Resource Allocation and Performance Analysis for Multiuser Video Transmission over Doubly Selective Channels

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

Abstract

We consider an uplink multicarrier system with multiple video users who want to send compressed video data to the base station. In the time domain, we model the time varying channel using Jakes' model, and in the frequency domain, each subcarrier is assumed to be independently fading. The video is scalably coded in units of group of pictures (GOP), and users have different video rate distortion (RD) functions. At the beginning of the GOP, the base station collects both the RD information and instantaneous channel state information (CSI) for subcarrier allocation purposes. We design a cross layer resource allocation algorithm to assign subcarriers to the users based on both the demand of the video and the quality of the channel. Once the resource allocation decision is made, the users then periodically adapt the modulation format of the subcarriers allocated according to the evolution of the CSI for the duration of the GOP. We show that our cross layer resource allocation robustly outperforms two baseline algorithms, each of which uses only one layer of information for resource allocation.

Index terms: Time-varying channel, Radio spectrum management, Multimedia communication, Multicarrier system, Wireless power allocation.

D. Wang, P.C. Cosman and L.B. Milstein are with the University of California, San Diego, Department of Electrical and Computer Engineering, 9500 Gilman Drive, Mail Code 0407, La Jolla, CA, 92037-0407, USA (email : daw017@ucsd.edu ; pcosman@ucsd.edu ; milstein@ece.ucsd.edu). L.Toni was with the University of California, San Diego. She is now with the École Polytechnique Fédérale de Lausanne (EPFL) (email : laura.toni@epfl.ch). This research is supported by the Intel-Cisco Video Aware Wireless Networks (VAWN) program and the National Science Foundation under grant No.CCF-0915727.

I. Introduction and Related Work

Over the past decade, the high demand for data rate for multimedia transmission and the limitation of communication bandwidth have become the bottleneck to further development of multimedia communications. For cellular and wireless local area network systems, various PHY/APP cross layer techniques have been studied to both improve video quality and increase cell capacity. Among them, point-to-point PHY/APP cross layer optimization seeks to exploit the unique characteristics of the video, and applies schemes like multiple description coding (MDP) to provide different protection levels and achieve higher end-to-end user QoE. Most research has focused on slow-varying channels [1]–[3].

To combat the uncertainty of a time-varying channel, video communication with automatic repeat request (ARQ) was proposed in [4] and [5]. Although ARQ is easy to implement, the time delay and uncertainty for exchanging the ARQ signals might not be suitable for delay sensitive video applications. More importantly, for a system with different Doppler spreads, the packet loss rate (PLR) varies dramatically with respect to channel estimation accuracy, which is determined jointly by the pilot spacing, the pilot power and the number of pilots used for interpolation [6]. Most papers on ARQ-based video communication oversimplify the PLR model. In [4], PLR is treated as a constant for all Doppler spreads. In [5], the authors study the performance of adaptive modulation with ARQ in a data communication system, and perfect channel estimation is assumed for choosing the modulation format and demodulation at the receiver. Channel estimation accuracy could be improved by reducing the interval between the pilots. However, the throughput loss due to pilot insertion might significantly reduce the number of video source bits delivered to the channel. Under the perfect channel state information (CSI) assumption, the critical tradeoff between the channel estimation accuracy and source encoding rate is missed in [5].

Forward Error Correction (FEC)-based video communication with no retransmission is often used for delay-sensitive video data. To achieve higher average image quality in systems with high mobility, [7] utilizes the coding diversity across both time and frequency, and analyzes the performance of progressive image transmission in the presence of inter-carrier interference and channel estimation error in an multi carrier setting. In [8], the authors study a joint link and source adaptation system, where the modulation and coding scheme at the PHY layer is chosen

according to instantaneous CSI and different importance levels of the packet, while the source rate at the APP layer is chosen based on the visibility of the packet and playback buffer status. Without any assumption of knowledge of the channel in the future, the adaptation scheme in [8] has the potential to be applicable for systems with arbitrary mobility. **Both [7] and [8] have been focusing on the point to point multimedia communication case and user is assumed to have fixed amount of resource regardless of the demand.**

Another technique is to exploit the relative diversity (both at the PHY layer and the APP layer) in a multiple access environment. To utilize multiple user channel diversity, multicarrier systems are widely used in cellular systems, as the resource allocation for multicarrier systems can be done flexibly. Multiple user subcarrier assignment with different user rate demand is investigated in [9], where the authors try to maximize the weighted sum of throughput for different user priority levels. In [10] and [11], user priority is further abstracted as a utility function, and the goal of the resource allocation is to maximize the sum of utilities. In [12], a multiple subcarrier assignment problem is presented, where both user rate distortion (RD) functions and CSI of multiple subcarriers are used for resource allocation to minimize the sum of distortions. Inspired by the equal-slope condition for video multiplexing [13], the authors derived a necessary condition for optimal spectrum sharing and designed an iterative algorithm for multiple subcarrier assignment. **In reference such as [10] on multiuser multicarrier resource allocation, the objective function (e.g. the utility function in [10]) is measured over some interval of time, and the channel is assumed to be stable for that duration.**

In this paper, we study a multiuser scalable video uplink system in a doubly selective environment. In the time domain, unlike [12], we do not assume any relation between the channel Doppler spread and video group of pictures (GOP) duration. For the physical layer, we use pilot symbol assisted modulation (PSAM) and allow users to adaptively change modulation format between resource allocation decisions. With the goal of minimizing average distortion, we design a PHY/APP cross layer resource allocation algorithm which takes into account both throughput loss and channel estimation error. As discussed above, for systems with high mobility, the end-to-end video performance is highly depends on the tradeoff between the channel estimation accuracy and video source encoding rate. To the best of our knowledge, this paper is the first work concerning video communication resource allocation with arbitrary user mobility.

The rest of the paper is organized as follows: In Section II, we describe the system model

for a multiple user multicarrier video communication system. We also provide details on the application layer scalable video codec and the physical layer doubly selective fading channel model. We then discuss the resource allocation framework and the PASM scheme for arbitrary user mobility in Section III. We present three resource allocation algorithms, each with a different degree of knowledge of the system information in Section IV, and we show performance results for the three algorithms in Section V. We conclude the paper in Section VI.

II. System Model

We consider a cellular multicarrier video communication system with a set of video users indexed by k, k=$\{1, 2, 3 ... K\}$. The system occupies a total frequency band of W (Hz) equally divided into M_c subcarriers indexed by m, m=$\{1, 2, 3 ...M_c\}$. We assume users experience the same Doppler spread and let f_{nd} represent the normalized Doppler spread for each user. We focus on an uplink system, and the task of the resource allocation is to assign subcarriers to users based on both RD functions and CSI to minimize the sum of video distortions.

A. Doubly Selective Channel Model

The system operates in a slotted manner, and the length of one time slot is T_s (sec), equal to both the video display time and the transmission duration of one GOP. We assume a block fading model in the frequency domain with coherence bandwidth $B_c = \Psi W/M_c$. **Here,** W/M_c **is the bandwidth of one subcarrier.** Ψ **is the number of the correlated subcarriers, whose fading realization will be identical.** Let $H_{k,m}[l]$ be the complex channel gain of user k for subcarrier m at the l-th symbol. The subcarrier assignment as well as the power allocation decision will be made on a slot-to-slot basis. Each subcarrier can only be used by one user, but it is possible for one user to get more than one subcarrier. Further, $H_{k,m}[l] = \gamma\alpha_{k,m}[l]$, where γ depends on the path-loss coefficient, the distance between the mobile user and the base station, and the shadowing caused by obstacles. The variable $\alpha_{k,m}[l]$ captures the multipath fading and is modeled as a zero-mean complex stationary Gaussian random process [14]. The magnitude of $\alpha_{k,m}[l]$ is Rayleigh distributed with a variance of unity for a non-line-of-sight system. The band-limited spectrum of $\alpha_{k,m}(t)$ is given by

$$S(f) = S(0)\left[1 - \frac{f}{f_d}\right]^{-1/2}, |f| < f_d \tag{1}$$

where f_d is the Doppler spread, and $S(0) = 2/\pi$. Define $f_{nd} = f_d T_0$ as the normalized Doppler spread, where T_0 as the symbol duration. The autocorrelation between two samples l and $l + \Delta l$ can be written as [14]

$$\mathbb{E}\left[\alpha_{k,m}[l]\alpha^*_{k,m}[l+\Delta l]\right] = J_0\left[2\pi \cdot f_{nd} \cdot (\Delta l)\right] \tag{2}$$

where $J_0(\cdot)$ is the zeroth order Bessel function of the first kind. The first zero crossing of the correlation function occurs when the product $f_{nd} \cdot (\Delta l)$ is about 0.4.

For a multicarrier system, the complex envelope of the transmitted signal for user k can be written as

$$x_k(t) = \sum_l \sum_{m=1}^{M_c} \sqrt{P_{k,m}} X_{k,m}[l] \exp\left(\frac{j2\pi mt}{T_0}\right) g(t - lT_0) \tag{3}$$

where $P_{k,m}$ and $X_{k,m}[l]$ are the transmission power and coded symbol with unit variance, respectively, of user k on subcarrier m. The power $P_{k,m}$ is assumed to be fixed for the duration of a time slot, and $P_{k,m} = 0$ if subcarrier m is not allocated to user k. Also, $g(t)$ is a zero-excess bandwidth Nyquist pulse, with $G(f) = \sqrt{T_0}, \forall f \in [-1/2T_0, 1/2T_0)$, and $G(f) = 0$ otherwise.

