Error-Resilient Video Encoding and Transmission in Multirate Wireless LANs

Antonios Argyriou, *Member, IEEE*

***Abstract*—In this paper, we present a cross-layer approach for video transmission in wireless LANs that employs joint source and application-layer channel coding, together with rate adaptation at the wireless physical layer (PHY). While the purpose of adopting PHY rate adaptation in modern wireless LANs like the IEEE 802.11a/b is to maximize the throughput, in this paper we exploit this feature to increase the robustness of wireless video. More specifically, we investigate the impact of adapting the PHY transmission rate, thus changing the throughput and packet loss channel characteristics, on the rate-distortion performance of a transmitted video sequence. To evaluate the video quality at the decoder, we develop a cross-layer modeling framework that considers jointly the effect of application-layer joint source-channel coding (JSCC), error concealment, and the PHY transmission rate. The resulting models are used by an optimization algorithm that calculates the optimal JSCC allocation for each video frame, and PHY transmission rate for each outgoing transport packet. The comprehensive simulation results obtained with the H.264/AVC codec demonstrate considerable increase in the PSNR of the decoded video when compared with a system that employs separately JSCC and PHY rate adaptation. Furthermore, our performance analysis indicates that the optimal PHY transmission rate calculated by the proposed algorithm, can be significantly different when compared with rate adaptation algorithms that target throughput improvement.**

***Index Terms*—Cross-layer design, joint source-channel coding, rate-adaptive PHY, real-time video encoding, video streaming, wireless LAN.**

I. Introduction

THE need for high-quality video services in the existing and next generation wireless networks dictates the design of video streaming algorithms that are flexible and highly adaptive to the rapid fluctuations of the wireless quality of service (QoS). Most of the existing video streaming applications rely on the support of wireless access systems that are designed to provide QoS support (e.g., 802.11e) [1]. However, the QoS parameters that can be enforced at the wireless medium access control/physical layer (MAC/PHY) are not comprehensive enough for capturing the highly complex dependencies that exist in encoded video. Furthermore, the classical layered protocol stack exacerbates this problem since it does not allow comprehensive information to be communicated across the involved protocols. It has been demonstrated that for the best performance, the functionality of protocols that belong to several layers should be jointly designed [2].

Manuscript received July 31, 2007; revised November 5, 2007. Published July 9, 2008 (projected). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Giacomo Morabito.

The author is with Philips Research, Eindhoven, 5656 AE, The Netherlands (e-mail: anargyr@ieee.org).

Digital Object Identifier 10.1109/TMM.2008.922776

In this paper, we focus on exploiting cross-layer interactions for improving the robustness of real-time encoded and transmitted video in wireless LANs. One way to improve robustness at the application layer is to allow the encoder to regulate the degree of the error resiliency that is added in the bitstream by appropriately selecting the source coding parameters [3], [4]. Another way to handle errors is either by adding redundancy with forward error correction (FEC) or by employing automatic repeat request (ARQ) [5], [6]. These mechanisms are also referred to as channel coding. When the system employs both error-resilient source coding and channel coding in a unified optimization setting, we come across a new class of techniques called the joint source-channel coding (JSCC). The objective of JSCC is to allocate the available channel rate between the source coder and the channel coder, so that video quality is maximized [7], [8]. For the JSCC to be optimal, it is imperative that the sender has an accurate estimate of the channel characteristics, like the available bandwidth, the round-trip time (RTT), and the packet loss rate. This information is usually transmitted through a feedback channel from the receiver. The more accurate the information is, the more efficient the resource allocation with JSCC will be.

In this paper, we extend the principle of JSCC for wireless video transmission by considering the channel data rate as another optimization parameter that can be adjusted. This is achieved by utilizing the multirate adaptation mechanism available at the 802.11a WLAN PHY. We claim that by employing JSCC at the application layer independently from the WLAN PHY rate adaptation algorithm, the optimal performance cannot be achieved for real-time video encoding and transmission. The reason is the conflicting objectives of JSCC and PHY rate adaptation that aim to maximize the rate-distortion (RD) performance of a transmitted video sequence and the effective throughput, respectively. To illustrate this concept more clearly, we present in Fig. 1 the convex hull of all feasible (rate, distortion) points for a video sequence transmitted over two different wireless channel realizations after JSCC has been applied. Channel 2 is not only characterized by higher available channel rate, but also suffers from higher packet error rate (PER), when compared to channel 1. These two channel realizations correspond actually to two different PHY transmission rates, but for the same channel signal-to-noise ratio (SNR) conditions. As expected, the optimal source-channel rate allocation is different in these two cases. The video bitstream transmitted in channel 2 can sustain a higher PER since it is encoded for increased error resiliency owing to the highest available bandwidth. This leads to a lower decoder distortion. Therefore, the question that has to be answered by the proposed cross-layer video transmission system is how to select the optimal PHY transmission rate and source-channel rate allocation that can lead the encoder

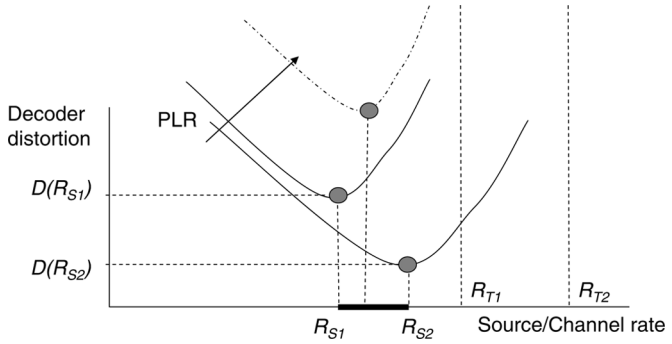

Fig. 1. Convex hull of the feasible RD points for encoded video transmission in a lossy channel.

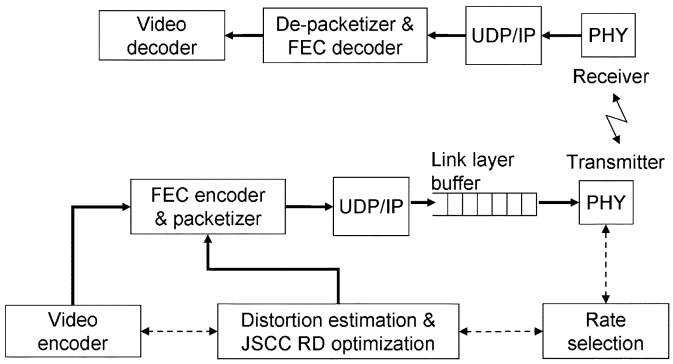

Fig. 2. Proposed cross-layer architecture for real-time video encoding and transmission in wireless LANs with adaptive PHY. Solid lines indicate data flow while dashed lines indicate cross-layer signaling.

to the operating point of minimum distortion, as shown in Fig. 1. Our end result is cross-layer optimization algorithm that dynamically selects the encoding mode of video macroblocks (MBs), application-layer FEC, and the PHY transmission rate to minimize the decoder distortion.

