

# Accurate Distortion Estimation and Optimal Bandwidth Allocation for Scalable H.264 Video Transmission Over MIMO Systems

Mohammad K. Jubran, Manu Bansal, Lisimachos P. Kondi, *Member, IEEE*, and Rohan Grover

**Abstract**—In this paper, we propose an optimal strategy for the transmission of scalable video over packet-based multiple-input multiple-output (MIMO) systems. The scalable extension of H.264/AVC that provides a combined temporal, quality and spatial scalability is used. For given channel conditions, we develop a method for the estimation of the distortion of the received video and propose different error concealment schemes. We show the accuracy of our distortion estimation algorithm in comparison with simulated wireless video transmission with packet errors. In the proposed MIMO system, we employ orthogonal space-time block codes (O-STBC) that guarantee independent transmission of different symbols within the block code. In the proposed constrained bandwidth allocation framework, we use the estimated end-to-end decoder distortion to optimally select the application layer parameters, i.e., quantization parameter (QP) and group of pictures (GOP) size, and physical layer parameters, i.e., rate-compatible turbo (RCPT) code rate and symbol constellation. Results show the substantial performance gain by using different symbol constellations across the scalable layers as compared to a fixed constellation.

**Index Terms**—Distortion estimation, MIMO systems, optimal bandwidth allocation, Scalable H.264, wireless video.

## I. INTRODUCTION

THE most recent third generation (3G) and upcoming fourth generation (4G) wireless technologies have made it possible to deliver multimedia services over high data rates. The latest H.264/AVC standard has already shown to provide superior compression efficiency and error-resilient transmission over varied networks [1]–[3], and the very recently proposed scalable extension of H.264/AVC, popularly known as SVC, inherits its error-resilient network adaptation layer (NAL) structure and provides layers of different importance depending on their contribution in the reconstructed video [4]–[7]. The scalability can be exploited to improve the video transmission over error-prone wireless networks by protecting the different

layers with unequal error protection (UEP). The UEP for the scalable data can be provided by using forward error correction (FEC) combined with an appropriate modulation technique. In many publications, it has been shown that jointly optimizing source and channel coding parameters for video transmission could improve the overall system performance [8]–[14].

Lots of research has been done in the field of scalable video coding, among the latest of which is the SVC, first proposed in October 2004 [4]–[7]. This codec uses and extends the NAL unit concept of the H.264/AVC standard, and also provides a base layer compatible with it. The most important feature of this codec is a combined scalability in the form of temporal scalability using a hierarchical prediction structure, fine granular quality scalability using progressive refinement slices and spatial scalability using interlayer prediction mechanisms. In this paper, we consider both temporal scalability and fine granular scalability (FGS); however, the same work can be extended to also include the spatial scalability. We consider the efficient transmission of these scalable layers over packet-based wireless networks, with optimization of both source and channel coding parameters. For that to be possible, a good knowledge of the total end-to-end decoder distortion should be available at the encoder. Various decoder distortion estimation algorithms have been proposed in the literature. In [15], a per-pixel based decoder distortion estimation algorithm, ROPE was proposed for the H.263+ codec. Using the ROPE algorithm, the first and second moments of the pixel values, which depend on packet loss probabilities, are recursively obtained to calculate the decoder distortion, which is further used for optimal (source coding) mode selection for a given target rate. In [16], the authors extend the ROPE algorithm to estimate the decoder distortion at the encoder for the scalable H.263+ codec for only SNR scalability. Also, in [17], the ROPE algorithm is further modified for different re-synchronization schemes for the transmission of non-scalable H.263 coded video over tandem channels. In [18], the mean as well as the variance of the end-to-end distortion is considered when allocating limited source and channel resources. In our paper, we develop a method for the accurate estimation of the SVC video distortion at the receiver for given channel conditions and also propose different error concealment schemes to handle packet losses for the SVC decoder. Our scalable decoder distortion estimation (SDDE) algorithm takes into account loss of both temporal (due to the hierarchical structure) and SNR scalable layers as well as error concealment at the decoder. We compare the performance of our distortion estimation algorithm with simulated video

Manuscript received February 20, 2007; revised September 01, 2008. First published November 18, 2008; current version published December 12, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Soheil Dianat.

M. K. Jubran is with the Department of Electrical Engineering, Birzeit University, Birzeit, Palestine (e-mail: mjubran@birzeit.edu).

M. Bansal is with the Intellectual Property Group, Goodwin Procter LLP, Boston, MA 02109 USA (e-mail: mbansal@goodwinprocter.com)

L. P. Kondi is with the Department of Computer Science, University of Ioannina, Ioannina 45110 Greece (e-mail: lkoni@cs.uoi.gr).

R. Grover is with the Radiospire Networks, Hudson, MA 01749 USA (e-mail: rgrover@radiospire.com)

Digital Object Identifier 10.1109/TIP.2008.2006600

transmission over wireless channels with packet errors. To the best knowledge of the authors, no algorithms for the estimation of the decoder distortion for video codecs based on hierarchical prediction structure and providing combined temporal and SNR scalability (such as SVC), have been proposed in the literature.