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

$$y_{k,m}(t) = \sqrt{P_{k,m}} H_{k,m}[l] X_{k,m}[l] \exp\left(\frac{j2\pi mt}{T_0}\right) + n_{k,m}(t) \tag{4}$$

where $n_{k,m}(t)$ is complex 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:

$$Y_{k,m} \triangleq \frac{1}{T_0} \int_0^{T_0} y_{k,m}(t) \exp(\frac{-j2\pi mt}{T_0}) dt \tag{5}$$

The noise power is given by $P_N = E[|N_{k,m}|^2] = 2N_0/T_0$, and the power for the desired signal is $P_{k,m}|H_{k,m}[l]|^2$. If the modulation format is adaptive QAM, from [15] and [16], the symbol error rate (SER) for an AWGN channel can be approximated as

$$SER \approx 4Q\left(\sqrt{\frac{3}{M-1}\frac{P_{k,m}|H_{k,m}[l]|^2}{P_N}}\right) \tag{6}$$

Here, M is the alphabet size of a QAM waveform, and for a given fixed SER_t, the information rate (number of bits each symbol can carry) $R_{k,m}(P_{k,m}, H_{k,m}[l])$ (in bits/symbol) can be written as a function of transmission power and channel response gain:

$$R_{k,m}(P_{k,m}, H_{k,m}[l]) = \min\{\lfloor \log_2\left[1 + \eta P_{k,m}|H_{k,m}[l]|^2\right] \rfloor, R_{max}\} \tag{7}$$

where $\eta = \frac{3}{P_N}[Q^{-1}(SER_t/4)]^{-2}$ and R_{max} is the largest alphabet size the system allows. The bit rate (in bits/sec) then can be written as $R_{k,m}(P_{k,m}, H_{k,m}[l])/T_0$. In the following sections, we will replace $H_{k,m}[l]$ by the estimate $\widetilde{H}_{k,m}[l]$ in (7) and use $R_{k,m}(P_{k,m}, \widetilde{H}_{k,m}[l])$ to determine the modulation alphabet size. Note that, because of the channel estimation error and floor operation in (7), the actual SER of the symbols might be different from the parameter SER_t. The effect of the channel estimation accuracy and the choice of SER_t will be investigated in the simulation section.

B. Scalable Video Codec

A scalable video codec is designed such that the decoder only needs a portion of the encoded bitstream (a substream) to display the video. The decoded fidelity of the video depends on the length of the substream, as well as the rate distortion characteristics of the video content. Scalable coding allows flexible adaptation to time-varying wireless channels and throughput variations in multiple-hop communication systems.

For each bitstream, the most important video information (e.g., motion vectors, frame indexes, macroblock IDs) is contained in a substream called the base layer (BL). One or more enhancement layers (EL) are added such that the mean square error (MSE) will decrease when more enhancement bits are received by the decoder. Previously, a fine granular-scalability (FGS) codec was proposed [17] [18] for accurate source-to-channel adaptation. FGS allows every successfully delivered video bit to improve the video quality, but the granularity of the scalability will sacrifice the video compression efficiency. The scalable extension of H.264/AVC (known as H.264/SVC) [19] with medium granular scalability (MGS) has emerged as a balanced solution for the tradeoff between compression efficiency and scalable granularity. An H.264/SVC MGS codec features temporal, spatial and quality scalability, and allows the flexibility of dropping a combination of substreams according to the communication channel. In this paper, we are interested in the quality scalable function, which packetizes the encoded bitstream according to the zonal location of the DCT coefficients and ranks the packets based on their importance in the GOP. The encoder assigns the highest priority for transmission to the packets which can most effectively reduce the compression distortion. If an error occurs in the transmission, the entire packet and all successive within that GOP packets will be dropped, but previous packets (all of which have higher priority) will be used for decoding the GOP.

To characterize the tradeoff between the compression fidelity and the number of bits used to describe the source, we model the rate distortion (RD) curve of the video using a parameterized function. Since the video is compressed in units of GOPs, this RD function is also measured on a GOP-by-GOP basis. Let $D_k(B)$ be the RD function of user k, where B is the number of bits in the substream (the length of the truncated bit stream). For each GOP, the MSE distortion can be approximated as [20]

$$D_k(B) = a_k + \frac{w_k}{B + v_k} \tag{8}$$

where a_k, v_k and w_k are constants which depend on the video content. When the picture is relatively spatially uniform, and the motion of the video is slow, the time and spatial redundancy can be easily compressed, and one would expect a relatively flat RD function. For a video with high complexity and fast motion, the RD function is normally steep and w_k is relatively large. The difference of the RD tradeoff between different users constitutes application layer diversity.

III. Cross Layer Video Uplink System with Arbitrary User Mobility

In a multiple user system, to minimize the sum of the MSEs across all users, the base station collects the RD information (coefficients of a_k, w_k and v_k in (8)) as well as the CSI of the subcarriers, and allocates the subcarriers jointly according to application layer and physical layer information. In [12], we studied video resource allocation for a system in which the channel varies slowly over the duration of a GOP. Under the condition of constant CSI for the entire GOP, we assumed that the modulation format remains unchanged for one GOP, and the throughput of each subcarrier could be perfectly estimated at the beginning of each GOP. The resource allocation problem then becomes a mixed integer programming problem, and allocation decisions are made on a GOP-by-GOP basis.

A. System Operation Overview

For mobile users operating in a high enough Doppler environment, the CSI estimated at the beginning of the GOP will be outdated prior to the end of the GOP. For subcarrier m, if the modulation format determined by the CSI at the beginning of the GOP is held constant for the entire GOP duration, it is likely that the video data will either be over-protected or under-protected. On the other hand, the resource allocation decision based on both RD information

and CSI is normally of high complexity [21] [22]. If we update the resource allocation decision every coherence time, the base station will need to collect the instantaneous CSI of all the subcarriers for all the users, as well as the amount of the GOP that has already been transmitted. If the allocator makes a resource allocation decision at each coherence time, it is not only computationally difficult, but also requires a large amount of information exchange. We thus propose a scheme having two phases which balances the computational complexity with the adaptation accuracy.

Phase I Cross Layer Resource Allocation: At the beginning of each GOP, user k submits the RD function $D_k(B)$ of the current GOP to the base station. The instantaneous CSI of the first symbol, $H_{k,m}[1], (k \in \{1, 2...K\}, m \in \{1, 2..M_c\})$ of every subcarrier for each user and uses the CSI jointly with the RD information to make an allocation decision. The allocation decision is fed back to the users and each user is allowed to access the subcarriers assigned to him for the entire GOP duration. The resource allocation algorithm and the information exchange for the RD function is conducted once per GOP.

Phase II Pilot Assisted Adaptive Modulation: After resource allocation, when each individual user knows the subset of assigned subcarriers, each user periodically sends a pilot symbol to the base station for channel estimation purposes. Based on the instantaneous CSI at each period, the base station updates the modulation format of each subcarrier and feeds the new modulation formats back to the corresponding users. The estimated CSI is also used for demodulation purposes. Since the modulation format is updated periodically, the number of information bits transmitted cannot be estimated accurately at the beginning of the GOP. The important bits will be transmitted first, and the actual number of bits transmitted is determined by the overall channel conditions over the duration of a GOP (T_s seconds). Note that in this phase, no information about the RD is exchanged, and the resource allocation decision is not updated. The details of the adaptive modulation scheme are discussed below.

B. Pilot Symbol Assisted Modulation (PSAM)

As depicted in Fig. 1, at most one pilot symbol will be sent from the user to the base station in each subcarrier every L_s symbols for channel estimation. Define a time epoch as a group of L_s ($L_s = T_m/T_0$) symbols, which is also the distance between two pilot symbols. The modulation format for every time epoch will be kept the same, and is determined by the CSI of

the pilot symbol. **We define λ as the ratio between the number of information symbols and total symbols. In most of this paper, we will focus on a system with independently faded subcarriers, i.e. $\Psi = 1$, so that, $\lambda = (L_s - 1)/L_s$.** In a system with coherence bandwidth $\Psi > 1$, since all the subcarriers within the coherence bandwidth will have the same fade realization, fewer pilot symbols need to be sent. For example, in Fig. 1, two subcarriers share the same fade ($\Psi = 2$), and one pilot symbol can be saved in each time epoch for every group of two subcarriers spanned by the same coherence bandwidth.

