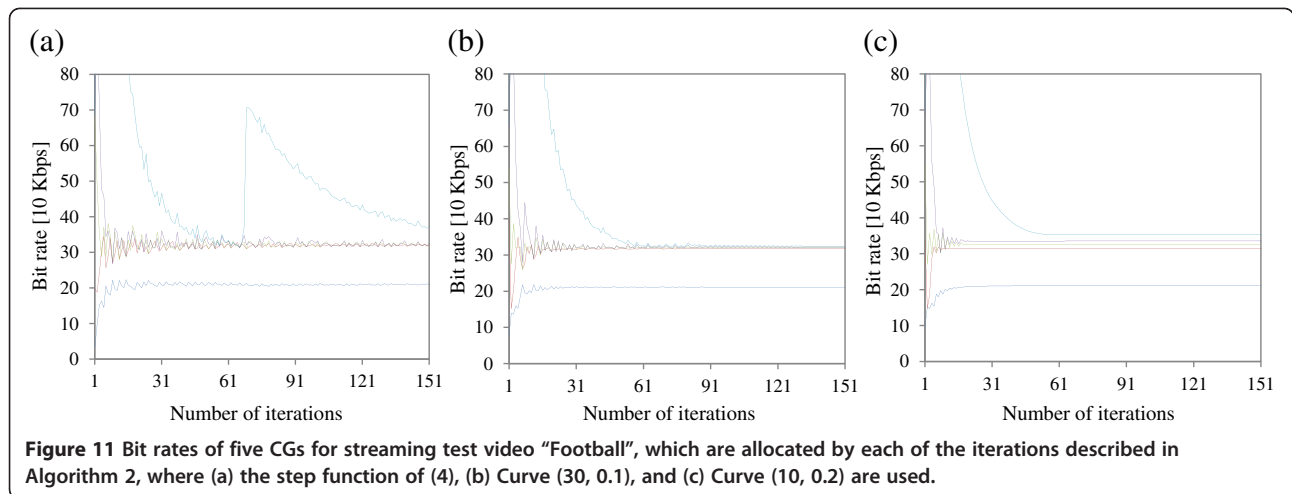
is higher than the maximum bit rate of 320 kbps. Therefore, the bit rate higher than the maximum is not allocated to each CG when the power is lower than 0.34 W. In this simulation, we can verify that the bit rate allocation scheme according to the required bit rates of the two types of the services can be achieved only by defining the service quality and the bit values according to (8)–(11).

As discussed in Section 4.1, we consider the concave approximation of the service quality. According to the service quality, the power and the subchannel resources are allocated according to the iterative resource allocation algorithm described in Algorithm 2. In order to achieve the concave function of the service quality, we defined the bit value function in that approximates the step functions as described in Figure 5. If we define the service quality according to the step function, the service quality function is not concave as discussed in Section 4.1, and therefore it is difficult to achieve the convergence by using the resource allocation. Hence, we defined the approximation function (7), which makes the service quality concave (e.g., in the simulation of Figure 10, Curve (30, 0.1) is used).

Figure 11 shows the bit rates of five CGs, allocated by each iteration described in Algorithm 2, where 15 users requesting "Football" are randomly distributed in the square of Figure 9, the Q is set to 5, and the P_{MAX} is set to 1 W. In this simulation, the step function, Curve (30, 0.1), and Curve (10, 0.2) are considered. When the step function is considered as shown in Figure 11a, each bit rate does not converge within 150 iterations. However, when Curve (30, 0.1) and Curve (10, 0.2) are considered as shown in Figures 11a,b, the convergence can be achieved within 130 and 60 iterations, respectively. That is, by adjusting the two parameters so the bit value changes slower with respect to the bit length (or bit rate) as shown in Figure 5, the convergence can be achieved within the fewer number of iterations. It is because that the TBVs described in Algorithm 2 change slowly over the iterations. However, as the bit value changes slower, the difference of the bit value between the approximation curve and the step function becomes larger, and therefore the performance degrades. For example, one of the bit rate curves in Figure 11c converges to 350 kbps which is higher than the maximum bit rate of 320 kbps. It means that the resources are wasted in terms of the PSNR, because the PSNR does not increase further when the bit rate is higher than 320 kbps as aforementioned. In the rest of the simulations, we considered Curve (30, 0.1).

