

Research Article

Cross-Layer Framework for Multiuser Real Time H.264/AVC Video Encoding and Transmission over Block Fading MIMO Channels Using Outage Probability

Slavche Pejосki and Venceslav Kafedziski

Faculty of Electrical Engineering and Information Technologies, University Cyril and Methodius, Skopje, Macedonia

Correspondence should be addressed to Slavche Pejосki; slavchep@feit.ukim.edu.mk

Received 25 June 2013; Accepted 29 August 2013; Published 2 January 2014

Academic Editor: Martin Reisslein

Copyright © 2014 S. Pejосki and V. Kafedziski. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present a framework for cross-layer optimized real time multiuser encoding of video using a single layer H.264/AVC and transmission over MIMO wireless channels. In the proposed cross-layer adaptation, the channel of every user is characterized by the probability density function of its channel mutual information and the performance of the H.264/AVC encoder is modeled by a rate distortion model that takes into account the channel errors. These models are used during the resource allocation of the available slots in a TDMA MIMO communication system with capacity achieving channel codes. This framework allows for adaptation to the statistics of the wireless channel and to the available resources in the system and utilization of the multiuser diversity of the transmitted video sequences. We show the effectiveness of the proposed framework for video transmission over Rayleigh MIMO block fading channels, when channel distribution information is available at the transmitter.

1. Introduction

The telecommunication industry is expected to benefit greatly from the transmission of video. Great portion of this video traffic will be real time, so, real time multiuser video transmission is expected to become a common service. The optimal real time video transmission over wireless channels is still an open problem. This is due to the complexity and requirements of the video coding process and the severity of the wireless channel. Namely, real time video requires variable bit rates and low delay, the wireless system has limited resources (bandwidth and power) and the wireless channel itself is time varying [1]. Previous research results [1–3] have shown that cross-layer adaptation can result in substantial performance improvement compared to treating the different layers as isolated entities. Here, we utilize the implementation of the cross-layer principle for real time H.264 single layer video encoding and transmission in systems that use capacity achieving codes presented in [4] and extend the concept to a multiuser MIMO TDMA system. It has been shown that the utilization of the multiuser principle can improve

system performance in a video streaming scenario [5, 6], but the joint cross-layer optimization for real time H.264/AVC single layer video encoding and transmission in a multiuser wireless system based on outage probability has not been addressed yet. The novelty of our work is in proposing a cross-layer framework for optimized real time transmission and encoding of video from multiple users in a TDMA wireless system using capacity achieving channel coding and single layer H.264/AVC video encoding in a MIMO block fading channel when channel distribution information is available at the transmitter. The similarities and differences to other publications will be explained in the next section.

The proposed cross-layer framework includes the application layer by changing the quantization level, the MAC layer by the allocation of the available time slots to different users and the physical layer by choosing the utilized number of bits per channel use. The video encoding and the transmission are carried out sequentially for every video frame. By taking into account the models utilized to describe different parts of the system, the proposed framework is most suitable for multiuser uplink video calls in a frequency division duplex

(FDD) system or downlink multiuser transmissions combined with transcoding. To characterize the framework, we give a detailed description and explanation of the procedures in the source coding part, the rate distortion modeling, the channel transmission modeling, the multiuser coordination, and the resource allocation procedure. The multiuser system has a central coordinator. The coordinator receives information about the video coding parameters and channel statistics of different users and allocates the available resources (time slots) to appropriate users in order to minimize the expected sum end-to-end distortion per video frame. The video coding parameters of the users consist of pairs of distortions and rates for allowed quantization parameters, the maximal distortion due to propagation, and the maximal distortion due to concealment. Those parameters are obtained as explained in [4] where a modified two stage video encoding is used, which naturally fits in the multiuser concept with a central coordinator. The first stage of the video encoding allows for the calculation of the video coding parameters and mitigates the influence of transmission errors in the past on the future video frames. In the second stage the actual encoding takes place and allows for adaptation to the available resources. The considered communication system is MIMO TDMA, where a slot can be occupied only by a single user. During the transmission of a video frame several slots are available and, the users, based on their video coding parameters and channel statistics will obtain an appropriate number of slots. The channel of every user is described by the probability density function (pdf) of the channel mutual information. For our system with Rayleigh MIMO block fading channel with channel distribution information at the transmitter, the pdf of the channel mutual information is approximated as Gaussian. If a user is allowed to transmit in several slots, the pdfs of the individual slots can be combined into a new pdf. For the considered system, this process becomes very simple and the new pdf is also Gaussian. The resource allocation is based on a modified Lagrangian optimization. It allocates the available slots to different users and decides on the video parameters used in the encoding of the current video frame. Then, the second stage of the video encoding and the actual transmission take place. We test the proposed framework in various simulation scenarios with different available resources and users in the system using a Rayleigh MIMO block fading channel and show its superior performance to the video transmission frameworks when multiuser principle is not utilized and when the video encoding is not carried out in real time.

The rest of the paper is organized as follows. In the second section we give an overview of related work. In the third section we explain the cross-layer video transmission framework. In Section 4 we present the simulation results of our framework. Section 5 concludes the paper.

2. Related Work

The work presented in this paper is based on the work from [4], where we proposed a cross-layer framework for video transmission for single user. In [4] the rate distortion

(RD) modeling of the video is carried out on a frame by frame basis by modifying the approach in [7], the end-to-end distortion is calculated based on [8] and the physical layer characterization is done in a similar manner as in [5] for an environment where capacity achieving channel codes are utilized. It should be noted that the work in [4] is an extension of our previous work [9, 10] by replacing the H.263 source encoder with a single layer H.264/AVC and considering only the single input single output scenario without utilizing diversity. Here we adopt the framework from [4] and extend it to a MIMO TDMA system. The new framework for multiuser transmission allows investigating the influence of the different parameters and possibilities in the joint source channel coding for real time video transmission over MIMO TDMA systems.

Multiuser video transmission in a wireless environment has been extensively investigated in the literature [5, 6, 11–31]. The publications in this area, when considering the video encoding, can be categorized into those that consider transmission of scalable video [11–14], transmission of preencoded single layer video [5, 6, 15–21], smoothing based [32, 33], video trace based system level simulations [22–24], testbed oriented approach [31] (preencoded video is transmitted but the implementation issues are the main interest), and those that consider change of the parameter of the single layer video encoders immediately prior to transmission [26–30]. Only the work from the last category can be strongly correlated to our work, but the publications in this category use rate distortion models applicable only for long term rate adaptation, unsuitable for real time frame by frame adaptation due to their low precision and large complexity or do not explain how the adaptation is carried out. Here we consider transmission of video where the encoding parameters are adapted in real time based on a model with precision and complexity suitable for real time transmission (for detailed comparison of the RD models we refer the interested reader to [4]). Also, we explain the adaptation process in detail. When considering the channel coding it should be noted that most of the publications in the field of multiuser video transmission do not consider transmission errors or use convolutional codes. Here we investigate the video transmission in a communication system that utilizes capacity achieving channel codes [5, 15, 19] where the outage probability is the main cause of packet errors.