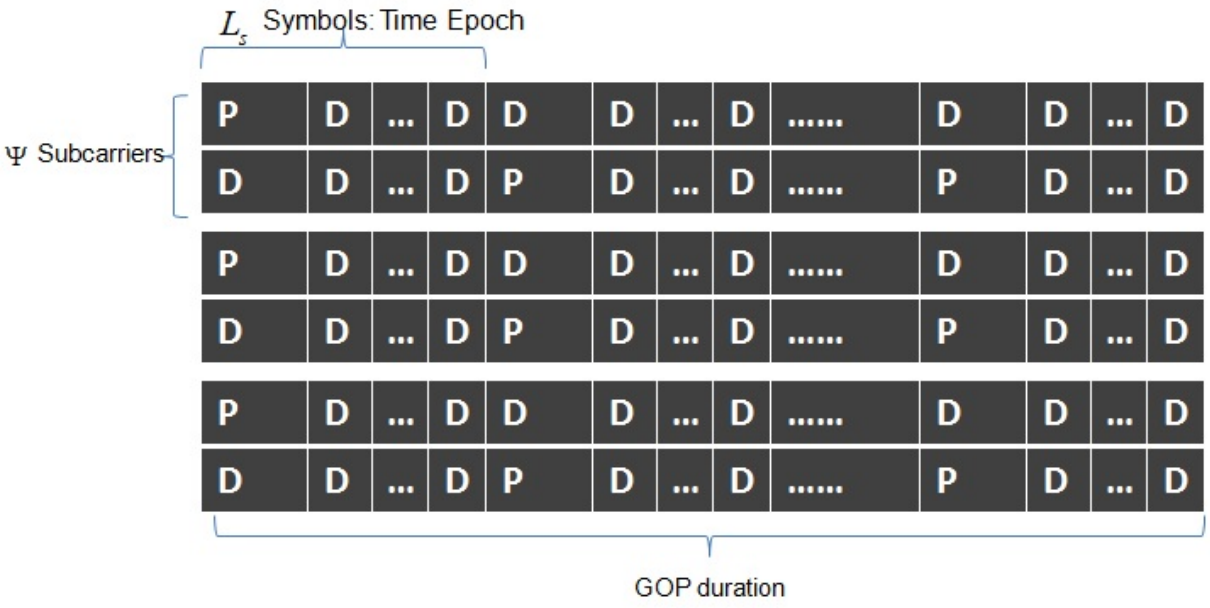

Figure 1. *Pilot assisted modulation. One pilot symbol is added for each subcarrier to estimate the channel for every time epoch. If the number of correlated subcarriers is larger than 1, only one pilot symbol is needed for every correlated band.*

Let $P_{k,m}$ be the average power for the k-th user on the m-th subcarrier, where $P_{k,m} = P > 0$ if the m-th subcarrier is allocated to the k-th user, and $P_{k,m} = 0$ otherwise. Let $\mu_{k,m}$ be the ratio between the power of the pilot symbols and average data power for the k-th user on the m-th subcarrier. The power of the pilot symbol, $P^p_{k,m}$, and the data power, $P^d_{k,m}$, are then given by

$$P^d_{k,m} = \frac{P_{k,m} L_s}{L_s - 1 + \mu_{k,m}} \tag{9}$$

$$P^p_{k,m} = \frac{P_{k,m} L_s \mu_{k,m}}{L_s - 1 + \mu_{k,m}} \tag{10}$$

We assume that the value of $\mu_{k,m}$ is the same for all subcarriers of all users, and hence we drop the indices k and m. The average power is given by

$$P_{k,m} = \mu \frac{1}{L_s} P^p_{k,m} + \frac{L_s - 1}{L_s} P^d_{k,m} \tag{11}$$

For a given SER_t, the modulation format for subcarrier m of user k is updated every L_s symbols based on $\widetilde{H}_{k,m}[iL_s + 1]$, which is the estimate of the channel response of the i-th pilot, $H_{k,m}[iL_s+1]$. The information rate (number of bits each symbol can carry in bits/symbol) for the i-th group of L_s symbols can be written as $R_{k,m}(P^d_{k,m}, \widetilde{H}_{k,m}[iL_s + 1])$ using (7). To estimate the channel response $H_{k,m}[iL_s+1]$, a Wiener filter with K_e pilots is used for interpolation. Since the decision of the modulation format needs to be fed back to the users immediately after the pilot symbol is sent, we can only use the pilots prior to the current one for channel estimation. In other words, to estimate $H_{k,m}[iL_s + 1]$, the pilots at indices $l = jL_s + 1, \{j = (i, i-1...(i-K_e+1))\}$ are used, where K_e is chosen to be even. To estimate the channel gain for the data symbols $H_{k,m}[iL_s + u], u = (2, 3...L_s)$, pilots from both sides can be used. That is, the pilots at time indices $l = jL_s + 1, \{j = ((i - K_e/2 + 1), (i - K_e/2 + 2)...(i + K_e/2))\}$ are jointly used to interpolate the channel gain.

From [6] [23], the channel estimation error $e = \widetilde{H}_{k,m}[l] - H_{k,m}[l]$ can be modeled as a Gaussian random variable with zero mean and variance equal to

$$\sigma^2_e = \sigma^2_u - \mathbf{w(l)}^{+}\mathbf{R^{-1}}\mathbf{w(l)}/(P^p_{k,m}\overline{|H_{k,m}|^2}) \tag{12}$$

where $\sigma^2_u = P^x_{k,m}\overline{|H_{k,m}|^2}$ is the average received power of the data/pilot symbols, and $+$ represents conjugate transpose. If we use K_e pilots for interpolation, $\mathbf{R}$ is a $K_e \times K_e$ matrix with the entry in the i-th row and j-th column given by

$$R_{ij} = P_N\delta_{ij} + P^p_{k,m}\overline{|H_{k,m}|^2}J_0(2\pi f_{nd}(i-j)L_s) \tag{13}$$

where δ is the Kronecker delta, and $\mathbf{w(l)}$ is a K_e column vector for the l-th channel sample. The v-th row of $\mathbf{w(l)}$ is given by

$$w_v(l) = P^p_{k,m}\overline{|H_{k,m}|^2}\rho_v(l) \tag{14}$$

where $\rho_v(l)$ is the correlation coefficient of the l-th channel sample and the channel estimate obtained from the v-th pilot. For example, as discussed above, to estimate the channel sample $H_{k,m}[l]$, if the symbol at time l is a data symbol and belongs to the i-th pilot time epoch, i.e.,

$iL_s + 2 \leq l \leq (i+1)L_s$, **the v-th pilot used for interpolation is at** $(i - K_e/2 + v)L_s + 1$**,** so that

$$\rho_v(l) = J_0(2\pi f_{nd}[l - ((i - K_e/2 + v)L_s + 1)]) \tag{15}$$

From (12), (13) and (14), we see that one of the crucial parameters for deciding the channel estimation error is the distance between pilots, L_s. For a given Doppler spread, a smaller L_s results in a larger overhead, and a larger L_s results in larger channel mismatch due to CSI outdating, but less in throughput loss. **Besides L_s, the variance of e is also jointly dependent on the number of pilot symbols for interpolation K_e, the average SNR, and the ratio μ between the pilot symbol power and data symbol power.**

C. Resource Allocation Problem Formulation

At the beginning of a GOP, the base station estimates the throughput of each subcarrier using (7), assuming that the adaptive QAM format will last for T_s seconds. The number of bits transmitted over subcarrier m of user k can then be written as $R_{k,m}(P^d_{k,m}, \widetilde{H}_{k,m}[1]) \cdot T_s/T_0$, where T_s/T_0 is the number of QAM symbols for a GOP. We denote by λ ($0 < \lambda \leq 1$) the fraction of data symbols, and $\lfloor (1-\lambda) \cdot T_s/T_0 \rfloor$ symbols will be pilots. To protect the data, a channel code of fixed rate u is added. If the channel stays constant, the number of information bits that the physical layer can support for user k across all M_c subcarriers is given by

$$B_k = \lfloor \sum_{m=1}^{M_c} u \cdot \lambda \cdot R_{k,m}(P^d_{k,m}, \widetilde{H}_{k,m}[1]) \cdot T_s/T_0 \rfloor \tag{16}$$

with $P^d_{k,m} = P^d$, if subcarrier m is assigned to user k, and $P^d_{k,m} = 0$ otherwise.

Although to obtain the performance of the resource allocation algorithm we will include the effects of channel errors and time varying modulation choices, for the allocation algorithm design, we ignore the effect of channel errors and assume that the modulation format is constant for the GOP duration. We use (16) as the channel throughput for our algorithm design. If we plug (16) into (8), then the MSE distortion for user k can be written as

$$D_k = a_k + \frac{b_k}{\sum_{m=1}^{M_c} \lambda R_{k,m}(P^d_{k,m}, \widetilde{H}_{k,m}[1]) + c_k} \tag{17}$$

Here, we have divided both the numerator and denominator by $u \cdot T_s/T_0$ for simplicity. So

$$b_k = \frac{w_k}{(u \cdot T_s/T_0)} \qquad c_k = \frac{v_k}{(v \cdot T_s/T_0)} \tag{18}$$

The base station needs to assign M_c subcarriers to K users at the beginning of each GOP, and users can access the subcarrier for the duration of the GOP. The allocation decision will be updated at the beginning of the next GOP as both CSI and RD are updated. Mathematically, our resource allocation goal is to minimize the sum of distortions among K users at each time slot. The optimization objective is

$$\min_{\underline{\underline{P}}} \sum_{k=1}^{K} \frac{b_k}{\sum_{m=1}^{M_c} \lambda R_{k,m}(P_{k,m}^d, \widetilde{H}_{k,m}[1]) + c_k} \tag{19}$$

where the entry in the k-th row and m-th column, $P_{k,m}^d$, is the power allocation of the m-th subcarrier for user k. The base station sends the allocation decision to the users. Note that we dropped the a_k term, as it is constant with respect to $\underline{\underline{P}}$. We assume that any subcarrier is used by one user exclusively, so the feasibility constraint for the optimization problem is as follow

For $m \in \{1, 2, 3...M_c\}$, $P_{k',m}^d P_{k,m}^d = 0$ and $P_{k,m}^d = \{0, P^d\}$

Mathematically, (19) is an NP-hard integer programming problem, and an exhaustive search approach would need K^{M_c} calculations. In the next section, we will propose a sub-optimal algorithm which gives priority to users with steep RD curvatures to access the subcarriers. We compare the performance of our algorithm with two baseline algorithms, each of which has limited information about the state of the channel and the state of the videos. Note that, in (16), the throughput of each subcarrier is estimated based on the instantaneous CSI at the beginning of the GOP. **The difference between the actual throughput and estimated throughput depends on the coherence time.**

IV. Resource Allocation Algorithms

Before we state the cross layer algorithm that solves the optimization problem (19), we first introduce two baseline algorithms, each of which uses only one layer of information when making the resource allocation decision.