5.2. Multicast gain of proposed algorithm
The simulations of Figures 12 and 13 are provided to verify the multicast gain of the proposed algorithm in terms of the average PSNR, where randomly distributed users in the square of Figure 9 request an MBMS of

Figure 11 Bit rates of five CGs for streaming test video "Football", which are allocated by each of the iterations described in Algorithm 2, where (a) the step function of (4), (b) Curve (30, 0.1), and (c) Curve (10, 0.2) are used.

test video "Football". In Figure 12, we considered fifteen users and tested three cases as follows, where the number of quality indices, Q, is diversely set as shown in Figure 12.

1. *Unicast case*: The BS unicasts the video stream to the all 15 users.
2. *Mixed case*: The BS unicasts and multicasts to the seven and eight users, respectively.
3. *Multicast case*: The BS multicasts to all the 15 users.

By comparing the three curves "Unicast", "Mixed ($Q = 5$)," and "Multicast ($Q = 5$)" in Figure 12, we can verify that the performance is improved by increasing the number of multicast users. It is because that the proposed algorithm benefits from both multicast gain and the diversity gin of the subchannel qualities by considering the configurations of the MCPs. That is, if we consider more MCPs by increasing the Q, the user diversity is improved (because more MCP candidates can be considered for each subchannel), and therefore the performance can be improved. For example, if only one CG is considered by setting the Q to 1, we can merely achieve the user diversity gain, and the performance degrades severely as shown in Figure 12. On the other hand, by increasing the Q to 10, we can improve the performance over the case that Q is set to 5.
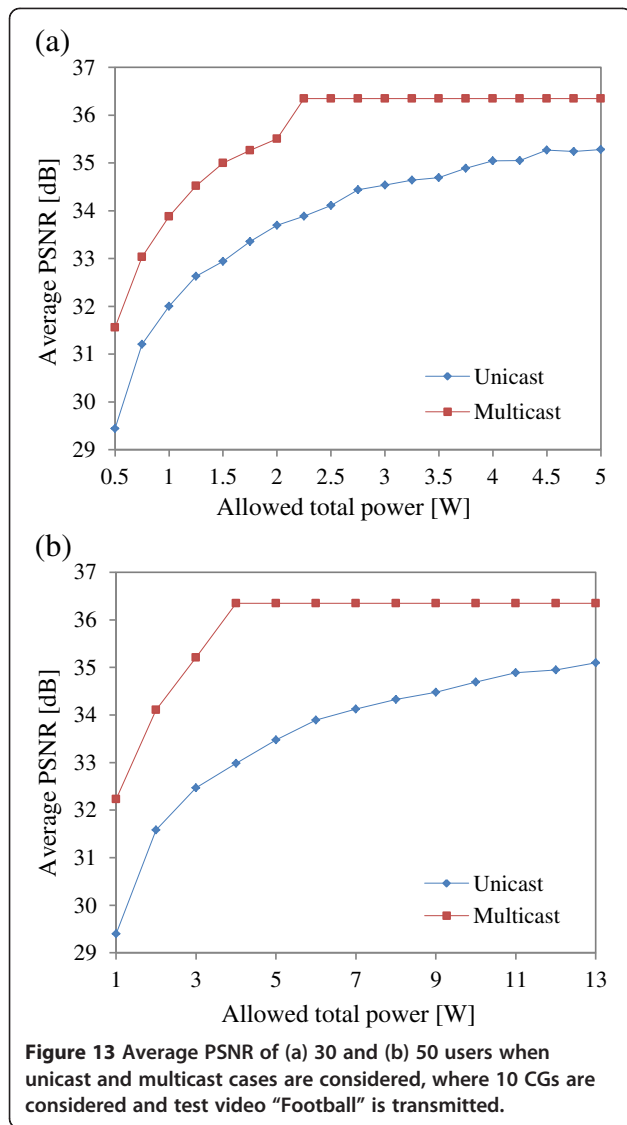
If we consider more users, the performance is improved further by using multicast. Figure 13 shows the performance of unicast case, i.e., the BS unicasts to all the users, and multicast case, i.e., the BS multicasts to all the users, where we tested two cases of 30 and 50 users. We assumed ten CGs by setting the Q to 10. In Figure 13, we can verify that, by using the multicast scheme, the PSNR performance can be improved by approximately 2 and 3 dB when we consider 30 and 50 users, respectively.
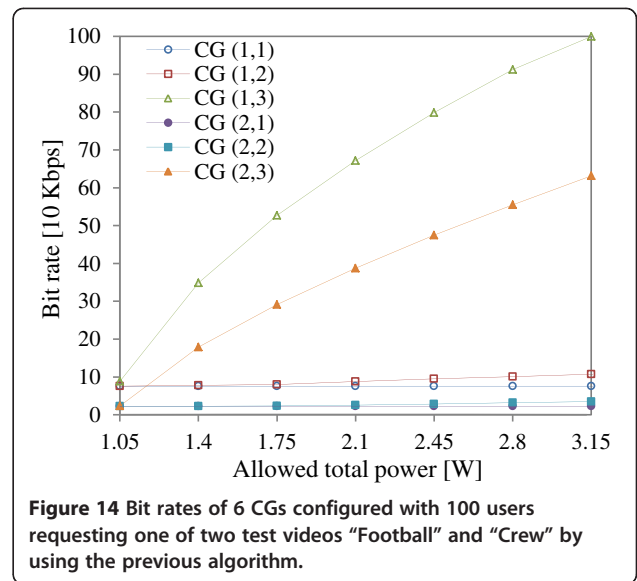
5.3. Performance comparison