Many authors consider video transmission over MIMO channels [10, 31, 34–39]. Most of the algorithms considered there are intended for layered or preencoded video or single layer video divided into parts with different importance that naturally fit in the unequal error protection concept [35–40] or even combine this concept with the cooperative communication concept [41]. Here we are interested in video transmission where the video rate is adapted on a frame by frame basis, where the concept of unequal protection is not important, and where the issue of cooperative communication is not addressed. Such approach can be found in [31, 34] for some practical MIMO schemes and in [10] as a theoretical concept compatible with the capacity achieving channel coding. The second concept is adopted here and it

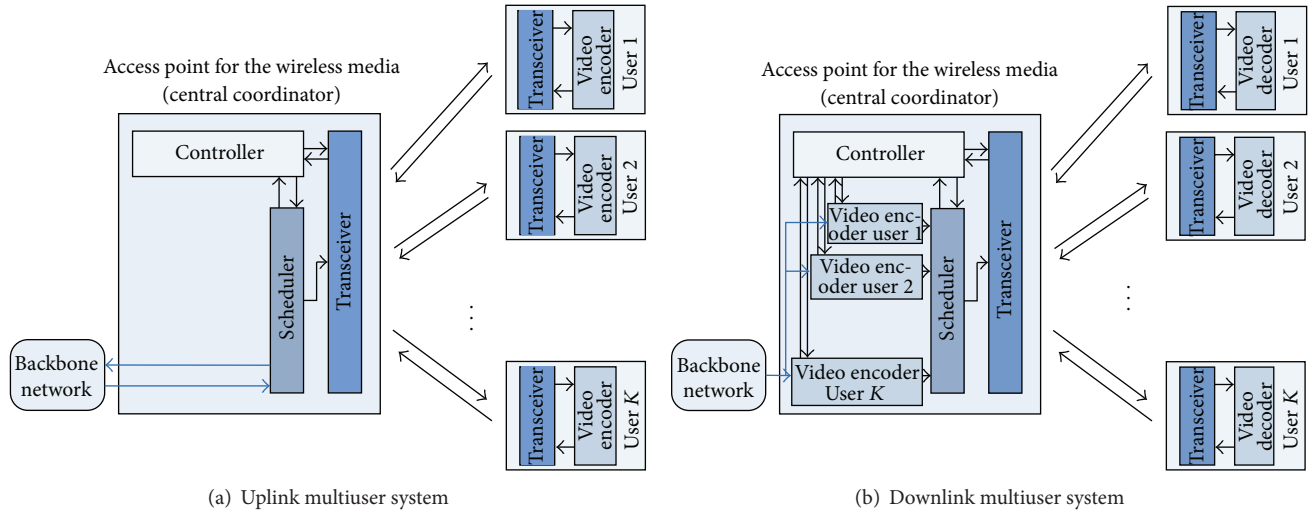


FIGURE 1: Multiuser video transmission systems of interest.

is extended to a multislot multiuser TDMA system and to H.264/AVC single layer video encoding.

3. Multiuser Cross-Layer Framework

The block diagram of both scenarios of the multiuser video transmission system considered here is shown in Figure 1.

The two scenarios of interest are the FDD uplink communications system presented in Figure 1(a) and downlink system with transcoding of uncoded video shown in Figure 1(b). In those scenarios the transmitting entity should encode the video sequence during the transmission process and only the channel statistics is known at the transmitting entity. The delay limitations are very strict and retransmissions are not included. Generally, retransmissions are considered to be inappropriate for real time content delivery, where data reception is significantly time-constrained [24], which is in line with our assumptions. Also, it is considered that the transmission over the backbone network is carried out without error. In both scenarios there exists a central controlling entity (central coordinator) at the base station that selects the encoder parameters and performs resource allocation for all the users. As seen in Figure 1 there is a feedback channel from the receiving to the transmitting entity to assist the video transmission by feedback information about the channel statistics, about the success of the reception of the video frames and about the transmission related parameters chosen by the controller for both scenarios and additionally, the information about the video coding parameters chosen by the controller for the uplink scenario. Next, we will explain the different parts of the system in a greater detail.

3.1. Video Source Encoder. The video coding standard used here is a single layer H.264/AVC. A video encoder is associated with every user in the system and must supply the controller with information about the resulting distortion and the rate needed for the video encoding procedure for the

appropriate user, and, after having the decision made by the controller, encode the video frame of that user according to the received information. In our framework this is solved by using the RD model based on the modified two pass approach from [7] for each user. In the first phase of this approach the video frame is passed through the motion estimation, mode decision, transform, quantization, dequantization, inverse transform, and reconstruction. We denote the quantization parameter in the first phase as Q_{p1} and the quantization parameter in the second phase as Q_{p2} . In the first phase the video frame is processed by setting the quantization parameter Q_{p1} equal to Q_{p2} of the previous video frame when no error is reported or to Q_{p2} of the previous video frame plus 3 if error is reported through the feedback channel. In this phase the accurate estimate of the header information and the Sum of Absolute Transformed Differences (SATD) can be obtained, and an accurate model for the number of bits used in the encoding procedure can be created. Then, in the second phase of the encoding process, a new value for the quantization parameter Q_{p2} that differs from the one used in the first phase by a maximum value of 3 is used:

$$\max |Q_{p1} - Q_{p2}| = 3. \tag{1}$$

The restriction in (1) is used since it leads to only a small degradation in overall coding efficiency and controls the quality variations between frames [7]. Both features are important. The first one allows for near optimal performance and the second one accounts for the temporal video quality variations which is impossible to be measured by the PSNR value but is considered as important in the subjective video assessment. The PSNR, and, more specifically, the Mean Squared Error (MSE), are used as optimization criteria in the proposed framework.

The motion vectors and macroblock coding modes determined in the first phase are also used in the second phase. For the intercoded parts of the frame the operations of requantization, coding, dequantization, inverse transform,

and reconstruction are carried out in the second phase. For the intracoded parts, the operations performed in the second phase are prediction, quantization, coding, dequantization, inverse transform, and reconstruction. These operations are performed only for the single mode chosen in the first phase. Having in mind that the highest complexity in the encoding process comes from macroblock mode decision and motion estimation, the extra computation only marginally increases the overall encoding complexity.