A. Application Layer Resource Allocation Algorithm

We assume that the base station only knows the application layer information (RD function) when making the allocation decision. Lacking the physical layer CSI, the base station will treat all the subcarriers the same. The resource allocation decision will only specify the number of

subcarriers each user is assigned. Define $L_k, (k = \{1, 2...K\})$ as the number of subcarriers allocated to user k, which is determined by the relative complexity of the RD functions $D_k(B)$.

Application Layer Optimization Algorithm:

Step 1: To measure the complexity of each RD curve, we first set a target MSE value D_t (the value is chosen such that the quality of the video is acceptable, say $D_t = 50$, which corresponds to $PSNR = 31dB$). To achieve this target MSE value, the rate required for user k is given by

$$r_k = \frac{b_k}{D_t - a_k} - c_k \tag{20}$$

We then use r_k as a measure of the complexity of the video for resource allocation purposes.

Step 2: Given the constraint of $\sum_{k=1}^{N} L_k = M_c$, the subcarriers will be split such that the number of subcarriers assigned to user k is proportional to r_k, or

$$L_k \approx M_c \frac{r_k}{\sum_{k=1}^{K} r_k} \tag{21}$$

$L_k \in \mathbb{Z}$. The reason for the approximation sign is as follows: We first calculate $\hat{L}_k = M_c \frac{r_k}{\sum_{k=1}^{K} r_k}$ for user k, where $\hat{L}_k$ can be any rational number. We then round each $\hat{L}_k$ to $\widetilde{L}_k \in \mathbb{Z}$. If $\sum_{k=1}^{K} \widetilde{L}_k = M_c$, we have $L_k = \widetilde{L}_k, \forall k$. If $\sum_{k=1}^{K} \widetilde{L}_k > M_c$, some of the $\hat{L}_k$'s which have been rounded up need to be rounded down. Let $j \triangleq \sum_{k=1}^{K} \widetilde{L}_k - M_c$, where j is an integer greater than zero. We choose those j $\hat{L}_k$'s, which have been rounded up and which are farthest from their respective integer ceilings, and round them down so that the sum of subcarriers across users equals M_c. A similar mechanism applies if $\sum_{k=1}^{K} \widetilde{L}_k < M_c$.

Since the CSI is not used in the allocation decision, the base station randomly chooses L_k subcarriers for user k. After the allocation, the user applies Phase II discussed in the previous section to update the modulation format using (7). For the application layer system, since the mechanism of Phase II, which includes the choice of the system parameters (L_s, μ, etc), is the same as the cross layer system, the performance difference of the two systems solely depends on the resource allocation decision.

B. Physical Layer Resource Allocation Algorithm

We now consider the case where only the instantaneous CSI $\widetilde{H}_{k,m}[1]$ is available at the base station. To maximize the sum throughput, a conventional multi-user diversity (MUD) algorithm will assign subcarrier m to user k^* with best channel gain, i.e., $k^* = \arg\max_k \left\{ \left| \widetilde{H}_{k,m}[1] \right|^2 \right\}$. To ensure fairness among users [24], we assign subcarrier m to the user k^* such that

$$k^* = \arg\max_k \left\{ \frac{\left| \widetilde{H}_{k,m}[1] \right|^2}{\overline{\left| \widetilde{H}_k[1] \right|^2}} \right\} \tag{22}$$

where $\overline{\left| \widetilde{H}_k[1] \right|^2} \triangleq \frac{1}{M_c} \sum_{m=1}^{M_c} \left| \widetilde{H}_{k,m}[1] \right|^2$ is the empirical average channel gain of user k at the beginning of the GOP.

We introduce the following definitions that will be used in the physical layer resource allocation algorithm.

Definitions I:

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 yet been assigned any subcarrier 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 yet been made.

d) Similar to the application layer optimization algorithm, let L_k be the number of subcarriers user k is assigned.

For a system with coherence bandwidth larger than the subcarrier bandwidth, i.e., $\Psi > 1$, the MUD based algorithm proposed in [25] [26] 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 very unbalanced and the average video performance will degrade. 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

$$\psi_n = \psi_{n-1} + \left\lceil \epsilon \left(\frac{M - \psi_{n-1}}{K - (n-1)} \right) \right\rceil \tag{23}$$

where, for $n = 1$, this expression reduces to $\psi_1 = \lceil \epsilon \frac{M_c}{K} \rceil$. 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_c/K. Assuming that one user has already been assigned the maximum of $\psi_1 = \lceil \epsilon \frac{M_c}{K} \rceil$ subcarriers, the average number of subcarriers for the remaining $(K-1)$ users is given by $(M_c - \psi_1)/(K-1)$ and the resource for any combination of two users is limited by $\psi_1 + \lceil \epsilon(M_c - \psi_1)/(K-1) \rceil$ 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 (23). 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 any 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. 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 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 $\Gamma = \{1, 2, ...M_c\}$.

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

$$(k^*, m^*) = \underset{k \in \Lambda, m \in \Gamma}{\arg\max} \left\{ \frac{\left| \widetilde{H}_{k,m}[1] \right|^2}{\left| \widetilde{H_k}[1] \right|^2} \right\} \tag{24}$$

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, $\left| \widetilde{H}_{k^*,m^*}[1] \right|^2$ stands for the best channel response of 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.

We then go back to Step 2 and repeat (24) to assign subcarriers until Γ is empty.

C. Cross Layer Resource Allocation Algorithm

When both the application layer RD and the physical layer CSI are used for resource allocation, a cross layer algorithm attempts to satisfy the two goals of giving more subcarriers to users with demanding RD curves and giving subcarriers to users with high channel gains. To solve the optimization problem (19), we propose an iterative algorithm as follows:

Definitions II:

a) We define $\theta_m^{(i)}$ to be the user who is assigned subcarrier m at the i-th iteration. For example, $\theta_1^{(2)} = 3$ means user 3 is assigned subcarrier 1 at the second iteration of the algorithm. Also, similar to the physical layer optimization algorithm, let Λ be the set of users who are eligible to be assigned additional subcarriers. We further define $|\Lambda|$ as the cardinality of the set.

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

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

Cross Layer Optimization Algorithm:

Step 1 Initialization: We initialize the resource allocation by assigning each subcarrier to the user, $\theta_m^{(0)} = \arg\max_k \left\{ \left|\widetilde{H}_{k,m}[1]\right|^2 / \overline{\left|\widetilde{H}_k[1]\right|^2} \right\}$, and let Λ be the set of all users, $\Lambda = \{1, 2...K\}$.

Step 2 Rate and Slope Calculation: Knowing the subcarriers in $A_k^{(i)}$, user k can calculate the anticipated number of information bits to be transmitted based on the current allocation decision

$$R_k^{(i)} = \lambda \sum_{A_k^{(i)}} R_{k,m}\left(P^d, \widetilde{H}_{k,m}[1]\right) \tag{25}$$

$R_k^{(i)}$ can be viewed as the aggregate rate of user k across all M_c subcarriers. The absolute value of the slope on the RD curve at $R_k^{(i)}$ bits comes from (17):

$$S_k^{(i)} = \frac{b_k}{\left(R_k^{(i)} + c_k\right)^2} \tag{26}$$

We then pick the user with the steepest slope $k^* = \arg\max_{k\in\Lambda}\{S_k^{(i)}\}$ and consider switching one subcarrier from some other user to user k^*.

Step 3 Subcarrier Reassignment: We now check every subcarrier m which is not currently assigned to user k^*, and consider the possibility of changing the assignment of m from user $\theta_m^{(i)}$ to user k^*. The loss of subcarrier m will cause the MSE value of user $\theta_m^{(i)}$ to increase by

$$\Delta_{\theta_m^{(i)},m} = \frac{b_{\theta_m^{(i)}}}{\lambda \sum\limits_{A_{\theta_m^{(i)}}^{(i)}\setminus m} R_{\theta_m^{(i)},m}\left(P^d, \widetilde{H}_{\theta_m^{(i)},m}[1]\right) + c_{\theta_m^{(i)}}} - \frac{b_{\theta_m^{(i)}}}{\lambda \sum\limits_{A_{\theta_m^{(i)}}^{(i)}} R_{\theta_m^{(i)},m}\left(P^d, \widetilde{H}_{\theta_m^{(i)},m}[1]\right) + c_{\theta_m^{(i)}}} \tag{27}$$

and allow the MSE of user k^* to decrease by

$$\Delta_{k^*,m} = \frac{b_{k^*}}{\lambda \sum\limits_{A_{k^*}^{(i)}} R_{k^*,m}\left(P^d, \widetilde{H}_{k^*,m}[1]\right) + c_{k^*}} - \frac{b_{k^*}}{\lambda \sum\limits_{A_{k^*}^{(i)}\cup m} R_{k^*,m}\left(P^d, \widetilde{H}_{k^*,m}[1]\right) + c_{k^*}} \tag{28}$$

We then find the subcarrier m^* which maximizes the difference between $\Delta_{k^*,m}$ and $\Delta_{\theta_m^{(i)},m}$.