A. Paper Organization

This paper is organized as follows. Related works in the area of cross-layer wireless video transmission and PHY rate adaptation are discussed in Section II. An overview of the main operating principles and functional components of our system are presented in Section III. In Section IV, we present one of the most crucial aspects of the system, that is the mathematical modeling of the 802.11a adaptive PHY. A detailed description of the video distortion model is given in Section V. In Section VI, we initially describe the RD optimization framework that incorporates all the previous models, and subsequently we propose an optimization algorithm. Simulation results are presented in Section VII. Finally, Section VIII concludes this paper.

II. Related Works

JSCC in the context of video transmission has been thoroughly studied in the literature [7]–[10]. In one of the earlier works that considered JSCC for image transmission, source coding was exercised by adjusting the step of the quantizer, whereas the entropy coder was adjusted according to the channel PER [7]. More recent designs were based on the use of source coding, together with application-layer FEC for channel coding [9], [10]. For wireless networks, JSCC has been thoroughly studied mostly for scalable pre-encoded video since source rate adaptation can be achieved by adding or dropping the layers of the bitstream [10]. One of the most recent works considers JSCC at the application layer by employing a real-time video encoder and a hybrid ARQ/FEC scheme for channel coding [8]. In this work, the authors design an adaptive error control system where it employs either ARQ or FEC, depending on the channel conditions. Even though JSCC is a well-studied topic, the majority of PHY rate adaptation schemes that have been proposed in the literature target elastic data transfers [11], [12]. For example, the work presented in [11] described a method for driving PHY rate adaptation in an 802.11a WLAN by using an analytical closed-form model of the achieved effective throughput. In a more recent work [12], the authors present an analytical study on the effects of the packet size and PHY rate selection on the throughput of a wireless system with adaptive PHY. The benefit of selecting a different PHY for wireless video transmission has also been demonstrated [2]. In that work, the authors developed an analytical performance model that can be used for PHY adaptation when scalable encoded video is employed at the sender. However, all these works lack a systematic methodology for determining the optimal PHY rate for a system where real-time encoding is employed.

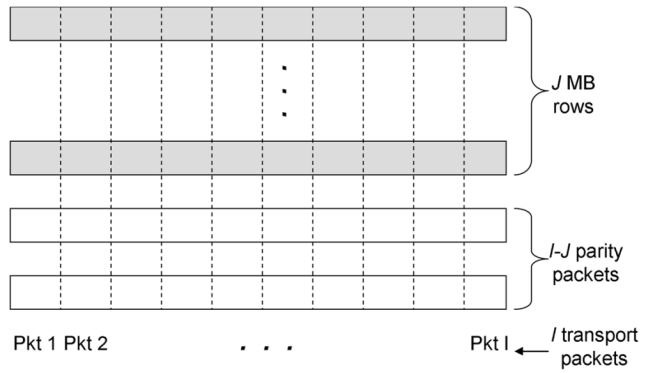

Fig. 3. Packetization with FEC across J source packets (MB rows). Each of the I transport packets can be encoded at a different PHY transmission rate r.

III. System Overview

Fig. 2 depicts the main components of the proposed video streaming system engineered to work with an adaptive WLAN PHY. The system consists of a real-time video encoder that also estimates the decoder distortion. Based on this estimate, the JSCC algorithm calculates the optimal source-channel rate allocation for each frame. Subsequently, it encodes each MB with the selected optimal encoding mode and generates the source video packets. Source packets are then sent to the application-layer FEC Reed Solomon (RS) encoder (Fig. 3). The RS encoder generates $I - J$ additional packets for J input source packets, based on the remaining rate from the source-channel allocation. Therefore, the overhead that is added with an $\mathrm{RS}(I, J)$ code is $I - J$ packets. FEC is applied across the source packets so that

each generated transport packet contains parts of both the source payload and the parity bits. This packetization strategy is crucial for efficient wireless LAN transmission, since most of the 802.11 software implementations discard link-layer frames that have even single-bit errors that cannot be corrected. In such a case, applying FEC along each source packet would mean that in the case of a link-layer frame loss, the complete source packet would be lost, and RS decoding would thus fail.

Subsequently, the I transport packets are packetized with the real-time transport protocol (RTP), and are placed in the application-layer transmission buffer. A UDP/IP header can be added for Internet transmission. This packet is finally sent to the 802.11 MAC layer, which appends a header (of 28 bytes), and creates a MAC protocol data unit (MPDU) for wireless transmission. Link-layer retransmissions are not used in this work, since we wanted to simplify the derived analytical models. Nevertheless, the modular modeling approach could be easily extended to take into account the link-layer retransmissions based on the existing works [6]. Subsequently, the actual PHY layer frame is created by encoding the MPDU with one of the eight available 802.11 transmission rates.

Once the packets are transmitted at the physical layer, they pass through the application-layer RS decoder at the receiver. The probability of RS decoding failure at the receiver is given by

$$\rho = 1 - \sum_{l=0}^{I-J} \binom{I}{l} p_w^l (1-p_w)^{I-l} \tag{1}$$

where p_w is the wireless PER before FEC error recovery. If RS decoding fails, the MBs that correspond to the partially received frames are not discarded but are placed in the decoder buffer (i.e., soft channel decoding). The decoder starts to decode video packets and display them, after an initial *startup delay* that is configured by the application. In case of packet loss, the decoder uses a simple temporal error concealment technique.

IV. Performance of 802.11a PHY With Rate Adaptation

Our system is based on adapting parameters both at the application layer (source-channel coding), and at the WLAN PHY layer. Therefore, it is crucial to express the key performance metrics analytically, so that cross-layer adaptation can be efficient. In this section, we focus on the 802.11a WLAN PHY, and more specifically the effects of rate adaptation on the PER and the throughput.

The IEEE 802.11 specification defines two functional layers [13]. The first one is the 802.11a PHY that integrates the modulation, demodulation, encoding, decoding, analog-to-digital conversion (ADC), digital-to-analog conversion (DAC), and filtering. These functions are always entirely implemented in hardware. The second one is the 802.11 MAC layer that is responsible for granting access to the wireless medium when many wireless stations are present. This is implemented by a combination of hardware and software. The precise split between these two layers depends on the particular device. Now, the 802.11a PHY is based on orthogonal frequency division multiplexing (OFDM) and provides eight different PHY transmission rates[1] with different modulation schemes and convolutional codes at the 5-GHz U-NII band that offer data transmission rates ranging from 6 to 54 Mb/s (Table I). For each transmission rate r, we denote the theoretical raw data rate as R_r. The lower-rate PHY rates are inherently more robust than the higher-rate ones, since they use modulation schemes with smaller constellation size. One of the most popular mechanisms for rate adaptation in practical 802.11 WLAN interfaces is through a mechanism called automatic rate fallback (ARF) [14]. This algorithm selects the transmission rate by counting the number of successfully transmitted or lost PHY frames, and is vendor-specific. Nevertheless, there are several research works that propose more intelligent adaptation schemes based on analytical performance models [11], [15]. This means that the derived model must be very accurate so that the cross-layer decisions are actually optimal.

