

# An Integrated Source and Channel Rate Allocation Scheme for Robust Video Coding and Transmission over Wireless Channels

**Jie Song**

*Media Connectivity Division, Agere Systems, Holmdel, NJ 07733, USA  
Email: jiesong@agere.com*

**K. J. Ray Liu**

*Electrical & Computer Engineering Department, University of Maryland, College Park, MD 20742, USA  
Email: kjrlui@eng.umd.edu*

*Received 22 November 2002; Revised 3 September 2003*

A new integrated framework for source and channel rate allocation is presented for video coding and transmission over wireless channels without feedback channels available. For a fixed total channel bit rate and a finite number of channel coding rates, the proposed scheme can obtain the near-optimal source and channel coding pair and corresponding robust video coding scheme such that the expected end-to-end distortion of video signals can be minimized. With the assumption that the encoder has the stochastic information such as average SNR and Doppler frequency of the wireless channel, the proposed scheme takes into account robust video coding, channel coding, packetization, and error concealment techniques altogether. An improved method is proposed to recursively estimate the end-to-end distortion of video coding for transmission over error-prone channels. The proposed estimation is about 1–3 dB more accurate compared to the existing integer-pel-based method. Rate-distortion-optimized video coding is employed for the trade-off between coding efficiency and robustness to transmission errors.

**Keywords and phrases:** multimedia communications, joint source and channel coding, wireless video.

## 1. INTRODUCTION

Multimedia applications such as video phone and video streaming will soon be available in the third generation (3G) wireless systems and beyond. For these applications, delay constraint makes the conventional automatic repeat request (ARQ) and the deep interleaver not suitable. Feedback channels can be used to deal with the error effects incurred in image and video transmission over error-prone channels [1], but in applications such as broadcasting services, there is no feedback channel available. In such cases, the optimal trade-off between source and channel coding rate allocations for video transmission over error-prone channels becomes very important. According to Shannon's separation theory, these components can be designed independently without loss in performance [2]. However, this is based on the assumption that the system has an unlimited computational complexity and infinite delay. These assumptions are not satisfied in delay-sensitive real-time multimedia communications. Therefore, it is expected that joint considerations of source and channel coding can provide performance improvement [3, 4].

Most of the joint source and channel coding (JSCC) schemes have been focusing on images and sources with ideal signal models [4, 5]. For video coding and transmission, many works still keep the source coding and channel coding separate instead of optimizing their parameters jointly from an overall end-to-end transmission point of view [6, 7]. Some excellent reviews about robust video coding and transmission over wireless channels can be found in [8, 9]. In [10], a JSCC approach is proposed for layered video coding and transport over error-prone packet networks. It presented a framework which trades video source coding efficiency off for increased bitstream error resilience to optimize the video coding mode selection with the consideration of channel conditions as well as error recovery and concealment capabilities of the channel codec and source decoder, respectively. However, the optimal source and channel rate allocation and corresponding video macroblock (MB) mode selection have to be selected through simulations over packet-loss channel models. In [11], a parameterized model is used for the analysis of the overall mean square error (MSE) in hybrid video coding for the error-prone transmission. Models for the video encoder, a bursty transmission channel, and error

propagation at the video decoder have been combined into a complete model of the entire video transmission system. However, the model for video encoder involves several parameters and the model is not theoretically optimal because of the use of random MB intramode updating, which does not consider the different motion activities within a video frame to deal with error propagation. Furthermore, the models depend on the distortion-parameter functions obtained through ad hoc numerical models and simulations over specific video sequences, which also involves a lot of simulation efforts and approximation. The authors of [12] proposed an operational rate-distortion (RD) model for DCT-based video coding incorporating the MB intra-refreshing rate and an analytic model for video error propagation which has relatively low computational complexity and is suitable for real-time wireless video applications. Both methods in [11, 12] focus on the statistical model optimization for general video sequence, which is not necessarily optimal for a specific video sequence because of the nonstationary behavior across different video sequences.

In this paper, we propose an integrated framework to obtain, the near-optimal source and channel rate allocation, and the corresponding robust video coding scheme for a given total channel bit rate with the knowledge of the stochastic characteristics of the wireless fading channel. We consider the video coding error (quantization and mode selection of MB), error propagation, and concealment effects at the receiver due to transmission error, packetization, and channel coding in an integrated manner. The contributions of this paper are the following. First, we present an integrated system design method for wireless video communications in realistic scenarios. This proposed method takes into account the interactions of fading channel, channel coding and packetization, and robust video coding in an integrated, yet simple way, which is an important system design issue for wireless video applications. Second, we propose an improved video distortion estimation which is about 1–3 dB peak signal-to-noise ratio (PSNR) more accurate than the original integer-pel-based method (IP) in [13] for half-pel-based video coding (HP), and the computational complexity in the proposed method is less than that in [13].

The rest of the paper is organized as follows. Section 2 describes first the system to be studied, then the packetization and channel coding schemes used. We also derive the integrated relation between MB error probability and channel coding error probability given the general wireless fading channel information such as average signal-to-noise ratio (SNR) and Doppler frequency. Section 3 presents the improved end-to-end distortion estimation method for HP-based video coding. Simulations are performed to compare the proposed method to the IP-based method in [13]. Then we employ RD-optimized video coding scheme to optimize the end-to-end performance for each pair of source and channel rate allocation. Simulation results are shown in Section 4 to demonstrate the accuracy of the proposed end-to-end distortion estimation algorithm under different channel characteristics. Conclusions are stated in Section 5.

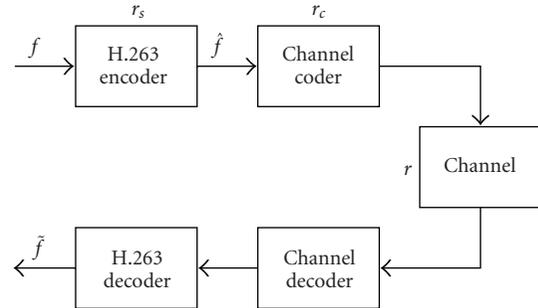


FIGURE 1: Joint source and channel video coding.

## 2. PROBLEM DEFINITION AND INTEGRATED SYSTEM STRUCTURE

The problem to be studied is illustrated in Figure 1 which can be specified by five parameters  $(r, r_c, \rho, f_d, F)$ :  $r$  is the total channel bit rate,  $r_c$  is the channel coding rate,  $\rho$  is the average SNR at the receiver,  $f_d$  is the Doppler frequency of the fading channel targeted, and  $F$  is the video frame rate. H.263 [14] is used for video coding. A video sequence denoted as  $f_l^s$ , where  $\mathbf{s} = (x, y)$ ,  $1 \leq x \leq X$ ,  $1 \leq y \leq Y$ , is the pixel spatial location and  $l = 1, \dots, L$  is the frame index, is encoded at the bit rate  $r_s = r \times r_c$  b/s and the frame rate  $F$  f/s with the MB error probability  $P_{Mb} = f(\rho, f_d, r_c)$  that will be detailed next. The resulted H.263 bitstream is packetized and protected by forward error correction (FEC) channel coding with the coding rate  $r_c$ . The resulted bitstream with rate  $r$  b/s is transmitted through wireless channels characterized by  $\rho$  and  $f_d$ . The receiver receives the bitstream corrupted by the channel impairment, then reconstructs the video sequence  $\tilde{f}_l^s$  after channel decoding, H.263 video decoding, and possible error concealment if residual errors occur. The end-to-end MSE between the input video sequence at the encoder and the reconstructed video sequence at the decoder is defined as

$$D_E(r_s, r_c) = \frac{1}{XYL} \sum_{x=1}^X \sum_{y=1}^Y \sum_{l=1}^L E \left\{ \left[ f_l^{(x,y)} - \tilde{f}_l^{(x,y)}(r_s, r_c) \right]^2 \right\}. \quad (1)$$

For the video system in Figure 1, there are two tasks to be performed with the five given system parameters  $(r, r_c, \rho, f_d, F)$ . First, we need to decide how to allocate the total fixed bit rate  $r$  to the source rate  $r_s = r \times r_c$  to minimize the end-to-end MSE of the video sequence. Furthermore, the video encoder should be able to, for a source/channel rate allocation  $(r_s, r_c)$  with residual channel decoding failure rate denoted as  $p_w(r_c)$ , select the coding mode and quantizer for each MB to minimize the end-to-end MSE of the video sequence. The goal is to obtain the source/channel rate pair  $(r_s^*, r_c^*)$  and the corresponding robust video coding scheme to minimize (1).

In practical applications, there are only finite number of source/channel pairs available. We can find the robust video encoding schemes for each rate pair  $(r_s, r_c)$  that minimizes (1) and denote the minimal end-to-end MSE obtained as

$D_E^*(r_s, r_c)$ , then the optimal source/channel rate pair  $(r_s^*, r_c^*)$  and the corresponding video coding scheme can be obtained as

$$(r_s^*, r_c^*) = \underset{(r_s, r_c)}{\operatorname{argmin}} D_E^*(r_s, r_c). \quad (2)$$

For each pair  $(r_s, r_c)$ , we use RD-optimized video coding scheme to trade off between the source coding efficiency and robustness to error propagation. An improved recursive method which takes into account the interframe prediction, error propagation, and concealment effect is used to estimate the end-to-end MSE frame by frame. In this paper, the wireless fading channel is modeled as a finite-state Markov chain (FSMC) model [15, 16, 17], and the Reed-Solomon (RS) code is employed for forward error coding.