If $\left(\Delta_{k^*,m^*} - \Delta_{\theta_{m^*}^{(i)},m^*}\right) > 0$, we reassign subcarrier m^* to user k^* at iteration $i+1$, i.e., $\theta_{m^*}^{(i)} = k^*$, and then go back to Step 2 to update k^*.

If $\left(\Delta_{k^*,m^*} - \Delta_{\theta_{m^*}^{(i)},m^*}\right) \leq 0$, which means that the sum of distortions will not be reduced by reassigning any subcarrier to user k^*, we exclude k^* from Λ, $\Lambda = \Lambda \setminus k^*$ and user k^* will not be assigned any additional resource. Next, we check the cardinality of $|\Lambda|$. If $|\Lambda| = 1$, meaning that there is no possibility of reassignment, we stop, otherwise we go back to Step 2 with iteration index i incremented.

The cross layer resource allocation algorithm is designed such that users with large slope are given priority to be assigned additional subcarriers. In the algorithm, we first assign subcarriers using a MUD algorithm and estimate the number of bits that the channel can support using (25). From the application layer point of view, it is most likely that the sum of distortions can be reduced by assigning more subcarriers to the user with the steepest slope, so we give the priority to the user with largest slope. We test the possible reassignment of subcarriers and use the CSI to find the subcarrier which can maximize the reduction of the sum of distortions through

reassignment. After reassignment, we update the user with the steepest slope and continue the iteration until we exhaust all possibilities for subcarrier switching.

V. System Parameter Optimization

We use a video with a resolution of 352×240. The video consists of 150 frames at 30 frames/second, and is organized into GOPs of 15 frames (IPPP). The content of the video includes both high motion segments and low motion segments. Each user is assigned the same video, but with random starting points. The simulation runs for one cycle of the entire video sequence, and users are then assigned another random set of starting points for the next cycle. By assigning random starting points of the same cyclic video to different users, we create instantaneous application layer diversity among users and yet have the same average complexity over time for the different users. Between cycles, different realizations of starting points will generate different levels of application diversity. We encode the video using H.264/SVC reference software JSVM version 9.19.12. For the MGS layer, the 4×4 DCT coefficients of each macroblock are split using MGS vector $[1, 1, 2, 2, 2, 8]$ [19]. To specify the a_k, b_k and c_k values of RD curves in (17), we extract bitstreams at 20 different encoding rates from 60 to 1200 kbps offline and use these operational points to find the RD function by non-linear regression. At the decoder side, the bit stream after the first channel error is discarded. In a very crowded system, it is possible that some users will not be allocated any subcarriers and the transmission rate for the GOP is zero. If that happens, the last frame of the previous GOP is held over for the duration of the current GOP.

We consider a single cell of radius equal to 50 meters. The bandwidth for each subcarrier is 100kHz. The channel response consists of both path loss and multipath fading, and the amplitude squared of the multiplicative channel coefficient $H_{k,m}[l]$ is given by

$$|H_{k,m}[l]|^2 = |\alpha_{k,m}[l]|^2 \cdot K_0 \cdot \left(\frac{d_0}{d_k}\right)^{\beta} \tag{29}$$

where $\alpha_{k,m}[l]$ follows the Jakes' model and is generated using the statistical model proposed by [27]. Also, d_k is the distance of user k to the base station. $d_0 = 10$m is the reference distance, and the path-loss model is accurate when $d_k > d_0$ [28]. The users are perfectly power controlled and the average received power is 17.8 dB per subcarrier. We set the path-loss exponent $\beta = 3$, and $K_0 = -30dB$ is a constant. For all three optimization schemes, we apply a rate $u = 1/2$

convolutional code with code generator polynomial [23, 35] in octal. The codeword length is equal to the length of the entire GOP bitstream, and the coded bits are interleaved across different subcarriers. The value of SER_t is set to 0.15. At the receiver, we use soft-decision decoding with eight reliability ranges. The MQAM modulation format can be $\{M = 4, 8, 16, 32, 64, 128, 256\}$ for all three algorithms. For the physical layer optimization algorithm, we set $\epsilon = 1.5$, so one user cannot be assigned more than 150% of the average number of subcarriers.

A. System With Different L_s

In Fig. 2, we show the performance of the three resource allocation algorithms with respect to different Doppler spreads. The pilot insertion spacing L_s equals 100 and the number of pilots used for channel estimation K_e is 24. The ratio μ between pilot and the power equals unity. The curves in Fig. 2 and 4−9 illustrate the video performance. The PSNR performance here consists of both the degradation caused by source compression and that caused by channel errors. The effects of packet loss, errors in RD curve fitting, and imperfection of encoder rate control are included in the simulation.

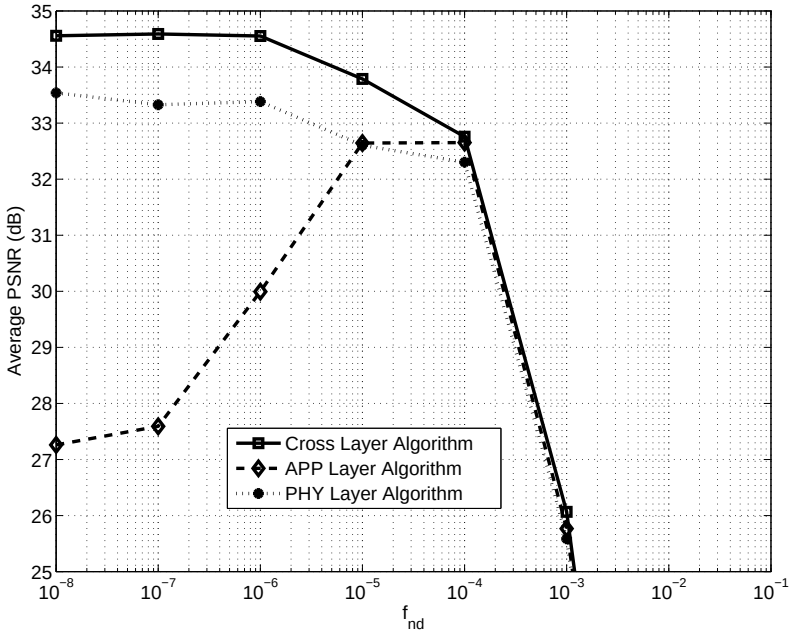

Figure 2. *$Ls = 100$, 16 Subcarriers, 3 Users, $SER_t = 0.15$, $\mu = 1$. The PSNR performance for a system with $f_{nd} > 10^{-2}$ is less than 15dB.*

In Fig. 3, we show the decoded BER, conditioned on the event that the user transmits, i.e., $R_{k,m}(P^d_{k,m}, \widetilde{H}_{k,m}[iL_s+1]) \geq 2$, **where the number '2' on the RHS corresponds to the alphabet size of QPSK**. When $1/f_{nd}$ is significantly larger than $L_s = 100$, the channel is relatively constant within each duration of L_s symbols. The decoded bit error rate for this scenario is in the range of 10^{-8} to 10^{-6}. Since we update the modulation format every L_s symbols, the correlation between the data symbols and pilot symbols is relatively high, and the channel estimation error given in (12) is small. The actual raw SER every group of 100 symbols is similar to the SER of an AWGN channel conditioned on the instantaneous channel fade, and is close to the SER estimated using (6). In other words, for $L_s = 100$ and $f_{nd} < 10^{-4}$ in Fig. 2, the proposed mechanism of PSAM (Phase II) can accurately adapt to the variation of the channel and properly control the channel error rate.

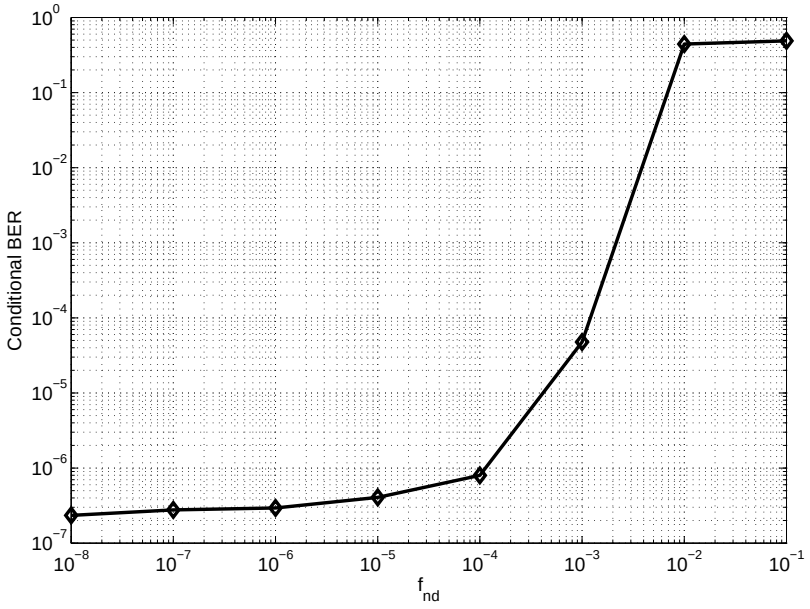

Figure 3. *Decoded BER vs* f_{nd}, $Ls = 100$, $SER_t = 0.15$

In Fig. 2, when $f_{nd} = 10^{-3}$, we see a large performance drop for all three algorithms. Since the channel estimates becomes increasingly outdated, the modulation format chosen based upon the CSI becomes increasingly meaningless. We see a significant increase of the decoder BER when f_{nd} reaches 10^{-3}, and the decoded PSNR is much worse than that at $f_{nd} = 10^{-4}$. When the normalized Doppler spread is 10^{-2}, the decoded BER is too large for the system to function.

