Uplink Resource Management for Multiuser OFDM Video Transmission Systems: Analysis and Algorithm Design

Dawei Wang, *Student member, IEEE,* Laura Toni, *Member, IEEE,* Pamela C. Cosman, *Fellow, IEEE,* and Laurence B. Milstein, *Fellow, IEEE*

Abstract—We consider a multiuser OFDM system in which users want to transmit videos via a base station. The base station knows the channel state information (CSI) as well as the rate distortion (RD) information of the video streams and tries to allocate power and spectrum resources to the users according to both physical layer CSI and application layer RD information. We derive and analyze a condition for the optimal resource allocation solution in a continuous frequency response setting. The optimality condition for this cross layer optimization scenario is similar to the equal slope condition for conventional video multiplexing resource allocation. Based on our analysis, we design an iterative subcarrier assignment and power allocation algorithm for an uplink system, and provide numerical performance analysis with different numbers of users. Comparing to systems with either only physical layer or only application layer information available at the base station, our results show that the user capacity and the video PSNR performance can be increased significantly by using cross layer design. Bit-level simulations which take into account the imperfection of the video coding rate control, the variation of RD curve fitting, as well as channel errors, are presented.

Index Terms—Radio spectrum management, multimedia communication, OFDMA, wireless power allocation, video coding rate control.

I. INTRODUCTION AND RELATED WORK

W E study video transmission in a cellular wireless communication system, where multiple mobile stations send compressed videos to a base station. We exploit both the characteristics of video content as well as the instantaneous wireless channel quality, and design a cross layer resource allocation algorithm to optimize video transmission performance.

Orthogonal Frequency Division Multiplexing (OFDM) is a promising solution to combat the problem of inter-symbol interference (ISI) in a wideband communication system. By allocating different subcarriers to users according to the user's channel state information (CSI) in a multiuser setting, Orthogonal Frequency Division Multiple Access (OFDMA) is a flexible and low-complexity way of managing communication resources. The problem of assigning resources in an OFDMA system was first studied in [3], where the authors formulated and solved a total transmission power minimization problem for different user quality-of-service (QoS) requirements. Research in [4]–[7] tried to solve the rate maximization problem, given power and spectrum constraints in different communication settings. Because of the complexity of the optimization problems, most of the work above proposed numerical algorithms instead of finding analytical solutions. Power allocation for an imperfect CSI case was explored in [8]. To reduce the complexity of resource allocation algorithms, chunk-based resource allocation, which makes allocation decisions on subcarriers in groups, was studied by [9], [10]. Results show that when the chunk bandwidth is smaller than the coherence bandwidth, the chunk-based resource allocation can significantly reduce the computational complexity while maintaining similar throughput performance compared to subcarrier-based resource allocation algorithms. Utility driven resource allocation was investigated by [11], [12] and most recently by [13], [14] in an information theoretic setting. In these papers, instead of maximizing the sum of the throughputs, the objective of the optimization is the overall utility, which is a function of throughput.

Regarding the application layer, video rate control algorithms as well as rate distortion (RD) analysis were studied by [1], [15]–[17]. The results of these papers show that the complexity (high or low motion) of a video stream can be reflected by its rate distortion curve. The diversity of different video RD curves provides us an opportunity to optimize the overall video quality when multiple video streams share the same resource pool, i.e., video multiplexing. In [18], the authors considered a multiple camera surveillance system, and exploited the difference between high complexity and low complexity videos. In [19], the economic concept of competitive equilibrium is used to allocate bit rate. The authors show that by trading bit rate between users across time, the video quality of each individual user improves. For

most of the literature on video multiplexing, the resource pool is either bits or bit rate, and the authors assume an error-free scenario. When multiplexing videos in a wireless mobile communication case, bit rates will be determined by the available bandwidth, transmission power and CSI. In this sense, multiplexing video streams in a wireless environment with a resource pool of power and bandwidth will be more challenging than conventional video multiplexing.

In a cellular wireless OFDMA video transmission system, the CSI as well as the complexity of video streams can be collected by the base station. Both the multiuser channel diversity and video complexity diversity could be used simultaneously to optimize the power and subcarrier assignment. In [20], the authors propose a joint uplink and downlink cross layer resource allocation framework with the resource being the channel access time duration. In [21] and [22], the authors study a subcarrier and power assignment problem in a downlink setting, where the subcarrier assignment and power allocation are treated as two independent steps. To better optimize the system, we propose an iterative algorithm which allows the application layer and physical layer to interact. Throughout this paper, we are interested in a cooperative setting for a slow fading scenario. Video streams with high complexity should be given more subcarriers with good channel gains, while streams with low complexity will get a relatively small number of subcarriers.

The rest of the paper is organized as follows: Section II first introduces the basic model of the OFDMA system and the video compression rate-distortion model. We then formulate the problem of cross layer optimization of an uplink cellular system. We derive an optimality condition for this problem in a continuous channel response setting in Section III. A resource allocation algorithm is presented in Section IV. We propose two baseline algorithms and compare the performance with our cross layer algorithm in Sections V and VI. Finally, we draw our conclusions in Section VII.

II. CROSS LAYER RESOURCE ALLOCATION SYSTEM MODEL

A. OFDMA System Description

Consider a cellular OFDMA video communication system with the set of users $k=\{1, 2, 3 ... K\}$. The system occupies a total frequency band of W (Hz) equally divided into M orthogonal subcarriers $m=\{1, 2, 3 ... M\}$. We assume that the channel gain within each subcarrier is flat. In our design, each subcarrier can only be used by one user, but it is possible for one user to get more than one subcarrier.

The system operates in a slotted manner and the length of one time slot is T_s (sec) for both downlink and uplink. One Group of Pictures (GOP) will be transmitted in one time slot. Let $H_k(s) = [H_{k,1}(s), H_{k,2}(s), ...H_{k,M}(s)]$ denote the complex channel gain of user k for the set of subcarriers in time slot s. In addition, we assume that the channel remains unchanged for the duration of one time slot. The subcarrier assignment as well as the power allocation decision will be made on a slot-by-slot basis. A block diagram of the transmitter is shown in Fig. 1. Let T be the data duration and T_{cp} be the length of the cyclic prefix. We define $T_0 = T + T_{cp}$ to be
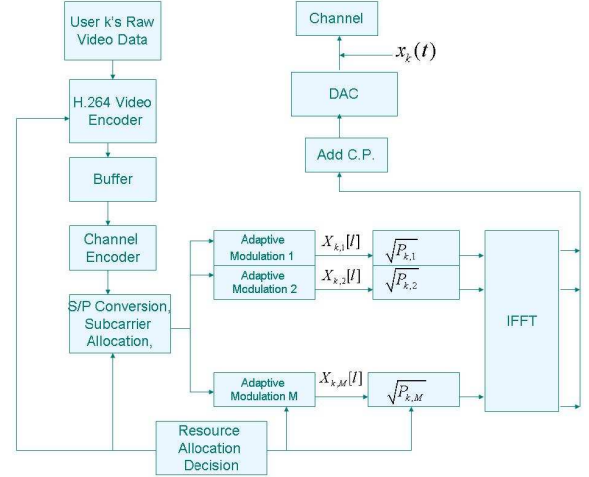

Fig. 1.   Cross-layer optimization system transmitter diagram.

the duration of an OFDM symbol. The baseband transmitted signal for user k can be written as

$$x_k(t) = \sum_l \sum_{m=1}^{M} \sqrt{P_{k,m}} X_{k,m}[l] \exp\left(\frac{j2\pi mt}{T}\right) \Pi(t - lT_0) \tag{1}$$

where $P_{k,m}$ and $X_{k,m}[l]$ are the transmission power and coded symbol with unit variance of user k on subcarrier m, respectively. Also, $\Pi(t) = 1, \forall t \in [0, T_0)$, and $\Pi(t) = 0$ otherwise.

Since we assume flat fading for each subcarrier, the lowpass equivalent received signal of user k on subcarrier m is given by

$$y_{k,m}(t) = \sqrt{P_{k,m}} H_{k,m} X_{k,m}[l] \exp\left(\frac{j2\pi mt}{T}\right) + n_{k,m}(t) \tag{2}$$

where $n_{k,m}(t)$ is Additive White Gaussian Noise (AWGN) with two-sided power spectral density N_0.

To detect the signal on subcarrier m, a correlation operation is performed $Y_{k,m} = \frac{1}{T} \int_0^T y_{k,m}(t) \exp(-j2\pi mt/T) dt$. The noise power can be calculated as $P_N = E[|N_{k,m}|^2] = 2N_0/T$ and the power for the desired signal is $P_{k,m}|H_{k,m}|^2$. If the modulation format is M-QAM, from [23], the symbol error rate (SER) can be approximated as

$$SER \approx 4Q\left(\sqrt{\frac{3}{M-1} \frac{P_{k,m}|H_{k,m}|^2}{P_N}}\right) \tag{3}$$

For a given target SER_t, the information rate (number of bits each symbol can carry) $R_{k,m}(P_{k,m}, H_{k,m})$ (in bits/symbol) can be written as a function of transmission power and channel response gain:

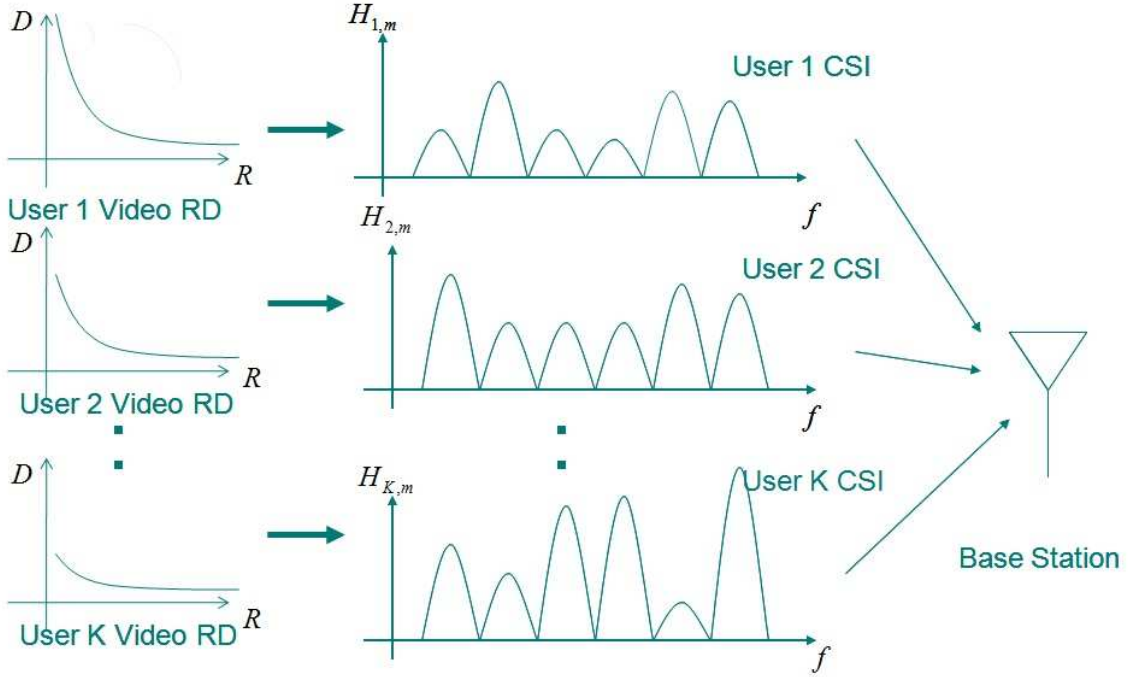$$R_{k,m}(P_{k,m}, H_{k,m}) = \min\{\lfloor \log_2\left[1 + \eta P_{k,m}|H_{k,m}|^2\right]\rfloor, R_{max}\} \tag{4}$$

Fig. 2. Uplink OFDMA video communication system.

where $\eta = \frac{3}{P_N} \left[Q^{-1} \left(SER_t/4 \right) \right]^{-2}$ and R_{max} is the largest alphabet size the system allows. The bit rate (in bits/sec) then can be written as: $R_{k,m}(P_{k,m}, H_{k,m})/T_0$.

B. Video Rate-Distortion Model

Let $D_k^s(B)$ be the rate distortion function of user k in time slot s, where B is the number of bits the encoder generated. For each GOP, the mean square error (MSE) distortion can be approximated as [24]

$$D_k^s(B) = a_k + \frac{w_k}{B + v_k} \qquad (5)$$

where a_k, v_k and w_k are constants which depend on the video content. For video with high complexity (e.g., high motion), w_k is relatively large. To protect the data, a channel code of fixed rate u is added. Since the channel slot time is equal to the duration of one GOP, the number of information bits which the physical layer can support for one user with a fixed

target symbol error rate is

$$B_k = \sum_{m=1}^{M} u \cdot R_{k,m}(P_{k,m}, H_{k,m}) \cdot T_s/T_0 \qquad (6)$$

For the purpose of resource allocation algorithm design, we ignore the effect of channel errors. In Sections III and IV, we use (6) as the channel throughput for our mathematical analysis and algorithm design. The effect of channel errors on the performance of the system will be evaluated by simulation in Section VI.

If we plug (6) into (5), then the MSE distortion for user k can be written as

$$a_k + \frac{b_k}{\sum\limits_{m=1}^{M} R_{k,m}(P_{k,m}, H_{k,m}) + c_k} \qquad (7)$$

Here we have divided both the numerator and denominator by $u \cdot T_s/T_0$ for simplicity, so $b_k = \frac{w_k}{(u \cdot T_s/T_0)}$, and $c_k = \frac{v_k}{(v \cdot T_s/T_0)}$.

C. Uplink Resource Allocation Formulation

In an uplink OFDMA system (Fig. 2), the mobile stations submit the RD values (a_k, b_k, and c_k) of the current GOP in their buffers. We assume that the base station has perfect CSI of each subcarrier for each user. Our resource allocation goal is to minimize the sum of distortions at each time slot s. The optimization objective is

$$\min_{\underline{\underline{P}}} \sum_{k=1}^{K} \frac{b_k}{\sum_{m=1}^{M} R_{k,m}(P_{k,m}, H_{k,m}) + c_k} \tag{8}$$

where $\underline{\underline{P}}$ is the power allocation matrix whose entry in the k-th row and m-th column, $P_{k,m}$, is the power allocation of the m-th subcarrier for user k. We drop the a_k term as it is constant with respect to $\underline{\underline{P}}$.

We assume that each user has a total power constraint of P over all subcarriers and any subcarrier is used by at most one user exclusively, so the feasible solutions for this problem satisfy the following two constraints:

(C1) For $m \in \{1, 2, 3...M\}$, if $\exists k'$ such that, if $P_{k',m} \neq 0$, then $P_{k,m} = 0$, $\forall k \neq k'$

(C2) $\sum_{m=1}^{M} P_{k,m} \leq P \ \forall k \in \{1, 2, 3...K\}$

For the optimization problem defined in (8), since constraint C1 is not a convex set, and this optimization problem is NP-hard, we propose an algorithm for a sub-optimal solution with two steps:

Step 1: The base station assigns subcarriers to different users according to channel conditions and rate-distortion curves;

Step 2: Given a subset of subcarriers, each user solves the optimization problem of minimizing its own distortion curve reduction under the power constraint;

We then iteratively update both the subcarrier assignments (according to the RD curve) and the power allocation strategy (based on the CSI). One of the major differences between our algorithm and those in [21] and [22] is that we allow application layer information and physical layer information to interact in our decision process. Before providing the details of the algorithm in Section IV, we first investigate a condition for the optimal solution in a continuous channel setting, where there can be variations within a subcarrier, as opposed to a block fading model. This condition inspires our algorithm.

III. CONTINUOUS FREQUENCY CHANNEL RESPONSE RESOURCE ALLOCATION ANALYSIS

We consider a system with only two users, $K = 2$, in a continuous channel setting. In this scenario, the allocator can divide the total frequency band B^{tot} into infinitely small bands for resource allocation. Note that $|H_k(f)|^2$ is the channel gain for user k at frequency f.

Let B_i be the frequency band assigned to user i. If we ignore the upper bound of the modulation alphabet size, the optimization problem becomes

$$\min_{\underline{\underline{P}}} \sum_{k=1}^{2} \frac{b_k}{\int_{\hat{B}_k} \log_2[1 + \eta P_k(f)|H_k(f)|^2]df + c_k} \tag{9}$$

subject to

(C1) $\hat{B}_1 \cap \hat{B}_2 = \emptyset, \left(\hat{B}_1 \cup \hat{B}_2\right) \subset B^{tot}$

(C2) $\int_{B_k} P_k(f)df \leq P, k = 1, 2$

Here, $P_k(f)$ is given by the water filling solution after the band allocation is decided. Given B^{tot}, the optimal band allocation can be viewed as a partition of the band $B^{tot} = B_1^{opt} \cup B_2^{opt} \cup B^{extra}$. Here, B_1^{opt} and B_2^{opt} are the optimal sets of frequency bands assigned to two users in the sense that the sum of distortions is minimized, and no frequency component in B_i^{opt} exceeds the water level of user i. The parameter B^{extra} is the set of bands not assigned to either user. We introduce the following definitions.

Definitions I:

a) Let $|.|$ be the bandwidth in Hz, e.g., $\left|B_1^{opt}\right|$ is the optimal bandwidth assigned to user 1.

b) Let $r_1 = \int_{B_1^{opt}} \log_2(1 + \eta P_1(f)|H_1(f)|^2)df$ and $r_2 = \int_{B_2^{opt}} \log_2(1 + \eta P_2(f)|H_2(f)|^2)df$ be the average optimal rate (in bits/sec) of two users. Here, $|H_1(f)|^2$ and $|H_2(f)|^2$ are the frequency channel responses of the two users. Let $P_1(f)$ and $P_2(f)$ be the power allocations which obey the water filling solution [25].

c) Define $W_1 = P_1(f) + \eta \frac{1}{|H_1(f)|^2}$ and $W_2 = P_2(f) + \eta \frac{1}{|H_2(f)|^2}$ to be the water levels for the two users at the optimal solution. See Fig. 3 and Fig. 4.

d) Let $\theta \in B_2^{opt}$ be an infinitesimally small band θ assigned to user 2. Note that $|H_1^\theta|^2$ and $|H_2^\theta|^2$ are the channel gains for user 1 and user 2 for band θ, respectively. They are constant since the band is infinitesimally small.

e) Let $\phi_i^\theta = \left(W_i - \eta \frac{1}{|H_i^\theta|^2}\right)^+$ be the non-negative distance between the water level of user i and the noise level of band $|H_i^\theta|^2$. By definition, $[x]^+ = x$ if $x > 0$ and $[x]^+ = 0$ if $x \leq 0$. For any frequency band θ of B_2^{opt}, $W_2 - \eta \frac{1}{|H_2^\theta|^2} > 0$. For user 2, the value of ϕ_2^θ is always positive.

Theorem 1: For a continuous frequency channel B^{tot}, the optimal band allocation of B_1^{opt} and B_2^{opt} for minimizing the sum of distortions should satisfy (10) at the top of next page, for any frequency band θ assigned to user 2. The proof of (10) can be found in Appendix A.

To find the optimal allocation in this cross layer problem, we wish to maximize the combination of application and physical layer metrics, which is the product of the absolute value of application layer RD slope, as seen in (10),

$$S_i = \frac{b_i}{(r_i + c_i)^2} \tag{11}$$

and the physical layer information given in (12).