In the rest of this section, we provide the simulation results to verify the performance of the proposed algorithm and the previous algorithm in [14] which is also designed for the multicast services by using SVC. We assumed that the Q is 3, and 100 users are randomly distributed in the square of Figure 9, where a half of the users request "Football" (service 1) and the other half of the users request "Crew" (service 2) in Figure 2. We want to note that considering the all possible MCPs is improper especially for the large number of users as discussed in Section 2.3, and therefore we apply the proposed configuration of MCP in Section 4.2 to the previous algorithm.

As discussed in Section 2.3, the previous algorithm is designed to maximize the total bit rate for the ELs,
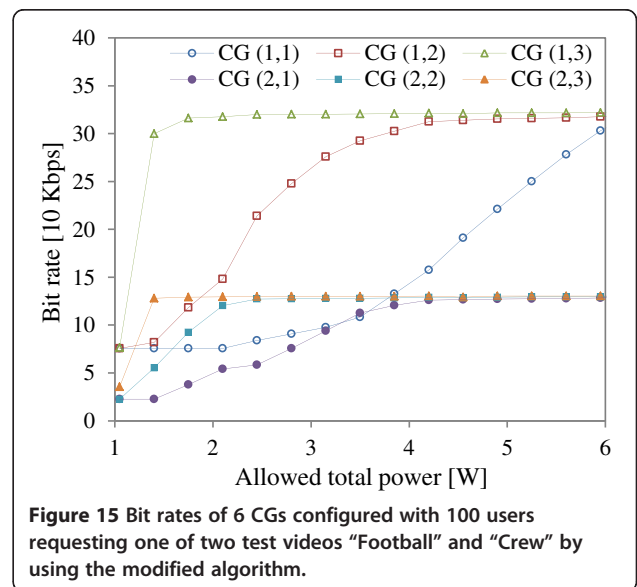

Figure 12 Average PSNR of 15 users with the various portions of multicast and unicast, and various setting of the number of quality indices, Q, where test video "Football" is transmitted by the proposed algorithm.

Figure 13 Average PSNR of (a) 30 and (b) 50 users when unicast and multicast cases are considered, where 10 CGs are considered and test video "Football" is transmitted.

Figure 14 Bit rates of 6 CGs configured with 100 users requesting one of two test videos "Football" and "Crew" by using the previous algorithm.

the received PSNR of the users in the CGs cannot be improved. Therefore, we need to consider another objective for subchannel group 2, and therefore we modified the previous algorithm by applying the proposed algorithm to subchannel group 2.

Figure 15 shows the bit rate performance of the modified algorithm. In Figure 15, we can verify that also the bit rates of CGs with the worse channel conditions are improved as the total power is increased.

The improvement of the average PSNR performance can be verified in Figure 16, where the previous, the modified, and the proposed algorithms are simulated in Figure 16. As shown in Figure 16, we can see that the objective of maximizing the total bit rate severely

i.e., MGS packets, and we discussed that the objective can be inefficient for transmitting the SVC bit streams. Figure 14 shows the bit rates allocated by the previous algorithm, where the "CG (s, q)" denotes the CG of service s and quality index q, i.e., $U_{s,q}$. As the previous algorithm is designed to guarantee the minimum bit rates, all the minimum bit rates for the BLs can be allocated to each CG when the total power is set to a certain power level, i.e., approximately 1.05 W in Figure 14. However, as discussed in Section 2.3, the previous algorithm finds the resource allocation that maximizes the total bit rate for subchannel group 2. Therefore, as the total power is increased over the power level, the previous algorithm overwhelmingly allocates the resources to the CGs with better channel conditions, i.e., CGs (1, 3) and (2, 3), in order to maximize the total bit rate. As a result, the bit rates of the rest CGs are not significantly improved, and