Any wireless transmission system suffers from environmental noise, fading and the bandlimited nature of the channel. Diversity techniques, including spatial, time and frequency domain diversity, have been suggested to help overcome these degradations by providing the receiver with multiple replicas of the transmitted signal over different channels [19]. As one of the diversity techniques, space-time coding (STC) over multiple antenna systems has been studied extensively [20]–[22]. It integrates antenna diversity with coding techniques to achieve a higher capacity and reduce co-channel interference in multiple access. Space-time block codes (STBC), which were first proposed by Alamouti [20] and later generalized by Tarokh *et al.* [22], are one of the STC techniques for broadband wireless communications. These codes exploit the orthogonality property of the code matrix to achieve the full diversity gain and have the advantage of having a low complexity maximum-likelihood (ML) decoder. We employ these codes in the proposed video transmission scheme using a multiple-input multiple-output (MIMO) system. The orthogonal STBC (O-STBC) used here guarantee independent transmission and decoding of each symbol in a given block code. This enables us to independently choose the elements of the codeword from different constellations (modulation schemes), and, hence, an additional unequal error protection can be provided for each element of the codeword. We consider two MIMO systems, one with multiple transmit and multiple receive antennas ( $2 \times 2$ ) and the other with multiple transmit and a single receive antenna ( $4 \times 1$ ). Both of these systems have the same diversity gain but the latter is more practical in case of space and power limitations at the receive (mobile) terminals.

In only a few publications such as [23] and [24], wireless video transmission using STC has been studied. In [23], an integrated system of data-partitioned video coding, layered space-time block coding, OFDM modulation and unequal error protection is proposed. It is shown that unequal error protection facilitates the interference cancellation and enhances the quality of reconstructed video, but no optimization for resource allocation is addressed. In [24], progressive video transmission over space-time differentially coded OFDM system is proposed with an UEP structure for optimal rate and power allocation among multiple layers. However, in all the above mentioned work, the orthogonal structure of STBC codes has not been exploited by independent transmission of the layered video over different symbols of the STBC code modulated with different constellations. In [25], an approach for using the SVC with unequal erasure protection (UXP) over a wireless IP networks is proposed. Temporal scalability is considered and their UXP scheme shows the advantage of protecting the layered video unequally over the single layer protection. In [26], the authors propose an SNR scalable version of H.264/AVC and use a hierarchical quadrature amplitude modulation (HQAM) technique for unequal protection of the scalable layers.

In this paper, we propose a system that integrates video coding with combined scalability, FEC through unequal channel coding, modulation schemes and spatial diversity for wireless video transmission. Temporal and quality scalable layers are obtained using the SVC and are unequally protected using rate-compatible punctured turbo (RCPT) [27] codes with cyclic redundancy check (CRC) [28] error detection. The channel coded layers are then modulated and encoded using O-STBC for transmission over multiple antennas. We address the problem of minimization of the expected end-to-end distortion by optimally selecting source coding parameters: the quantization parameter (QP) and the group of picture (GOP) size, and the channel coding and the physical layer parameters: RCPT channel coding rate and the symbol constellation choice for the MIMO transmission. The optimization is constrained on the total available bandwidth for transmission. The accurate estimation of the decoder distortion using the SDDE algorithm plays a key role in this optimization problem.

The rest of the paper is organized as follows. Section II discusses the coding structure of SVC and explains in detail the proposed estimation algorithm for decoder distortion. Also in the same section, we propose different error concealment schemes. In Section III, we describe the MIMO systems used in this work. In Section IV, we define and address the optimal bandwidth allocation problem. Experimental results are discussed in Section V. Finally, conclusions are drawn in Section VI.

## II. SCALABLE H.264 CODEC AND DECODER DISTORTION ESTIMATION

### A. Scalable Extension of H.264/AVC (SVC)

SVC is based on a hierarchical prediction structure and has base layer compatibility with H.264/AVC. It provides three types of scalability: spatial, temporal and quality. In each of the spatial scalable layers, a hierarchical prediction structure is used to code a GOP. The hierarchical coding structure that provides temporal scalability is shown in Fig. 1. The first picture of the video sequence is always intracoded and is called a key picture. A GOP consists of a key picture and all other pictures temporally located between the key picture and the previously encoded key picture. The key pictures are either encoded in intra or inter mode, with only previously encoded key pictures as the reference pictures. These key pictures collectively form the lowest temporal resolution of the video sequence and are called temporal level zero (TL0). The other pictures encoded in each GOP define different temporal levels (TL1, TL2, so on) and always use the pictures from the lower temporal levels as reference pictures. Each of these pictures is represented by a non-scalable base layer that includes the corresponding motion and an approximation of the intra and residual data, and zero or more quality scalable enhancement representations (or layers). These quality scalable layers are produced in a fine-granular scalability (FGS) manner by using finer quantization step sizes and finding the residual between the original base layer and the reconstructed base representation. From here on in this paper, the base and enhancement layers will be referred to as the SNR