In the model from [7], the number of source coding bits per frame required to encode the residual signal and the distortion due to source coding for each user are calculated as follows:

$$R_{\text{src}}(Q) = \frac{\alpha \text{SATD}_c(Q)}{Q^{p_r}}, \quad (2)$$

$$D_s(Q) = D_{\text{skip}}(Q) + \beta \text{SATD}_c(Q) Q^{p_d}, \quad (3)$$

where α and β are model parameters that are specific to the video sequence and are calculated by the least squares approach based on the last five video frames, p_r and p_d are parameters specific to the video encoder and they are different for intra- and inter-coded video frames, $\text{SATD}_c(Q)$ is the SATD of the 4×4 blocks that will be encoded at a quantization level Q , and $D_{\text{skip}}(Q)$ is the distortion of the 4×4 blocks that will not be quantized and coded at this quantization level. Q is the quantization level associated with the quantization parameter Q_p . For further details we direct the interested reader to [7].

During the macroblock mode decision in the first phase of the encoding process the correct number of header bits (R_{header}) must be known and sent to the controller for further optimization. The estimation of R_{header} can be carried out by encoding it or by using a model as the one in [7]. R_{header} is constant for different values of Q , so the model for the required number of bits is calculated as in [7]:

$$R(Q) = R_{\text{header}} + R_{\text{src}}(Q). \quad (4)$$

To consider video transmission over channels with errors, we split the overall distortion of each user as in [8]:

$$\begin{aligned} D &= E \left\{ (f_n^i - \tilde{f}_n^i)^2 \right\} = (1 - p_n) E \left\{ (f_n^i - \tilde{f}_n^i)^2 \right\} \\ &\quad + (1 - p_n) E \left\{ (\tilde{f}_{\text{ref}}^j - \tilde{f}_{\text{ref}}^j)^2 \right\} + p_n E \left\{ (f_n^i - \tilde{f}_{n-1}^k)^2 \right\} \quad (5) \\ &= (1 - p_n) D_s + (1 - p_n) D_p + p_n D_c, \end{aligned}$$

where p_n is a constant corresponding to the error probability of video frame n . In (5), D is the overall end-to-end distortion, D_s is the source coding distortion, D_p is the distortion due to error propagation, and D_c is the distortion due to error concealment. The estimated distortion is calculated at a pixel or block level, but, by combining the appropriate parts it can be used at different scales. During the macroblock mode selection it will be calculated at a macroblock level and during the optimization of the quantization parameter and the utilized number of bits per channel use it will be computed at a frame level so the source coding distortion

in this case will be equal to the one calculated in (3). In (5) \tilde{f}_n^i stands for the value of the i th coded pixel of the n th video frame at the encoder, obtained by encoding the uncoded value f_n^i of the i th pixel of the n th video frame, \tilde{f}_{ref}^j and \tilde{f}_{ref}^j stand for the values of the j th pixel from the (ref)th video frame at the encoder and the decoder used as reference pixel during the encoding of the i th pixel of the n th video frame. \tilde{f}_{n-1}^k is the k th pixel from the $(n-1)$ th video frame used to conceal the i th pixel in an error event. \tilde{f}_{n-1}^k and \tilde{f}_{ref}^j are not known at the encoder, and, thus, they are random variables that depend on the error probability of transmission of the video frames with number lower than n . Equation (5) is based on the assumption of additivity of the distortion due to source coding at the encoder and the distortion due to transmission errors, evaluated at the decoder. The distortion due to error propagation and error concealment can be further decomposed until the values of all pixels needed to calculate the distortion are known.

The mode selection procedure for every macroblock is based on the Lagrangian optimization:

$$\min J = \min (D + \lambda R). \quad (6)$$

The choice of λ is based on the conclusions in [8, 42]; that is, we set $\lambda = (1 - p_n)\lambda_0$, where λ_0 is the value of the Lagrangian multiplier for error free environments. In order to make the video coding resilient to transmission errors we use the distortion from (5) in (6), which increases its complexity. This complexity increase comes from the recursive nature of the algorithm for calculating the distortion due to propagation and error concealment, and depends on the number of reference frames and the method (per pixel or per block) used in the specific implementation. Since D_c is independent of the chosen mode for the current macroblock and p_n is constant, according to [8], (6) can be simplified to $\min J$ with $J = D_s + D_p + \lambda_0 R$, where λ_0 is the value of the Lagrangian multiplier for error free environments. For the motion compensation procedure $D = D_s$ is used. If additional complexity can be allowed in the system; then (6) can be also used in the motion compensation procedure.

At the end of the first phase of the video coding procedure, the video source encoder of each user obtains the pairs $(R(Q), D_s(Q))$ for Q associated with $Q_p \in \{Q_{p1} - 3, Q_{p1} - 2, \dots, Q_{p1} + 2, Q_{p1} + 3\}$ and the values of D_p and D_c . These are the video coding parameters that are sent to the controller. After the optimization, the video encoder receives the value of Q_{p2} . Then, the second phase of the encoding process takes place. The bit stream obtained from this second phase is sent to the transmitter.

3.2. Transceiver. The transceiver includes all the functionalities of the physical layer. Since feedback information is required for proper functioning of the system this system block should be capable of two-way communication. The communication part where only feedback information is sent is error free. This, for information precision close to perfect, can be achieved by using very low number of bits per channel use and since the feedback information has much lower rate

than the video information this will not affect the system efficiency in a significant manner. We consider a system that utilizes capacity achieving codes. In such a system, according to [5] the outage probability is the main cause of packet errors, so that the frame error rate, that is, the error probability, is equal to the outage probability. In the transmitter, the channel mutual information can be described by its pdf $\rho(I)$. Having determined the pdf of the channel mutual information, the outage probability, when utilizing r bits per channel use, can be calculated as

$$P_o(r) = P(I < r) = \int_0^r \rho(I) dI. \quad (7)$$

We consider a TDMA MIMO system, where the base station communicates with a single user during a time slot. Also, we are interested in a transmission over a MIMO Rayleigh block fading channel where all users have N_u antennas and the base station has N_B antennas and only one user is in communication with the base station. In such case during each block, the channel between the communicating entities with M transmit and N receive antennas is described by $N \times M$ channel matrix \mathbf{H} with independent and identically distributed zero mean circularly symmetric complex Gaussian entries with variance σ^2 . The utilization of a FDD uplink or a downlink system allows us to assume that only the channel distribution is known at the transmitter, and no information about the instantaneous values of the elements in \mathbf{H} is available at the transmitter. It should be noted that there are U such channel matrices describing the channels between the U active users whose videos are transmitted in the system and the base station, but only one such matrix is of importance in one slot. For the considered channel equal power allocation to all transmit antennas is the optimal solution [43]. The mutual information of such a MIMO channel with additive white Gaussian noise (AWGN) can be evaluated as

$$I = \sum_{i=0}^{k-1} \log_2 \left(1 + \frac{\text{SNR}}{M} \lambda_i^2 \right), \quad (8)$$

where $k = \min\{N, M\}$, P is the total transmit power available to all the transmit antennas of the MIMO system, N_0 is the power spectral density of the AWGN, and λ_i are the singular values of the matrix \mathbf{H} . For this channel we can define the channel signal to noise ratio as $\text{SNR} = P\sigma^2/N_0$. The elements of \mathbf{H} and also λ_i and I are random. In [44] it is shown that the pdf of the mutual information can be approximated using a Gaussian approximation:

$$p(I, M, N, \text{SNR}) \approx \frac{1}{\sqrt{2\pi}\sigma_I} e^{-(I-\mu_I)^2/2\sigma_I^2}. \quad (9)$$