For video RD characteristics in the form of $D_i = a_i + \frac{b_i}{r_i + c_i}$, (11) is the absolute value of the slope of the RD curve for user i at rate r_i. In this sense, for an allocation scheme to be optimal, the application layer contribution to the overall metric should be the slope of the curve instead of the distortion value [1], [2]. To solve the optimization problem of Section II, the algorithm should give priority to the user with the steepest slope. On the other hand, (12)

$$\frac{\frac{b_1}{(r_1+c_1)^2}\left\{\ln\left(1+\eta\phi_1^\theta|H_1^\theta|^2\right)-\int\limits_{B_1^{opt}}\frac{\eta|H_1(f)|^2}{|B_1^{opt}|(1+\eta P_1(f)|H_1(f)|^2)}\phi_1^\theta df\right\}}{\frac{b_2}{(r_2+c_2)^2}\left\{\ln\left(1+\eta\phi_2^\theta|H_2^\theta|^2\right)-\int\limits_{B_2^{opt}}\frac{\eta|H_2(f)|^2}{|B_2^{opt}|(1+\eta P_2(f)|H_2(f)|^2)}\phi_2^\theta df\right\}} \leq 1 \qquad (10)$$

$$\left\{\ln\left(1+\eta\phi_i^\theta|H_i^\theta|^2\right)-\int\limits_{B_i^{opt}}\frac{\eta\left|H_i(f)\right|^2}{\left|B_i^{opt}\right|\left(1+\eta P_i(f)\left|H_i(f)\right|^2\right)}\phi_i^\theta df\right\} \qquad (12)$$
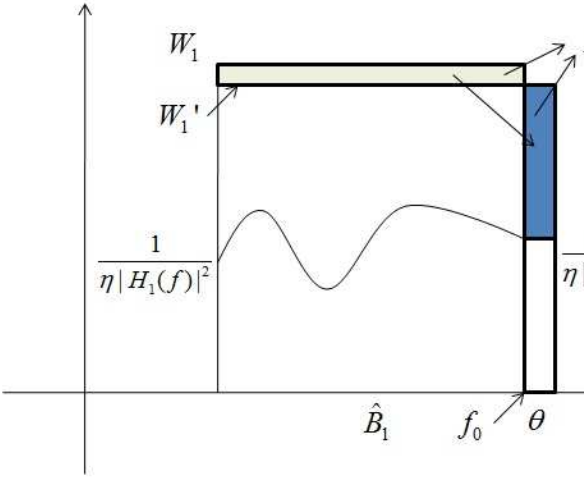

Fig. 3.   Water level change for user 1 gaining one band. The water level drops from W_1 to W_1' after user 1 gains one additional band θ.
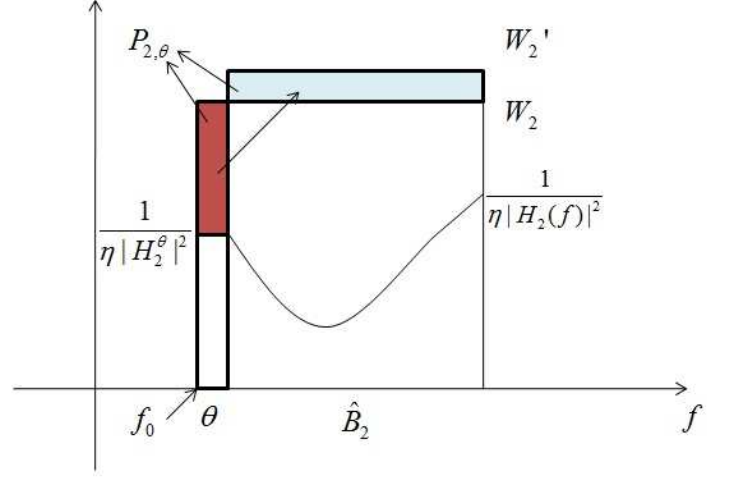

Fig. 4.   Water level change for user 2 loses one band. The water level raises from W_2 to W_2' after user 2 losing one band θ.

is an explicit relation between the physical layer rate (in bits/sec) and channel state information. As the bandwidth of θ becomes infinitesimally small, (12) can be considered as the marginal rate change (either increase or decrease) of switching a band from one user to the other. More specifically, one may treat $\ln\left(1+\eta\phi_i^\theta|H_i^\theta|^2\right)$ as the direct rate change caused by gaining or losing θ, and $\int\limits_{B_i^{opt}}\frac{\eta|H_i(f)|^2}{|B_i^{opt}|(1+\eta P_i(f)|H_i(f)|^2)}\phi_i^\theta df$ as the corresponding rate decrease or increase due to the effect of water level change. The optimal cross layer allocation would assign band θ to the user who has the maximum physical layer marginal rate increase given by (12) weighted by the slope of the RD curve.

To solve (8), given finite subcarrier bandwidths, the physical layer metric expression of (12) would not be valid, as the frequency bands are modeled as experiencing block fading. We thus design an iterative subcarrier allocation algorithm in the next section. Similar to the optimal condition derived in (10), the application layer metric is the slope of the RD curve. We will give users with steep slope priority to access

subcarriers. In the continuous channel response allocation analysis, the increment considered for switching between users was infinitesimal, whereas in the algorithm, the increment is the bandwidth of a single subcarrier.

IV. UPLINK RESOURCE ALLOCATION ALGORITHM

To find a solution to the problem defined in (8), we design an iterative algorithm which allows physical layer CSI and application layer RD information to interact. This algorithm first assigns the subcarriers purely based on channel conditions. However, it is possible that the overall performance (from an average distortion perspective) might be better if we assign some subcarriers to a user with worse channel conditions, but who might need a greater bit rate. We then try to reassign one subcarrier to the user with the steepest distortion curve slope. To solve a conventional video multiplexing bit rate allocation problem, a condition for a global optimum is that users operate at a rate with the same slope of their corresponding RD curves [1] [19] [26]. Note that at each iteration we only change the assignment of one
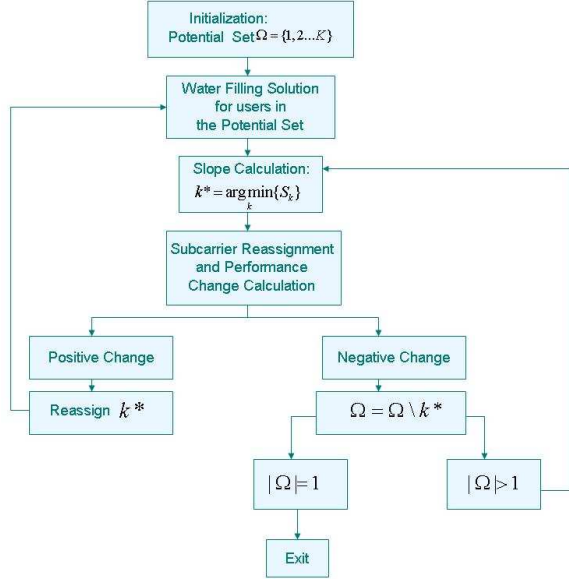
Fig. 5.   Uplink optimization algorithm flow chart.

subcarrier through a search process, and for every subcarrier which is not assigned to the user with the steepest slope, the calculations of distortion loss for the user losing that subcarrier and the performance improvement for the user with the steepest slope gaining that subcarrier is of low complexity. We then make the reassignment of the subcarrier that can most effectively reduce the overall distortion. We repeat this procedure iteratively until we run out of the possibility of reassigning subcarriers. We introduce the following definitions that will be used in the algorithm.

Definitions II:

a) Let $\rho_m^{(i)}$ denote the user who is assigned subcarrier m at the i-th iteration. For example, $\rho_2^{(1)} = 3$ means user 3 is assigned subcarrier 2 at the first iteration of the algorithm.

b) Define $A_k^{(i)}$ to be the set of subcarriers assigned to user k at the i-th iteration.

c) Define the potential set Ω as the set of users that have the potential to improve the average performance by receiving extra subcarriers, and define $|\Omega|$ as the cardinality of the potential set.

d) Define $\Delta_{k,m} \geq 0$ as the absolute value of the video distortion change of user k by gaining or losing subcarrier m.

Iterative Cross Layer Resource Allocation Algorithm:

Step (1) Initialization:

Initialize $\rho_m^{(0)} = \arg\max_k \{|H_{k,m}|^2\}$ for $m \in \{1,2,3...M\}$. Initialize the potential set $\Omega = \{1,2,3...K\}$. We first assign each subcarrier to the user who has the best channel response, and let the potential set be the total set.

Step (2) Water Filling and Slope Calculation:

After subcarrier assignment, each user tries to solve a MSE distortion minimization problem as follows:

$$\min_{P_{k,m}} \frac{b_k}{\sum_{m \in A_k^{(i)}} \log_2[1 + \eta P_{k,m}|H_{k,m}|^2] + c_k} \tag{13}$$

$$s.t. \sum_{m \in A_k^{(i)}} P_{k,m} \leq P \tag{14}$$

The optimization problem can be further simplified as

$$\max_{P_{k,m}} \sum_{m \in A_k^{(i)}} \log_2[1 + \eta P_{k,m}|H_{k,m}|^2] \tag{15}$$

The solution to this problem is the conventional power water filling allocation [25]

$$P_{k,m}^* = [\frac{1}{\lambda_k} - \frac{1}{\eta|H_{k,m}|^2}]^+, \forall m \in A_k^{(i)} \tag{16}$$

where λ_k can be found numerically to make the total power equal to P. Let

$$r_k^* = \sum_{m \in A_k^{(i)}} \log_2[1 + \eta P_{k,m}^*|H_{k,m}|^2] \tag{17}$$

be the optimal rate (in bits/symbol) user k gets using water filling. Then $S_k = \left. \frac{d \frac{b_k}{r_k + c_k}}{dr_k} \right|_{r_k = r_k^*} = -\frac{b_k}{(r_k^* + c_k)^2}$ is the slope of the k-th user's RD curve evaluated at the rate that user k is assigned. Let $k^* = \arg\min_{k \in \Omega} \{S_k\}$ be the user with the steepest slope in the potential set. This is the user who stands to benefit the most from receiving an increment of rate.

Step (3) Subcarrier Reassignment:

We consider taking one subcarrier away from some other users in Ω and reassigning it to user k^*, as user k^* has the largest marginal performance increment in the potential set. We consider each subcarrier $m \in \{1,2,3...M\}\backslash A_{k^*}^{(i)}$, which is not currently assigned to user k^*. We calculate the MSE performance change $-\Delta_{\rho_m^{(i)},m} < 0$ of user $\rho_m^{(i)}$ from losing one subcarrier, and the performance gain of the user k^*, $\Delta_{k^*,m} > 0$. The rate change for switching subcarriers is similar to the derivation for (12) in a continuous channel case, and the details of the calculation are in Appendix B. Since we only take one subcarrier from one user each time, the MSE performance loss and gain can be found analytically. We then find $m^* = \arg\max_{m \in \{1,2,3...M\}\backslash A_{k^*}^{(i)}} (\Delta_{k^*,m} - \Delta_{\rho_m^{(i)},m})$, which maximizes the performance change.