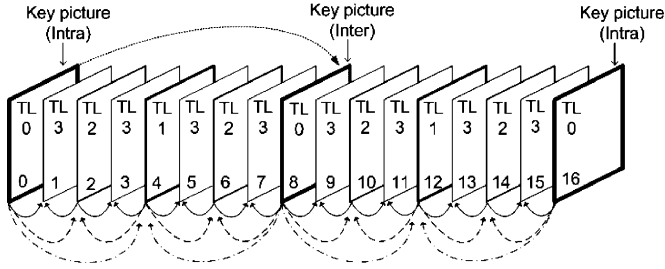


Fig. 1. Hierarchical prediction structure for SVC for a GOP size of 8.

scalable layers, where the base layer of each frame is associated with a particular temporal level. In our work, we limit ourselves to temporal and SNR scalability such that priority for the base layer (FGS0) of each temporal level increases from the lowest to the highest temporal level, i.e., base layer packets from TL0 are considered more important than those from TL1 and so on. And then each FGS layer for all the frames is considered as a single layer, again in the decreasing order of importance such that FGS1 is more important than FGS2.

We consider the transmission of the layers defined above over packet-based wireless networks. In any typical wireless network, packet losses are unavoidable because of noisy channels and congestion. These losses translate into degraded video quality at the receiver and result in video distortion due to the channel. This is in addition to the deterministic distortion due to source coding, which is the result of quantization. Both of these distortions define the total end-to-end distortion of the transmitted video sequence at the decoder.

In our proposed system, for efficient transmission of video over wireless networks, we have to optimally select certain system parameters at the transmitter. To do so, at the encoder we should have a good knowledge of the total decoded video distortion. For this purpose, we have developed a method for the accurate estimation of the video distortion at the receiver for given channel conditions. This method, as discussed next, takes into account loss of temporal and quality (SNR) scalable layers as well as error concealment at the decoder.

## B. Error Concealment

In our video transmission system, each layer of each frame is packetized into constant size packets, which are transmitted over a lossy wireless network. At the receiver, any unrecoverable errors in each packet would result in dropping (erasing) of that packet and, hence, would mean loss of the layer (of that particular frame) to which the packet belongs. In this system, we assume that the base layer of all the key pictures are received error free. It should be emphasized here that the SVC encoding and decoding are done on a GOP basis (using the hierarchical structure), which makes it possible to use the frames within a GOP for error concealment purposes. In the event of losing a frame, temporal error concealment at the decoder is applied as follows.

- Scheme 1: The lost frame is replaced by the previous frame in the decreasing sequential order, e.g., in a GOP if frame  $f_n$  is lost, it is concealed using frame  $f_{n-1}$  given that it was

received error free, otherwise  $f_{n-2}$  is used for concealment and so on, till the start of GOP is reached.

- Scheme 2: The lost frame is replaced by the nearest available frame in the decreasing as well as increasing sequential order. We start towards the GOP end closer to the frame being concealed, e.g., in a GOP of eight frames, if frame  $f_6$  is lost, the order in which the frames of this GOP are used for concealment is  $f_7$ ,  $f_5$ , and finally  $f_8$ .
- Scheme 3: The lost frame is replaced by the previous reconstructed frame in the sequential order that leads faster towards the GOP end, and only using the frames from lower or same temporal levels (defined according to the hierarchical structure), e.g., in a GOP of eight frames, if frame  $f_5$  is lost, the order in which the frames are used for concealment is  $f_6$  and then  $f_8$ . However, if frame  $f_3$  is lost the order in which the frames are used for concealment is  $f_2$  and  $f_0$ . For the frame in the center of the GOP (like  $f_4$ ), the key picture at the start of the GOP is used for concealment.
- Scheme 4: The lost frame is replaced by the nearest available frame in the decreasing as well as increasing sequential order from only lower or same temporal levels. We start towards the frame that has a temporal level closer to the temporal level of the lost frame, e.g., in a GOP of eight frames, if frame  $f_6$  is lost, the order in which the frames are used for concealment is  $f_4$  and then  $f_8$ . As in Scheme 3, the center picture of the GOP is concealed using the starting key picture.

In both Scheme 1 and Scheme 2, the frame to-be-concealed and the frame used for concealment (concealing frame) could be consecutive frames, and, hence, we could assume that the texture and the motion information do not change significantly. However, for both the schemes, the to-be-concealed frame might also be a reference frame to the concealing frame. This implies that the loss of the frame to-be-concealed might affect the decoding of the concealing frame. On the other hand, for both Scheme 3 and Scheme 4, the to-be-concealed frame will never be a reference frame to the frames used for concealment, because the frame used for concealment is always chosen from the lower temporal levels. Hence, the loss of the frame to-be-concealed will not affect the decoding of the concealing frame. Also, according to the decoding order, the concealing frame will already be decoded by the time the decoder starts decoding the frame to-be-concealed, and, thus, for the concealment process, the decoder does not have to wait for the complete GOP to be decoded. However, because of the hierarchical structure of the codec, the to-be-concealed and the concealing frames will never be the consecutive frames and they will be separated by at least one frame.