TABLE I
THE 802.11 TRANSMISSION RATES

mode	Modulation	Code rate	Data rate (Mbps)
1	BPSK	1/2	6
2	BPSK	3/4	9
3	QPSK	1/2	12
4	QPSK	3/4	18
5	16-QAM	1/2	24
6	16-QAM	3/4	36
7	64-QAM	2/3	48
8	64-QAM	3/4	54

In the following subsections, we will calculate the average PER p_w and the effective throughput T as a function of the 802.11a PHY transmission rate r for a Rayleigh fading channel. We will also consider a multipath frequency-selective fading channel [16].

A. PER for a Rayleigh Channel

The objective of adopting the Rayleigh fading wireless channel model is to explore the use of analytical formulas for fast real-time calculations. Our assumptions in this case include a frequency-flat fading wireless link that remains invariant per PHY frame, but may vary from frame to frame. For slow-varying flat fading channels, the channel quality can be captured by the average received SNR γ of the wireless link. Since the channel varies from frame to frame, the Nakagami-η fading model is adopted for describing γ [16]. This means that the received SNR per frame is a random variable, where we assume $\eta = 1$ for Rayleigh fading. SNR estimation can be done from the received signal strength indication (RSSI) at the receiver as in [16]. Furthermore, we also assume that the noise over the wireless spectrum is additive white Gaussian noise (AWGN). Now if $Q(x)$ is the Gaussian Q function, then for BPSK modulation in a Rayleigh channel, the instantaneous bit error rate (BER) is given by [17]

$$\epsilon = Q(\sqrt{2\gamma}) \tag{2}$$

[1]Even though the appropriate term is transmission mode, we do not use it to avoid confusion with the encoding mode of video MBs.

The bit error probability for a Rayleigh fading channel and for M-ary QAM is given by [18]

$$\epsilon = 2\left(\frac{\sqrt{M}-1}{\sqrt{M}}\right)\left(\frac{1}{\log_2 M}\right)Q\left(\sqrt{\frac{3\gamma}{M-1}}\right) \quad (3)$$

The above two equations provide the BER for all the 802.11a modulation schemes, including QPSK, which is equivalent to 4-QAM. For calculating the PER, the effect of convolutional coding must considered. For binary convolutional coding and hard decision Viterbi decoding, an upper bound on the packet error probability was derived in [19]. This bound for an L-bit packet is given as

$$p_w(L,r) \leq 1-(1-p_u(r))^{8L} \quad (4)$$

where $p_u(r)$ is the union bound of the first-event error probability corresponding to PHY transmission rate r. This is equal to

$$p_u(r) = \sum_{d=d_{\text{free}}}^{\infty} \alpha_d P_d. \quad (5)$$

Now given that the bit error probability for the transmission rate r is ϵ, then for hard decision Viterbi decoding, P_d is given by [17]

$$P_d = \begin{cases} \sum_{i=(d+1)/2}^{d} \binom{d}{i} \epsilon^i (1-\epsilon)^{d-i} \text{ if } d \text{ is odd} \\ \frac{1}{2}\binom{d}{d/2}\epsilon^{d/2}(1-\epsilon)^{d/2} \\ +\sum_{i=d/2+1}^{d}\binom{d}{i}\epsilon^i(1-\epsilon)^{d-i} \text{ if } d \text{ is even.} \end{cases} \quad (6)$$

By using the above three expressions, we can calculate the PER p_w very accurately.

B. Throughput for a Rayleigh Channel

For calculating the effective throughput at the sender, we assume that no other station is transmitting and therefore there is no packet loss due to MAC-layer collisions. Considering MAC-layer contention and its impact on throughput is out of the scope of this paper. We also assume that the transmission of acknowledgments on the reverse path is considered error-free, which is something that can be easily achieved by applying strong error correcting codes. Therefore, if the video payload consists of S_d bytes, and the combined protocol overheads is S_{hdr} bytes, then the *effective throughput* for PHY rate r is given by

$$T(r,\gamma) = \frac{S_d * R_r}{S_d + S_{hdr} + S_{dcf}(r)} * (1 - p_w(L,r)) \quad (7)$$

where S_{dcf} is the overhead for accessing the channel with the 802.11 DCF mechanism even when no other stations contend for the wireless medium. This value is calculated by considering the 802.11 PHY overhead between the transmission of two successive PHY frames [11], [13]. The above equation is not directly used by any algorithm in the proposed system. However, the system that we used for comparison during our simulations selects the optimal PHY data rate based on this formula. We will later explain why this is a valid comparison. In our system, we are more concerned with the raw application-layer data rate that can be achieved. To obtain this quantity, packet losses in (7) have to be ignored, which makes the maximum possible application data rate

$$T_{\max}(r) = \frac{S_d * R_r}{S_d + S_{hdr} + S_{dcf}(r)}. \quad (8)$$

Actually, $T_{\max}$ for a particular PHY rate r corresponds to the effective throughput for $\gamma \to \infty$, where packet loss is negligible. Owing to the nature of the Rayleigh channel model, there is always a PHY mode for which the throughput is increased sharply within a given SNR range. At the same time, for a specific transmission rate the PER is decreased as channel SNR is increased.

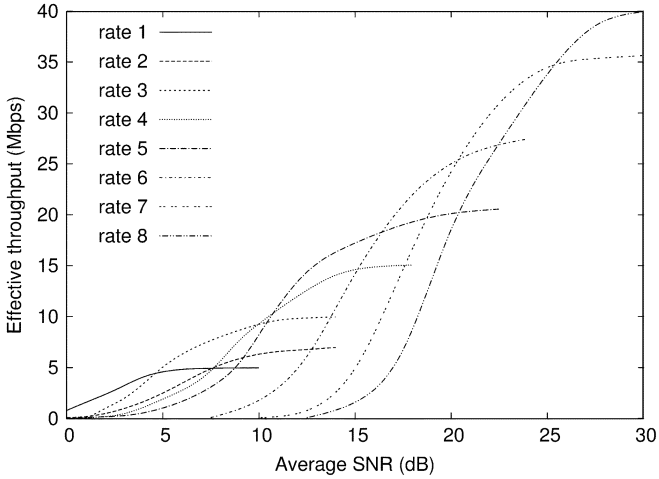

Fig. 4. Effective throughput versus channel SNR for all the different transmission rates of 802.11a and a multipath frequency-selective fading channel. Application payload is 2000 bytes.