### 2.1. Modeling fading channels using finite-state Markov chain

Gilbert and Elliott [15, 16] studied a two-state Markov channel model, where each state corresponds to a specific channel quality. This model provides a close approximation for the error rate performance of block codes on some noisy channels. On the other hand, when the channel quality varies dramatically such as in a fast Doppler spread, the two-state Gilbert-Elliott model becomes inadequate. Wang and Moayeri extended the two-state model to an FSMC model for characterizing the Rayleigh fading channels [17]. In [17], the received SNR is partitioned into a finite number of intervals. Denote by  $0 = A_0 < A_1 < A_2 < \dots < A_K = \infty$  the SNR thresholds of different intervals, then if the received SNR is in the interval  $[A_k, A_{k+1})$ ,  $k \in \{0, 1, 2, \dots, K-1\}$ , the fading channel is said to be in state  $S_k$ . It turns out that if the channel changes slowly and is properly partitioned, each state can be considered as a steady state, and a state transition can only happen between neighboring states. As a result, a fading channel can be represented using a Markov model if given the average SNR  $\rho$  and Doppler frequency  $f_d$ .

### 2.2. Performance analysis of RS code over finite-state Markov channel model

RS codes possess maximal minimum distance properties which make them powerful in correcting errors with arbitrary distributions. For RS symbols composed of  $m$  bits, the encoder for an  $RS(n, k)$  code groups the incoming bitstream into blocks of  $k$  information symbols and appends  $n - k$  redundancy symbols to each block. So the channel coding rate is  $r_c = k/n$ . For an  $RS(n, k)$  code, the maximal number of symbol errors that can be corrected is  $t = \lfloor (n - k)/2 \rfloor$ . When the number of symbol errors is more than  $t$ , RS decoder reports a flag to notify that the errors are uncorrectable. The probability that a block cannot be corrected by  $RS(n, k)$ , denoted as a decoding failure probability  $p_w(n, k)$ , can be calculated as

$$p_w(n, k) = \sum_{m=t+1}^n P(n, m), \quad (3)$$

where  $P(n, m)$  denotes the probability of  $m$  symbol errors within a block of  $n$  successive symbols. The computation of  $P(n, m)$  for FSMC channel model has been studied before (see [16, 18]).

### 2.3. Packetization and macroblock error probability computation

We use baseline H.263 video coding standard for illustration. H.263 GOB/slice structure is used where each GOB/slice is encoded independently with a header to improve resynchronization. Denoting by  $N_s$  the number of GOB/slice in each frame, the  $RS(n, k)$  code block size  $n$  (bytes) is set to

$$n = \left\lceil \frac{r}{8 \cdot F \cdot N_s} \right\rceil \quad (4)$$

such that each GOB/slice is protected by an RS codeword in average, where  $\lceil x \rceil$  is the smallest integer larger than  $x$ . No further alignment is used. In case of decoding failure of an RS codeword, the GOBs (group of blocks) covered by the RS code will be simply discarded and followed by error concealment. If a GOB is corrupted, the decoder simply drop the GOB and performs a simple error concealment as follows: the motion vector (MV) of a corrupted MB is replaced by the MV of the MB in the GOB above. If the GOB above is also lost, the MV is set to zero, then the MB is replaced by the corresponding MB at the same location in the previous frame. To facilitate error concealment at the decoder when errors occur, the GOBs which are indexed by even numbers are concatenated together, followed by concatenated GOBs indexed by odd numbers. By using this alternative GOB organization, the neighboring GOBs are normally not protected within the same RS codeword. Thus, when a decoding failure occurs in one RS codeword, the neighboring GOBs will not be corrupted simultaneously, which helps the decoder to perform error concealment using the neighboring correctly received GOB.

In order to estimate the end-to-end distortion, we need to model the relation between video MB error probability  $P_{MB}(n, k)$  and  $RS(n, k)$  decoding failure probability  $p_w(n, k)$ , that is,

$$P_{MB}(n, k) \approx \alpha \cdot p_w(n, k). \quad (5)$$

Since no special packetization or alignment is used, one RS codeword may contain part of one GOB/slice or overlap more than one GOB/slice. It is difficult to find the exact relation between  $P_{MB}(n, k)$  and  $p_w(n, k)$  because the length of GOB in each frame is varying. Intuitively,  $\alpha$  should be between 1 and 2. Experiments are performed to find the suitable  $\alpha$ . Figure 2 shows the experiment results of RS codeword failure probability and GOB error probability over Rayleigh fading channels. It turns out that  $\alpha \approx 1.5$  is a good approximation in average. For a source and channel code pair  $(r_s, r_c)$  or  $RS(n, k)$ , the channel code decoding failure probability  $p_w(n, k)$  can be derived from  $\rho$  and  $f_d$  as described in Sections 2.1 and 2.2, then we have the corresponding video MB error probability  $P_{MB}(n, k)$  from (5). Based on the derived MB error rate  $P_{MB}(n, k)$ , a recursive estimation method

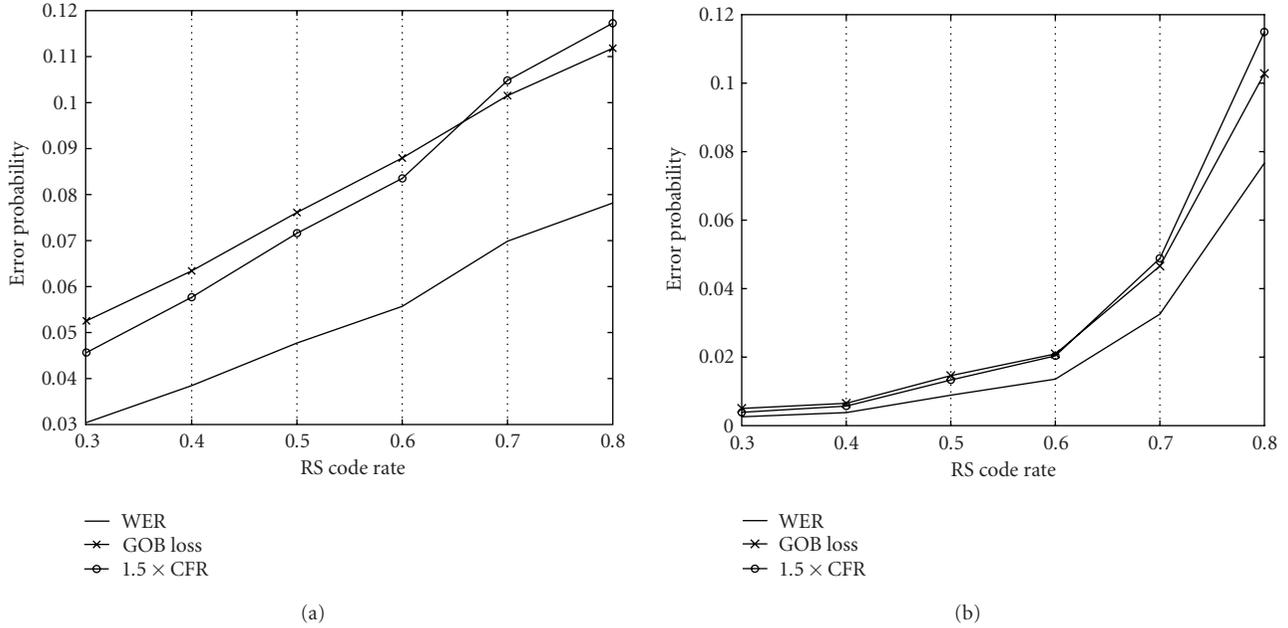


FIGURE 2: Simulated RS codeword failure rate (CFR), GOB loss rate, and the values of 1.5 + WER. (a) Rayleigh fading, SNR = 18 dB, QPSK, and  $fd = 10$  Hz. (b) Rayleigh fading, SNR = 18 dB, QPSK,  $fd = 100$  Hz.

and an RD-optimized scheme are employed to estimate the minimal end-to-end MSE of the video sequence and obtain the corresponding optimized video coding scheme, which is to be described in detail in the next section.

### 3. OPTIMAL DISTORTION ESTIMATION AND MINIMIZATION

We first describe the proposed distortion estimation method for both HP- and IP-based video coding over error-prone channels. Simulations are performed to demonstrate the improved performance of the proposed method. Then an RD framework is used to select the coding mode and quantizer for each MB to minimize the estimated distortion, given the source rate  $r_s$ ,  $P_{MB}$  which is derived as in Section 2, and the frame rate  $F$ .

#### 3.1. Optimal distortion estimation

Recently, modeling of error propagation effects have been considered in order to optimally select the mode for each MB to trade off the compression efficiency and error robustness [11, 13, 19]. In particular, a recursive optimal per-pixel estimate (ROPE) of decoder distortion was proposed in [13] which can model the error propagation and quantization distortion more accurately than other methods. But the method in [13] is only optimal for the IP-based video coding. For the HP case, the computation of spatial cross correlation between pixels in the same and different MBs is needed to obtain the first and second moments of bilinear interpolated HPs, the process is computationally prohibitive. Most of the current video coding use the HP-based method to improve

the compression performance. We propose a modified recursive estimate of end-to-end distortion that can take care of both IP- and HP-based video coding.