If $(\Delta_{k^*,m^*} - \Delta_{\rho_m^{(i)},m^*}) > 0$, we reassign subcarrier m^* to user k^* at iteration $i+1$, $\rho_m^{(i+1)} = k^*$, and return to Step (2) to update k^*.

If $(\Delta_{k^*,m^*} - \Delta_{\rho_m^{(i)},m^*}) < 0$, which means that the overall performance will not be enhanced by reassigning any subcarrier to user k^*, we update the potential set $\Omega = \Omega\backslash\{k^*\}$. User k^* is dropped from the potential set. User k^* will keep the subcarriers already assigned to him, but will not be assigned any additional subcarriers. Next, we check the cardinality of Ω. If $|\Omega| = 1$, we stop, otherwise, i is incremented, and we go back to step (2) to update k^*.

Based on our analysis in Section III, from the perspective of minimizing the sum of distortions, the subcarrier assignment balances both the application layer metric (the slope of the RD curve) and the physical layer metric. The initialization step, which is purely based on the physical layer metric, will most likely mismatch the optimal criteria we described in (12). The idea of reassignment is that, when we are using the iterative method to allocate limited resources, the user operating at the steepest rate distortion curve has the priority to be assigned extra subcarriers.

V. BASELINE ALGORITHMS

We compare the performance of our cross layer optimization algorithm to two baseline algorithms, one with only application layer RD information and the other with only physical layer CSI available for resource allocation at the base station.

A. Application Layer Optimization Algorithm

The application layer optimization allocates subcarriers purely based on the RD information of the video streams. Since CSI is not used, the allocator will treat all subcarriers the same when making the allocation decision. As we will see in the numerical results, to determine the number of subcarriers assigned to each user, we first choose a PSNR value (e.g., PSNR=28dB). To achieve this PSNR, user k needs video coding rate r_k based on his RD information. The number of subcarriers assigned to the k-th user is proportional to the k-th user's rate, or

$$L_k \sim M \cdot \frac{r_k}{\sum_i r_i} \qquad (18)$$

where M is the total number of subcarriers in the system. Subcarriers are then randomly assigned to users.

After being informed of the resource allocation decision, we assume that users know their CSI, and can use it to select their modulation and coding scheme. In other words, CSI is not used for resource allocation, but is used to determine the transmitted waveform. Similar to the cross layer and algorithm, user k conducts a water filling calculation for transmission power assignment, and the modulation format is chosen based on (4) for each subcarrier. The source encoding rate is then determined using (6).

B. Physical Layer Optimization Algorithm

Suppose $[H_{1,m}, H_{2,m}...H_{K,m}]$ is the vector of channel gains of users $\{1,2...K\}$ at subcarrier m. Similar to conventional resource allocation based on multi-user diversity (MUD), we assign subcarrier m to the user k^*, where

$$k^* = \arg\max_k \left\{ \frac{|H_{k,m}|^2}{\overline{|H_k|}^2} \right\} \qquad (19)$$

And, $\overline{|H_k|}^2 = \frac{1}{M} \sum_{m=1}^{M} |H_{k,m}|^2$. After subcarrier assignment, every user would apply water-filling to allocate power to each assigned subcarrier.

Define B_c to be the coherence bandwidth of the system. For simplicity, we assume that the coherence bandwidth is always an integer multiple of the subcarrier bandwidth, i.e., $B_c = \Psi W/M, \Psi \in \mathbb{Z}^+$ in the simulation. Further, we assume that channel gains are identical within the coherence bandwidth, but independent between different coherence bands.

For a system with coherence bandwidth larger than the subcarrier bandwidth, i.e., $\Psi > 1$, an MUD based algorithm proposed by [9] [10] allocates subcarriers in chunks, i.e., if a given user is assigned a particular subcarrier, that user will also get all the other subcarriers in the chunk. For a system using MUD with large Ψ, since individual users could get multiple chunks with large bandwidth, the resource allocation might be unbalanced and the average video performance will suffer a large degradation. To avoid a scenario where a small set of users dominate the use of the subcarriers, we design an algorithm that limits the number of subcarriers assigned to each user.

Definitions III:
a) Define Λ as the set of users who are eligible for being assigned additional subcarriers;
b) Define Θ as the set of users who have not been assigned any subcarrier yet in the iteration. We design the algorithm such that each user will get at least one subcarrier;
c) Define Γ as the set of subcarriers whose allocation decision has not been made yet;
d) Similar to the application layer optimization algorithm, let L_k be the number of subcarriers user k is assigned. To control the degree of imbalance in the number of subcarriers that users receive, we impose set of thresholds of $\psi_n, n = 1, 2...K - 1$, such that the sum of subcarriers for any group of n users will not exceed ψ_n.

We set ψ_n, for $1 \leq n \leq K - 1$, equal to

$$\psi_n = \psi_{n-1} + \left\lceil \epsilon \left(\frac{M - \psi_{n-1}}{K - (n-1)} \right) \right\rceil \qquad (20)$$

where, for $n = 1$, this expression reduces to $\psi_1 = \left\lceil \epsilon \frac{M}{K} \right\rceil$. In (20), the parameter ϵ is chosen to be greater than or equal to 1, and controls the imbalance of the resource allocation. A larger value of ϵ means that the resource allocation decision will be more unbalanced, biased to the users who have larger channel gains. For each individual user, the number of subcarriers threshold ψ_1 is set to be ϵ times larger than the average number of subcarriers per user M/K subcarriers. Assuming that one user has already been assigned the maximum of $\psi_1 = \left\lceil \epsilon \frac{M}{K} \right\rceil$ subcarriers, the average number of subcarriers for the remaining $(K - 1)$ users is given by $(M - \psi_1)/(K - 1)$ and the resource for any combination of two users is limited by $\psi_1 + \lceil \epsilon(M - \psi_1)/(K - 1) \rceil$ subcarriers. We repeat this process iteratively for $n \leq (K - 1)$, and the total number of subcarriers assigned to any group of n users can be found iteratively using (20). As a specific example, consider a system with 1000 subcarriers, 3 users, and $\epsilon = 1.5$. The threshold would be $\psi_1 = 500$ subcarriers for each individual user, and $\psi_2 = 875$ subcarriers for any group of two users. When the coherence bandwidth is equal to the entire bandwidth, the user with the strongest channel gain will get 500 subcarriers. Any group of two users cannot get more than 875 subcarriers, so the user with second best channel gain gets 375 subcarriers. The remaining 125 subcarriers are assigned to the third user. When the coherence bandwidth becomes smaller, it will be
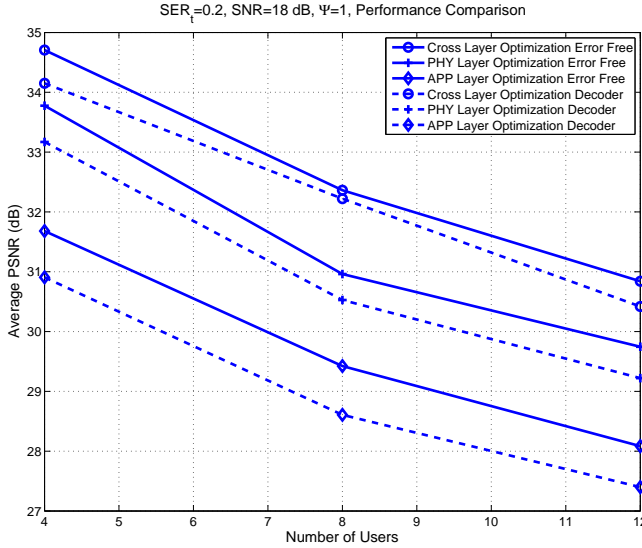
Fig. 6.   Video PSNR performance vs. number of users. 0.2 target error, 16 subcarriers, $\Psi=1$, average SNR=18dB if only one subcarrier is assigned, users for PHY layer optimization are limited to be assigned at most 1.5 times the average remaining resources.

increasingly unlikely that the total number of subcarriers for a group of n users will reach the threshold of ψ_n.

Physical Layer Optimization Algorithm:

Step 1 Initialization: We initialize Λ and Θ as the complete set of users, i.e. $\Lambda = \{1, 2, ...K\}$, $\Theta = \{1, 2, ...K\}$, Γ as the complete set of subcarriers $\Gamma = \{1, 2, ...M\}$ and ψ_n is given by (20).

Step 2 Subcarrier Assignment: We choose the best channel gain from all the possible assignments,

$$(k^*, m^*) = \arg\max_{k \in \Lambda, m \in \Gamma} \left\{ \frac{|H_{k,m}|^2}{\overline{|H_k|}^2} \right\} \qquad (21)$$

with $\overline{|H_k|}^2 = \frac{1}{M} \sum_{m=1}^{M} |H_{k,m}|^2$, and assign subcarrier m^* to user k^*. We update $\Gamma = \Gamma \setminus m^*$. If $k^* \in \Theta$, we update $\Theta = \Theta \setminus k^*$, meaning that user k^* has been assigned at least one subcarrier. Here, $|H_{k^*,m^*}|^2$ stands for the best channel response in all possible subcarrier assignment combinations at the current step.

Step 3 Status Update: We check the remaining resource and conduct the following two updates:

1) For every n, $(1 \leq n \leq K-1)$, we compare the sum of subcarriers for all groups of n users with ψ_n. If the sum is equal to ψ_n for any group, all the users in that group will be excluded from Λ.

2) We then check the relation between the number of subcarriers left and the cardinality of Θ. To ensure that each user can get at least one subcarrier, if $|\Gamma| = |\Theta|$, we will terminate the algorithm by assigning exactly one of the remaining unallocated subcarriers to each of the users who has no subcarrier yet using (21). We then go back to Step.2 and repeat (21) to assign subcarriers until Γ is empty.