Figure 15 Bit rates of 6 CGs configured with 100 users requesting one of two test videos "Football" and "Crew" by using the modified algorithm.
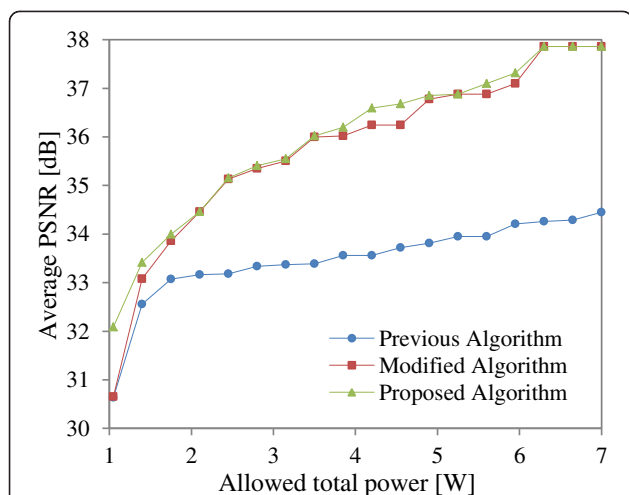
Figure 16 Average PSNR of 6 CGs configured with 100 users requesting two test videos "Football" and "Crew" by using the previous, the modified, and the proposed algorithms.

degrades the average PSNR. In Section 2.3, we also discussed that the resource can be wasted when the previous algorithm divides the subchannels into the two groups. Therefore, we do not determine the subchannels for the BLs before allocating the resources. Instead, we set the bit values of the BLs significantly higher than those of the MGS packets. As a result, the average PSNR performance of the proposed algorithm is improved by up to 1.3 dB over the modified algorithm.

6. Conclusion

In this article, we considered an OFDM-based cellular system that serves the regular services and the MBMSs. First, we discussed that the MCPs have to be limited so each user can receive the bit stream properly and the information overheads for the users to understand the position of the received bits in the bit streams can be reduced. Second, in order to utilize the resource allocation more efficiently, we defined the service quality according to the bit rate requirements and bit rate–PSNR behavior of the SVC streams. Third, we introduced the resource allocation algorithm for maximizing the total service quality. From the computer simulation results, we verified that the three objectives can be achieved, which are summarized as follows: First, the bit rate of each regular service has to be guaranteed. Second, the minimum bit rate of each MBMS for achieving the minimum PSNR has to be guaranteed. Third, the average PSNR of the MBMS users has to be maximized. Also, we could verify that the performance of the proposed algorithm is superior to that of the previous algorithm designed to maximize the total bit rate while guaranteeing the minimum bit rates.

Endnote

[a]In this article, we consider the case that any BL cannot be transmitted to the corresponding users as a system failure. Therefore, we do not define the BL for the bit rate lower than the minimum bit rate required for the BLs.

Abbreviations
ACQ: average subchannel quality; BL: base layer; CAF: subchannel allocation function; CGS: coarse grain scalability; CIF: common intermediate format; EL: enhancement layer; Fps: frames per second; GOP: group of pictures; JSVM: joint scalable video model; LTE: long-term evolution; MBMS: multimedia broadcast multicast service; MCP: multicast patterns; MCS: modulation and coding scheme; MGS: medium grain scalability; OFDM: orthogonal frequency division multiplexing; PAF: power allocation function; PSNR: peak signal-to-noise ratio; QL: quality layer; SVC: scalable video coding; TBV: total bit value.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Advanced Wireless Communication Research Center, The University of Electro-communications, Tokyo, Japan. [2]Department of Radio Communications, Kyung Hee University, Kyung-Gi, South Korea.