The expected end-to-end distortion for the pixel  $f_l^s$  at  $\mathbf{s} = (x, y)$  in frame  $l$  is

$$\begin{aligned} d_l^s &= E\{(f_l^s - \tilde{f}_l^s)^2\} \\ &= E\{(f_l^s - \hat{f}_l^s + \hat{f}_l^s - \tilde{f}_l^s)^2\} \\ &= (f_l^s - \hat{f}_l^s)^2 + 2(f_l^s - \hat{f}_l^s)E\{\hat{f}_l^s - \tilde{f}_l^s\} + E\{(\hat{f}_l^s - \tilde{f}_l^s)^2\}, \end{aligned} \quad (6)$$

where  $\hat{f}_l^s$  is the quantized pixel value at  $\mathbf{s}$  in frame  $l$ . Denote  $e_l^s = f_l^s - \hat{f}_l^s$  as the quantization error,  $\hat{e}_l^{s,v} = \hat{f}_l^s - \hat{f}_{l-1}^{s-v}$  as the motion compensation error using MV  $\mathbf{v}$ , and  $\tilde{e}_l^s = \hat{f}_l^s - \tilde{f}_l^s$  as the transmission and error-propagation error. Assuming that  $\tilde{e}_l^s$  is an uncorrelated random variable with a zero mean, which is a reasonable assumption when  $P_{MB}$  is relatively low as will be shown in the simulations later, we have

$$d_l^s = (e_l^s)^2 + E\{(\tilde{e}_l^s)^2\}. \quad (7)$$

We derive a recursive estimate of  $E\{(\tilde{e}_l^s)^2\}$  for intra-MB and inter-MB as follows.

#### Intramode MB

The following three cases are considered.

- (1) With the probability  $1 - P_{MB}$ , the intra-MB is received correctly and then  $\hat{f}_l^s = \tilde{f}_l^s$ . As a result,  $\tilde{e}_l^s = 0$ .

(2) With the probability  $(1 - P_{MB})P_{MB}$ , the intra-MB is lost but the MB above is received correctly. Denoting by  $\mathbf{v}_c = (x_c, y_c)$  the MV of the MB above, two cases of error concealment are considered depending on whether  $\mathbf{v}_c$  is at the HP location or not.

(i) If  $\mathbf{v}_c$  is at the IP location, we have  $\tilde{f}_l^s = \tilde{f}_{l-1}^{s-\mathbf{v}_c}$ . Then after error concealment,

$$\begin{aligned}\bar{e}_l^s &= \hat{f}_l^s - \tilde{f}_l^s \\ &= \hat{f}_l^s - \tilde{f}_{l-1}^{s-\mathbf{v}_c} \\ &= \hat{f}_l^s - \hat{f}_{l-1}^{s-\mathbf{v}_c} + \hat{f}_{l-1}^{s-\mathbf{v}_c} - \tilde{f}_{l-1}^{s-\mathbf{v}_c} \\ &= \hat{e}_l^{s,\mathbf{v}_c} + \tilde{e}_{l-1}^{s-\mathbf{v}_c}.\end{aligned}\quad (8)$$

The clipping effect is ignored in the computation.

(ii) If  $\mathbf{v}_c$  is at the HP location, without loss of generality, assume that  $\mathbf{v}_c = (x_c, y_c)$  is at HP location that is interpolated from four neighbouring IP locations with MVs:  $\mathbf{v}_c^1 = (\lfloor x_c \rfloor, \lfloor y_c \rfloor)$ ,  $\mathbf{v}_c^2 = (\lceil x_c \rceil, \lfloor y_c \rceil)$ ,  $\mathbf{v}_c^3 = (\lfloor x_c \rfloor, \lceil y_c \rceil)$ , and  $\mathbf{v}_c^4 = (\lceil x_c \rceil, \lceil y_c \rceil)$ , where  $\lfloor x_c \rfloor$  and  $\lceil x_c \rceil$  denote the largest integer that is smaller than  $x_c$  and the smallest integer larger than  $x_c$ , respectively. We have  $\tilde{f}_l^s = (\tilde{f}_{l-1}^{s-\mathbf{v}_c^1} + \tilde{f}_{l-1}^{s-\mathbf{v}_c^2} + \tilde{f}_{l-1}^{s-\mathbf{v}_c^3} + \tilde{f}_{l-1}^{s-\mathbf{v}_c^4})/4$  and

$$\begin{aligned}\bar{e}_l^s &= \hat{f}_l^s - \tilde{f}_l^s \\ &= \frac{\hat{f}_l^s - \tilde{f}_{l-1}^{s-\mathbf{v}_c^1}}{4} + \frac{\hat{f}_l^s - \tilde{f}_{l-1}^{s-\mathbf{v}_c^2}}{4} \\ &\quad + \frac{\hat{f}_l^s - \tilde{f}_{l-1}^{s-\mathbf{v}_c^3}}{4} + \frac{\hat{f}_l^s - \tilde{f}_{l-1}^{s-\mathbf{v}_c^4}}{4} \\ &= \frac{\hat{e}_l^{s,\mathbf{v}_c^1} + \hat{e}_l^{s,\mathbf{v}_c^2} + \hat{e}_l^{s,\mathbf{v}_c^3} + \hat{e}_l^{s,\mathbf{v}_c^4}}{4} \\ &\quad + \frac{\tilde{e}_{l-1}^{s-\mathbf{v}_c^1} + \tilde{e}_{l-1}^{s-\mathbf{v}_c^2} + \tilde{e}_{l-1}^{s-\mathbf{v}_c^3} + \tilde{e}_{l-1}^{s-\mathbf{v}_c^4}}{4}.\end{aligned}\quad (9)$$

(3) With the probability  $P_{MB}^2$ , both the current MB and the MB above are lost. The MB in the previous video frame at the same location is repeated, that is,  $\mathbf{v}_c = \mathbf{0} = (0, 0)$ :

$$\begin{aligned}\bar{e}_l^s &= \hat{f}_l^s - \tilde{f}_l^s \\ &= \hat{f}_l^s - \tilde{f}_{l-1}^s \\ &= \hat{f}_l^s - \hat{f}_{l-1}^s + \hat{f}_{l-1}^s - \tilde{f}_{l-1}^s \\ &= \hat{e}_l^{s,\mathbf{0}} + \tilde{e}_{l-1}^s.\end{aligned}\quad (10)$$

Combining all of the cases together, we have the following results.

(1) If  $\mathbf{v}_c$  is at the IP location, then

$$\begin{aligned}E_{\text{Intra}}\{(\bar{e}_l^s)^2\} &= (1 - P_{MB})P_{MB}\{(\hat{e}_l^{s,\mathbf{v}_c})^2 + E[(\tilde{e}_{l-1}^{s-\mathbf{v}_c})^2]\} \\ &\quad + P_{MB}^2\{(\hat{e}_l^{s,\mathbf{0}})^2 + E[(\tilde{e}_{l-1}^s)^2]\}.\end{aligned}\quad (11)$$

(2) If  $\mathbf{v}_c$  is at the HP location in both  $x$  and  $y$  dimensions, then

$$\begin{aligned}E_{\text{Intra}}\{(\bar{e}_l^s)^2\} &= (1 - P_{MB})P_{MB} \\ &\quad \times \left\{ \left( \frac{\hat{e}_l^{s,\mathbf{v}_c^1} + \hat{e}_l^{s,\mathbf{v}_c^2} + \hat{e}_l^{s,\mathbf{v}_c^3} + \hat{e}_l^{s,\mathbf{v}_c^4}}{4} \right)^2 \right. \\ &\quad \left. + \frac{E[(\tilde{e}_{l-1}^{s-\mathbf{v}_c^1})^2] + E[(\tilde{e}_{l-1}^{s-\mathbf{v}_c^2})^2] + E[(\tilde{e}_{l-1}^{s-\mathbf{v}_c^3})^2] + E[(\tilde{e}_{l-1}^{s-\mathbf{v}_c^4})^2]}{16} \right\} \\ &\quad + P_{MB}^2\{(\hat{e}_l^{s,\mathbf{0}})^2 + E[(\tilde{e}_{l-1}^s)^2]\}.\end{aligned}\quad (12)$$

The cases when  $\mathbf{v}_c$  has only  $x_c$  or  $y_c$  at the HP location can be obtained similarly.

#### Intermode MB

When an intermode MB is correctly received with the probability  $1 - P_{MB}$ , the motion compensation error  $\hat{e}_l^s = \hat{f}_l^s - \hat{f}_{l-1}^{s-\mathbf{v}}$  and the MV  $\mathbf{v}$  are received correctly and are reconstructed from the previous reconstructed frame at the decoder. We again consider two cases depending on whether  $\mathbf{v} = (x, y)$  is at the IP or HP location.