C. Multipath Fading Channel

Previous works that dealt with the problem of optimal rate adaptation for increasing the effective throughput have revealed significant discrepancies between the calculated rate adaptation thresholds for Rayleigh and multipath frequency-selective fading channels [12]. Therefore, we decided to use a multipath frequency-selective fading channel for further investigation. The model we selected for evaluating our system is widely used for indoor wireless environments [20]. With this model, the channel impulse response is captured through a tapped delay line, in which the distribution of the path amplitude is characterized by a Rayleigh fading path-delay profile. Also, the average power of the different taps declines exponentially as delay increases. Using the above channel model, we simulated several realizations of the wireless channel with the help of an existing simulator [21]. Subsequently, we obtained the performance curves of BER versus SNR for different PHY transmission rates of 802.11a. Therefore, the PER can be easily calculated as

$$p_w(L,r) = 1-(1-\epsilon(r,\gamma))^{8L}. \quad (9)$$

Since (7) is not related to the channel model, it can also be used for calculating the throughput with the PER results that correspond to the different channel models. Throughput results from our simulations can be seen in Fig. 4 for payloads of 2000 bytes.

We can see that six of the eight transmission rates are actually the optimal choice for a specific SNR range. Furthermore, the differences in throughput are reduced between "neighboring" PHY transmission rates for a particular SNR range. Therefore, the choice of the optimal transmission rate is not straightforward as in the case of the Rayleigh channel, especially if another application-specific metric has to be optimized besides throughput. We will later analyze in detail how the proposed JSCC mechanism can exploit this behavior for increasing the robustness of wireless video transmission.

V. Distortion Estimation at the Encoder

Up to this point, we have expressed the system throughput and PER as a function of the selected PHY transmission rate. The next step is to correlate the derived results for the wireless channel throughput and residual PER with the video distortion at the decoder.

Many research works have dealt with distortion estimation for hybrid motion-compensated video coding and transmission over lossy channels [3], [22], [23]. In this type of video coder, each video frame is represented in block-shaped units of the associated luminance and chrominance samples (16×16 pixel region) called macroblocks (MBs). In H.264, MBs can be either intra- or inter-coded from the samples of previous frames [24]. Intra-coding is performed in the spatial domain by referring to neighboring samples of previously coded blocks, which are on the left or above the block to be predicted. Nine possible modes exist that use subblocks of a 4×4 size, whereas a 16×16 mode is possible for coding smoother areas in the frame. Inter-coding is performed with temporal prediction from samples of previous frames: MBs are first partitioned into 16×16, 16×8, 8×16, or 8×8 blocks. Further partitioning is possible by defining subblocks of 8×4, 4×8, and 4×4.

It is understandable that many options exist for coding a single MB, where each of them provides different RD characteristics. Therefore, we adopt a model that estimates the decoder distortion as a function of the previously defined MB encoding modes. The reason is that we want to apply error-resilient source coding by selecting the encoding mode of each particular MB. This requirement is crucial since it allows the encoder to tradeoff bit rate with error resiliency at the MB level. We also decided to design our encoder so that distortion is calculated recursively across frames [4], [25]. This practically means that the estimate of the expected distortion for a frame currently being encoded is derived by considering the total distortion introduced in the previous frames.

Now before we start the derivation of the closed-form recursive model, we will define the necessary notation. Let $f_i(n)$ denote the value of pixel i in frame n, $\hat{f}_i(n)$ the reconstructed pixel at the encoder, and $\tilde{f}_i(n)$ the encoder's estimate of the reconstructed pixel at the decoder. If the distortion of a single pixel $f_i(n)$ is denoted as $d(i,n)$, then according to the previous definition it can be written as

$$d(i,n) = [f_i(n) - \tilde{f}_i(n)]^2 \tag{10}$$

But since we follow a frame-level recursive approach, the overall expected distortion can also be written as the sum of the three contributing components, namely source, error propagation, and channel distortion [4], [25]. Therefore, the overall distortion can now be expressed as

$$d(i,n) = (1-\rho)d_s(i,n) + (1-\rho)d_{\rm ep}(i,n) + \rho d_{\rm ec}(i,n). \tag{11}$$

Recall that the residual end-to-end PER ρ after FEC is given by (1) and can be readily used in the above formula. When the distortion of particular MBs is calculated according to (11), it is stored in a structure called distortion map. The reason is that when calculating recursively the distortion for subsequent frames as given in (11), the appropriate distortion contribution of specific subblocks in previous frames must be identified and added to the overall expected distortion. Since H.264 supports the encoding modes even for 4×4 pixel areas, a 4×4 subblock is the minimum-sized element required for storing the calculated distortion components of the previous frames. Furthermore, if multi-frame prediction is activated, several subblock distortion components have to be averaged to get the distortion for the current block and frame.

According to our previous analysis, (l0) can now be written for a 4×4 subblock b

$$d(b,n) = (1-\rho)[d_s(b,n) + d_{\rm ep}(b,n)] + \rho d_{\rm ec}(b,n). \tag{12}$$

Next, we will derive a recursive analytical formula for each of the components in (l2).

The source distortion is attributed to the use of different H.264 encoding modes. It is easily calculated at the encoder as

$$d_s(b,n) = \frac{1}{16}\sum_{i=1}^{16}[f_i - \hat{f}_i]^2 \tag{13}$$

for a single subblock in accordance with the previously defined notation. Note that since this component is different for each encoding mode (including the used reference frame), it has to be frequently calculated by the MB encoding mode selection algorithm. Even though it is computationally expensive, the calculation of this distortion component is straightforward, unlike the other two.

The second component in (l2) is the error concealment distortion. We consider a simple error concealment mechanism at the decoder according to which the concealment motion vector (MV) for a lost MB is copied from the MB that resides in the same spatial location in the previous frame. If the MB from the previous frame is also lost, then the concealment MV is set to zero. In accordance with our recursive notation, the error concealment distortion can be expressed as the sum of other components [25]

$$d_{\rm ec}(b,n) = d_{\rm ec\_e}(b,n) + d_{\rm ep}(b',n-1). \tag{14}$$

The first component in the previous formula ($d_{\rm ec\_e}$), is the error concealment distortion at the encoder. This is the distortion that will be introduced by the event that a subblock b' from frame $n-1$ will be copied in the spatial location of the lost subblock b in the current frame. The selection of this subblock is determined

by the previously described error concealment approach. The distortion for this operation and pixel i can be calculated as

$$d_{\mathrm{ec\_}e}(i,n) = [f_i(n) - \hat{f}_i(n-1)]^2. \quad (15)$$

On the other hand, $d_{\mathrm{ep}}(b', n-1)$ is the error propagation distortion that exists in subblock b' on the frame that will be used for error concealment (i.e., $n-1$). This last component is also recursively calculated and is described in the next sections.