In (9) μ_I and σ_I^2 are the mean and the variance of the mutual information, so in our system the channel of a user is represented by a pair (μ_I, σ_I) . The values of μ_I and σ_I^2 can be calculated as [43]

$$\mu_I = \int_0^\infty \log \left(1 + \frac{\lambda \text{SNR}}{M} \right) K(\lambda, \lambda) d\lambda,$$

$$\begin{aligned} \sigma_I^2 = & \int_0^\infty \log^2 \left(1 + \frac{\lambda \text{SNR}}{M} \right) K(\lambda, \lambda) d\lambda \\ & - \int_0^\infty \log \left(1 + \frac{\lambda_1 \text{SNR}}{M} \right) \\ & \times \log \left(1 + \frac{\lambda_2 \text{SNR}}{M} \right) K^2(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2, \end{aligned} \quad (10)$$

where $K(x, y) = \sum_{i=0}^{k-1} \tilde{\phi}_i(x) \tilde{\phi}_i(y)$, $\tilde{\phi}_i(\lambda) = \sqrt{i!/(i+d)!} L_i^d(\lambda) \lambda^{d/2} e^{-\lambda/2}$ and $L_i^d(\lambda) = (1/i!) e^{\lambda} \lambda^{-d} (d^i (e^{-\lambda} \lambda^{d+i}) / d! \lambda^i)$ is the associated Laguerre polynomial of order i . In the formulas above $d = \max\{M, N\} - k$. According to [44] this Gaussian approximation fits the actual mutual information pdf quite well and in our framework it is suitable for describing the channels since only a pair of parameters (μ_I, σ_I) or the value of the SNR for each user should be sent to the controller in order to optimize the transmission of the video.

Since we assume a block fading channel model, the channel is constant during one block and changes to an independent value in the next block. Here we consider that the duration of the block is equal to the duration of one time slot that can be allocated by the controller, but the analysis can be easily extended to the case when multiple blocks are contained in a slot. If l slots are allocated to a user, then the channel mutual information when transmitting in those l slots, according to the principles explained in [9], can be also approximated by the Gaussian approximation, but the new parameters of the distribution will be $\mu_{I,l} = \mu_I$ and $\sigma_{I,l} = \sigma_I/\sqrt{l}$. Since, only the channel distribution information is available at the transmitter, the positions of the slots are not important; that is, only their number is what matters.

After the optimization, parameters characterizing all the layers included in the transmission, in our case the chosen values of r , the allocated number of slots l , and the positions of the slots, will be sent from the controller to the transmitting and receiving entities. The positions of the allocated time slots will be also sent to the scheduler.

3.3. Controller. The controller receives information from the video source encoders and the transceivers of all users. The main purpose of the controller is to achieve minimal sum end-to-end video distortion. This is carried out by allocating the available time slots to the users and selecting the appropriate Q_p parameter (which is related to the utilized source rate) for the second phase of the encoding process of every user. Since the system is intended for real time transmission, the allowed delay is equal to the time between the encoding of two consecutive video frames. It should be noted that this stringent delay makes the utilization of smoothing techniques [32, 33] in such system unsuitable. During this time there are L time slots available for transmission. Combining the end-to-end distortion from (5) with the outage probability of every user from (7) and considering that U users are present

in the system, the controller will carry out the following constrained optimization:

$$\begin{aligned} \min_{\mathbf{Q}, l} \quad & \sum_{u=1}^U \left\{ \left(1 - P_{o,u} \left(\frac{R_u(Q_u)}{l_u N_s} \right) \right) (D_{su}(Q_u) + D_{pu}) \right. \\ & \left. + P_{o,u} \left(\frac{R_u(Q_u)}{l_u N_s} \right) D_{cu} \right\} \\ \text{s.t.} \quad & Q_1 \in \mathbf{Q}_{1a} \\ & \vdots \\ & Q_U \in \mathbf{Q}_{Ua} \\ & \sum_{u=1}^U l_u = L, \end{aligned} \quad (11)$$

where N_s is the total number of channel uses (number of symbols) per slot, and, thus, the utilized number of bits per channel use is $R_u(Q_u)/(l_u N_s)$, $\mathbf{Q} = [Q_1 \cdots Q_U]^T$, $l = [l_1 \cdots l_U]^T$, \mathbf{Q}_{ua} represents the allowed values for the Q parameter of user u according to (1). R_u , D_{su} , D_{pu} , and D_{cu} are the video coding parameters for user u and $P_{o,u}$ is the outage probability of user u when l_u slots are allocated to her/him. The first U constraints are due to the RD model and (1), and the last constraint comes from the available resources. We would like to emphasize that (11) accounts for frame drops in the system. Namely, if no slot is allocated to a user then its $P_{o,u}$ will be equal to 1 and its video frame will be dropped.

The optimization of (11) is not an easy task since the distortion of a single user u , $D_u(Q_u, l_u) = (1 - P_{o,u}(R_u(Q_u)/l_u N_s))(D_{su}(Q_u) + D_{pu}) + P_{o,u}(R_u(Q_u)/l_u N_s)D_{cu}$ is neither a convex nor a concave function on Q_u or l_u . Generally it can be solved by using exhaustive search but the computational complexity of such solution is very high. To reduce the complexity we propose a solution that uses Lagrangian relaxation in order to decouple (11) to U smaller problems connected by the Lagrangian parameter λ_g that comes from the last constraint in (11):

$$\begin{aligned} \min_{Q_u, l_u} \quad & \left\{ \left(1 - P_{o,u} \left(\frac{R_u(Q_u)}{l_u N_s} \right) \right) (D_{su}(Q_u) + D_{pu}) \right. \\ & \left. + P_{o,u} \left(\frac{R_u(Q_u)}{l_u N_s} \right) D_{cu} + \lambda_g l_u \right\} \\ \text{s.t.} \quad & Q_u \in \mathbf{Q}_{ua}. \end{aligned} \quad (12)$$