(1) If  $\mathbf{v}$  is at the IP location, then  $\tilde{f}_l^s = \hat{e}_l^s + \tilde{f}_{l-1}^{s-\mathbf{v}}$  and

$$\begin{aligned}\bar{e}_l^s &= \hat{f}_l^s - \tilde{f}_l^s = \hat{f}_l^s - (\hat{e}_l^s + \tilde{f}_{l-1}^{s-\mathbf{v}}) \\ &= \underbrace{\hat{f}_l^s - \hat{f}_{l-1}^{s-\mathbf{v}} - \hat{e}_l^s}_0 + \hat{f}_{l-1}^{s-\mathbf{v}} - \tilde{f}_{l-1}^{s-\mathbf{v}} \\ &= \tilde{e}_{l-1}^{s,\mathbf{v}}.\end{aligned}\quad (13)$$

(2) If  $\mathbf{v} = (x, y)$  is at the HP location in both  $x$  and  $y$  dimensions, the prediction is interpolated from four pixels with MVs:  $\mathbf{v}^1 = (\lfloor x \rfloor, \lfloor y \rfloor)$ ,  $\mathbf{v}^2 = (\lceil x \rceil, \lfloor y \rceil)$ ,  $\mathbf{v}^3 = (\lfloor x \rfloor, \lceil y \rceil)$ , and  $\mathbf{v}^4 = (\lceil x \rceil, \lceil y \rceil)$ . Then  $\tilde{f}_l^s = \hat{e}_l^s + (\tilde{f}_{l-1}^{s-\mathbf{v}^1} + \tilde{f}_{l-1}^{s-\mathbf{v}^2} + \tilde{f}_{l-1}^{s-\mathbf{v}^3} + \tilde{f}_{l-1}^{s-\mathbf{v}^4})/4$ . We have

$$\begin{aligned}\bar{e}_l^s &= \hat{f}_l^s - \tilde{f}_l^s \\ &= \hat{f}_l^s - \left( \hat{e}_l^s + \frac{\tilde{f}_{l-1}^{s-\mathbf{v}^1} + \tilde{f}_{l-1}^{s-\mathbf{v}^2} + \tilde{f}_{l-1}^{s-\mathbf{v}^3} + \tilde{f}_{l-1}^{s-\mathbf{v}^4}}{4} \right) \\ &= \left( \frac{\hat{f}_{l-1}^{s-\mathbf{v}^1} + \hat{f}_{l-1}^{s-\mathbf{v}^2} + \hat{f}_{l-1}^{s-\mathbf{v}^3} + \hat{f}_{l-1}^{s-\mathbf{v}^4}}{4} + \hat{e}_l^s \right) \\ &\quad - \left( \hat{e}_l^s + \frac{\tilde{f}_{l-1}^{s-\mathbf{v}^1} + \tilde{f}_{l-1}^{s-\mathbf{v}^2} + \tilde{f}_{l-1}^{s-\mathbf{v}^3} + \tilde{f}_{l-1}^{s-\mathbf{v}^4}}{4} \right) \\ &= \frac{\tilde{e}_{l-1}^{s,\mathbf{v}^1} + \tilde{e}_{l-1}^{s,\mathbf{v}^2} + \tilde{e}_{l-1}^{s,\mathbf{v}^3} + \tilde{e}_{l-1}^{s,\mathbf{v}^4}}{4}.\end{aligned}\quad (14)$$

The results of the other MB loss cases are the same as that of the intra-MB. We have the following two results.

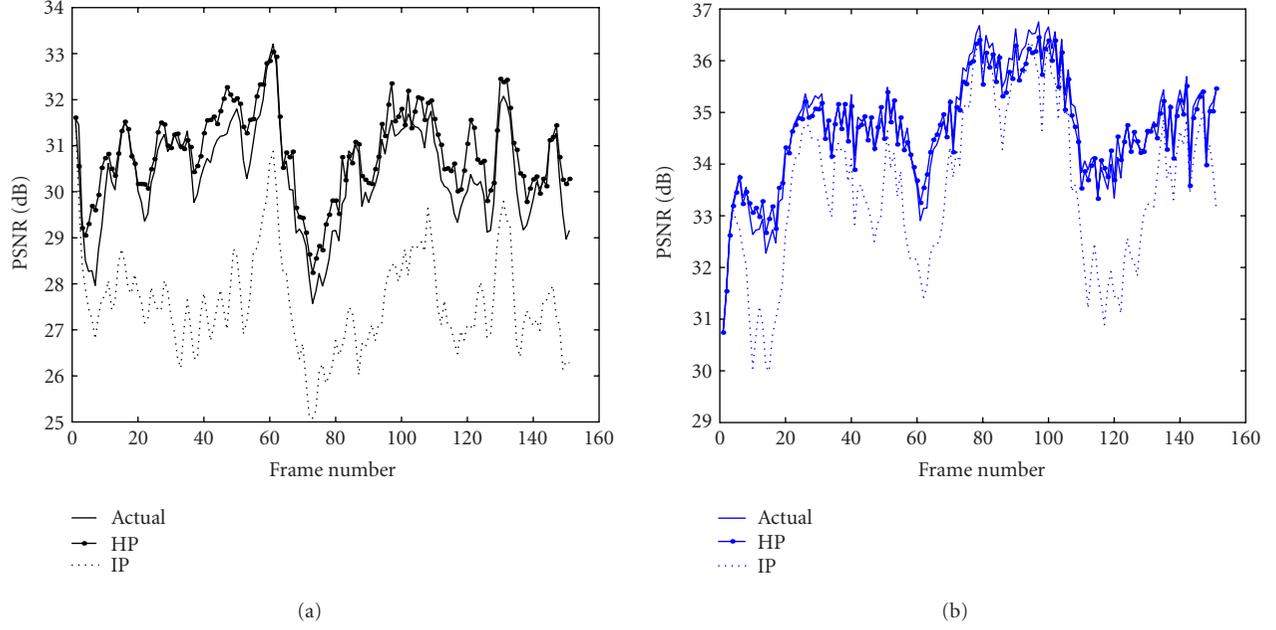


FIGURE 3: Comparison between HP- and IP-based distortion estimation in the HP video coding case (a) Foreman:  $r = 300$  kbps,  $f = 30$  f/s, and  $P_{MB} = 0.1$ . (b) Salesman:  $r = 300$  kbps,  $f = 30$  f/s, and  $P_{MB} = 0.1$ .

(1) If both  $\mathbf{v}$  and  $\mathbf{v}^c$  are at the IP location, then

$$\begin{aligned}
 E_{\text{Inter}}\{(\hat{e}_i^s)^2\} &= (1 - P_{MB})E[(\hat{e}_{i-1}^{s-v})^2] \\
 &+ (1 - P_{MB})P_{MB}\{(\hat{e}_i^{s,v^c})^2 + E[(\hat{e}_{i-1}^{s-v^c})^2]\} \\
 &+ P_{MB}^2\{(\hat{e}_i^{s,0})^2 + E[(\hat{e}_{i-1}^s)^2]\}.
 \end{aligned} \tag{15}$$

(2) If  $\mathbf{v}$  and  $\mathbf{v}^c$  are at the HP location in both  $x$  and  $y$  dimensions, then

$$\begin{aligned}
 E_{\text{Inter}}\{(\hat{e}_i^s)^2\} &= (1 - P_{MB}) \\
 &\times \left\{ \frac{E[(\hat{e}_{i-1}^{s-v^1})^2] + E[(\hat{e}_{i-1}^{s-v^2})^2] + E[(\hat{e}_{i-1}^{s-v^3})^2] + E[(\hat{e}_{i-1}^{s-v^4})^2]}{16} \right\} \\
 &+ (1 - P_{MB})P_{MB} \\
 &\times \left\{ \left( \frac{\hat{e}_i^{s,v^1} + \hat{e}_i^{s,v^2} + \hat{e}_i^{s,v^3} + \hat{e}_i^{s,v^4}}{4} \right)^2 \right. \\
 &\quad \left. + \frac{E[(\hat{e}_{i-1}^{s-v^1})^2] + E[(\hat{e}_{i-1}^{s-v^2})^2] + E[(\hat{e}_{i-1}^{s-v^3})^2] + E[(\hat{e}_{i-1}^{s-v^4})^2]}{16} \right\} \\
 &+ P_{MB}^2\{(\hat{e}_i^{s,0})^2 + E[(\hat{e}_{i-1}^s)^2]\}.
 \end{aligned} \tag{16}$$

The encoder can use the above procedures to recursively estimate the expected distortion  $d_i^s$  in (7), based on the accumu-

lated coding and error propagation effects from the previous video frames and current MB coding modes and quantizers. To implement the HP-based estimation, the encoder needs to store an image for  $E\{(\hat{e}_i^s)^2\}$ ; for the locations in which either  $x$  or  $y$  is at HP, the value is obtained by scaling the sum of the neighboring two values by 1/4, and for locations in which both  $x$  and  $y$  are at HP precision, it is obtained by scaling the sum of the neighboring four values by 1/16. It should be noted that the scaling by 1/4 or 1/16 can be done by simple bitshift. Both IP- and HP-based estimations need the same memory size to store either two IP images,  $E\{f_i\}$  and  $E\{f_i^2\}$ , or one HP image,  $E\{(\hat{e}_i^s)^2\}$ , but  $E\{(\hat{e}_i^s)^2\}$  requires smaller bandwidth/pel since it is an error signal instead of a pixel value. The HP-based computational complexity is less than the IP-based method since it only needs to compute  $E\{(\hat{e}_i^s)^2\}$  instead of computing both  $E\{f_i\}$  and  $E\{f_i^2\}$  in the IP-based estimate.