The distortion due to error propagation depends primarily on the selected reference frame that was used at the encoder for motion estimation. For achieving higher compression efficiency, H.264 allows an arbitrary reference frame to be selected, instead of just the previous frame [24]. Therefore, for the distortion component caused by error propagation from the reference frame ref, we can write the following expression:

$$d_{\mathrm{ep}}(b,n) = (1-\rho)d_{\mathrm{ep}}(b\prime, \mathrm{ref}) + \rho d_{\mathrm{ep}}(b'', n-1). \quad (16)$$

Essentially, the above formula captures the gradual propagation of errors from frame to frame. The first term is the error propagated distortion of the subblock b' pointed by the MVs in the reference frame. In case of correct transmission of b, this subblock will be the only component that has to be included. However, if the subblock b' that belongs to the reference frame is lost, then d_{ep} must include the distortion that will be propagated from the subblock b'', which will be used for error concealment (i.e., $d_{\mathrm{ep}}(b'', n-1)$).

Finally, we have to write the expression for the total expected distortion that will be introduced if a complete network transport packet is lost. This is necessary, since both the FEC mechanism and the PHY transmission rate are selected for each transport packet, which consists of several MBs. Therefore, for the expected distortion of the transport packet k we will have

$$D_k(n) = \sum_{b=1}^{B} d(b,n) \quad (17)$$

where B is the total number of subblocks contained in transport packet k. With this last expression, we have derived all the necessary analytical recursive formulas that will be needed for estimating at the encoder the expected decoder distortion. The only parameter that these formulas require is the residual PER $\rho(p_w)$. This parameter is updated after each frame is encoded, since p_w is also continuously updated.

VI. RD Framework for JSCC and PHY Rate Selection

In this section, we present the JSCC rate selection (JSCCRS) RD framework for jointly optimizing the selection of the source encoding mode of individual MBs, channel coding with application-layer FEC, and PHY transmission rate. The JSCC mechanism is responsible for optimally allocating the available channel rate between the source and FEC to minimize the video distortion at the decoder for the given throughput and PER channel conditions. However, the proposed JSCCRS scheme strives to identify a PHY transmission rate that is optimal for different JSCC allocations. This is the problem that must be solved, which is essentially a JSCC allocation problem but with an additional parameter, that is the actual PHY transmission rate.

A. Problem Formulation

The optimization problem that we described can be formally expressed as follows. Let $\mathbf{S}$ be the set of all the available source coding options for an H.264 MB. Let also $\mathbf{C}$ be the set of the available channel coding (FEC) parameters. The vector of the candidate source and channel coding parameters for the specific frame n is denoted as $\mu(n)$ and $c(n)$. Finally, let $\mathbf{M}$ be the set of 802.11a PHY transmission rates. The objective is to minimize the decoder distortion

$$\min_{\mu\in\mathbf{S}, c\in\mathbf{C}, r\in\mathbf{M}} E[d_{e2e}(\mu, c, r)] \quad (18)$$

such that the following rate constraint is satisfied:

$$R(k,n) \le T_{\max}(r_k). \quad (19)$$

The above constraint means that the transmitted transport packet k should not exceed the available maximum raw application data rate under transmission rate r. Our approach is to solve the above problem by considering a set of PHY transmission rates that correspond to the throughput $T(r)$, instead of using a fixed channel rate constraint like the existing RD optimization techniques.

By using Lagrangian relaxation, we convert the previous constrained optimization problem in an unconstrained one. The total Lagrangian cost for frame n can be expressed as

$$J_n = d_{e2e}(n) + \lambda_n \sum_{k=1}^{K} (R_k(n) - T_{\max}(r_k)) \quad (20)$$

where λ is the Lagrange multiplier. Since the contribution of each MB in the overall cost has been shown to be additive, the Lagrangian cost is minimized for each MB individually [3]. The selection of the Lagrange multiplier is critical for the RD performance of the transmitted video signal. It is well known that when λ varies from zero to infinity, the solution to (18) traces the convex hull of the operational RD function (see Fig. 1). It is evident from the nature of the operational RD function, that bisection search is the most suitable way of calculating the Lagrange multiplier for each frame. With this simple algorithm we observed very fast convergence to the global minimum.

B. Solution Algorithm

The first important feature of our algorithm is that the optimal transmission rate r^* does not change across frames, unless there is a change in the channel SNR γ since this might translate to different system throughput and PER[2]. By using this idea, considerable reduction in the processing requirements can be achieved at the encoder. In case there was a change in SNR, the sender calculates the transmission rate $r_{\max}$ that provides the highest throughput under the new channel SNR conditions. This step is used so that the algorithm can reduce the search space for possible optimal transmission rates in the neighborhood of $r_{\max}$. For each possible FEC protection code and transmission rate r, the encoder estimates the residual PER ρ and the throughput. Subsequently, the encoder calculates for each possible choice of H.264 encoding mode μ, the number of bits needed for the

[2]It is more critical to fine-tune this algorithm when a Rayleigh channel is assumed since a small variation in the channel SNR can result in significant variations in both throughput and PER.

current macroblock and also the associated distortion. Therefore, the encoding modes of each MB are calculated exhaustively. After the cost is estimated for a single MB, the algorithm proceeds to the same calculations for the complete frame. Exhaustive search can be applied because of the restricted number of both FEC coding options and transmission rates.

At the end of one execution run of this algorithm, the optimal values for $\mu^*(n)$, $c^*(n)$, and $r^*(n)$ for each MB of frame n are derived. Two additional constraints require that the FEC code c_m and the PHY transmission rate r_m for each MB m of the transport packet k are the same since these MBs correspond to one transport packet. Now after the current frame is encoded, the corresponding distortion map of the current frame is constructed according to (16) for the encoding of subsequent frames. Therefore, for the next frame, the encoder has the distortion estimates of the previous frame readily available.

C. Complexity Considerations

From the previous analysis, it is evident that many distortion components need to be calculated for different candidate encoding modes. However, the recursive frame-level distortion estimation algorithm shares the same complexity with related works on distortion estimation [3], [4]. More specifically, according to (l2), the source distortion component d_s must be calculated for each possible MB encoding mode represented by the set $\mathbf{S}_{\rm enc}$, and reference frame option represented by $\mathbf{S}_{\rm ref}$. The error propagation distortion $d_{\rm ep}$ depends on the reference frame selected and hence it is calculated only when the reference frame in the encoding mode changes. Regarding the error concealment component $d_{\rm ec}$, it is calculated once during the calculation of d_{e2e} for a single MB, since irrespectively of the selected encoding mode the option for error concealment is only one. Finally, the candidate MB encoding modes that must be tested are increased by a factor of two only when there is a significant change in channel SNR, since they have to be calculated for each of the two possible PHY transmission rates. Based on this analysis, the worst-case storage requirements for the array that holds the MB cost is $O(|\mathbf{S}_{\rm enc}| \times |\mathbf{S}_{\rm ref}| \times |\mathbf{C}| \times 2)$. However, the actual cost is even lower because not all the MB encoding options can be used for a particular FEC code available in $\mathbf{C}$, as the bit-rate requirements exceed the allowed maximum data rate $T_{\max}$.