The problem (12) can be solved using an exhaustive search. This should not be computationally too complex, since the cardinality of \mathbf{Q}_{ua} is 7 and L values of l_u are used for evaluation of (12). Thus, the complexity of the evaluation of a single value of λ_g for all users is equal to $7LU$. The estimation of the value of λ_g can be done using one of the well-known methods, such as the cutting-plane or the subgradient method. We propose to solve the optimization using the bisectional method [45]. The solution of the Lagrangian optimization is always on the convex operational

rate distortion (ORD) curve of the solved problem. If the optimum point is not on the ORD curve we propose the use of a solution of the Lagrangian optimization with maximum $\sum_{u=1}^U l_u$ but such that $\sum_{u=1}^U l_u < L$ and then allocate the rest of the available slots, one by one in a greedy manner according to

$$\max_u \left\{ \max_{Q_u} (D_u(Q_u, l_u) - D_u(Q_u, l_u + 1)) \right\}. \quad (13)$$

In (11) and (12) the error probability $P_{o,u}(r)$ is calculated according to the type of channel in use. For our system this value can be calculated as

$$P_{o,u}(r) = \int_0^r \frac{1}{\sqrt{2\pi}\sigma_{I,l_u}} e^{-(x-\mu_{I,l_u})^2/2\sigma_{I,l_u}^2} dx. \quad (14)$$

Considering that μ_{I,l_u} is several σ_{I,l_u} 's apart from the value 0, the value of $P_{o,u}(r)$ can be very closely approximated as

$$P_{o,u}(r) \approx \frac{1}{2} \operatorname{erfc} \left(\frac{\mu_{I,l_u} - r}{\sqrt{2}\sigma_{I,l_u}} \right). \quad (15)$$

After the optimization is carried out, that is, the values of Q_u and l_u are obtained, those values are sent back to the transmitters, the receivers, the encoders of the appropriate users, and the scheduler in the base station. Then, the second phase of the encoding can take place. In order to keep the encoding rate as close to the one selected from the RD model, (in this phase) the rate control algorithm without bit allocation described in [7] is used at the encoders of different users. The bits obtained in the encoding procedure are sent to the appropriate transmitter and then transmitted in the wireless channel.

One interesting observation can be made about the power allocation to different antennas, that is, the antenna power profile. Here we assume that when slots are allocated to a user, it transmits at its peak allowed power level in each slot and uses equal power at each antenna. We can relax the constraint of using an equal power at each antenna and allow the controller to choose the antenna power profile. In such case, during the optimization (12) the controller would be able to choose this antenna power profile together with l_u and Q_u . At this point the antenna power profile will depend on l_u and Q_u only through the utilized number of bits per channel use $r = R_u(Q_u)/l_u N_s$. If we fix l_u and Q_u , that is, fix r , as long as

$$D_{cu} - D_{pu} - D_{su}(Q_u) > 0, \quad (16)$$

the antenna power profile that minimizes the outage probability at the utilized number of bits per channel use $r = R_u(Q_u)/l_u N_s$ is the optimal one. For our observed channel according to [43] the desired power profile is the one using equal power at each individual antenna. If (16) does not hold then it is better for the system not to transmit anything, so any chosen antenna power profile will lead to a point that is not on the ORD curve and cannot be chosen in the Lagrangian optimization procedure. Obviously, in our case, the antenna power profile is independent of l_u and Q_u . This discussion can be used to find the antenna power profile for

an arbitrary channel, when channel information is available at the transmitter and in the general case it will depend on l_u and Q_u .

An interesting situation arises if the feedback channel is able to obtain some partial information about the channel state at the transmitter. We will consider the case when there is at least one channel block between the beginning of the optimization and the transmission of the first block. This additional information by itself will not improve the performance of the system since the controller in the current setting is not aware of it. Additionally, the controller makes an estimate of the pdf of the channel mutual information before this information is available and as long as the antenna power profile remains the same, the pdf of the channel mutual information will also remain the same. If we want to improve the performance of the system, the transmitter should adapt the antenna power profile to the available partial channel state information and the influence of this adaptation must be calculated in the pdf of the channel mutual information. Namely, in such a scenario the antenna power profile in each block will depend on the partial channel state information and so will the pdf of the channel mutual information. Then, the controller should average the pdf of the channel mutual information over all the possible values of the partial channel state information according to its distribution and their associated antenna power profiles. This averaged pdf will depend on N_u , N_B and the SNR of the users, but in the general case it cannot be easily approximated by a Gaussian function as in (9). In order to obtain the pdf of the channel mutual information over l_u slots the controller should carry out $l_u - 1$ convolutions of the same averaged pdf of the channel mutual information. Then this pdf of the channel mutual information over l_u slots can be used in (11) and (12). The computation of the averaged pdf in the general case will be very involved and cannot be easily carried out by on-line methods.

4. Simulation

In order to evaluate the performance of our framework we used the JM 17.0 reference software [46] and made the necessary modifications to it. We used only luma encoding. The encoded video sequences were Foreman (user 1), Carphone (user 2), News (user 3), Stefan (user 4), Silent (user 5), and Mother and daughter (user 6) QCIF at a frame rate of 30 fps. The first video frame of each video sequence was encoded as I video frame and all the other frames were encoded as P video frames. We used the restricted intraencoding mode where the intraencoded macroblock is not predicted from inter encoded macroblocks. This allows avoiding the error propagation in the system when using intra encoding. We used the value of p_r equal to 1.0 [7] for the intercoded parts of the video frame and we changed its value from 0.8 [7] to 0.55 for the intracoded parts for better estimation precision. We used the values of p_d equal to 1.0 and 1.2 for the inter- and intracoded parts of the video frame, respectively [7]. The information about a frame been received or not is delayed for a single video frame; that is, when encoding

the current video frame the encoder has no information if the previous video frame has been correctly received. For all video frames before the previous one this information was available. The results presented in this section, if not stated differently, were obtained by averaging the distortion at the decoder, calculated for 100 different channel realizations. In all figures the performance comparison among different users is shown using the PSNR which was calculated after computing the average distortion, and the average overall performance is obtained by averaging the PSNRs of different users and is labeled as average PSNR. In all simulations the base station and the users were equipped with 4 antennas. The channel statistics were considered to be equal for all users. The channels were Rayleigh block fading channels and the duration of the block was considered to be equal to a single time slot. As explained in Section 3.3 the transmitter had only channel distribution information.