VI. RESULTS

A. Simulation Setup

We study an uplink OFDM system with 16 subcarriers, each with a bandwidth of 50 kHz. We evaluate performance by the Peak-Signal-to-Noise Ratio (PSNR), defined as $PSNR = 10\log_{10}\frac{255 \times 255}{MSE}$. For all three optimization algorithms, the modulation decision will be rounded down to a valid integer value corresponding to a modulation format of MQAM, with M=4, 8, 16, 32, 64, 128 or 256 based on (4). For example, if the cross layer allocation assigns a rate of any real value $R_{1,4} \in [3, 4)$ for user 1 on subcarrier 4, the actual alphabet size would be 8-QAM. The channel response consists of both the path loss and multi-path fading, and the magnitude square of the channel can be written as $|H_{k,m}|^2 = \alpha^2 \cdot K_0 \cdot (\frac{d_0}{d_k})^\gamma$ [25], where $\gamma = 2.4$ is the path-loss exponent [25]. d_k is the distance between user k and the base station, and d_0 is a reference distance set to 10m [25]. In addition, α is assumed to be a Rayleigh random variable, and K_0 is a constant of -24dB. We assume that the distance d_k between user k and the base station is a random variable, and follows a uniform distribution between $[30, 120]$ meters. For the user who is 75 meters away from the base station, the average SNR is assumed to be 18 dB if only one subcarrier is assigned. Unless otherwise stated, the subcarriers are assumed to fade independently. For the physical layer optimization algorithm, we set $\epsilon = 1.5$, which means that one user cannot be assigned more than 150% of the average number of subcarriers.

For all three optimization schemes, we use a rate 1/2 convolutional code with code generator polynomial [23, 35] in octal, and the coded bits are interleaved across different subcarriers. For example, if one user gets three subcarriers, the first coded bit goes to the first subcarrier, the second coded bit goes to the second subcarrier, etc. We use log-likelihood ratio demodulation to detect each bit of the QAM symbol. We then decode the bitstream using soft-decision decoding with eight reliability ranges.

We use a sequence of CIF videos of total length 50 seconds at 30 frames per second. Compression is by the baseline profile of H.264/AVC reference software JM 11.0 [27]. The GOP size is 15 frames (I-P-P-P) and the frames inside one GOP are encoded using H.264 rate control. We encode each GOP at rates of 80, 100, 120, 140, 160, 180, 200, 220, 240, 280, 300, 340, 380, 420, 460, 500 and 600 kbps, and use these operational points to fit the rate distortion function $D(R) = a_k + b_k/(R + c_k)$ by nonlinear regression. We randomly assign different starting points of the same video to different users, and the resource allocation decision is done in every GOP. The video in the simulation is a travel documentary which consists of both high motion and low motion GOP's. By assigning random starting points of the same cyclic video to different users, we create application layer diversity among users and yet have the same average complexity over time for different users. Each video is encoded at 10 slices per frame, and any channel error will make the system lose the entire slice. At the decoder side, slice copying conceals losses.

B. Systems with Different SER_t

As discussed in Section II, the uplink resource allocation algorithm needs a target error rate; we used $SER_t = 0.2$, and varied the number of users from 4 to 12 in the system of 16 subcarriers. Fig. 6 shows the performance of the three optimization algorithms. The solid lines represent the numeric results obtained from the RD curves. That is, the resource allocator decides the rate for each user, and the distortion is calculated directly from $D(R)$. This can be considered the error-free distortion, or distortion at the encoder side. The dashed lines are the distortion results measured at the decoder; the videos are reconstructed from the bitstream corrupted by the channel. The effects of packet loss, errors in RD curve fitting, and imperfection of encoder rate control are included in the simulation.

With SER_t set to 0.2, we find that the decoded bit error rate is small, and distortion curves at the encoder and decoder are close. Comparing the performance of these three algorithms, we see that when the number of users in the system is small, the physical layer optimization outperforms the application layer optimization algorithm, and the gap between the cross layer and the physical layer algorithms is relatively small. When the system has abundant resources so each user can be assigned several subcarriers, both the cross layer optimization algorithm and the physical layer optimization algorithm will allow users to operate at a high data rate, or in the flat region of the convex RD curve. Utilizing application RD information in the resource allocation will thus not benefit the overall users' performance by much. Fig. 7 shows a sample of the performance for individual users in systems with different numbers of users employing the cross layer algorithm. In the first row of the plot, we see that all four users are operating near the right end of their RD curves, and the slopes of users are relatively small. When the average resource for each user gets smaller, the users are forced to operate at steeper parts of the RD curves (see the second and third rows of Fig. 7). As we increase the number of users in the system to 8 and 12, the gap between the cross layer and the physical layer algorithms widens. We conjecture this is because source characteristics play a more important role when many users compete for the available resources. For a system with large number of users, it becomes important to combine the information of CSI and RD in the system design for a resource-scarce system, as most users operate on the steep slope of their individual RD curve. Mismatch of the physical layer resource with the RD curve would cause a large loss of system performance.

As seen in Fig. 6, when the system has 12 users, cross layer optimization outperforms physical layer optimization by about 1.25 dB, and the gap to application layer optimization is even larger. For a system with average PSNR of 30.5 dB, the cross layer scheme can support 12 users, compared to 8 users for physical layer optimization and less than 5 users for application layer optimization. In this sense, the cross layer algorithm can almost increase the capacity (the number of users a system can support) by 50%.

We now change the value of SER_t to 0.1 and 0.25 (Fig. 8 and 9, respectively). When we set $SER_t = 0.1$, the
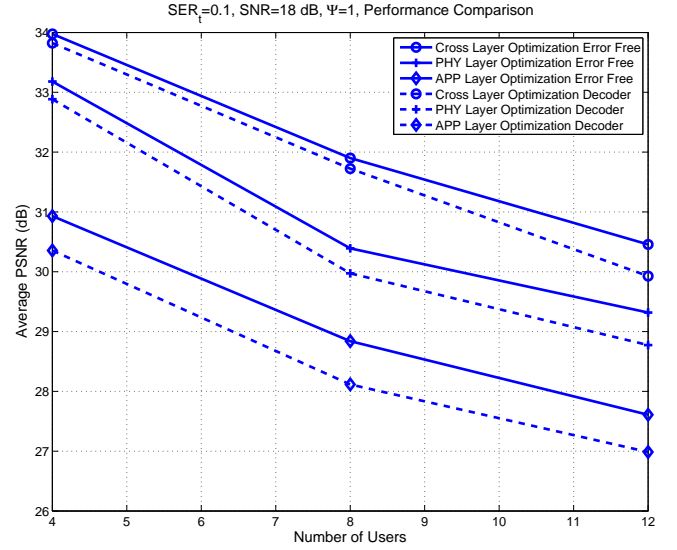

Fig. 8. Video PSNR performance vs. number of users. 0.1 target error, 16 subcarriers, $\Psi=1$, average SNR=18dB if only one subcarrier is assigned.
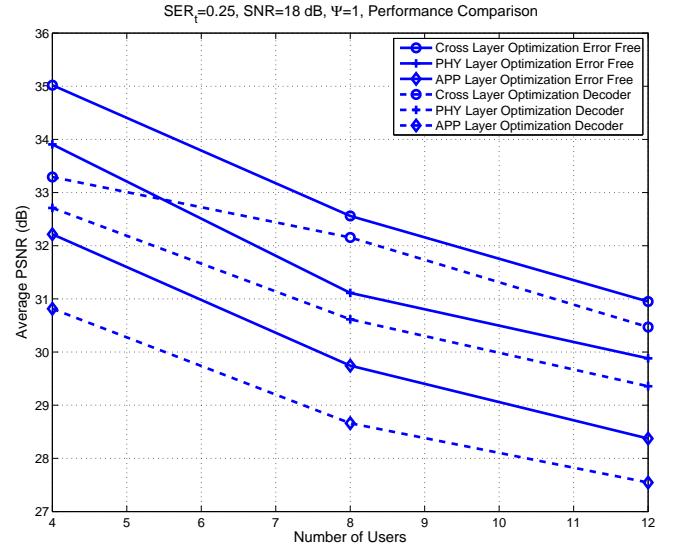

Fig. 9. Video PSNR performance vs. number of users. 0.25 target error, 16 subcarriers, $\Psi=1$, average SNR=18dB if only one subcarrier is assigned.

modulation alphabet size will be chosen more conservatively and thus force the video source encoding rate to be smaller than for $SER_t = 0.2$. On the other hand, a high SER_t value will lead to a relatively large gap between the error free curves and curves for PSNR performance at the decoder side, and we see that for $SER_t = 0.25$, the impact of channel errors has significantly decreased the throughput of the system and the PSNR of the video from the error free scenarios. Comparing the performance of the three algorithms, we see a similar performance gain of adopting cross layer optimization, and the capacity gain by adopting the cross layer algorithm is still around 1.5.
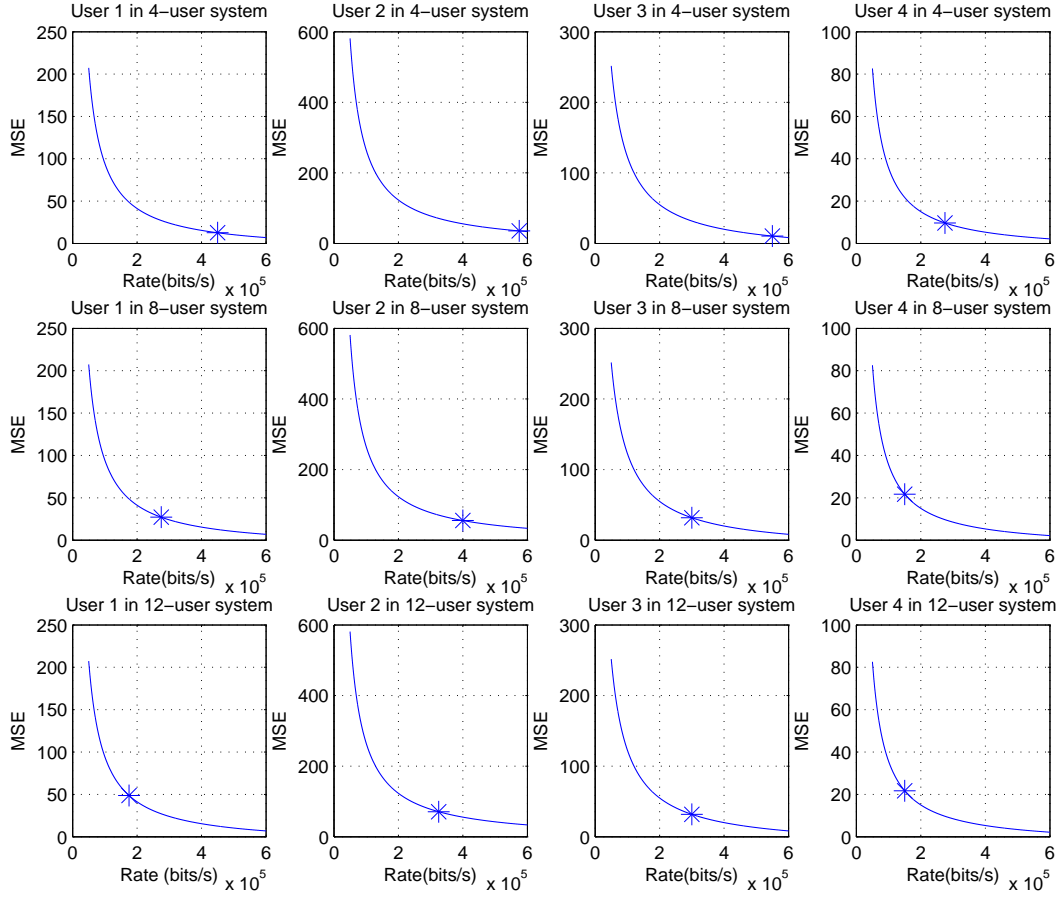
Fig. 7. Individual user's performance in systems with different number of users. Each column indicates the same user's RD relations in systems with four, eight and twelve users.