VII. Performance Analysis and Simulation Results

In this section, we present the simulation results that highlight the performance of the proposed cross-layer video encoding and transmission system for IEEE 802.11a WLANs. For our experiments, we used the JVT reference software implementation of H.264 [26], although a few software performance optimizations have been implemented. The luminance component of the QCIF sequences Akiyo, Coastguard, Container, Foreman, and Silent were used for real-time encoding. We used only I and P frames and we set the GOP size to 128 frames. The QP was set equal to 14 to get constant video quality, at the cost of VBR. However, the average bit-rate throughout the sequence was 512 kbps. Furthermore, the maximum source video packet size was set to 1200 bytes. With this frame size, the encoded P frames are packetized into three source packets. Rate control is not implemented in the video streaming systems under test. This last assumption practically means that every frame has the same transmission delay over the network, which is set equal to the inter-frame playback delay $(1/\text{fps})$. FEC parity blocks are generated for each frame individually (see Fig. 3), while three columns of MBs correspond to one transport packet, and therefore a QCIF frame corresponds to three packets. For application-layer FEC, we chose the following set of possible RS codes $\mathbf{C} = \{(9,9), \ldots, (18,9)\}$. The channel was simulated according to the parameters we analyzed in Section IV.

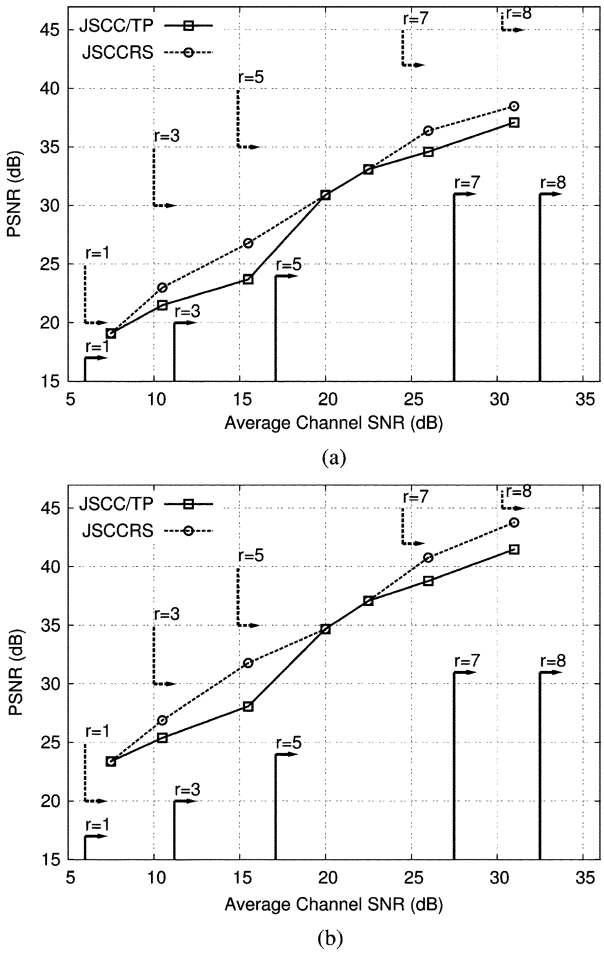

Fig. 5. PSNR for JSCCRS and JSCC/TP with a Rayleigh fading wireless channel model. (a) Frame rate is 30 fps. (b) Frame rate is 15 fps.

The proposed scheme that exercises JSCC jointly with the PHY transmission rate selection is named JSCCRS. For comparison, we also implemented the JSCC/TP scheme, where JSCC is applied independently of the PHY rate adaptation algorithm. With JSCC/TP, the rate selection algorithm uses our analytical throughput formula in Section IV and attempts to maximize the throughput independently of JSCC.

A. Average PSNR

In this subsection, we present results for the average PSNR measured at the receiving decoder for the Foreman sequence. The experimental results were averaged for 50 realizations of the wireless channel when the simulator was used.

1) Rayleigh Fading Channel: We demonstrate the advantage of the proposed system in terms of PSNR for a Rayleigh fading channel in Fig. 5. The vertical lines in this figure indicate the wireless channel SNR for which the optimal transmission rate

was changed for the two systems under test. The vertical lines in the lower part of the figure correspond to JSCC/TP, whereas the lines in the upper part correspond to JSCCRS. The first striking observation from this figure is that once the optimal PHY transmission rate is selected for either system, it remains constant until the next change. This is something to be expected for the JSCC/TP algorithm since the rate that maximizes the throughput is always selected. Even though with JSCCRS the sender can select the transmission rate of each individual outgoing PHY frame, similar behavior should be expected. The reason is that when a transmission rate has been selected as currently optimal, an increasing channel SNR, leads to improvements both in both throughput and PER that maintain the transmission rate as the optimal.

Another interesting observation is that the proposed system switches to a higher rate at a lower channel SNR than the JSCC/TP algorithm. With the extra available bandwidth from the selection of a higher PHY rate, a stronger FEC code is used and compensates for increased packet losses. The error concealment algorithm at the decoder is also partly responsible for the increase in PSNR. We also observed that a drop in the residual PER below 10% can be handled more effectively from the error concealment algorithm. For example, in Fig. 5(a) and for a target frame rate of 30 fps, when channel SNR is close to 10 dB, JSCC/TP uses RS(11,9). However, JSCCRS has already switched to rate $r3$ that provides double data rate, but lower effective throughput because of higher packet loss. For the same SNR, JSCCRS uses RS(16,9) but it has also increased the source coding rate. With respect to the actual video quality results in terms of PSNR, in Fig. 5(a) we observe that JSCCRS can achieve a higher average PSNR when compared with the JSCC/TP system. If we look more carefully into the same figure, we observe that JSCCRS outperforms JSCC by a constant margin that varies from 1 to 4.5 dB depending on the channel SNR γ. The performance gain is slightly higher when switching from $r1 \rightarrow r3$, $r3 \rightarrow r5$, and $r5 \rightarrow r7$ as the raw data rate is doubled when compared with the case of $r7 \rightarrow r8$.

Now in Fig. 5(b), we see PSNR results but for a lower frame rate of 15 fps. The objective of this experiment is to evaluate the effect of an increased bit budget per frame, since the average bit budget allocated to each frame is given by $T_{\max}(r)/\text{fps}$. The increased bit budget corresponds to six transport packets per frame. In this case, when the channel PER increases (for decreasing γ), then the PSNR curve for the proposed system deviates from the JSCC/TP at a much higher degree when compared with the situation in Fig. 5(a). One of the reasons is that the low bit budget per frame in Fig. 5(a) restricts the use of channel coding with FEC, since most of the available channel bandwidth is needed for source coding. However, the PHY transmission rate switching thresholds do not change since we are using the same bitstream that is characterized by the same RD characteristics. The higher bit budget means that a lower minimum distortion can be achieved for either JSCC/TP or JSCCRS.