References
1. S Sadr, A Anpalagan, K Raahemifar, Radio resource allocation algorithms for the downlink of multiuser OFDM communication systems. IEEE Surv. Tutor Commun. 11(3), 92–106 (2009). doi:10.1109/SURV.2009.090307
2. J Huang, VG Subramanian, R Agrawal, RA Berry, Downlink scheduling and resource allocation for OFDM systems. IEEE Trans. Wirel. Commun. 8(1), 288–296 (2009). doi:10.1109/T-WC.2009.071266
3. D Huang, Z Shen, C Miao, C Leung, Resource allocation in MU-OFDM cognitive radio systems with partial channel state information. EURASIP J. Wirel. Commun. Netw. 2010(6) (2010). doi:10.1155/2010/189157
4. X Ji, J Huang, M Chiang, F Catthoor, Downlink OFDM scheduling and resource allocation for delay constraint SVC streaming (International Conference on Communication (ICC'08), Beijing, China, 2008), pp. 2512–2518. doi:10.1109/ICC.2008.476
5. T Le-Ngoc, N Damji, Y Xu, Dynamic resource allocation for multimedia services over OFDM downlink in cellular systems, 5th edn. (Vehicular Technology Conference (VTC'04), Milan, Italy, 2004), pp. 1864–1868. doi:10.1109/VETECS.2004.1390596
6. N Damji, T Le-Ngoc, Dynamic resource allocation for delay-tolerant services in downlink OFDM wireless cellular systems, 5th edn. (International Conference on Communication (ICC'05), Seoul, Korea, 2005), pp. 3095–3099. doi:10.1109/ICC.2005.1494968
7. G Su, Z Han, M Wu, A scalable multiuser framework for video over OFDM networks: fairness and efficiency. IEEE Trans. Circuits Syst. Video Technol. 16(10), 1217–1231 (2006). doi:10.1109/TCSVT.2006.883513
8. GPP TS 36.440 V9.1.0, Technical Specification Group Radio Access Network, Evolved Universal Terrestrial Radio Access Network (E-UTRAN); General aspects and principles for interfaces supporting Multimedia Broadcast Multicast Service (MBMS) within E-UTRAN, Release 9, 2010
9. B Wu, J Shen, H Xiang, Resource allocation with minimum transmit power in multicast OFDM systems. J. Syst. Eng. Electron. 21(3), 355–360 (2010). doi:10.3969/j.issn.1004-4132.2010.03.002
10. W Xu, K Niu, J Lin, Z He, Resource allocation in multicast OFDM systems: lower/upper bounds and suboptimal algorithm. IEEE Trans. Commun. Lett. 15(7), 722–724 (2011). doi:10.1109/LCOMM.2011.051911.110203
11. J Kim, T Kwon, D Cho, Resource allocation scheme for minimizing power consumption in OFDM multicast systems. IEEE Trans. Commun. Lett. 11(6), 486–488 (2007). doi:10.1109/LCOMM.2007.070007
12. H Schwarz, D Marpe, T Wiegand, Overview of the scalable video coding extension of the H.264/AVC standard. IEEE Trans. Circuits Syst. Video Technol. 17(9), 1103–1120 (2007). doi:10.1109/TCSVT.2007.905532

13. K Seoshin, S Hanbyeol, L Byeong-Gi, *Suitability-based subcarrier allocation for multicast services employing layered video coding in wireless OFDM systems* (Vehicular Technology Conference (VTC'07), Maryland, USA, 2007), pp. 1752–1756. doi:10.1109/VETECF.2007.370

14. D Hui, T Xiaoming, L Jianhua, *QoS-aware resource allocation for mixed multicast and unicast traffic in OFDMA networks* (Vehicular Technology Conference (VTC'11), San Francisco, USA, 2011), pp. 1–5. doi:10.1109/VETECF.2011.6093180

15. I Amonou, N Cammas, S Kervadec, S Pateux, Optimized rate-distortion extraction with quality layers in the scalable extension of H.264/AVC. IEEE Trans. Circuits Syst. Video Technol **9**, 1186–1193 (2007). doi:10.1109/TCSVT.2007.906870

16. JSVM, *JSVM Software Manual," ver. JSVM 9.19 (CVS tag: JSVM_9_19)*, 2009

17. *Xiph.org*, 2012. http://media.xiph.org/video/derf, Accessed 30 October

18. H Mansour, P Nasiopoulos, V Krishnamurthy, Rate and distortion modeling of CGS coded scalable video content. IEEE Trans. Multimed. **13**(2), 165–180 (2011). doi:10.1109/TMM.2010.2099648

19. COST Secretariat, *European Commission* (Brussels, Belgium, 1999)

20. P Mogensen, N Wei, IZ Kovacs, F Frederiksen, A Pokhariyal, KI Pedersen, T Kolding, K Hugl, M Kuusela, *LTE Capacity compared to the Shannon bound* (Vehicular Technology Conference (VTC'07), Dublin, Ireland, 2007), pp. 1234–1238. doi:10.1109/VETECS.2007.260