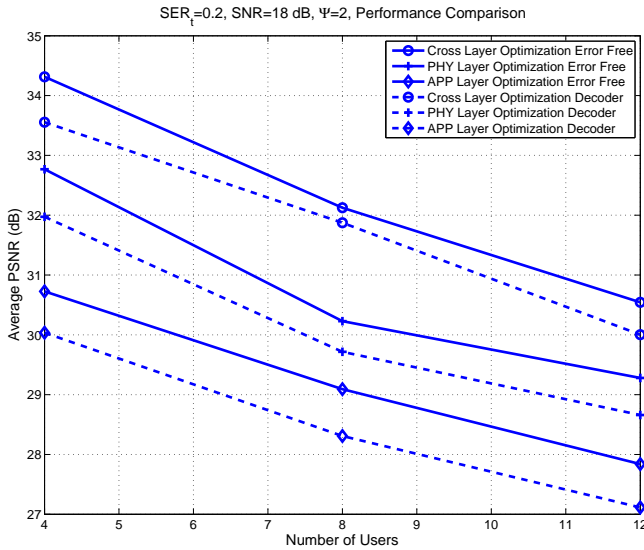## C. Systems with Different Coherence Bandwidths

Fig. 10. Video PSNR performance vs. number of users. 0.2 target error, 16 subcarriers, $\Psi=2$, average SNR=18dB if only one subcarrier is assigned.

In Fig. 10, we set $\Psi = 2$. For simplicity in the simulation, we assume that two adjacent subcarriers have the same realization, and the correlation coefficient between different coherence bands is zero. We observe a very slight performance degradation for both the cross layer and the application layer optimization algorithms. From the results shown in Fig. 6, which are shown in (11) and (12), since the cross layer optimization algorithm exploits both physical layer multiuser channel diversity and application layer RD diversity, increasing the coherence bandwidth will not affect the cross layer optimization's ability to utilize the application layer diversity. Similarly, for application layer optimization, increasing the coherence bandwidth will not change the number of subcarriers assigned to each user, and the performance loss is very limited. On the other hand, compared to the scenario of $\Psi = 1$, we see a large performance degradation for the physical layer optimization. As subcarriers will have the same fading realization in groups of two, we lose half of the frequency diversity. Since physical layer optimization does not exploit any application layer diversity, losing frequency diversity at the physical layer will have a big impact on the system performance. If we further increase Ψ to four,

as shown in Fig. 11, the performance for the physical layer optimization will further decrease, while the performance of the proposed cross layer algorithm is still robust. Comparing the performance between different algorithms, we see that when the cross layer optimization can support 12 users with an average PSNR of about 29.5dB, the baseline algorithm can at most support 7 users.

D. Complexity of Iterative Water Filling Algorithm

To show the path of performance improvement from initialization to convergence of the cross layer algorithm, in Fig. 12 we plot the average MSE for the systems with different numbers of users versus the iteration number. To obtain the plot, we observe the MSE values after each iteration for each individual user and average over the entire video sequence and all the users. The iteration number equal to one corresponds to the performance of the initialization step. Because of the greediness of the algorithm, the biggest performance improvement occurs in the first few iterations, and MSE curves appear to be convex. After the eighth step, we see a very small performance improvement. As shown in Appendix B, since we can find the performance improvement switching subcarrier at each iteration analytically, the overall complexity of the proposed algorithm is much lower than that of an exhaustive search.

VII. Conclusion

We proposed a cross layer resource allocation framework for transmitting video in an uplink OFDMA setting, and derived an optimality condition for the bandwidth allocation in continuous frequency response channel. The power allocation and subcarrier assignment strategy are jointly decided by each user's CSI and RD characteristics. Our analytical results show that the optimal allocation is achieved only if the product of the RD slope and a physical layer metric related to the water filling solution, given by (12), is minimized for each band and each user. With a similar technique of switching bandwidth increments as in the analysis, we designed an iterative resource allocation algorithm. At each iteration, our algorithm first evaluates the application layer metric defined by (11), and then greedily updates the resource allocation decision jointly according to (11) and (12). Compared to a resource allocation using either only application layer or only physical layer information, for the same video performance, the cross layer optimization significantly increased the capacity of the system, and resulted in robust performance as the coherence bandwidth changed, over the range of parameter values considered in our numerical results.

Appendix A
Optimality Condition for Continuous Channel Allocation Solution

If an assignment is optimal, any reassignment will not decrease the sum of distortions. Let B_1^{opt}, B_2^{opt} be the optimal assignment, and let $B_1^{opt} \cup \theta$, $B_2^{opt} - \theta$ be a new assignment
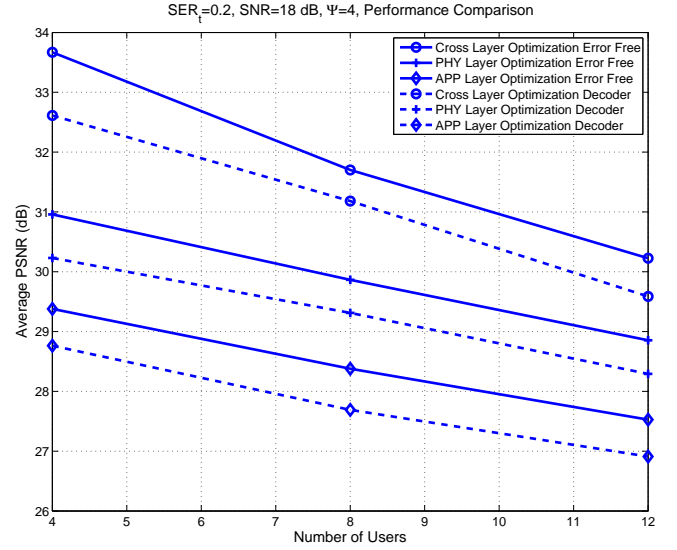

Fig. 11. Video PSNR performance vs. number of users. 0.2 target error, 16 subcarriers, $\Psi=4$, average SNR=18dB if only one subcarrier is assigned.
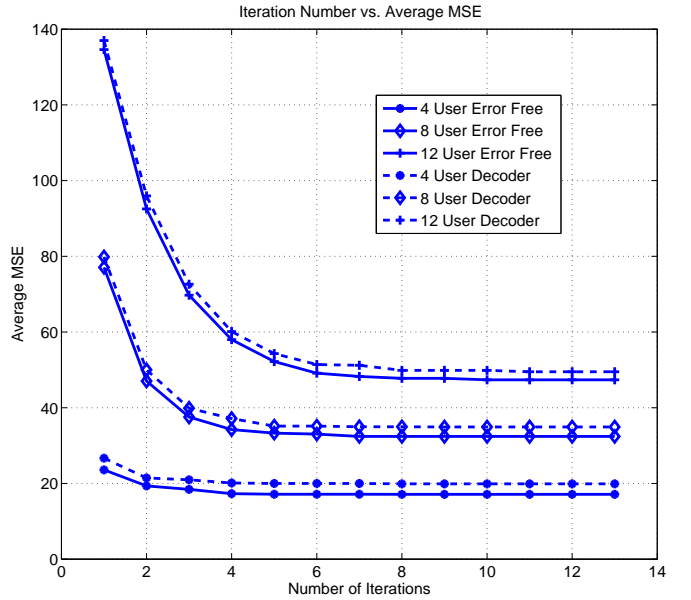

Fig. 12. Average MSE vs. number of iterations.

which reassigns band θ to user 1. If an assignment is optimal, then

$$\frac{b_1}{r_1 + c_1} + \frac{b_2}{r_2 + c_2} \leq \frac{b_1}{(r_1 + \Delta r_1) + c_1} + \frac{b_2}{(r_2 - \Delta r_2) + c_2} \tag{A.1}$$

where Δr_1 and Δr_2 are the rate changes caused by switching band θ. We have two scenarios.