2) Multipath Fading Channel: For the experiments with a multipath frequency-selective fading channel, we used the throughput simulation results depicted earlier in Fig. 4. PSNR results are now illustrated in Fig. 6. With this channel model, the results are even more intriguing, since we observe that the rate switching thresholds are not so sharply defined as in our previous experiment. This behavior can be explained if we

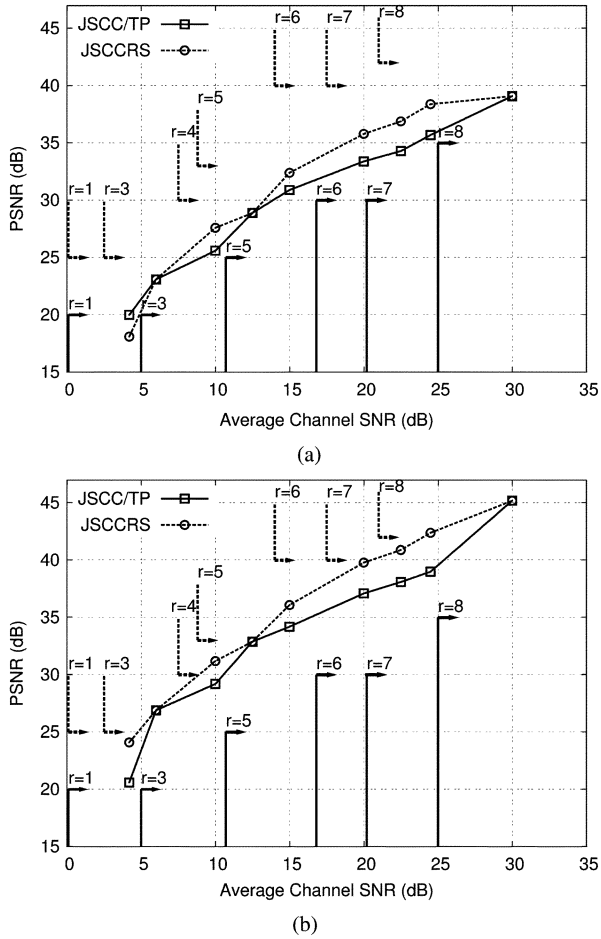

Fig. 6. Average PSNR for JSCCRS and JSCC/TP with a multipath fading wireless channel model. (a) Frame rate is 30 fps. (b) Frame rate is 15 fps.

recall the throughput results in Fig. 4. In Fig. 4, the difference between the achieved throughput of the different transmission rates is not so significant as with a Rayleigh channel. Practically, this means that the JSCCRS system can select a higher rate at even lower SNR, since both the packet loss rate and the throughput are at a level that can be compensated by the JSCC algorithm at the encoder. In Fig. 6(a), the PSNR results show that the actual difference between JSCC/TP and JSCCRS is increased when compared with the previous experiment. For the lower frame rate of 15 fps, the additional results in Fig. 6(b) present a similar trend but with a higher average value as expected. Note that for the same channel SNR, there is no significant performance difference for the two channel models that we considered. This means that the multipath channel model affects primarily the PHY rate switching thresholds.

B. Source-Channel Rate Allocation

To obtain a better understanding of the inner workings of the JSCCRS algorithm, we also traced the allocation of the available bandwidth between source and channel coding. For the simulations we analyzed in the previous subsection, Fig. 7 illustrates this allocation versus the channel SNR. This figure depicts the percentage of the bandwidth allocated to source and channel coding respectively. The dark areas correspond to the additional percentage of the bandwidth that is used for source

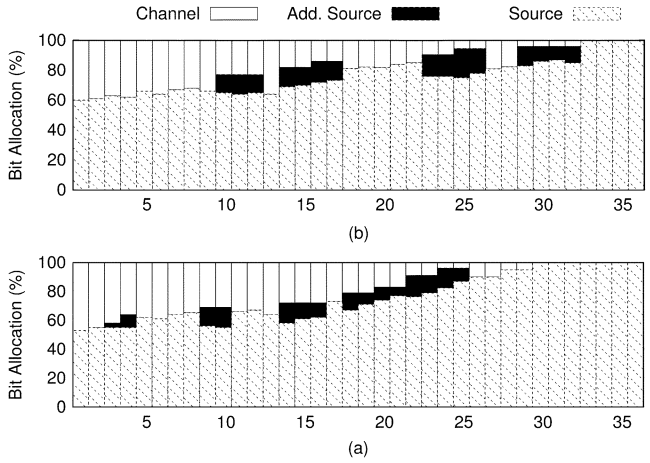

Fig. 7. Source-channel rate allocation for the JSCCRS and JSCC/TP systems with (a) Rayleigh and (b) multipath frequency-selective fading channel models.

TABLE II
PSNR (dB) FOR SIMULATIONS IN A MULTIPATH FADING CHANNEL

Sequence	Algorithm	PSNR(dB) for different channel SNR γ(dB)			Intra MBs
		10	20	24.5	(%)
Akiyo	JSCC/TP	30.44	40.19	43.23	4.44%
	JSCCRS	33.49	43.48	46.28	11.20%
Silent	JSCC/TP	26.91	35.53	38.22	14.60%
	JSCCRS	29.61	38.44	40.92	25.30%
Coastguard	JSCC/TP	22.43	29.61	31.85	54.44%
	JSCCRS	24.67	32.03	34.10	86.74%
Container	JSCC/TP	28.83	38.07	40.95	8.62%
	JSCCRS	31.72	41.19	43.84	19.80%

coding by JSCC/TP, since JSCCRS is using more aggressive FEC codes. When the Rayleigh channel is used, we observe in Fig. 7(a) that the allocation for both JSCC/TP and JSCCRS is the same for most areas in the overall SNR range. However, for specific values of SNR, JSCC/TP and JSCCRS use a different PHY transmission rate according to the used algorithm, and therefore the source-channel rate allocation is different. In these cases, the bigger part of the bandwidth is needed for channel coding (FEC). For example, for an average SNR γ equal to 22 dB, both systems operate under transmission rate $r5$. However, when γ is increased at a value of 25 dB, JSCCRS reallocates the available bandwidth and reduces the percentage allocated to source coding. The multipath frequency-selective fading channel exhibits the same behavior in Fig. 7(b), although the JSCCRS scheme is used in a wider SNR range. Furthermore, the difference between the allocations of the two systems is not so significant because of the smoother variation in both the throughput and the PER.