We now compare the accuracy of the proposed HP-based estimation to the original IP-based method (ROPE) in [13]. In the simulation, each GOB is carried by one packet. So the packet loss rate is equivalent to the MB error probability  $P_{MB}$ . A memoryless packet loss generator is used to drop the packet at a specified loss probability. QCIF sequences *Foreman* and *Salesman* are encoded by the Telenor H.263 encoder with the intra-MB fresh rate set to 4, that is, each MB is forced to be intramode coded if it has not been intracoded for consecutive four frames. The HP- and IP-based estimates are compared to the actual decoder distortion averaged over 50 different channel realizations.

In Figure 3a, the sequence *Foreman* of 150 frames is encoded with HP motion compensation at a bit rate of

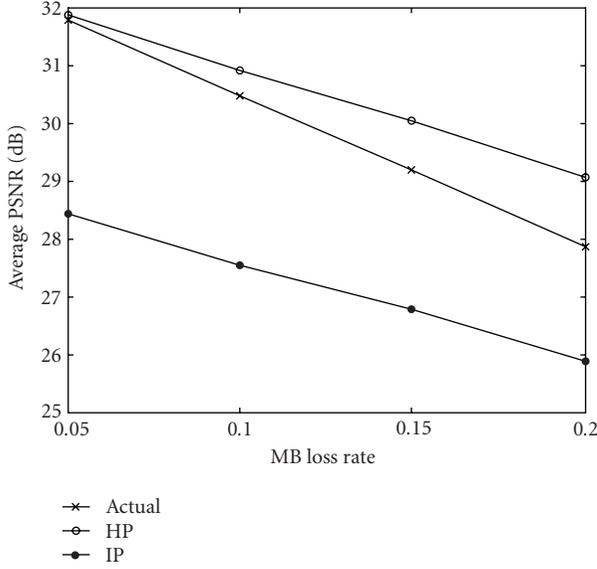


FIGURE 4: Average PSNR versus MB loss rate for HP- and IP-based distortion estimation; Foreman:  $r = 300$  kb/s and  $f = 30$  f/s.

300 Kbps, frame rate of 30 f/s, and MB loss rate of 10%. In Figure 3b, the sequence Salesman is encoded in the same way. It can be noted that the HP-based estimation is more accurate to estimate the actual distortion at the decoder compared to the IP-based estimation. Figure 4 also shows the average PSNR of the 150 coded frames with respect to MB loss rates from 5% to 20%. When MB loss rate is as small as 5%, the HP-based estimation is almost the same as the actual distortion, while the IP-based method has about 3 dB difference. The results is as expected since there is about 2–4 dB PSNR difference between HP- and IP-based video coding efficiency given the same bit rate. As the MB loss rate increases as large as 20%, the HP-based estimation is about 1 dB better than the actual distortion, while the IP-based estimation is about 2 dB worse. So the HP-based method is still 1 dB more accurate than the IP-based method. The reason is that the error propagation effects play a more significant role when MB loss rate gets larger, so the coding gain of the HP-based motion compensation is reduced. Also, the assumption in HP-based method that the transmission and propagation errors are not correlated and zero mean may become loose. For practical scenarios, it is demonstrated that the HP-based estimation outperforms the original IP-based method by about 1–3 dB.

### 3.2. Rate-distortion-optimized video coding

The quantizer step size and code mode for each MB in a frame is optimized by an RD framework. Denote by  $b_{i,j,l}$  the MB at location  $(i, j)$  in frame  $l$ , where  $i \in \{1, 2, \dots, H\}$  and  $j \in \{1, 2, \dots, V\}$ . Let  $q_{i,j,l} \in Q$ ,  $Q = \{1, 2, \dots, 32\}$ , be the quantizer parameter for  $b_{i,j,l}$ , and let  $m_{i,j,l} \in M$  be the encoding mode for  $b_{i,j,l}$ , where  $M = \{\text{intra, inter, skip}\}$  is the set of all admissible encoding modes. Denote by  $c_{i,j,l} =$

$[q_{i,j,l}, m_{i,j,l}] \in C$  the encoding vector for  $b_{i,j,l}$ , where  $C = Q \times M$  is the set of all admissible encoding vectors. For each source/channel pair  $(r_s, r_c)$ , we have the corresponding  $P_{\text{MB}}(n, k)$  from (5). The encoder needs to determine the coding mode and quantizer for each MB in total  $L$  frames to minimize the end-to-end MSE  $D_E(r_s, P_{\text{MB}})$  of the video sequence, which is defined as

$$D_E(r_s, P_{\text{MB}}) = \sum_{l=1}^L D_l(R_l, P_{\text{MB}}), \quad (17)$$

where  $R_l$  is the number of bits used to encode frame  $l$ , its maximal value is denoted as  $R_l^{\text{max}} = r_s/F + \Delta_l$  which is the maximal number of bits available to encode frame  $l$  provided by a frame level rate control algorithm with average  $r_s/F$  and buffer related variable  $\Delta_l$ . Moreover  $D_l(R_l, P_{\text{MB}})$  is the estimated end-to-end MSE of frame  $l$ ,  $l = 1, 2, \dots, L$ , which can be obtained as

$$D_l(R_l, P_{\text{MB}}) = \sum_{i=1}^H \sum_{j=1}^V D(c_{i,j,l}, P_{\text{MB}}), \quad (18)$$

where  $D(c_{i,j,l}, P_{\text{MB}})$  is the end-to-end MSE of MB  $b_{i,j,l}$  using encoding vector  $c_{i,j,l}$  and  $D(c_{i,j,l}, P_{\text{MB}})$  can be computed from  $d_l^s$  as

$$D(c_{i,j,l}, P_{\text{MB}}) = \sum_{s \in b_{i,j,l}} d_l^s. \quad (19)$$

Since there is dependency between neighboring interframes because of the motion compensation, the optimal solution of (17) has to be searched over  $C^{H \times V \times L}$ , which is computationally prohibitive. We use greedy optimization algorithm, which is also implicitly used in most JSCC video coding methods such as [10, 11, 13], to find the coding modes and quantizers for MBs in frame  $l$  that minimize  $D_l(R_l, P_{\text{MB}})$ , then find coding modes and quantizers for MBs in frame  $l+1$  that minimize  $D_{l+1}(R_{l+1}, P_{\text{MB}})$  based on the previous optimized frame  $l$ , and so on. The optimal pair  $(r_s^*, r_c^*)$  and the corresponding optimal video coding scheme can be found such that

$$(r_s^*, r_c^*) = \underset{(r_s, r_c)}{\operatorname{argmin}} D_E^*(r_s, P_{\text{MB}}). \quad (20)$$

The goal now is to optimally select the quantizers and encoding modes on the MB level for a specific MB error rate  $P_{\text{MB}}$  and frame rate  $R_l^{\text{max}}$  to trade off the source coding efficiency and robustness to error. The notation of  $P_{\text{MB}}$  and  $(r_s, r_c)$  is dropped from now on unless needed.

The optimal coding problem for frame  $l$  can be stated as

$$\min_{C^{H \times V}} D_l = \sum_{i=1}^H \sum_{j=1}^V D(c_{i,j,l}) \quad (21)$$

subject to

$$R_l = \sum_{i=1}^H \sum_{j=1}^V R(c_{l,i,j}) \leq R_l^{\max}. \quad (22)$$

Such RD-optimized video coding schemes have been studied for noiseless and noisy channels recently [19, 20, 21, 22, 23, 24]. Using Lagrangian multiplier, we can solve the problem by minimizing

$$J_l(\lambda) = D_l + \lambda R_l, \quad (23)$$

where  $\lambda \geq 0$ . For video coding over error-prone channels, GOB coding structure is used for H.263 video coding over noisy channels with each GOB encoded independently. Therefore, if the transmission errors occur in one GOB, the errors will not propagate into other GOBs in the same video frame.

For video coding over noiseless channels, the independent GOB structure leads to the fact that the optimization of (23) can be performed for each GOB separately. However, when considering RD-optimized video coding for noisy channels, the MB distortion  $D_{i,j}(c_{i,j}, P_{MB})$  depends not only on the mode and quantizer of the current MB but also on the mode of the MB above to take into account error concealment distortion. Therefore, there is a dependency between neighboring GOBs for this optimization problem. We use greedy optimization algorithm again to find the solution by searching the optimal modes and quantizers from the first GOB to the last GOB in each frame.

## 4. SIMULATION RESULTS

We first use a simple two-state Markov chain model for simulation to show the performance of the integrated source and channel rate allocation and robust video coding scheme, where the given channel stochastic knowledge is accurate. Then simulations over Rayleigh fading channel is performed to verify the effectiveness of the proposed scheme for practical wireless channels.