In Fig. 4, we decrease the value of L_s to 25. Recall that the fraction of data is $\lambda = (L_s - 1)/L_s$. Decreasing the value of L_s will increase the amount of overhead (pilot symbols) used for channel estimation. However, decreasing the value of L_s will help the system to achieve better estimation accuracy at high Doppler. Compared to a system with $L_s = 100$ in Fig. 2, we see that a pilot spacing $L_s = 25$ has better performance when $f_{nd} = 10^{-3}$, despite the drop of the source rate. If we further decrease the value of L_s to 5 as in Fig. 5, the loss of the source data further increases to 20%, but the benefits of accurate channel estimation and modulation adaptation allow the system to operate at a reasonable PSNR value at $f_{nd} = 10^{-2}$.

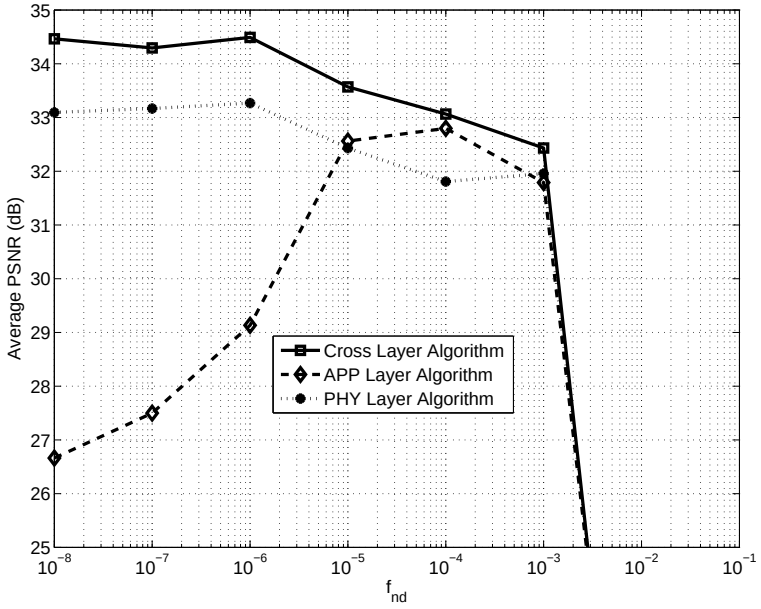

Figure 4. $Ls = 25$, *16 Subcarriers, 3 Users,* $SER_t = 0.15$, $\mu = 1$

B. Comparison of the Three Algorithms

Comparing the systems with different L_s, we see that the performance follows a very similar trend. For both the physical layer and cross layer algorithms, we see that PSNR decreases when f_{nd} increases from 10^{-6} to 10^{-3}. Since we use instantaneous CSI $\widetilde{H}_{k,m}[1]$ for resource allocation, the multiuser diversity will allow both physical layer and cross layer optimization algorithms to assign subcarriers that are, with high probability, experiencing a strong channel compared to average channel gain. In other words, if a subcarrier m^* is assigned to a user k^*, the instantaneous

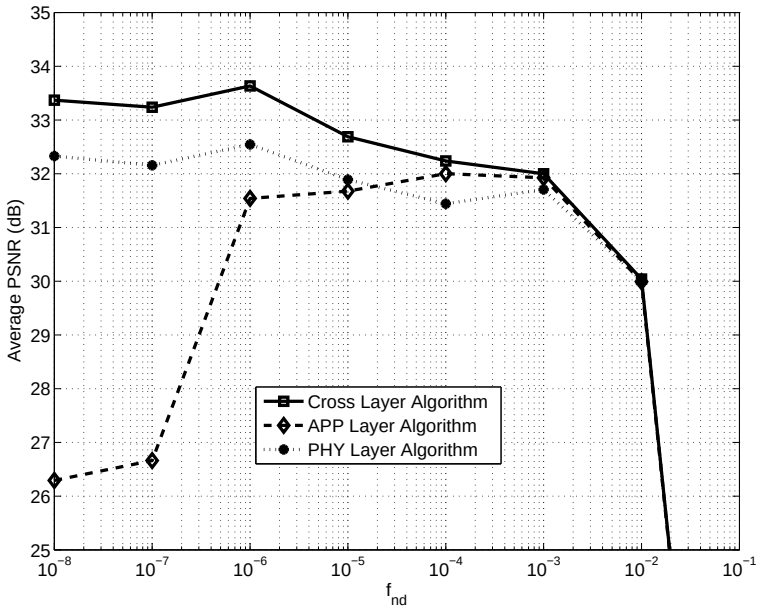

Figure 5. *$Ls = 5$, 16 Subcarriers, 3 Users, $SER_t = 0.15$, $\mu = 1$*

CSI $\widetilde{H}_{k^*,m^*}[1]$ at the beginning of the GOP has a high probability of being better than the average CSI. If the coherence time of the channel is long, the CSI will be relatively constant over the duration of the GOP. In this particular system, if f_{nd} is in the range of 10^{-8} to 10^{-6} the channel is not varying much and the modulation format changes little over a GOP. As f_{nd} increases, at some point, we expect to see different fades of the channel over a GOP. The average of these different fades with high probability will be worse than the CSI at the beginning of the GOP. In particular, the initial CSI will be outdated very soon for systems with f_{nd} in the vicinity of 10^{-3} to 10^{-1}.

In Fig. 2, for the application layer algorithm, we see that the performance improves significantly when f_{nd} increases from 10^{-8} to 10^{-4}. From (2), we see that the first zero-crossing of the correlation function occurs when the product of Δl and f_{nd} is about 0.4. When f_{nd} is 10^{-4}, the correlation between the first and the 4000th symbol is roughly 0. Since a GOP spans 5×10^4 symbols, we would expect to see about 10 independent fades during one GOP. In (7), the modulation format for a group of L_s symbols is determined by the estimate of the instantaneous CSI $\widetilde{H}_{k,m}[iL_s+1]$. Note that when the channel coherence time is much smaller than the GOP duration, since we assume the channel is both stationary and ergodic, every subcarrier

will ultimately (i.e., for high enough f_{nd}) tend to approach the same alphabet size given by $\mathbb{E}_{H_{k,m}}[R_{k,m}(P^d_{k,m}, \widetilde{H}_{k,m}[iL_s+1])]$.

Comparing the application layer and physical layer optimization algorithms, we see that the physical layer algorithm has better performance when f_{nd} is small, and the application layer algorithm wins when f_{nd} increases. That is, it is more important to exploit the multiuser channel diversity and allocate the resources based on the channel realization in a slowly varying environment. As the channel will stay relatively constant for a GOP, the resource allocation made based on the CSI at the beginning of the GOP will be meaningful over the entire GOP. On the other hand, when the coherence time is sufficiently small, the CSI will become outdated and the throughput of each subcarrier will be very similar regardless of the initial state of the channel. Thus, when f_{nd} is high, it is more important to allocate resources based on RD information, and the application layer algorithm has better performance. Note that the cross layer optimization algorithm is robust with respect to different normalized Doppler spreads and always outperforms the two baselines, **until the Doppler spread is too large for the system to track the fades, at which point all three algorithms fail.**

C. Systems with Different Resources

We now study the performance for different parameter values. In Fig. 6, we increase the number of users to 4. In Fig. 7(a) and Fig. 7(b), we show the performance of a system of 24 subcarriers with 3 and 4 users, respectively. Comparing Figures 2, 6, 7(a) and 7(b), we see that the system with 24 subcarriers and 3 users not only has the best performance, but also has the largest gap between the cross layer and physical layer optimization algorithms. This is because when the system has 24 subcarriers and 3 users, the cross layer algorithm has more degrees of freedom to allocate the resources among the users.