## C. SDDE Derivation

For the SDDE algorithm, we will consider a base layer and two FGS layers. The same algorithm can simply be generalized to any number of FGS layers. Let  $f_n^i$  denote the original value of pixel  $i$  in frame  $n$  and  $\hat{f}_n^i$  denote its encoder reconstruction. The reconstructed pixel value at the decoder is denoted by  $\tilde{f}_n^i$ . The mean square error for this pixel is

$$\begin{aligned} d_n^i &= \mathbb{E} \left\{ \left( f_n^i - \tilde{f}_n^i \right)^2 \right\} \\ &= \left( f_n^i \right)^2 - 2f_n^i \mathbb{E} \left\{ \tilde{f}_n^i \right\} + \mathbb{E} \left\{ \left( \tilde{f}_n^i \right)^2 \right\} \end{aligned} \quad (1)$$

where  $d_n^i$  is the distortion per pixel. To calculate  $d_n^i$  we need to get both the first moment,  $\mathbb{E}\{\tilde{f}_n^i\}$  and the second moment,  $\mathbb{E}\{(\tilde{f}_n^i)^2\}$ . As mentioned earlier, the base layer of all the key pictures are guaranteed to be received error free. The first and the second moment of the pixels of the key pictures are as follows:

$$\begin{aligned} \mathbb{E} \left\{ \left( \tilde{f}_n^i \right) \right\} &= P_{nE1} \left( \hat{f}_{nB}^i \right) \\ &\quad + (1 - P_{nE1}) P_{nE2} \left( \hat{f}_{n(B+E1)}^i \right) \\ &\quad + (1 - P_{nE1})(1 - P_{nE2}) \left( \hat{f}_{n(B+E1+E2)}^i \right) \end{aligned} \quad (2)$$

$$\begin{aligned} \mathbb{E} \left\{ \left( \tilde{f}_n^i \right)^2 \right\} &= P_{nE1} \left( \hat{f}_{nB}^i \right)^2 \\ &\quad + (1 - P_{nE1}) P_{nE2} \left( \hat{f}_{n(B+E1)}^i \right)^2 \\ &\quad + (1 - P_{nE1})(1 - P_{nE2}) \left( \hat{f}_{n(B+E1+E2)}^i \right)^2 \end{aligned} \quad (3)$$

where  $\hat{f}_{nB}^i, \hat{f}_{n(B+E1)}^i, \hat{f}_{n(B+E1+E2)}^i$  are the reconstructed pixel values at the encoder of only the base layer, the base along with the first FGS layer and the base layer with both of the FGS layers of frame  $n$ , respectively.  $P_{nE1}$  and  $P_{nE2}$  are the probability of losing the first and the second FGS layer of frame  $n$ , respectively.

For all the frames except the key pictures of a GOP, let us denote  $\hat{f}_{nB-u,v}^i$  as the  $i$ th pixel value of the base layer of frame  $n$  reconstructed at the encoder. Frames  $u (< n)$  and  $v (> n)$  are the reference pictures used in the hierarchical prediction structure for the reconstruction of frame  $n$ . In the decoding process of SVC, the frames of each GOP are decoded in the order starting from the lowest to the highest temporal level. At the decoder.

- If frame  $u$  is not available as the reference picture for frame  $n$  (where frame  $n$  does not belong to the highest temporal level), then frame  $u'$  is selected as the new reference picture such that  $u' < n$  and  $\text{TL}(u') \leq \text{TL}(n)$  where  $\text{TL}(\cdot)$  is the temporal level to which the corresponding frame belongs. For the frames in the highest temporal level,  $u' < n$  and  $\text{TL}(u')$  is strictly less than  $\text{TL}(n)$ . Let us define  $\mathbf{L}_n$  as the set consisting of frame  $u$  and all the possible choices of  $u'$  for frame  $n$ . In other words, if  $f_n$  does not belong to the highest temporal level,  $\mathbf{L}_n$  is defined as

$$\mathbf{L}_n = \{ \hat{f}_{n-j} : \text{TL}(\hat{f}_{n-j}) \leq \text{TL}(\hat{f}_n) \} \quad (4)$$

otherwise

$$\mathbf{L}_n = \{ \hat{f}_{n-j} : \text{TL}(\hat{f}_{n-j}) < \text{TL}(\hat{f}_n) \} \quad (5)$$

$$j = 1, 2, \dots, n.$$

- If frame  $v$  is not available as the reference picture for frame  $n$ , then frame  $v'$  is selected as the new reference picture such that  $v' > n$  and  $\text{TL}(v') < \text{TL}(n)$ . In this case, we define  $\mathbf{R}_n$  as the set consisting of frame  $v$  and all the possible choices of  $v'$  for frame  $n$ . So, it is defined as

$$\mathbf{R}_n = \{ \hat{f}_{n+j} : \text{TL}(\hat{f}_{n+j}) < \text{TL}(\hat{f}_n) \} \quad (6)$$

$$j = 1, 2, \dots, N - n + 1$$

where  $N$  is the GOP size.