C. Performance for Additional Sequences

Representative PSNR results for the other four sequences encoded under a target frame rate of 30 fps and a multipath fading model can be seen in Table II. The proposed algorithm outperforms JSCC/TP for all the tested sequences. In particular, it outperforms JSCC/TP with a margin of at least 1 dB for channel conditions of 15 dB or more. Thus, in all of our tests, the JSCC/TP scheme always provides the lower performance bound for JSCCRS since the selection of transmission rate is suboptimal. Nevertheless, the RD performance of the proposed method is still considerably better than that of JSCC/TP for significant operational regions of the channel SNR range.

The most interesting results are presented in the last column of Table II. In that column, we present the percentage of intra-coded MBs used for each combination of sequence/algorithm. Essentially, this result breaks down the source rate allocation between intra- and inter-coded MBs. These results indicate the degree of error resiliency that the encoder needs to add in the bitstream, when the two algorithms are employed. We observe that for the sequences Akiyo and Container, which have relatively low motion, the use of intra-coding modes is limited with JSCC/TP. Even with the proposed JSCCRS scheme, the majority of the MBs are inter-coded. The interesting result is that these are also the sequences that have the biggest improvement in quality when we compare the two systems under test. For the sequences Silent and Coastguard—that are characterized by medium and high motion respectively—the PSNR quality gains decrease even though the used of intra-coded MBs is more aggressive.

VIII. Conclusion

In this paper, we presented a cross-layer approach for enhancing the error resiliency of video streaming applications in wireless LANs that support a multi-rate PHY. Our system employs error-resilient source encoding, application-layer channel coding with FEC, error concealment, and PHY rate adaptation. We developed analytical models that consider all the parameters that affect the performance of the aforementioned mechanisms, and subsequently we defined an RD optimization framework. Finally, we implemented an algorithm that calculates the optimal encoding mode of individual MBs, application-layer FEC, and the 802.11a PHY data transmission rate. We provided extensive simulation results to quantify the advantage of our approach. Our comparative performance analysis showed that video performance in terms of PSNR is considerably enhanced with the proposed scheme for realistic channel models. Furthermore, our analysis indicates that significant gains can be achieved, if the PHY transmission rate is selected according to the RD characteristics of the transmitted video sequence.

References

[1] M. van der Schaar and P. Chou, *Multimedia Over IP and Wireless Networks.*. New York: Academic, 2007.
[2] M. van der Schaar and S. Shankar, "Cross-layer wireless multimedia transmission: Challenges, principles, and new paradigms," *IEEE Wireless Commun.*, pp. 50–58, Aug. 2005.
[3] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 966–976, Jun. 2000.
[4] 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 Trans. Multimedia*, vol. 9, no. 3, pp. 445–454, Apr. 2007.
[5] F. Hartanto and H. R. Sirisena, "Hybrid error control mechanism for video transmission in the wireless IP networks," in *Proc. 10th IEEE Workshop on Local and Metropolitan Area Networks (LANMAN)*, 1999.

[6] A. Argyriou, "Cross-layer adaptive ARQ for uplink video streaming in tandem wireless/wireline networks," in *Proc. IEEE Multimedia Signal Processing Workshop (MMSP)*, , Oct. 2007.
[7] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: An approach to combined source-channel coding," *IEEE Trans. Inform. Theory*, vol. IT-38, no. 4, pp. 827–838, Jul. 1987.
[8] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Rate-distortion optimized hybrid error control for real-time packetized video transmission," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 40–53, Jan. 2006.
[9] G. Cheung and A. Zakhor, "Bit allocation for joint source/channel coding of scalable video," *IEEE Trans. Image Process.*, vol. 9, no. 3, pp. 340–356, Mar. 2000.
[10] Q. Zhang, W. Zhu, and Y. Zhang, "Channel-adaptive resource allocation for scalable video transmission over 3G wireless network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 8, pp. 1049–1063, Aug. 2004.
[11] D. Qiao, S. Choi, and K. G. Shin, "Goodput analysis and link adaptation for IEEE 802.11a wireless LANs," *IEEE Trans. Mobile Comput.*, vol. 1, no. 4, Oct.–Dec. 2002.
[12] S. Choudhury and J. D. Gibson, "Payload length and rate adaptation for multimedia communications in wireless LANs," *IEEE J. Select. Areas Commun.*, vol. 25, no. 4, pp. 796–807, May 2007.
[13] *Part 11: Wireless LAN Medium Access Control and Physical Layer (PHY) Specifications*, IEEE 802.11-1999 Standard, Aug. 1999.
[14] A. Kamerman and L. Monteban, "WaveLAN 2: A high-performance wireless LAN for the unlicensed band," *Bell Labs Tech. J., Tech. Rep.*, 1997.
[15] J. Choi, K. Park, and C. K. Kim, "Cross-layer analysis of rate adaptation, DCF and TCP in multi-rate WLANs," in *Proc. IEEE INFOCOM*, Anchorage, AK, May 2007.
[16] G. L. Stuber, *Principles of Mobile Communication*. Norwell, MA: Kluwer, 2001.
[17] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.
[18] M. Alouini and A. Goldsmith, "Adaptive modulation over Nakagami fading channels," *J. Wireless Commun.*, vol. 13, no. 1–2, pp. 119–143, May 2000.
[19] M. B. Pursley and D. J. Taipale, "Error probabilities for spread-spectrum packet radio with convolutional codes and Viterbi decoding," *IEEE Trans. Commun.*, vol. COM-35, no. 1, p. 112, Jan. 1987.
[20] "Tentative criteria for comparison of modulation methods," *IEEE P802.11-97/96*, Sept. 1997.
[21] J. Heiskala and J. Terry, *OFDM Wireless LANs: A Theoretical and Practical Guide.*. Indianapolis, IN: Sams, 2001.
[22] G. Cote, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error prone networks," *IEEE J. Select. Areas .Commun.*, vol. 18, no. 6, pp. 952–965, Jun. 2000.
[23] A. Argyriou, "Real-time and rate-distortion optimized video streaming with TCP," *Signal Process.: Image Commun.*, vol. 22, no. 4, pp. 374–388, Apr. 2007.
[24] "Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification," , ITU-T Rec. H.264—ISO/IEC 14496-10 AVC, May 2003.
[25] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 511–523, Jun. 2002.
[26] JVT Reference Software [Online]. Available: http://bs.hhi.de/suehring/tml/download/

Antonios Argyriou (S'00–M'07) received the Diploma in electrical and computer engineering from Democritus University of Thrace, Thrace, Greece, in 2001, and the M.S. and Ph.D. degrees in electrical and computer engineering (as a Fulbright scholar) from the Georgia Institute of Technology, Atlanta, in 2003 and 2005, respectively.

Currently, he is a Research Scientist at Philips Research, Eindhoven, The Netherlands. From 2004 to 2005, he was also a Research Engineer at Soft.Networks, Atlanta, GA, and was involved in the design of MPEG-2 and H.264 video compression algorithms. His current research interests include wireless networks and networked multimedia systems. He has published more than 30 papers in international journals and conferences.