We compare our proposed framework for multiuser video transmission (in the figures labeled to as ‘‘Proposed’’) with two other frameworks. The first one is called EqRes framework since all the users utilize equal resources, that is, equal number of time slots. The video encoder uses the same principle as in the proposed framework. It is intended to show the gain from utilization of the multiuser diversity of the video content. This framework does not need to communicate its parameters to the central entity and can locally decide on the optimal parameters for the video encoding. This algorithm also gives the performance limit of the algorithms that use constant pairs of error probability and bit rate from [4] in a multiuser scenario. The second framework used for comparison is called PreERes. It uses preencoded video obtained with Variable Bit Rate (VBR) encoding of the video sequence using a fixed quantization parameter and allocates the available slots to the different users according to the procedure used in our proposed framework, that is, the resource allocation is optimized. In order to allow error resilience, during the encoding, 9 macroblocks are intraencoded in every video frame of every user. This represents the algorithms from [5, 6, 15–21], but instead of dividing the video frame in smaller slices, the whole frame is considered as a single slice. This slightly degrades the performance but is still an appropriate representative of this framework. It should be noted that this framework has the complexity of a basic video encoding algorithm which is lower compared to the proposed framework.

The first comparison of different frameworks is shown in Figure 2, where each time slot consists of 150 symbols (channel uses), $L = 9$ time slots are available for transmission of a single video frame, and 3 users are simultaneously transmitting video. The average SNR of each user is equal to 13 dB.

As shown in Figure 2 the proposed framework outperforms the models used for comparison. Performance gain compared to the EqRes framework is observed for all users and the average improvement is around 2 dB. This PSNR increase comes from utilization of the multiuser diversity due to the varying video content. The difference compared to the PreERes framework is more dramatic; that is, it is 10 dB on the average. This is due to inability of the VBR video encoding

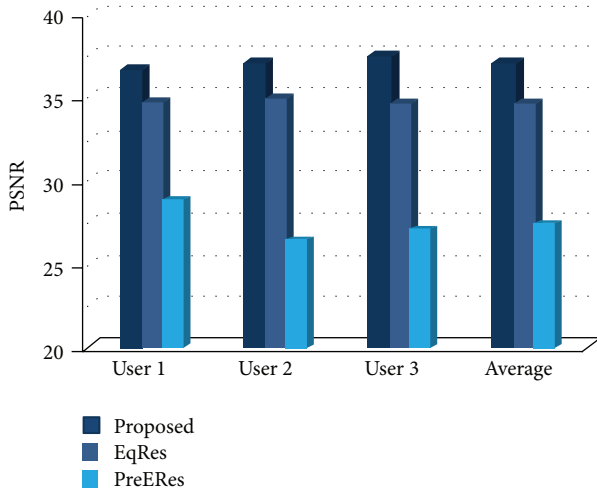


FIGURE 2: Comparison of the performance of different frameworks.

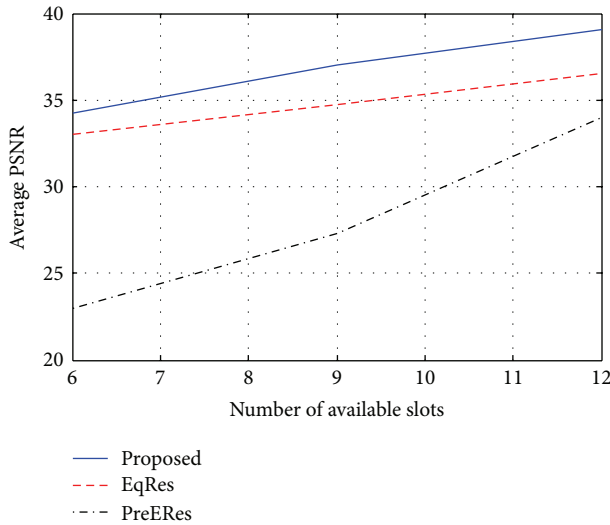


FIGURE 3: Comparison of the performance of different frameworks in terms of the number of available slots.

with fixed quantization parameter to adapt its rate to the channel. Also, in case of transmission errors, only the random intraencoded macroblocks are available for mitigating the error propagation, which is obviously not flexible enough.

In Figure 3 the comparison of the average performance of different frameworks is shown for the case when all 3 users experience average channel SNRs of 13 dB, the number of symbols in a slot is set to 150 and the number of available slots is varied. As the number of available slots is increased, the performance is improved for all frameworks. The performance gain for the proposed framework compared to EqRes remains approximately the same, and the performance gain compared to PreERes is always higher than 10 dB in the given region of L . Similar behavior can be observed in Figures 4 and 5. In Figure 4 the number of available slots, is set to $L = 9$ and the number of symbols in the slot is varied and in Figure 5 the SNR value is varied for fixed number of users,

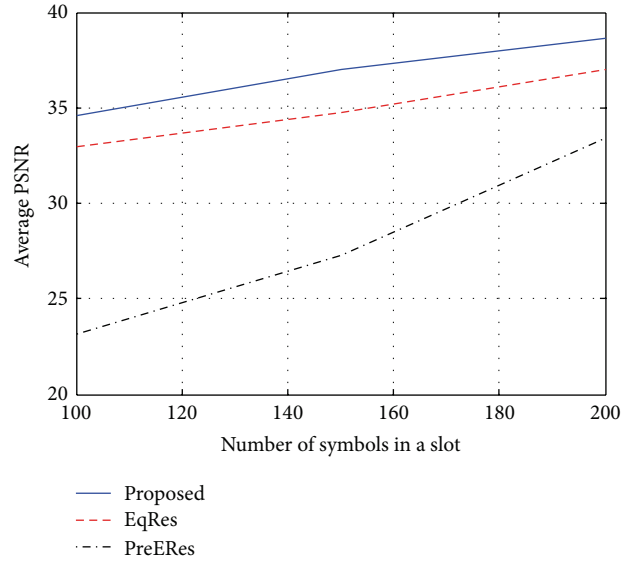


FIGURE 4: Comparison of the performance of different frameworks in terms of the number of symbols in a slot.

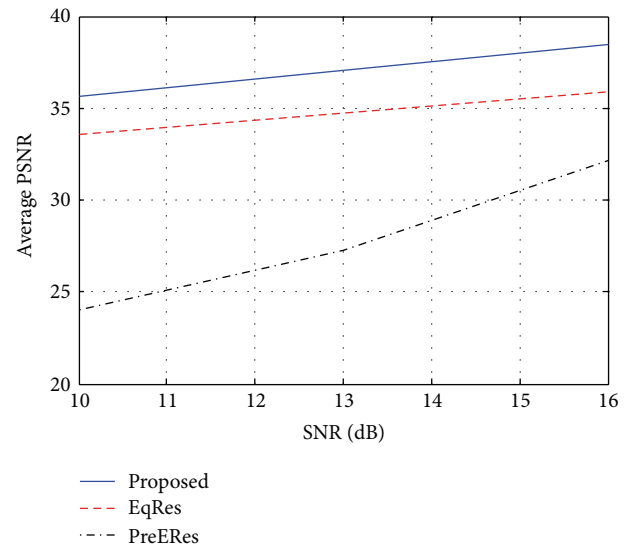


FIGURE 5: Comparison of the performance of the different frameworks in terms of SNR values.

available slots, and number of symbols in a slot. This shows the universality of the performance gain of the proposed framework compared to the other two frameworks.