The  $s$ th moment of the  $i$ th pixel of frame  $n$  when at least the base layer is received at the decoder correctly,  $\mathbb{E}\{(\tilde{f}_n^i(\mathbf{L}_n, \mathbf{R}_n))^s\}$  is defined as

$$\begin{aligned} &\mathbb{E} \left\{ \left( \tilde{f}_n^i(\mathbf{L}_n, \mathbf{R}_n) \right)^s \right\} \\ &= \sum_{j=1}^{|\mathbf{L}_n|} \sum_{k=1}^{|\mathbf{R}_n|} (1 - P_{\mathbf{L}_n(j)}) \\ &\quad \times (1 - P_{\mathbf{R}_n(k)}) \times \prod_{c=1}^{j-1} P_{\mathbf{L}_n(c)} \prod_{d=1}^{k-1} P_{\mathbf{R}_n(d)} \mathbb{E} \\ &\quad \times \left\{ \left( \hat{f}_{n-\mathbf{L}_n(j)\mathbf{R}_n(k)}^i \right)^s \right\} \end{aligned} \quad (7)$$

where

$$\begin{aligned} &\mathbb{E} \left\{ \left( \hat{f}_{n-\mathbf{L}_n(j)\mathbf{R}_n(k)}^i \right)^s \right\} \\ &= P_{nE1} \left( \hat{f}_{nB-\mathbf{L}_n(j)\mathbf{R}_n(k)}^i \right)^s \\ &\quad + P_{nE2} (1 - P_{nE1}) \left( \hat{f}_{n(B+E1)-\mathbf{L}_n(j)\mathbf{R}_n(k)}^i \right)^s \\ &\quad + (1 - P_{nE2})(1 - P_{nE1}) \\ &\quad \times \left( \hat{f}_{n(B+E1+E2)-\mathbf{L}_n(j)\mathbf{R}_n(k)}^i \right)^s \end{aligned} \quad (8)$$

where  $P_{\mathbf{L}_n(j)}$  and  $P_{\mathbf{R}_n(k)}$  are the probabilities of losing the base layer of the reference frames  $j$  and  $k$  from the sets  $\mathbf{L}_n$  and  $\mathbf{R}_n$ , respectively.

Now to get the distortion per-pixel after error concealment, we will define  $\mathbf{Q}$ , a set of frames the first member of which is frame  $n$  whose distortion is to be calculated and the rest of the frames in this set define the sequence of frames used for the concealment of  $f_n$ . This sequence is decided according to one of the error concealment schemes described above. So,  $\mathbf{Q} = \{f_n, f_{q1}, f_{q2}, f_{q3}, \dots, f_{\text{GOPend}}\}$  where  $f_{q1}$  is the first frame,  $f_{q2}$  is the second frame to be used for concealment, and so on till one of the GOP ends is reached. The  $s$ th moment of the  $i$ th pixel using the set  $\mathbf{Q}$  is defined as  $\mathbb{E}\{(\tilde{f}_n^i)^s\}$  [see (9), shown at the bottom of the next page], where  $P_n$  and  $P_{qz}$  are the probabilities of losing the base layer of frame  $n$  and  $qz$ , respectively;  $\mathbf{L}_n \setminus \{f_w\}$  is the set of all the reference frames  $\mathbf{L}_n$  excluding frame  $f_w$ , and  $\mathbf{R}_n \setminus \{f_w\}$  is the set of all the reference frames  $\mathbf{R}_n$

excluding frame  $f_w$ . The SDDE algorithm (for each GOP) is summarized as follows.

- 1) The reconstructed video pixel values are obtained using the SVC encoder for a specific QP value and GOP size of interest.
- 2) For the given values of  $P_{nE1}$  and  $P_{nE2}$ , the first and second moments of the reconstructed pixel values at the decoder of the key pictures are calculated as in (2) and (3), respectively.
- 3) The set of frames  $\mathbf{Q}$  is defined based on the error concealment scheme.
- 4) For the given set of probabilities of losing each of the layer in a GOP, the first and second moment of the reconstructed pixel values at the decoder of the nonkey pictures are calculated based on (7)–(9).
- 5) Having the first and the second moments of the pixel values, the distortion of each pixel (in the mean square sense (MSE)) is calculated using (1). The MSE of all the frames is obtained by averaging the corresponding calculated pixels' distortion.