### 4.1. Two-state Markov chain channel

Simulations have been performed using base mode H.263 to verify the accuracy of the proposed integrated scheme. In the simulations, the total channel signaling rate  $r$  equals 144 kbps, which is a typical rate provided in the 3G wireless systems. Video frame rate is  $F = 10$  f/s. The video sequence used for simulation is Foreman in QCIF format. RS code over  $GF(2^8)$  is used for FEC. The channel coding rate used are  $\{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$ . The source and channel coding rates  $r_s, r_c$  and the corresponding RS code  $(n, k)$  are listed in Table 1. A two-state Markov channel model [16] is used, where the state transition is at the RS symbol level. The two states of the model are denoted by  $G$  (good) and  $B$  (bad). In state  $G$ , the symbols are received correctly ( $e_g = 0$ ) whereas in state  $B$ , the symbols are erroneous ( $e_b = 1$ ). The model is fully described by the transition probabilities  $p$  from state  $G$  to state  $B$ , and  $q$  from state  $B$  to state  $G$ . We use the probabil-

ity of state  $B$ :

$$P_B = \frac{p}{p+q}, \quad (24)$$

and the average bursty length:

$$L_B = \frac{1}{q}, \quad (25)$$

which is the average number of consecutive symbol errors to model the two-state Markov model [11, 16].

The simulations are performed through the following steps.

- (i) For each channel coding rate  $r_c$  (or  $RS(n, k)$ ) in each column of Table 1, the RS code decoding failure rate  $p_w(n, k)$  is computed using (3) for a given two-state Markov channel model. The results for different  $r_c$  and channel models are shown in Table 2.
- (ii) The corresponding video MB error rate  $P_{MB}(r_c)$  is obtained using (5), where  $\alpha = 1.5$ .
- (iii) For each source rate  $r_s = r \times r_c$  and the corresponding  $P_{MB}(r_c)$ , the RD-optimized H.263 video coding is employed while estimating the end-to-end MSE  $D_E^*(r_s, r_c)$ .
- (iv) The H.263 bitstream is packetized and protected using  $RS(n, k)$ , and then transmitted over the two-state Markov channel model.
- (v) The receiver receives the bitstream, reconstructs the video sequence after the FEC decoding, and performs the H.263 decoding and possible error concealment if errors occur. The distortion for each simulation run between the original video sequence and the reconstructed video sequence at the receiver is also computed.

The average estimated PSNR,  $PSNR_E$ , of video signals is used to measure the performance:

$$PSNR_E(r_s, r_c) = \frac{1}{L} \sum_{l=1}^L PSNR_E^l(r_s, r_c), \quad (26)$$

where

$$PSNR_E^l(r_s, r_c) = 10 \log_{10} \frac{255^2}{D_E^*(r_s, r_c)} \quad (27)$$

is the estimated average PSNR between the original frame  $l$  and the corresponding reconstruction at the decoder using the pair  $(r_s, r_c)$ , and  $D_E^*(r_s, r_c)$  is the minimal estimated end-to-end MSE from (17) through RD-optimized video coding. The average PSNR of  $N$  runs of simulation is defined as

$$PSNR_S(r_s, r_c) = \frac{1}{N} \sum_{n=1}^N \frac{1}{L} \sum_{l=1}^L PSNR_S^{(n,l)}(r_s, r_c), \quad (28)$$

where  $PSNR_S^{(n,l)}(r_s, r_c)$  is the PSNR between the original

TABLE 1: The source and channel rates used in simulation.

$r$ (kbps)	144	144	144	144	144	144	144
$RS(n, k)$	(200, 40)	(200, 60)	(200, 80)	(200, 100)	(200, 120)	(200, 140)	(200, 160)
$r_c = k/n$	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$r_s = r \times r_c$ (kbps)	28.8	43.2	57.6	72.0	86.4	100.8	115.2

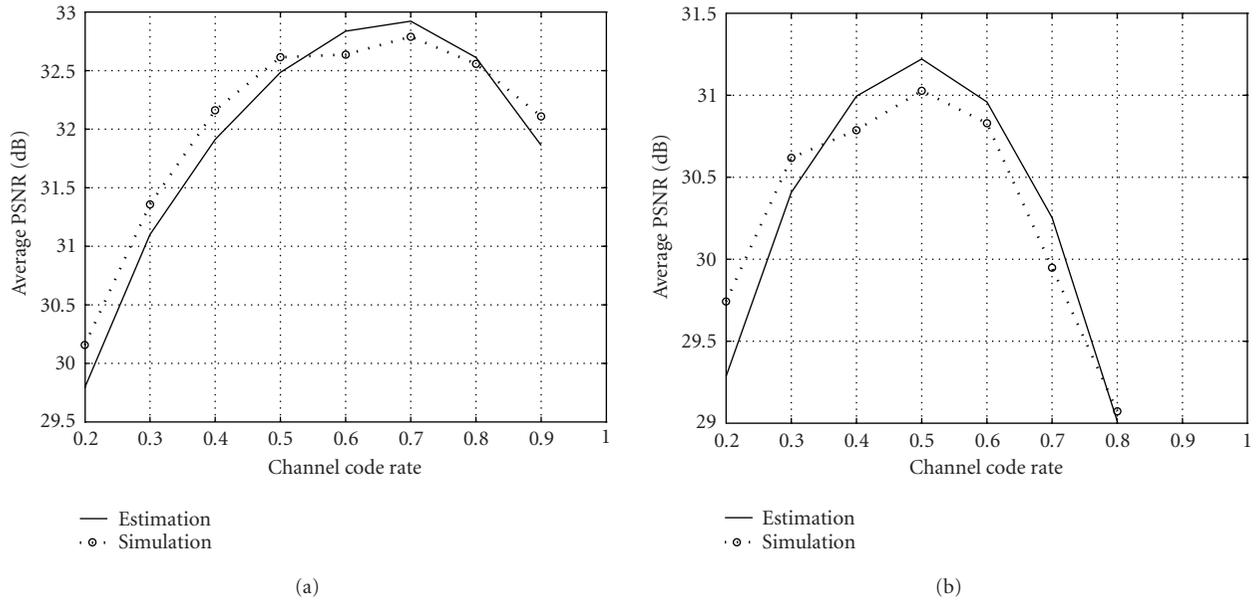


FIGURE 5: Average PSNR obtained by estimation versus simulation using two-state Markov model. (a) Symbol error rate = 0.01, where  $P_B = 0.01$  and  $L_B = 16$ . (b) Symbol error rate = 0.05, where  $P_B = 0.05$  and  $L_B = 16$ .

frame  $l$  and the corresponding reconstruction at the decoder in the  $n$ th simulation using the source/channel rate pair  $(r_s, r_c)$ .

Figure 5a shows the average estimated PSNR<sub>E</sub> of the optimal rate allocation and robust video coding for different channel code rates when the symbol error rate is  $P_B = 0.01$  and the bursty length  $L_B = 16$  symbols, and the corresponding average simulated PSNR<sub>S</sub> is of 50 times video transmission. Figure 5b also shows the same comparison when the symbol error rate is  $P_B = 0.05$ . It can be noted that the estimated PSNR<sub>E</sub>, which is obtained at the encoder during RD-optimized video encoding, matches the simulated PSNR<sub>S</sub> very well. The optimal source and channel rate pair can also be found through Figures 5a and 5b for different channel characteristics. The corresponding channel decoding failure rate of the optimal channel coding rates in Figures 5a and 5b are 0.018 and 0.034, respectively.

We also compare the performance when the knowledge of channel model used at video encoder does not match the real channel used in simulations. Figure 6 shows two cases of channel mismatch. In Figure 6a, the video stream, which is encoded based on  $P_B = 0.01$  and  $L_B = 16$  two-state Markov channel, is simulated using two-state Markov channel with  $P_B = 0.01$  and  $L_B = 8$ . The simulated average PSNR is bet-

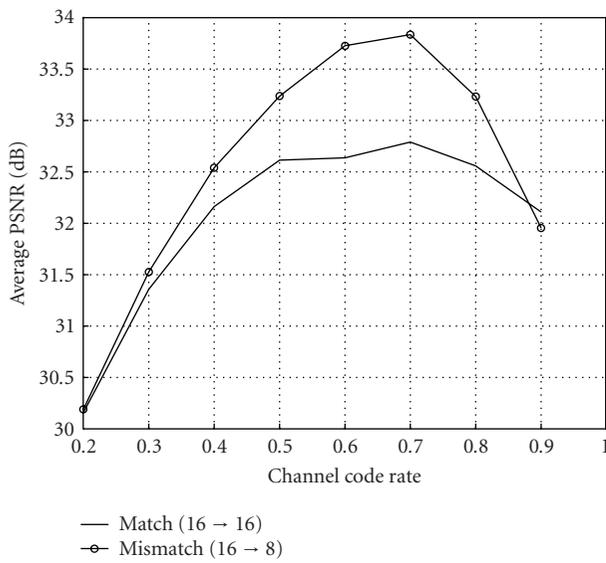
ter than the average PSNR estimated at the encoder during encoding because the channel model used in estimation is worse than the model used in simulation. On the other hand, when the video stream, which is encoded based on  $P_B = 0.01$  and  $L_B = 8$  two-state Markov channel, is simulated using two-state Markov channel with  $P_B = 0.01$  and  $L_B = 16$ , the simulated average PSNR is much worse than the average PSNR estimated at the encoder as shown in Figure 6b. Furthermore, the optimal source and channel coder pair obtained at the encoder is not optimal when the channel condition used in simulation is worse than the channel information used at the encoder. This simulation result suggests that the optimal rate allocation and video coding should be focused on the worse channel conditions for broadcasting services.