In Figure 6 the performance of the three frameworks with different number of active users is shown. The number of available slots is set to $L = 12$, the average SNR is set to 13 dB, and the available number of symbols per slot is set to 150. As the number of user increases, the average PSNR in the system decreases, which is expected since all the users share the same resources. As expected, in the case when sufficient resources are available, the performance of PreERes is close to the maximum which is 35.88 dB for the chosen settings, but as

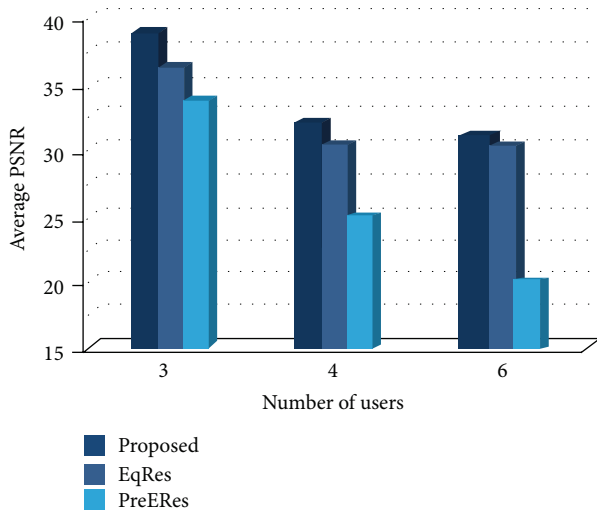


FIGURE 6: Comparison of the performance of different frameworks in terms of the number of users.

the resources become insufficient the performance difference is more obvious.

This behavior is due to the increased number of errors and the inability of the PreERes framework to adjust its source rate according to the available resources.

5. Conclusion

A new framework for multiuser cross-layer transmission and encoding of video signals using single layer H.264/AVC in a MIMO system is proposed. The cross-layer optimization is controlled by a central entity that receives information about the channel statistics and the video encoding parameters from every user. The framework thoroughly explains the process of the source coding adaptation, the MIMO channel description, and the optimization procedure. The framework is intended for real time uplink FDD TDMA MIMO system or downlink TDMA MIMO system with transcoding. The investigated wireless system utilizes capacity achieving channel codes. The performance of the proposed framework is superior to the performance of a system with equally allocated resources to all the users and a system with channel optimized allocation of the available resources, but using a preencoded video.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of the paper.

References

- [1] Y.-M. Hsiao, J.-F. Lee, J.-S. Chen, and Y.-S. Chu, "H.264 video transmissions over wireless networks: challenges and solutions," *Computer Communications*, vol. 34, no. 14, pp. 1661–1672, 2011.
- [2] M. van der Schaar and N. S. Shankar, "Cross-layer wireless multimedia transmission: challenges, principles, and new

- paradigms," *IEEE Wireless Communications*, vol. 12, no. 4, pp. 50–58, 2005.
- [3] M. M. Khan, "Cross-layer design: a survey," *International Journal of Computer Applications*, vol. 53, no. 8, pp. 44–49, 2010.
- [4] S. Pejovski and V. Kafedziski, "Cross-layer framework for real time H.264/AVC video transmission over wireless channels using outage probability," in *Proceedings of the 9th International Symposium on Telecommunications (BIHTEL '12)*, pp. 1–6, Sarajevo, Bosnia and Herzegovina, October 2012.
- [5] E. Maani, P. V. Pahalawatta, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Resource allocation for downlink multiuser video transmission over wireless lossy networks," *IEEE Transactions on Image Processing*, vol. 17, no. 9, pp. 1663–1671, 2008.
- [6] S. Pejovski and V. Kafedziski, "Improved multiuser wireless video streaming by considering influence on subsequent frame," in *Proceedings of the TELFOR*, pp. 942–945, November 2010.
- [7] D.-K. Kwon, M.-Y. Shen, and C.-C. Jay Kuo, "Rate control for H.264 video with enhanced rate and distortion models," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 5, pp. 517–529, 2007.
- [8] Y. Zhang, W. Gao, Y. Lu, Q. Huang, and D. Zhao, "Joint source-channel rate-distortion optimization for H.264 video coding over error-prone networks," *IEEE Transactions on Multimedia*, vol. 9, no. 3, pp. 445–454, 2007.
- [9] S. Pejovski and V. Kafedziski, "Video transmission on slowly fading channels using diversity," in *Proceedings of the TELFOR*, pp. 307–310, November 2009.
- [10] V. Kafedziski and S. Pejovski, "Optimization of video transmission on MIMO channels using outage probability," in *Proceedings of the 52nd International Symposium (ELMAR '10)*, pp. 273–276, Zadar, Croatia, September 2010.
- [11] S. Parakh and A. K. Jagannatham, "Game theory based dynamic bitrate adaptation for H.264 scalable video transmission in 4G wireless systems," in *Proceedings of the International Conference on Signal Processing and Communications (SPCOM '2012)*, pp. 1–5, Bangalore, India, July 2012.
- [12] Z. Han, G.-M. Su, A. Kwasinski, M. Wu, and K. J. R. Liu, "Multiuser distortion management of layered video over resource limited downlink multicode-CDMA," *IEEE Transactions on Wireless Communications*, vol. 5, no. 11, pp. 3056–3067, 2006.
- [13] H. Zhang, Y. Zheng, M. A. Khojastepour, and S. Rangarajan, "Cross-layer optimization for streaming scalable video over fading wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 3, pp. 344–353, 2010.
- [14] V. Joseph and G. de Veciana, "Joint optimizing multi-user rate adaptation for video transport over wireless systems: mean-fairness-variability tradeoffs," *International Journal of Computer Applications*, vol. 53, no. 8, pp. 44–49, 2012.
- [15] P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "Content-aware resource allocation and packet scheduling for video transmission over wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 4, pp. 749–759, 2007.
- [16] L.-U. Choi, W. Kellerer, and E. Steinbach, "On cross-layer design for streaming video delivery in multiuser wireless environments," *EURASIP Journal on Wireless Communications and Networking*, vol. 2006, no. 2, Article ID 060349, p. 55, 2006.
- [17] F. Fu and M. van der Schaar, "A systematic framework for dynamically optimizing multi-user wireless video transmission," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 3, pp. 308–320, 2010.