In Table I, we show a typical example of the performance evolution of a three-user system with 16 subcarriers and $f_{nd} = 10^{-6}$. The resource allocation is done based on the estimate of the first sample of the GOP, and we show the change of the estimated MSE for each individual user at each step of the iteration. Among the three users, the first and second users have demanding RD curves and the corresponding b_k of the RD information is much larger than that of the third user. For the cross layer algorithm initialization, the average PSNR is equal to 32.48 dB, and subcarriers are assigned evenly to all the users. The algorithm will converge in 3 steps (3

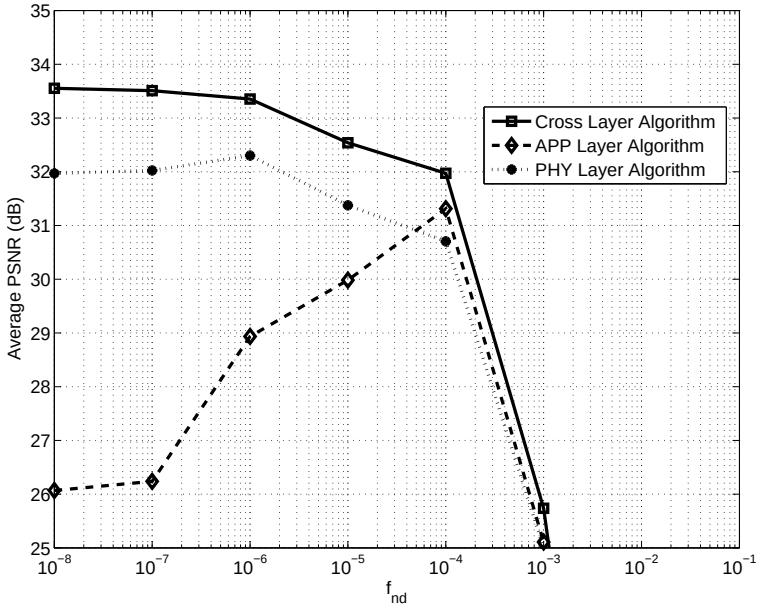

Figure 6. $Ls = 100$, *16 Subcarriers, 4 Users,* $SER_t = 0.15$, $\mu = 1$

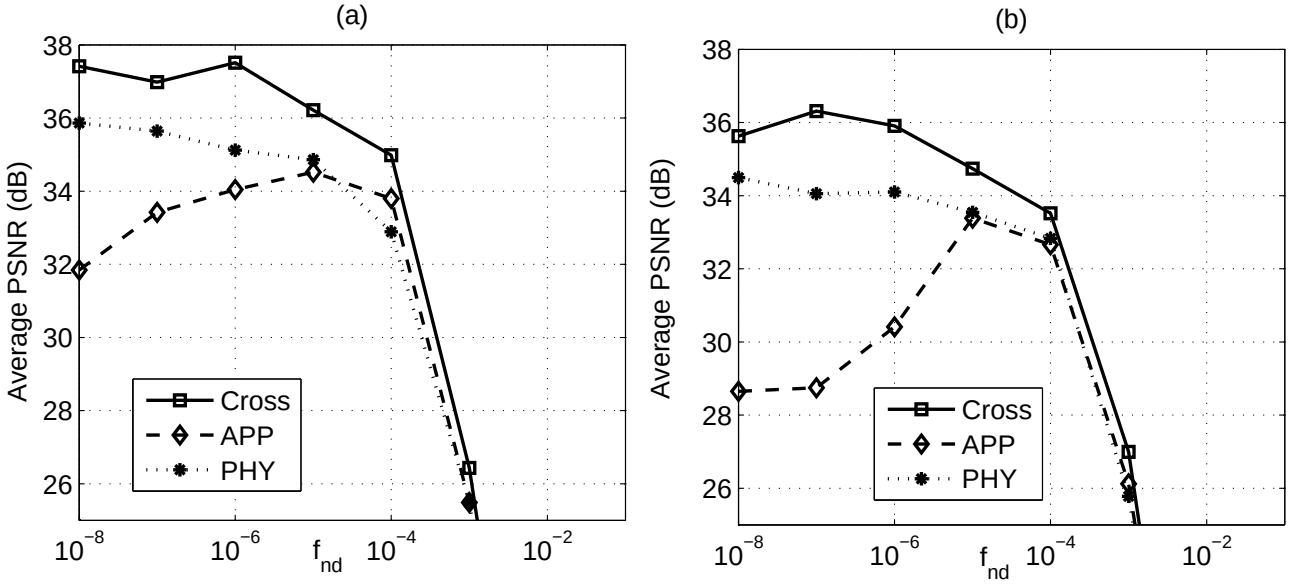

Figure 7. $Ls = 100$, *24 Subcarriers,* $SER_t = 0.15$, $\mu = 1$*; (a) 3 Users, (b) 4 Users.*

iterations) and the final subcarrier allocation is 6 and 7 for the two demanding users and 3 for the third user. The average PSNR value of the cross layer algorithm is 33.16 dB. In this case, the cross layer optimization improves the initial performance of the system by only 0.7 dB.

In Table II, we study a system with 24 subcarriers and 3 users. To compare with the previous

	U1 MSE	U1 NSub	U2 MSE	U2 NSub	U3 MSE	U3 NSub	Avg. MSE	PSNR
Initialization	53.68	5	56.52	5	~ 0	6	36.73	32.48
Step 1	53.68	5	45.89	6	~ 0	5	33.19	32.97
Step 2	44.82	6	45.89	6	4.66	4	31.79	33.10
Step 3	44.82	6	40.35	7	9.08	3	31.41	33.16

Table I

Evolution of Performance: 16 Subcarriers, U1=User 1, U2=User 2, U3=User 3. NSub=# Subcarriers Allocated.

	U1 MSE	U1 NSub	U2 MSE	U2 NSub	U3 MSE	U3 NSub	Avg. MSE	PSNR
Initialization	34.38	7	29.94	8	~ 0	9	21.43	34.82
Step 1	29.41	8	29.94	8	~ 0	8	19.78	35.17
Step 2	25.27	9	29.94	8	~ 0	7	18.40	35.48
Step 3	25.27	9	25.31	9	~ 0	6	16.86	35.86
Step 4	23.11	10	25.31	9	~ 0	5	16.14	36.05
Step 5	23.80	9	23.31	10	~ 0	5	15.71	36.17

Table II

Evolution of Performance: 24 Subcarriers, U1=User 1, U2=User 2, U3=User 3. NSub=# Subcarriers Allocated.

example fairly, the channel realizations for 16 out of the 24 subcarriers are the same as the realizations of the previous example. Realizations for the remaining 8 subcarriers are generated using the same mechanism as for the first 16 subcarriers. For the 24-subcarrier system, the initialization of the allocation is (7,8,9) subcarriers for the three users, and the corresponding average PSNR is 34.82 dB. It takes the cross layer algorithm 5 switches of subcarriers, and the cross layer algorithm assigns (9,10,5) subcarriers to the three users. The resulting average PSNR performance is 36.17dB, and so the cross layer algorithm improves the performance by 1.35 dB. Going from 16 to 24 subcarriers, we see that it takes the cross layer algorithm more steps to converge and the PSNR improvement is larger.

D. Systems With Different μ

Besides lowering the value of L_s, another way of achieving accurate channel estimation is to allocate more power to pilot symbols. In Fig. 8, we show the system performance versus the pilot-to-data power ratio μ for the cross layer algorithm when $f_{nd} = 10^{-3}$. As expected, for small values of μ, the pilot power is too small, and the PSNR is correspondingly small. As μ increases, the PSNR increases up to a point where it levels off, and eventually (as μ continues to

increase) it will decrease, because as more power is put into the pilot tones, the allocated to the data symbols is decreased. Note that the $L_s = 25$ curve performs better the $L_s = 100$ curve for all values of μ, because the normalized Doppler spread of 10^{-3} results in too much outdating when pilots are spaced every 100 symbols apart.

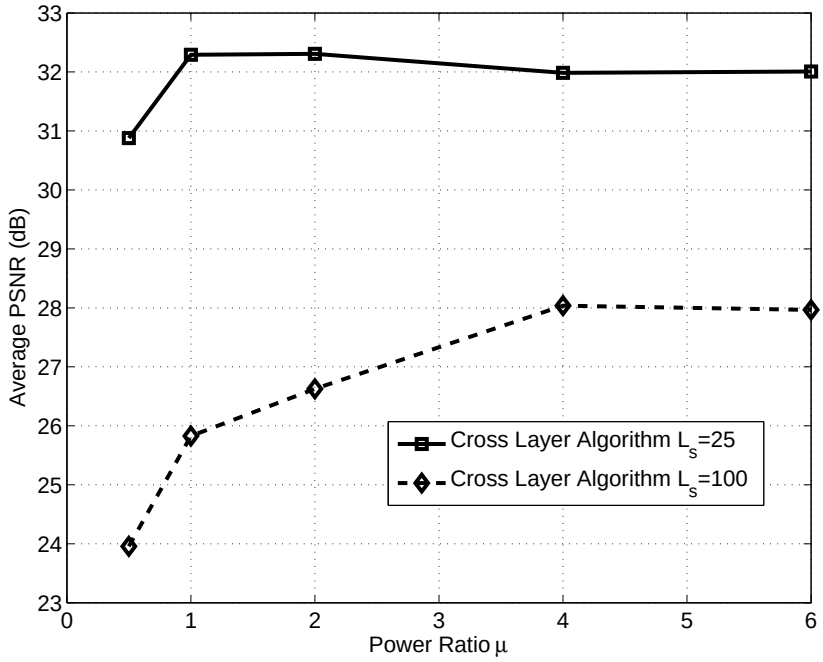

Figure 8. *16 Subcarriers,* $SER_t = 0.15$*,* $f_{nd} = 10^{-3}$.