#### 4.2. Rayleigh fading channel

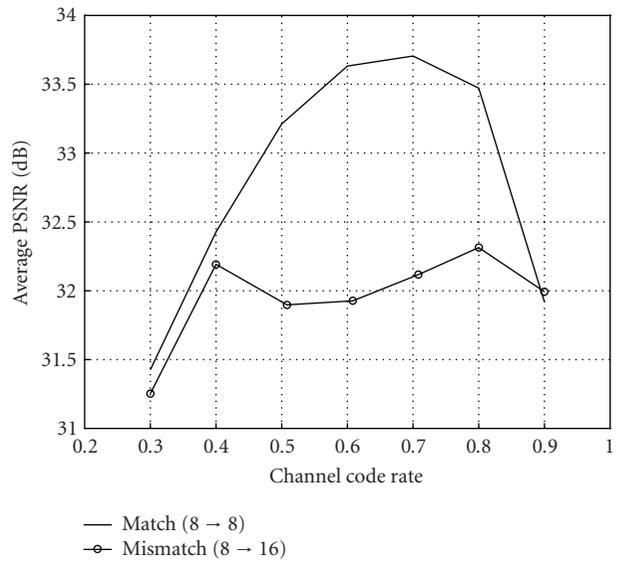
The simulation over the Rayleigh fading channel is also performed to verify the effectiveness of the proposed scheme over realistic wireless channels. In the simulation, QPSK with coherent demodulation is used for the sake of simplicity. The channel is a frequency-nonselective Rayleigh fading channel. An FSMC with  $K = 6$  states is used to model the Rayleigh fading channel. The SNR thresholds for the  $K$  states are

TABLE 2: The Reed-Solomon code decoding failure rate.

Rate	$P_B = 0.01, L_B = 16$	$P_B = 0.05, L_B = 16$	$P_B = 0.01, L_B = 8$	$P_B = 0.05, L_B = 8$
0.1	0.00028	0.00266	0.00000	0.00006
0.2	0.00058	0.00521	0.00001	0.00023
0.3	0.00117	0.00998	0.00003	0.00079
0.4	0.00233	0.01871	0.00011	0.00259
0.5	0.00462	0.03435	0.00042	0.00803
0.6	0.00909	0.06170	0.00155	0.02342
0.7	0.01776	0.10840	0.00559	0.06384
0.8	0.03445	0.18603	0.01976	0.16098
0.9	0.06635	0.31135	0.06829	0.36890



(a)



(b)

FIGURE 6: Average PSNR obtained in channel mismatch cases. (a) Error burst is shorter than that used in estimation. (b) Error burst is longer than that used in estimation.

selected in such a way that the probability that the channel gain is at state  $s_k, k = 0, 1, \dots, K - 1$ , is

$$\begin{aligned}
 p_0 &= \frac{2}{K(K+1)}, \\
 p_k &= kp_0, \quad k = 1, 2, \dots, K - 1.
 \end{aligned}
 \tag{29}$$

The FSMC state transition is described at the RS codeword symbol level (8-bit RS symbol) with the assumption that the four QPSK modulation symbols within an RS codeword symbol stay in the same FSMC state. Given the average SNR  $\rho$  and the Doppler frequency  $f_d$ , we can obtain the parameters such as steady state probability  $p_k$ , RS symbol error probability  $e_k$ , and state transition rates [17]. Then following the procedures described in Section 2.1, we are able to analyze the RS code performance over Rayleigh fading channels. Table 3

shows the estimated RS code decoding failure probability using FSMC model and the simulation values when the SNR is 18 dB and the Doppler frequency is 10 Hz and 100 Hz, respectively. The RS codeword error rate obtained by the FSMC matches the simulation results very well when  $f_d$  is 10 Hz. When  $f_d$  is 100 Hz, the FSMC-based estimate is not as accurate as the results when  $f_d$  is 10 Hz, but is still within acceptable range compared to the simulated values.

Figure 7a shows the average estimated  $PSNR_E$  and simulated  $PSNR_S$  of the video coding after optimal rate allocation and robust video coding for different channel code rates when the SNR is 18 dB and  $f_d$  is 10 Hz. Figure 7b also shows the comparison when the  $f_d$  is 100 Hz. Even though it can be noted that there are about 1 dB difference between the estimated  $PSNR_E$  and the simulated  $PSNR_S$ , the near-optimal source and channel rate allocation (or the channel code rate

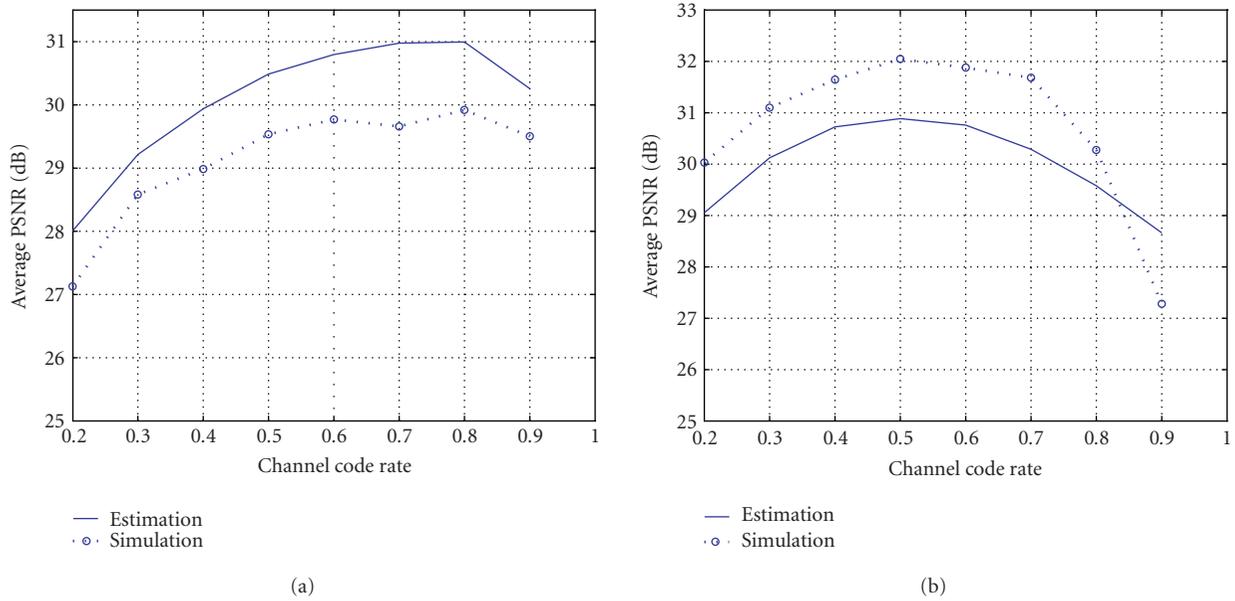


FIGURE 7: Average end-to-end PSNR obtained by estimation versus simulation for Rayleigh fading channels: (a) SNR = 18 dB,  $f_d = 10$  Hz, and (b) SNR = 18 dB,  $f_d = 100$  Hz.

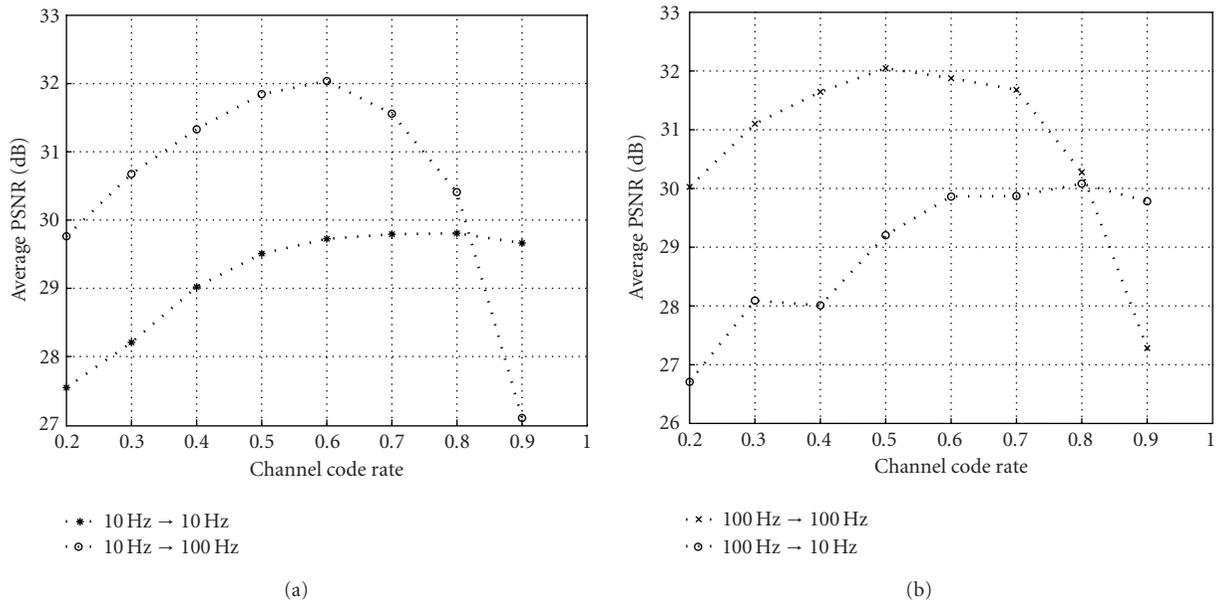


FIGURE 8: Average end-to-end PSNR over Rayleigh fading Channels: (a) SNR = 18 dB,  $f_d$  used for estimation at the encoder is 10 Hz, and (b) SNR = 18 dB,  $f_d$  used for estimation at the encoder is 100 Hz.