- [18] F. Li, G. Liu, and L. He, "Cross-layer approach to multiuser H.264 video transmission over wireless networks," *Journal of Multimedia*, vol. 5, no. 2, pp. 110–117, 2010.
- [19] S. Pejowski and V. Kafedziski, "Downlink video streaming for users nonequidistant from base station," *Radioengineering*, vol. 21, no. 2, pp. 772–780, 2012.
- [20] S. Karachontzitis, T. Dagiuklas, and L. Dounis, "Fair cross-layer scheme for heterogeneous H.264/AVC video streams over LTE-based broadband systems," in *Proceedings of the 8th International Wireless Communications & Mobile Computing Conference (IWCMC '12)*, pp. 1006–1010, Limassol, Cyprus, August 2012.
- [21] J. Huang, Z. Li, M. Chiang, and A. K. Katsaggelos, "Joint source adaptation and resource allocation for multi-user wireless video streaming," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 5, pp. 582–594, 2008.
- [22] D.-E. Meddour, A. Abdallah, T. Ahmed, and R. Boutaba, "A cross layer architecture for multicast and unicast video transmission in mobile broadband networks," *Journal of Network and Computer Applications*, vol. 35, no. 5, pp. 1377–1391, 2012.
- [23] L. Superiori, M. Wrulich, P. Svoboda et al., "Content-aware scheduling for video streaming over HSDPA networks," in *Proceedings of the 2nd International Workshop on Cross Layer Design (IWCLD '09)*, pp. 1–5, Palma de Mallorca, Spain, June 2009.
- [24] C. Bouras, N. Kanakis, V. Kokkinos, and A. Papazois, "Application layer forward error correction for multicast streaming over LTE networks," *International Journal of Communication Systems*, vol. 26, no. 11, pp. 1459–1474, 2012.
- [25] J. Xing, W. Fan, and Z. Lu, "A cross-layer scheduling scheme for video streaming based on fuzzy decision-making," in *Proceedings of the IEEE 13th International Conference on Communication Technology (ICCT '11)*, pp. 84–87, Jinan, China, September 2011.
- [26] L. Zhou, X. Wang, W. Tu, G.-M. Muntean, and B. Geller, "Distributed scheduling scheme for video streaming over multi-channel multi-radio multi-hop wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 3, pp. 409–419, 2010.
- [27] W. Saesue, C. T. Chou, and J. Zhang, "Cross-layer QoS-optimized EDCA adaptation for wireless video streaming," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 2925–2928, Hong Kong, September 2010.
- [28] R. Chen, Z. Mi, and L. Jiao, "Parameter-aware cross-layer scheme for wireless video," *Journal of Information Science and Engineering*, vol. 28, no. 2, pp. 379–391, 2012.
- [29] D. Yuan, Z. Guan, and H. Zhang, "Optimal and fair resource allocation for multiuser wireless multimedia transmissions," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 801613, pp. 1–12, 2009.
- [30] W. Ma, W. Zheng, H. Zhang, X. Wen, and Z. Lu, "Distortion-guaranteed energy saving power allocation for downlink OFDMA multiuser video transmission," *Journal of Information & Computational Science*, vol. 9, no. 11, pp. 3011–3019, 2012.
- [31] A. Chan, H. Lundgren, and T. Salonidis, "Video-aware rate adaptation for MIMO WLANs," in *Proceedings of the 19th IEEE International Conference on Network Protocols (ICNP '11)*, pp. 321–330, British Columbia, Canada, October 2011.
- [32] G. van der Auwera and M. Reisslein, "Implications of smoothing on statistical multiplexing of H.264/AVC and SVC video streams," *IEEE Transactions on Broadcasting*, vol. 55, no. 3, pp. 541–558, 2009.
- [33] J. Rexford, "Performance evaluation of smoothing algorithms for transmitting prerecorded variable-bit-rate video," *IEEE Transactions on Multimedia*, vol. 1, no. 3, pp. 302–312, 1999.
- [34] M. Hafeez, S. Jangsher, and S. A. Khayam, "A cross-layer architecture for motion-adaptive video transmission over MIMO channels," in *Proceedings of the IEEE International Conference on Communications (ICC '11)*, pp. 1–5, Kyoto, Japan, June 2011.
- [35] A. Khalek, C. Caramanis, and R. Heath, "Video-aware MIMO precoding with packet prioritization and unequal modulation," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO '12)*, pp. 1905–1909, Bucharest, Romania, August 2012.
- [36] Y. P. Fallah, H. Mansour, S. Khan, P. Nasiopoulos, and H. M. Alnuweiri, "A link adaptation scheme for efficient transmission of H.264 scalable video over multirate WLANs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 7, pp. 875–887, 2008.
- [37] M. Bansal, M. Jubran, and L. P. Kondi, "Scalable H.264 wireless video transmission over MIMO-OFDM channels," in *Advanced Concepts for Intelligent Vision Systems*, vol. 6475 of *Lecture Notes in Computer Science*, pp. 34–45, Springer, Berlin, Germany, 2010.
- [38] J. Park, U. Jang, T. Oh, S. Lee, and A. C. Bovik, "Optimal power allocation for minimizing visual distortion over MIMO communication systems," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '09)*, pp. 1833–1836, Cairo, Egypt, November 2009.
- [39] S. Mahajan and A. K. Jagannatham, "Hierarchical DWT based optimal diversity power allocation for video transmission in MIMO wireless systems with quantized feedback," in *Proceedings of the Australasian Telecommunication Networks and Applications Conference (ATNAC '11)*, pp. 1–7, Victoria, Australia, November 2011.
- [40] S. Zhao, Y. Zhang, and L. Gui, "Optimal resource allocation for video delivery over MIMO OFDM wireless systems," in *Proceedings of the 53rd IEEE Global Communications Conference (GLOBECOM '10)*, pp. 1–5, Miami, Fla, USA, December 2010.
- [41] O. H. Salim and W. Xiang, "A novel unequal error protection scheme for 3-D video transmission over cooperative MIMO-OFDM systems," *EURASIP Journal of Wireless Communication and Networking*, vol. 2012, article 269, 2012.
- [42] L. Liu and X. Zhuang, "Fast motion estimation and mode decision for H.264 video coding in packet loss environment," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '08)*, pp. 781–784, Hanover, Germany, April 2008.
- [43] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*, Cambridge University Press, New York, NY, USA, 2004.
- [44] Z. Wang and G. B. Giannakis, "Outage mutual information of space-time MIMO channels," *IEEE Transactions on Information Theory*, vol. 50, no. 4, pp. 657–662, 2004.
- [45] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.
- [46] H.264/AVC reference software, <http://iphome.hhi.de/suehring/tml/>.