#### D. Performance Analysis

The algorithm explained above is implemented by modifying the SVC codec and its performance is evaluated by comparing it with the actual decoder distortion averaged over 200 channel realizations. Different video sequences (QCIF format) encoded at different rates are used in packet-based video transmission simulations. Figs. 2 and 3 compare the actual decoder distortion with the SDDE algorithm results for different error concealment schemes proposed above. In these figures, the results are shown for the “Foreman” video sequence encoded at 30 fps, QP = 40 and GOP size of eight frames (four temporal levels and two FGS enhancement layers), which resulted in the source coding rate of 195 kbps. Each of these layers is considered to be affected with different loss rates as follows:  $P_{TL0} = 0\%$ ,  $P_{TL1} = 10\%$ ,  $P_{TL2} = 20\%$ ,  $P_{TL3} = 30\%$ ,  $P_{E1} = 50\%$ , and  $P_{E2} = 60\%$  where  $P_{TLx}$  is the probability of losing the base layer of a frame that belongs to  $TLx$  and  $P_{E1}$  and  $P_{E2}$  are the probabilities of losing FGS1 and FGS2 of each frame, respectively. It is evident that the proposed SDDE algorithm provides an accurate estimate of the SVC decoder distortion at the encoder. Similar results are also shown in Figs. 4 and 5 using the “Akiyo” video sequence encoded at 30 fps, QP = 25, GOP = 16 and the resultant source coding rate equal to 316 kbps. The packet error rates considered in this case are  $P_{TL0} = 0\%$ ,  $P_{TL1} =$

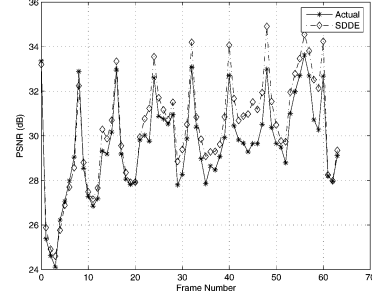


Fig. 2. Comparison between the actual and estimated decoder PSNR, “Foreman” sequence using error concealment scheme 3.

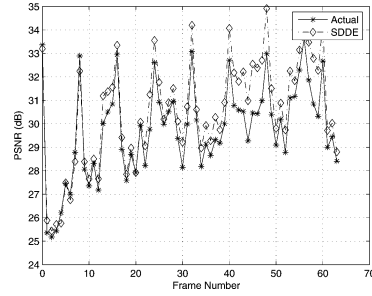


Fig. 3. Comparison between the actual and estimated decoder PSNR, “Foreman” sequence using error concealment scheme 4.

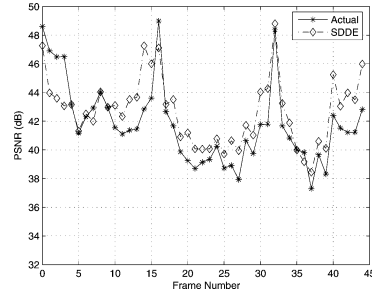


Fig. 4. Comparison between the actual and estimated decoder PSNR, “Akiyo” sequence using error concealment scheme 1.

10%,  $P_{TL2} = 15\%$ ,  $P_{TL3} = 20\%$ ,  $P_{TL4} = 25\%$ ,  $P_{E1} = 30\%$ , and  $P_{E2} = 40\%$ . Also, comparison of average PSNRs (160 frames) for both the actual and SDDE are listed in Table I. It can be observed that error concealment scheme 2 provides better PSNR values than those obtained by using other schemes, and, hence, for the rest of the paper, we will use scheme 2 for error concealment at the decoder.

$$\begin{aligned}
 E \left\{ \left( \tilde{f}_n^i \right)^s \right\} &= (1 - P_n) E \left\{ \left( \tilde{f}_n^i(\mathbf{L}_n, \mathbf{R}_n) \right)^s \right\} \\
 &+ P_n (1 - P_{q1}) E \left\{ \left( \tilde{f}_{q1}^i(\mathbf{L}_{q1} \setminus \{f_n\}, \mathbf{R}_{q1} \setminus \{f_n\}) \right)^s \right\} \\
 &+ P_n P_{q1} (1 - P_{q2}) E \left\{ \left( \tilde{f}_{q2}^i(\mathbf{L}_{q2} \setminus \{f_n, f_{q1}\}, \mathbf{R}_{q2} \setminus \{f_n, f_{q1}\}) \right)^s \right\} + \dots \\
 &+ P_n \prod_{z=1}^{|\mathbf{Q}|-2} P_{qz} E \left\{ \left( \tilde{f}_{\text{GOPend}}^i \right)^s \right\}
 \end{aligned} \tag{9}$$

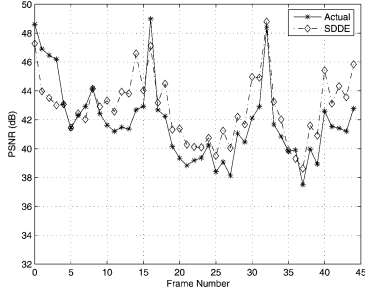


Fig. 5. Comparison between the actual and estimated decoder PSNR, “Akiyo” sequence using error concealment scheme 2.