Scenario A: $W_1 > \eta \frac{1}{|H_1^\theta|^2}$ or $\phi_1^\theta > 0$

In this case, user 1 would get positive rate gain by acquiring the additional band θ. In other words, $\Delta r_1 > 0$ and $\Delta r_2 > 0$. Continuing from (A.1), we can go one step further and get

$$\frac{b_1}{\left(r_1 + c_1\right)^2 + \Delta r_1 \left(r_1 + c_1\right)}\Delta r_1$$

$$\leq \frac{b_2}{\left(r_2 + c_2\right)^2 - \Delta r_2 \left(r_2 + c_2\right)}\Delta r_2 \quad \text{(A.2)}$$

Because $D(R) = a_k + \frac{b_k}{R + c_k}$ is strictly convex, (A.1) must be satisfied as we take $|\theta| \to 0$. It is easy to see that as $|\theta| \to 0$, $\frac{\Delta r_i}{r_i + c_i} \to 0$ for $i = 1$ and 2. We thus can drop the $\Delta r_i \left(r_i + c_i\right)$ terms, as they will be small compared to the squared term. So the optimal condition is negligible.

$$\frac{b_1}{\left(r_1 + c_1\right)^2}\Delta r_1 \leq \frac{b_2}{\left(r_2 + c_2\right)^2}\Delta r_2 \quad \text{(A.3)}$$

Now, we are interested in finding $\lim_{|\theta| \to 0} \frac{\Delta r_1}{\Delta r_2}$, which is the ratio of rate change as $|\theta| \to 0$. Again, in the new frequency assignment, user 1 gets $B_1^{opt} \cup \theta$ and user 2 gets $B_2^{opt} - \theta$. Fig. 3 shows the power redistribution after switching band θ. $P_{1,\theta}$ is the total power user 1 will put over band θ after the reassignment. Since we consider $|\theta| \to 0$, $P_{1,\theta}$ is collected uniformly from B_1^{opt} and redistributed uniformly over band θ. Note that $W_1 = P_1(f) + \frac{1}{|H_1(f)|^2}$ is the water level of user 1 before reallocation, and W_1' is the level after reallocation. We then have

$$W_1 - \frac{P_{1,\theta}}{\left|B_1^{opt}\right|} = \eta\frac{1}{|H_1^\theta|^2} + \frac{P_{1,\theta}}{|\theta|} = W_1' \quad \text{(A.4)}$$

where $|H_1^\theta|^2 = |H_1(f_0 + \frac{|\theta|}{2})|^2$ is the channel response over the band θ, and f_0 is the left limit of θ. To go further, we have

$$P_{1,\theta} = \left(W_1 - \eta\frac{1}{|H_1^\theta|^2}\right)\left(\frac{\left|B_1^{opt}\right||\theta|}{|\theta| + \left|B_1^{opt}\right|}\right) \quad \text{(A.5)}$$

Before reallocation, the rate for user 1 is: $\int_{B_1^{opt}} \log_2\left(1 + \eta P_1(f)|H_1(f)|^2\right) df$. After getting θ, the new rate is given by

$$\int_{B_1^{opt}} \log_2\left(1 + \eta\left(P_1(f) - \frac{P_{1,\theta}}{\left|B_1^{opt}\right|}\right)|H_1(f)|^2\right) df +$$

$$|\theta| \log_2\left(1 + \eta\frac{P_{1,\theta}}{|\theta|}|H_1^\theta|^2\right) \quad \text{(A.6)}$$

We then can calculate the rate difference as

$$\Delta r_1 = \int_{B_1^{opt}} \log_2\left(1 - \frac{\eta P_{1,\theta}|H_1(f)|^2}{\left|B_1^{opt}\right|\left(1 + \eta P_1(f)|H_1(f)|^2\right)}\right) df +$$

$$|\theta| \log_2\left(1 + \eta\frac{P_{1,\theta}}{|\theta|}|H_1^\theta|^2\right) \quad \text{(A.7)}$$

Similar to the setting for user 1, $P_{2,\theta}$ is the power allocation for band θ and $|H_2^\theta|^2$ is the frequency response of user 2 over this band. Fig. 4 shows the power redistribution after reallocation for user 2, and we can calculate $P_{2,\theta}$,

$$W_2 - \eta\frac{1}{|H_2^\theta|^2} = \frac{P_{2,\theta}}{|\theta|} \quad \text{(A.8)}$$

as well as the absolute value of the rate change given by (A.9) at the top of next page.

We are interested in finding the ratio between the rate changes of these two user expressed as (A.10). We then use L'Hopital's rule and obtain (A.11).

From (A.3) and (A.11), for a two-user uplink video transmission scenario, the optimal frequency and power allocation scheme should satisfy (A.12) for any frequency band θ assigned to user 2.

Scenario B: $W_1 \leq \eta\frac{1}{|H_1^\theta|^2}$ or $\phi_1^\theta = 0$

This condition means that when we try to switch band θ from user 2 to user 1, the frequency response of user 1 over this band does not exceed the original water level, and the optimal solution will not put any power into this band. In this case $\Delta r_1 = 0$, $\Delta r_2 > 0$, and $\phi_1^\theta = 0$. If we plug $\phi_1^\theta = 0$ into the numerator of (A.11), we have (A.13).

Combining both scenarios, for a two-user uplink video transmission scenario, the optimal frequency and power allocation scheme should satisfy (A.12).

Similarly, in a system with an arbitrary number of users, it is easy to conclude that, for frequency band θ to be assigned to user j, the (A.14) condition must be satisfied for any user $i \neq j$.

APPENDIX B
PERFORMANCE CHANGE CALCULATION FOR REASSIGNING SUBCARRIERS

Consider a user k who gets assigned a set of $A_k^{(i)}$ subcarriers. As discussed in Step (2) of Section IV, the optimal power allocation scheme is

$$P_{k,m}^* = \left[\frac{1}{\lambda_k} - \frac{1}{\eta|H_{k,m}|^2}\right]^+ \quad \text{(B.1)}$$

$\forall m \in A_k^{(i)}$. We want to find the video performance degradation of user k after losing a subcarrier $\hat{m}$, $\hat{m} \in A_k^{(i)}$, $\frac{1}{\lambda_k} > \frac{1}{\eta|H_{k,\hat{m}}|^2}$. For the scenario that all the subcarriers' frequency responses are below the water level, or $\frac{1}{\lambda_k} > \frac{1}{\eta|H_{k,m}|^2}$, the operating rate (in bits/symbol) of user k is given by

$$r_k^* = \sum_{m \in A_k^{(i)}} \log_2\left[1 + \eta|H_{k,m}|^2\left(\frac{1}{\lambda_k} - \frac{1}{\eta|H_{k,m}|^2}\right)\right] \quad \text{(B.2)}$$

Note that we start the resource allocation by assigning the subcarrier to the user with the best response, so we expect that the condition of $\frac{1}{\lambda_k} > \frac{1}{\eta|H_{k,m}|^2}$ holds for most of the subcarriers at the beginning of the iterations. The video distortion is

$$D_k = a_k + \frac{b_k}{\sum_{m \in A_k^{(i)}} \log_2\left[1 + \eta|H_{k,m}|^2\left(\frac{1}{\lambda_k} - \frac{1}{\eta|H_{k,m}|^2}\right)\right] + c_k} \quad \text{(B.3)}$$

After losing subcarrier $\hat{m}$, the water level will increase by

$$\left(\frac{1}{\lambda_k} - \frac{1}{\eta|H_{k,\hat{m}}|^2}\right)\Bigg/\left(\left|A_k^{(i)}\right| - 1\right) \quad \text{(B.4)}$$

and the updated video distortion is expressed as (B.5) on the next page.

We can then calculate the performance change of user k for losing subcarrier $\hat{m}$ as $-\Delta_{k,\hat{m}} = \widehat{D}_k - D_k$. If a user k

$$\Delta r_2 = |\theta| \log_2\left(1 + \eta \frac{P_{2,\theta}}{|\theta|}|H_2^\theta|^2\right) - \int_{B_2^{opt}} \log_2\left(1 + \frac{\eta P_{2,\theta}|H_2(f)|^2}{|B_2^{opt}|(1 + \eta P_2(f)|H_2(f)|^2)}\right) df \tag{A.9}$$

$$\lim_{|\theta|\to 0} \frac{\Delta r_1}{\Delta r_2} = \lim_{|\theta|\to 0} \frac{\int_{B_1^{opt}} \log_2\left(1 - \frac{\eta P_{1,\theta}|H_1(f)|^2}{|B_1^{opt}|(1+\eta P_1(f)|H_1(f)|^2)}\right) df + |\theta| \log_2\left(1 + \eta \frac{P_{1,\theta}}{|\theta|}|H_1^\theta|^2\right)}{|\theta| \log_2\left(1 + \eta \frac{P_{2,\theta}}{|\theta|}|H_2^\theta|^2\right) - \int_{B_2^{opt}} \log_2\left(1 + \frac{\eta P_{2,\theta}|H_2(f)|^2}{|B_2^{opt}|(1+\eta P_2(f)|H_2(f)|^2)}\right) df} \tag{A.10}$$

$$\lim_{|\theta|\to 0} \frac{\Delta r_1}{\Delta r_2} = \frac{\ln\left(1 + \eta \phi_1^\theta |H_1^\theta|^2\right) - \int_{B_1^{opt}} \frac{\eta|H_1(f)|^2}{|B_1^{opt}|(1+\eta P_1(f)|H_1(f)|^2)}\phi_1^\theta df}{\ln\left(1 + \eta \phi_2^\theta |H_2^\theta|^2\right) - \int_{B_2^{opt}} \frac{\eta|H_2(f)|^2}{|B_2^{opt}|(1+\eta P_2(f)|H_2(f)|^2)}\phi_2^\theta df} \tag{A.11}$$

$$\frac{\frac{b_1}{(r_1+c_1)^2}\left\{\ln\left(1 + \eta \phi_1^\theta |H_1^\theta|^2\right) - \int_{B_1^{opt}} \frac{\eta|H_1(f)|^2}{|B_1^{opt}|(1+\eta P_1(f)|H_1(f)|^2)}\phi_1^\theta df\right\}}{\frac{b_2}{(r_2+c_2)^2}\left\{\ln\left(1 + \eta \phi_2^\theta |H_2^\theta|^2\right) - \int_{B_2^{opt}} \frac{\eta|H_2(f)|^2}{|B_2^{opt}|(1+\eta P_2(f)|H_2(f)|^2)}\phi_2^\theta df\right\}} \leq 1 \tag{A.12}$$

$$\frac{\frac{b_1}{(r_1+c_1)^2}\left\{\ln\left(1 + \eta \phi_1^\theta |H_1^\theta|^2\right) - \int_{B_1^{opt}} \frac{\eta|H_1(f)|^2}{|B_1^{opt}|(1+\eta P_1(f)|H_1(f)|^2)}\phi_1^\theta df\right\}}{\frac{b_2}{(r_2+c_2)^2}\left\{\ln\left(1 + \eta \phi_2^\theta |H_2^\theta|^2\right) - \int_{B_2^{opt}} \frac{\eta|H_2(f)|^2}{|B_2^{opt}|(1+\eta P_2(f)|H_2(f)|^2)}\phi_2^\theta df\right\}} = 0 < 1 \tag{A.13}$$

$$\frac{\frac{b_i}{(r_i+c_i)^2}\left\{\ln\left(1 + \eta \phi_i^\theta |H_i^\theta|^2\right) - \int_{B_i^{opt}} \frac{\eta|H_i(f)|^2}{|B_i^{opt}|(1+\eta P_i(f)|H_i(f)|^2)}\phi_i^\theta df\right\}}{\frac{b_j}{(r_j+c_j)^2}\left\{\ln\left(1 + \eta \phi_j^\theta |H_j^\theta|^2\right) - \int_{B_j^{opt}} \frac{\eta|H_j(f)|^2}{|B_j^{opt}|(1+\eta P_j(f)|H_j(f)|^2)}\phi_j^\theta df\right\}} \leq 1 \tag{A.14}$$

$$\widehat{D}_k = a_k + \frac{b_k}{\sum_{m\in\left(A_k^{(i)}-\hat{m}\right)} \log_2\left[1 + \eta |H_{k,m}|^2\left(\frac{\left|A_k^{(i)}\right|}{\left(\left|A_k^{(i)}\right|-1\right)\lambda_k} - \frac{1}{\eta|H_{k,m}|^2} - \frac{1}{\eta|H_{k,\hat{m}}|^2\left(\left|A_k^{(i)}\right|-1\right)}\right)\right] + c_k} \tag{B.5}$$

is given one subcarrier, the performance improvement for that user can be found in a similar way.