E. Systems With Different Number of Users

We now study the system performance when the number of users increases. We fix the number of subcarriers to 16 and change the number of users in the system from 3 to 9. In Fig. 9(a), we show the average performance of all users when $f_{nd} = 10^{-6}$ and $L_s = 100$. We see that in this case, when f_{nd} is small, the PSNR gain of using the cross layer scheme is very large compared to the application layer scheme. With 9 users, the users will normally get a small number of subcarriers. For the application layer algorithm, the subcarriers are randomly assigned, and because the number of subcarriers for each user is small or zero, it is very likely that at least one user will be in a very bad situation. The user will thus have to function with, at best, a low rate, and this has a large impact on the average distortion of the group.

Compared to the physical layer algorithm, the PSNR gain of the cross layer scheme is 1dB in a three-user system and more than 2 dB in a 6 or 9-user system. Alternatively, rather than fixing the number of users, if we fix the average PSNR at, say, 30 dB, the capacity gain of the cross layer algorithm is a factor of three times that of the application layer algorithm, and about a factor of 1.5 times that of the physical layer algorithm.

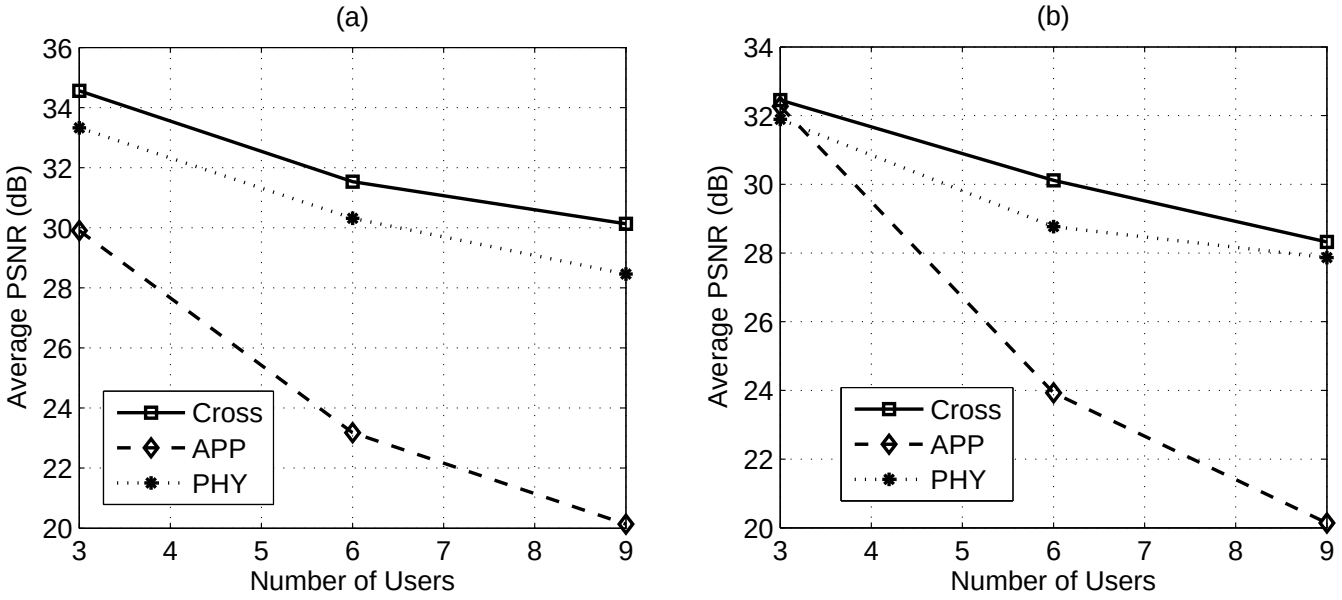

Figure 9. *16 Subcarriers,* $SER_t = 0.15$*,* $\mu = 1$*: (a)* $Ls = 100$*,* $f_{nd} = 10^{-6}$*, (b)* $Ls = 25$*,* $f_{nd} = 10^{-3}$*.*

In Fig. 9(b), we let $f_{nd} = 10^{-3}$ and $L_s = 25$. We see that the cross layer scheme outperforms the physical layer and application layer schemes by less than 0.5 dB, respectively, with 3 users. The gap between the cross layer and the physical layer algorithms is in the range from 0.6 to 1.3 dB when f_{nd} increases, which is smaller than that for $f_{nd} = 10^{-6}$. This is because when f_{nd} increases, the CSI used for resource allocation is outdated more quickly, and this affects the ability of the cross layer algorithm to balance the resources among the users. In other words, at high f_{nd}, the overall system performance is more determined by the number of subcarriers assigned to each user, and the effect of the CSI of the first symbol of the GOP has less impact than that for a slow fading system.

Comparing the cross layer and application layer optimization algorithms, we see that the performance of the application layer algorithm decreases sharply when the number of users increases, because some of the less demanding users in the system will not be allocated any

subcarriers. The performance of these users will be determined by holding the last frame of the previous GOP, resulting in a low average PSNR value.

VI. Conclusion

In this paper, we study a multicarrier uplink video communication system over a channel with arbitrary normalized Doppler spread. We use both the application layer RD information and the physical layer CSI to allocate subcarriers to video users. After the resource allocator assigns the subcarriers to the users, each user continues to send pilot symbols and update the modulation format of each subcarrier based on feedback from the base station. The critical parameter L_s controls the tradeoff between the channel outdating and source rate. Our key results can be summarized as follows: For a system that has to function well over a large range of Doppler spread, robustness in performance is arguably the key characteristic that it should exhibit. The cross layer design we presented in this paper satisfies this criterion. In particular, it was seen that if only a single layer design is employed, for some values of the Doppler spread the application layer algorithm was superior to the physical layer algorithm, with the opposite result holding for the other ranges of the Doppler spread. However, the cross layer algorithm outperformed both single layer algorithms over the range of Doppler spread that allowed meaningful performance.

References

[1] A. Mohr, E. Riskin, and R. Ladner, "Unequal loss protection: graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 819–828, Jun. 2000.

[2] Y. Chan, P. Cosman, and L. Milstein, "A multiple description coding and delivery scheme for motion-compensated fine granularity scalable video," *IEEE Trans. Image Processing*, vol. 17, no. 8, pp. 1353–1367, Aug. 2008.

[3] G.-M. Su, Z. Han, M. Wu, and K. Liu, "Joint uplink and downlink optimization for real-time multiuser video streaming over WLANs," *IEEE J. Select. Areas Commun.*, vol. 1, no. 2, pp. 280 –294, Aug. 2007.

[4] M. Van der Schaar and D. Turaga, "Cross-layer packetization and retransmission strategies for delay-sensitive wireless multimedia transmission," *IEEE Trans. Multimedia*, vol. 9, no. 1, pp. 185–197, Jan. 2007.

[5] Q. Liu, S. Zhou, and G. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Trans. Commun.*, vol. 3, no. 5, pp. 1746–1755, Sep. 2004.

[6] X. Cai and G. Giannakis, "Adaptive PSAM accounting for channel estimation and prediction errors," *IEEE Trans. on Wireless Commun.*, vol. 4, no. 1, pp. 246 – 256, Jan. 2005.

[7] L. Toni, Y. Chan, P. Cosman, and L. Milstein, "Channel coding for progressive images in a 2-D time-frequency OFDM block with channel estimation errors," *IEEE Trans. Image Processing*, vol. 18, no. 11, pp. 2476–2490, Nov. 2009.

[8] A. Khalek, C. Caramanis, and R. Heath, "A cross-layer design for perceptual optimization of H.264/SVC with unequal error protection," *IEEE J. Select. Areas Commun.*, vol. 30, no. 7, pp. 1157–1171, Jul. 2012.

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

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

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

[12] D. Wang, L. Toni, P. Cosman, and L. Milstein, "Uplink resource management for multiuser OFDM video transmission systems: Analysis and algorithm design," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 2060–2073, May 2013.

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

[14] W. Jakes, *Microwave Mobile Communications*. Wiley, 1974.

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

[16] Q. Qu, L. Milstein, and D. Vaman, "Cognitive radio based multi-user resource allocation in mobile Ad hoc networks using multi-carrier CDMA modulation," *IEEE J. Select. Areas Commun.*, vol. 26, no. 1, pp. 70 –82, Jan. 2008.

[17] F. Wu, S. Li, and Y. Zhang, "A framework for efficient progressive fine granularity scalable video coding," *IEEE Trans. on Circuits and Syst. for Video Tech.*, vol. 11, no. 3, pp. 332 –344, Mar. 2001.

[18] M. van der Schaar and H. Radha, "Adaptive motion-compensation fine granular-scalability (AMC-FGS) forwireless video," *IEEE Trans. on Circuits and Syst. for Video Tech.*, vol. 12, no. 6, pp. 360 –371, Jun. 2002.

[19] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. on Circuits and Syst. for Video Tech.*, vol. 17, no. 9, pp. 1103 –1120, Sep. 2007.

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

[21] G.-M. Su, Z. Han, M. Wu, and K. Liu, "A scalable multiuser framework for video over OFDM networks: Fairness and efficiency," *IEEE Trans. on Circuits and Syst. for Video Tech.*, 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. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels [mobile radio]," *IEEE Trans. on Veh. Technol.*, vol. 40, no. 4, pp. 686 –693, Nov. 1991.

[24] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[25] 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.

[26] ——, "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.

[27] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 920 – 928, Jun. 2003.

[28] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.