TABLE I  
ACTUAL AND SDDE AVERAGE PSNR VALUES (DB)

Error Concealment	Foreman Actual	Foreman SDDE	Akiyo Actual	Akiyo SDDE
Scheme 1	30.27	31.12	42.22	43.68
Scheme 2	30.73	31.39	42.95	43.80
Scheme 3	29.64	30.34	41.65	42.63
Scheme 4	29.78	30.56	41.76	43.14

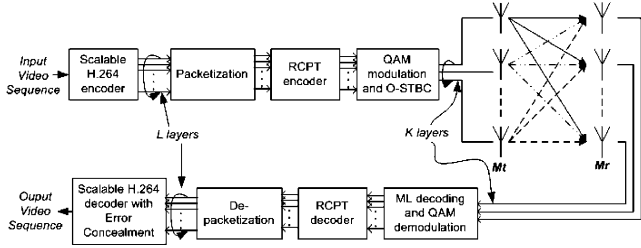


Fig. 6. Block diagram for the SVC coded Video Transmission over MIMO systems.

### III. SYSTEM DESCRIPTION

In our packet-based video transmission system, we consider use of channel encoder followed by orthogonal space-time block codes (O-STBC) as shown in Fig. 6. After the scalable encoding of the video, the base and FGS layers of each frame are divided into packets of constant size  $\gamma$ . These constant size source packets are then channel encoded using 16-bit CRC for error detection and rate-compatible punctured Turbo (RCPT) codes for unequal error protection. These channel encoded packets are further encoded using O-STBC for transmission over MIMO wireless systems. A Rayleigh flat-fading channel with AWGN is considered between each transmitter and each receiver. At the receiver, maximum-likelihood (ML) decoding is used to detect the transmitted symbols, which are then demodulated and channel decoded for error correction and detection. If a packet is not detected to be error-free, the layer of the corresponding frame to which the packet belongs is dropped. All the error-free packets for each frame are buffered and then fed to the source decoder with error concealment for video reconstruction.

For the MIMO wireless communication system, we consider  $M_t$  transmit antennas and  $M_r$  receive antennas and as mentioned earlier, O-STBC are used for transmission over the MIMO systems. In our work, we have considered two different MIMO systems.

- MIMO System 1

In MIMO System 1, we consider the O-STBC design proposed by Alamouti  $G_2(x_1, x_2)$  [20] of rate 1 for  $M_t = M_r = 2$ , where  $x_1$  and  $x_2$  are the symbols that can be chosen independently from either the same or different constellations. These two symbols are transmitted in  $T = 2$  time slots (symbol period)

$$G_2(x_1, x_2) = \begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix} \quad (10)$$

where  $*$  denotes complex conjugate.

- MIMO System 2

In this system, we have considered  $M_t = 4$  and  $M_r = 1$ , and the O-STBC design proposed by Tarokh *et al.* [21], [22]  $G_4(x_1, x_2, x_3)$  of rate 3/4 is used, where  $x_1, x_2$  and  $x_3$  are the symbols that can be chosen from either the same or different constellations. These three symbols are transmitted in  $T = 4$  time slots

$$G_4(x_1, x_2, x_3) = \begin{bmatrix} x_1 & x_2 & x_3 & 0 \\ -x_2^* & x_1^* & 0 & x_3 \\ -x_3^* & 0 & x_1^* & -x_2 \\ 0 & -x_3^* & x_2^* & x_1 \end{bmatrix}. \quad (11)$$

Clearly, this code gives us more flexibility by allowing us to use three different constellations in the same block code compared to only two constellations as in  $G_2(x_1, x_2)$ .

For both MIMO systems, the received signal can be modeled as

$$Y = \sqrt{\frac{\rho}{M_t}} CH + N \quad (12)$$

where  $C = \{c_t^i : 1 \leq t \leq T, 1 \leq i \leq M_t\}$  is the  $T \times M_t$  transmitted signal matrix. It is given as

$$C = \sqrt{\frac{T}{K}} G_{M_t}(x_1, x_2, \dots, x_K) \quad (13)$$

where  $K$  is the number of different symbols in a codeword. Each element  $c_t^i$  is the signal transmitted at antenna  $i$  at time  $t$ .  $H = \{h_{i,j}\}$  is the channel coefficient matrix of size  $M_t \times M_r$  where  $h_{i,j}$  is the channel coefficient from transmit antenna  $i$  to receive antenna  $j$ ;  $Y = \{y_t^j : 1 \leq t \leq T, 1 \leq j \leq M_r\}$  is the received signal matrix of size  $T \times M_r$  where  $y_t^j$  is the signal received at antenna  $j$  at time  $t$  and  $N = \{n_t^j\}$  is the noise matrix of size  $T \times M_r$ , where  $n_t^j$  is the additive noise at time slot  $t$  on the receiver antenna  $j$ . The noise samples and the elements of  $H$  are independent samples of a zero-mean complex Gaussian random variable with variance 1. The fading coefficients are assumed to remain constant during one codeword transmission period and change independently from one codeword to the other. Such a channel is called quasi-static. The factor  $\sqrt{(\rho)/(M_t)}$  in (12) is to ensure that  $\rho$  is the SNR at each receiver antenna and is independent of  $M_t$ . The energy of transmission codeword is normalized to the constraint  $E\{\|C\|_F^2\} = M_t T$  where  $\|C\|_F^2$  is the Frobenius norm of  $C$ . If perfect channel state information is known at the receiver, then the ML decoding is used to minimize the decision metric  $\min_C \|Y - \sqrt{(\rho)/(M_t)} CH\|_F^2$  for detecting the transmitted codeword. For all O-STBC, each of the symbols in the codeword, i.e.,  $x_1, x_2, \dots, x_K$ , can be decoded independently using ML decoding.