REFERENCES

[1] A. Ortega, K. Ramchandran, and M. Vetterli, "Optimal trellis-based buffered compression and fast approximations," *IEEE Trans. Image Process.*, vol. 3, no. 1, pp. 26–40, Jan. 1994.

[2] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers [speech coding]," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 9, pp. 1445–1453, Sep. 1988.

[3] C. Y. Wong, R. Cheng, K. Lataief, and R. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[4] Y. Yao and G. Giannakis, "Rate-maximizing power allocation in OFDM based on partial channel knowledge," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1073–1083, May 2005.

[5] W. Yu and J. Cioffi, "FDMA capacity of Gaussian multiple-access channels with ISI," *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 102–111, Jan. 2002.

[6] Z. Shen, J. Andrews, and B. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2726–2737, Nov. 2005.

[7] D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Trans. Wireless Commun.*, vol. 2, no. 6, pp. 1150–1158, Nov. 2003.

[8] I. Wong and B. Evans, "Optimal resource allocation in the OFDMA downlink with imperfect channel knowledge," *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 232–241, Jan. 2009.

[9] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems-Part I: Chunk allocation," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2734–2744, Sep. 2009.

[10] ——, "Chunk-based resource allocation in OFDMA systems-Part II: Joint chunk, power and bit allocation," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 499–509, Feb. 2012.

[11] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks-Part I: Theoretical framework," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 614–624, Mar. 2005.

[12] ——, "Cross-layer optimization for OFDM wireless networks-Part II: Algorithm development," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 625–634, Mar. 2005.

[13] Y. Su and M. van der Schaar, "Multiuser multimedia resource allocation over multicarrier wireless networks," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 2102–2116, May 2008.

[14] A. Parandeh Gheibi, A. Eryilmaz, A. Ozdaglar, and M. Medard, "On resource allocation in fading multiple-access channels-An efficient approximate projection approach," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4417–4437, Sep. 2010.

[15] G. Cook, J. Prades-Nebot, Y. Liu, and E. Delp, "Rate-distortion analysis of motion-compensated rate scalable video," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2170–2190, Aug. 2006.

[16] Y. Liu, J. Prades-Nebot, P. Salama, and E. Delp, "Rate distortion analysis of leaky prediction layered video coding using quantization noise modeling," in *Proc. Int. Conf. Image Process. (ICIP)*, Oct. 2004, vol. 2, pp. 801–804.

[17] Z. Chen and K. N. Ngan, "Recent advances in rate control for video coding," *Signal Process.: Image Commun.*, vol. 22, no. 1, pp. 19–38, 2007.

[18] X. Zhu, E. Setton, and B. Girod, "Rate allocation for multi-camera surveillance over an ad-hoc wireless network," in *Proc. Picture Coding Symp. (PCS)*, 2004, pp. 1–6.

[19] M. Tiwari, T. Groves, and P. Cosman, "Competitive equilibrium bitrate allocation for multiple video streams," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 1009–1021, Apr. 2010.

[20] G.-M. Su, Z. Han, M. Wu, and K. Liu, "Joint uplink and downlink optimization for real-time multiuser video streaming over WLANs," *IEEE Trans. Sel. Topics Signal Process.*, vol. 1, no. 2, pp. 280–294, Aug. 2007.

[21] ——, "A scalable multiuser framework for video over OFDM networks: Fairness and efficiency," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 10, pp. 1217–1231, Oct. 2006.

[22] H. Ha, C. Yim, and Y. Y. Kim, "Cross-layer multiuser resource allocation for video communication over OFDM networks," *Comput. Commun.*, vol. 31, pp. 3553–3563, Sep. 2008.

[23] J. Proakis, *Digitial Communications*. McGraw-Hill, 2000.

[24] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 1012–1032, June 2000.

[25] A. Goldsmith, *Wireless Communications*. Cambridge, 2005.

[26] M. Kalman and B. Girod, "Optimal channel-time allocation for the transmission of multiple video streams over a shared channel," in *Proc. IEEE 7th Workshop Multimedia Signal Process.*, 2005, pp. 1–4.

[27] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.
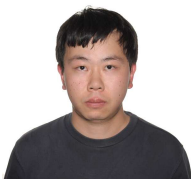
Dawei Wang (S'11) received the B.Eng. degree in electronic engineering (First Class Honors) from the Hong Kong University of Science and Technology (HKUST), Kowloon, Hong Kong SAR, China, in 2008, and the M.S. degree, in 2011, from the University of California, San Diego (UCSD), La Jolla, CA. He is currently working towards his Ph.D. at UCSD. He was also a visiting student at the University of Minnesota, Minneapolis.

His industry experience includes internship at Intel Corporation, Hillsboro, OR, in 2011. His research interests are in the areas of communication theory and video processing.

Laura Toni (S'06-M'09) received the M.S. degree (with honors) in electrical engineering and the Ph.D. degree in electronics, computer science and telecommunications from the University of Bologna, Italy, in 2005 and 2009, respectively.

In 2005, she joined the Department of Electronics, Informatics and Systems at the University of Bologna, to develop her research activity in the area of wireless communications. During 2007, she was a visiting scholar at the University of California at San Diego, CA, working on video processing over wireless systems. Since 2009, she has been a frequent visitor to UCSD, working on joint source and channel coding for wireless communication systems.

From June 2009 to November 2011, she worked at the Tele-Robotics and Application (TERA) Department at the Italian Institute of Technology (IIT) as a Post-doctoral Fellow. From November 2011 to November 2012, she was a Post-doctoral fellow in the Electrical and Computer Engineering Department at the University of California, San Diego. Since December 2012, she has been a Post-doctoral fellow in the Signal Processing Laboratory (LTS4) at the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. Her research interests are in the areas of image and video processing, wireless communications, and underwater communications.

Pamela C. Cosman (S'88-M'93-SM'00-F'08) obtained her B.S. with Honors in electrical engineering from the California Institute of Technology in 1987, and her M.S. and Ph.D. in electrical engineering from Stanford University in 1989 and 1993, respectively.

She was an NSF postdoctoral fellow at Stanford University and a Visiting Professor at the University of Minnesota during 1993–1995. In 1995, she joined the faculty of the Department of Electrical and Computer Engineering at the University of California, San Diego, where she is currently a Professor and Vice Chair. She was the Director of the Center for Wireless Communications from 2006 to 2008. Her research interests are in the areas of image and video compression and processing, and wireless communications.

Dr. Cosman is the recipient of the ECE Departmental Graduate Teaching Award, a Career Award from the National Science Foundation, a Powell Faculty Fellowship, and a Globecom 2008 Best Paper Award. She was a guest editor of the June 2000 special issue of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS on "Error-resilient image and video coding," and was the Technical Program Chair of the 1998 Information Theory Workshop in San Diego. She was an associate editor of IEEE COMMUNICATIONS LETTERS (1998–2001), and an associate editor of IEEE SIGNAL PROCESSING LETTERS (2001–2005). She was the Editor-in-Chief (2006–2009), as well as a Senior Editor (2003–2005, 2010-present), of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. She is a member of Tau Beta Pi and Sigma Xi.

Laurence B. Milstein (S'66-M'68-SM'77-F'85) received the B.E.E. degree from the City College of New York, New York, in 1964, and the M.S. and Ph.D. degrees in electrical engineering from the Polytechnic Institute of Brooklyn, Brooklyn, NY, in 1966 and 1968, respectively.

From 1968 to 1974, he was with the Space and Communications Group, Hughes Aircraft Company, and from 1974 to 1976, he was a member of the Department of Electrical and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY. Since 1976, he has been with the Department of Electrical and Computer Engineering, University of California, San Diego, where he is currently the Ericsson Professor of Wireless Communications Access Techniques and former Department Chair, working in the area of digital communication theory with special emphasis on spread-spectrum communication systems. He has also been a consultant to both government and industry in the areas of radar and communications.

Dr. Milstein was an Associate Editor for Communication Theory for the IEEE TRANSACTIONS ON COMMUNICATIONS, an Associate Editor for Book Reviews for the IEEE TRANSACTIONS ON INFORMATION THEORY, an Associate Technical Editor for *IEEE Communications Magazine*, and the Editor-in-Chief of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He was the Vice President for Technical Affairs in 1990 and 1991 of the IEEE Communications Society and was a former chair of the IEEE Fellows Selection Committee. He was a recipient of the 1998 Military Communications Conference Long Term Technical Achievement Award, an Academic Senate 1999 UCSD Distinguished Teaching Award, an IEEE Third Millennium Medal in 2000, the 2000 IEEE Communication Society Armstrong Technical Achievement Award, and various prize paper awards, including the 2002 MILCOM Fred Ellersick Award. He was also the recipient of the IEEE Communication Theory Technical Committee (CTTC) Service Award in 2009, and the CTTC Achievement Award in 2012.