$r_c$ ) obtained from the estimation (0.8 and 0.5 as shown in Figure 7) still has the maximal simulated end-to-end PSNR over Rayleigh fading channels. The simulation results verify the effectiveness of the proposed scheme to obtain the optimal source and channel coding pair when given a fixed total bit rate for wireless fading channels.

Experiments are also performed when the knowledge of channel Doppler frequency used at the video encoder does not match the actual Doppler frequency used in simulations. Figure 8 shows two cases of channel mismatch. In Figure 8a, the video bitstream which is encoded based on  $f_d = 10$  Hz is simulated over fading channels with Doppler frequency

TABLE 3: Analysis and simulation values of the RS code decoding failure probability for the Rayleigh fading channel with SNR equals 18 dB and the Doppler frequency  $f_d$  equals 10 and 100 Hz.

Code rate	$f_d = 10$ Hz		$f_d = 100$ Hz	
	FSMC model	Simulation	FSMC model	Simulation
0.2	0.0430	0.0391	0.0098	0.0044
0.3	0.0482	0.0464	0.0170	0.0072
0.4	0.0536	0.0555	0.0282	0.0119
0.5	0.0593	0.0650	0.0450	0.0181
0.6	0.0653	0.0772	0.0692	0.0280
0.7	0.0717	0.0915	0.1032	0.0526
0.8	0.0815	0.1085	0.1499	0.1110
0.9	0.1240	0.1333	0.2131	0.2856

of 10 Hz and 100 Hz, separately. In Figure 8b, the video bitstream which is encoded based on  $f_d = 100$  Hz is simulated over fading channels with Doppler frequency of 100 Hz and 10 Hz, separately. In both scenarios, the video quality would be better if the actual condition in terms of MB loss rate is smaller than the knowledge used at the encoder, and would be worse otherwise. Furthermore, the optimal source and channel coder pair obtained at the encoder is not optimal when the channel condition used in simulation is worse than the channel information used at the encoder. This simulation result again suggests that the optimal rate allocation and video coding should be focused on the worse channel conditions for broadcasting services.

## 5. CONCLUSION

We have proposed an integrated framework to find the near-optimal source and channel rate allocation and the corresponding robust video coding scheme for video coding and transmission over wireless channels when there is no feedback channel available. Assuming that the encoder has the stochastic channel information when the wireless fading channel is modeled as an FSMC model, the proposed scheme takes into account the robust video coding, packetization, channel coding, error concealment, and error propagation effects altogether. This scheme can select the best source and channel coding pair to encode and transmit the video signals. Simulation results demonstrated the optimality of the rate allocation scheme and accuracy of end-to-end MSE estimation obtained at the encoder during the process of robust video encoding.

## REFERENCES

- [1] B. Girod and N. Farber, "Feedback-based error control for mobile video transmission," *Proceedings of the IEEE*, vol. 87, no. 10, pp. 1707–1723, 1999.
- [2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Communications. John Wiley & Sons, NY, USA, 1991.
- [3] J. Modestino and D. G. Daut, "Combined source-channel coding of images," *IEEE Trans. Communications*, vol. 27, pp. 1644–1659, November 1979.
- [4] N. Tanabe and N. Farvardin, "Subband image coding using entropy-coded quantization over noisy channels," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 5, pp. 926–943, 1992.
- [5] N. Farvardin and V. Vaishampayan, "On the performance and complexity of channel-optimized vector quantizers," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 155–160, 1991.
- [6] G. Cheung and A. Zakhor, "Bit allocation for joint source/channel coding of scalable video," *IEEE Trans. Image Processing*, vol. 9, no. 3, pp. 340–356, 2000.
- [7] M. Bystrom and J. W. Modestino, "Combined source-channel coding schemes for video transmission over an additive white Gaussian noise channel," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 880–890, 2000.
- [8] R. E. V. Dyck and D. J. Miller, "Transport of wireless video using separate, concatenated, and joint source-channel coding," *Proceedings of the IEEE*, vol. 87, no. 10, pp. 1734–1750, 1999.
- [9] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Processing Magazine*, vol. 17, no. 4, pp. 61–82, 2000.
- [10] M. Gallant and F. Kossentini, "Rate-distortion optimized layered coding with unequal error protection for robust internet video," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 357–372, 2001.
- [11] K. Stuhlmüller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1012–1032, 2000.
- [12] 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 and Systems for Video Technology*, vol. 12, no. 6, pp. 511–523, 2002.
- [13] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 966–976, 2000.
- [14] ITU-T, "H.263, video coding for low bitrate communication," February 1998.
- [15] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell System Technical Journal*, vol. 39, no. 5, pp. 1253–1265, 1960.
- [16] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell System Technical Journal*, vol. 42, pp. 1977–1997, September 1963.
- [17] H. Wang and N. Moayeri, "Finite-state Markov channel—a useful model for radio communication channels," *IEEE Trans. Vehicular Technology*, vol. 44, no. 1, pp. 163–171, 1995.
- [18] J. Lu, K. B. Letaief, and M. L. Liou, "Robust video transmission over correlated mobile fading channels," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 5, pp. 737–751, 1999.
- [19] G. Cote, S. Shirani, and F. Kossentini, "Robust H.263 video communication over mobile channels," in *IEEE Proc. of International Conference on Image Processing*, pp. 535–539, Kobe, Japan, October 1999.
- [20] G. M. Schuster and A. K. Katsaggelos, "Fast and efficient mode and quantizer selection in the rate distortion sense for H.263," in *Proc. SPIE: Visual Communications and Image Processing*, vol. 2727, pp. 784–795, Orlando, USA, February 1996.
- [21] H. Sun, W. Kwok, M. Chien, and C. H. J. John, "MPEG coding performance improvement by jointly optimizing coding mode decisions and rate control," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, no. 3, pp. 449–458, 1997.
- [22] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 23–50, 1998.

- [23] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 74–90, 1998.
- [24] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. Mitra, "Rate-distortion optimized mode selection for very low bit-rate video coding and the emerging H.263 standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 6, no. 2, pp. 182–190, 1996.

**Jie Song** received his B.S. degree from Beijing University, Beijing, China, in 1990, his M.S. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 1993, and his Ph.D. degree from the University of Maryland, College Park, Md, in 2000, all in electrical engineering. From April 1993 to June 1996, he was a Lecturer and a Researcher in the Information Engineering Department at Beijing University of Posts and Telecommunications. He worked for Fujitsu Labs of America, Calif, in the summer of 1997. From November 1997 to February 1999, he was a part-time Consultant of multimedia technologies in Odyssey Technologies Inc., Jessup, Md, where he was involved in the projects of H.323/H.324 videophone, portable multimedia terminal design, and multichannel video capturing systems. Since August 2000, he has been working on research, design, and implementation for broadband and satellite communication systems at Agere Systems (formerly Microelectronics Group, Lucent Technologies). His research interests include signal processing for digital communication and multimedia communications.



**K. J. Ray Liu** received the B.S. degree from the National Taiwan University in 1983, and the Ph.D. degree from UCLA in 1990, both in electrical engineering. He is a Professor at the Electrical and Computer Engineering Department and Institute for Systems Research of the University of Maryland, College Park. His research contributions encompass broad aspects of signal processing algorithms and architectures; multimedia communications and signal processing; wireless communications and networking; information security; and bioinformatics, in which he has published over 300 refereed papers. Dr. Liu is the recipient of numerous honors and awards including IEEE Signal Processing Society 2004 Distinguished Lecturer, the 1994 National Science Foundation Young Investigator Award, the IEEE Signal Processing Society's 1993 Senior Award (Best Paper Award), IEEE 50th Vehicular Technology Conference Best Paper Award, Amsterdam, 1999. He also received the George Corcoran Award in 1994 for outstanding contributions to electrical engineering education and the Outstanding Systems Engineering Faculty Award in 1996 in recognition of outstanding contributions in interdisciplinary research, both from the University of Maryland. Dr. Liu is a Fellow of IEEE. Dr. Liu is the Editor-in-Chief of IEEE Signal Processing Magazine and was the founding Editor-in-Chief of EURASIP Journal on Applied Signal Processing. Dr. Liu is a Board of Governor and has served as Chairman of Multimedia Signal Processing Technical Committee of IEEE Signal Processing Society.