#### IV. OPTIMAL BANDWIDTH ALLOCATION FOR WIRELESS TRANSMISSION

To achieve a good performance in a wireless system, a global optimization is required that takes into consideration the source encoder, channel encoder and the channel conditions. To account for this, we consider the minimization of the expected end-to-end distortion by optimally selecting the quantization parameter (QP) and the GOP size for the source encoder, and the RCPT channel coding rate and the symbol constellation choice for the MIMO transmission. The optimization is constrained on the total available bandwidth  $B_{\text{budget}}$ . The scalable source encoder produces a layered bitstream where each layer is of different importance, and by protecting these layers unequally using the channel parameters, we can ensure efficient bandwidth allocation between all the layers and then for each of the layers, between the source and the channel coding.

The SVC codec used here encodes and decodes on a GOP basis, which results in various temporal layers, and on top of the temporal scalability, quality scalability is applied to get FGS layers. We consider this combined scalability by defining a total of  $L$  layers. The first  $L - 2$  layers ( $\mu_1, \dots, \mu_{L-2}$ ) are the base layers (FGS0) of the frames associated with the lowest to the highest temporal level in decreasing order of importance for video reconstruction. The other two FGS layers (FGS1 and FGS2) of all the frames in a GOP are defined as individual layers ( $\mu_{L-1}, \mu_L$ ) of even lesser importance.

The bandwidth allocation problem described above can be formulated as: Given an overall transmission bandwidth  $B_{\text{budget}}$  the goal is to allocate the bandwidth  $B_{s+c}$  between the source and the channel coding and among all the layers such that the total expected distortion is minimized

$$\{\text{GOP}^*, \text{QP}^*, \mathbf{R}_c^*, \mathbf{M}^*\} = \underset{\{\text{GOP}, \text{QP}, \mathbf{R}_c, \mathbf{M}\}}{\text{arg min}} E\{D_{s+c}\} \quad \text{s.t. } B_{s+c} \leq B_{\text{budget}} \quad (14)$$

where  $E\{D_{s+c}\}$  is the total expected end-to-end distortion due to source and channel coding which, for given source coding parameters, channel conditions and error concealment, is accurately estimated using the SDDE algorithm as explained in Section II.  $\text{GOP}$ ,  $\text{QP}$ ,  $\mathbf{R}_c$  and  $\mathbf{M}$  are the admissible set of values for GOP size, QP values, RCPT coding rates and symbol constellations, respectively.  $\text{GOP}^*$ ,  $\text{QP}^*$ ,  $\mathbf{R}_c^* = \{R_{c,\mu_1}, \dots, R_{c,\mu_L}\}$  and  $\mathbf{M}^* = \{M_{\mu_1}, \dots, M_{\mu_L}\}$  define the GOP size, QP value, RCPT coding rates and the symbol constellations, respectively after optimization. Here,  $\text{QP} = \text{QP}_{\mu_{L-2}}$  is the quantization parameter value for the base layer (FGS0) of the highest temporal level. All the other quantization parameters  $\{\text{QP}_{\mu_{L-3}}, \dots, \text{QP}_{\mu_1}\}$  depend on  $\text{QP}_{\mu_{L-2}}$  as follows:

$$\begin{aligned} \text{QP}_{\mu_{L-3}} &= \text{QP}_{\mu_{L-2}} - \alpha_{\mu_{L-3}} \Delta_{\text{QP}} \\ \text{QP}_{\mu_{L-4}} &= \text{QP}_{\mu_{L-3}} - \alpha_{\mu_{L-4}} \Delta_{\text{QP}} \\ &\vdots \\ \text{QP}_{\mu_1} &= \text{QP}_{\mu_2} - \alpha_{\mu_1} \Delta_{\text{QP}} \end{aligned}$$

where  $\Delta_{\text{QP}} = 1$  and  $\alpha_{\mu_{L-3}}, \alpha_{\mu_{L-4}}, \dots, \alpha_{\mu_1} \in \{1, 2\}$ . The quantization parameter values for the frames of FGS1 and FGS2

is less than the corresponding frames of FGS0 by  $6\Delta_{\text{QP}}$  and  $12\Delta_{\text{QP}}$ , respectively.

According to the O-STBC structure defined in (10) and (11), we restrict the number of optimal constellation choices for the  $L$  layers to a maximum of  $k$  different constellations, where  $k$  as defined before is the number of different symbols in the O-STBC codeword.

As mentioned before, the expected distortion  $E\{D_{s+c}\}$  is obtained using the SDDE algorithm. It is clear from (9) that the accurate calculation of the decoder distortion depends on the individual probabilities of losing each of the FGS layers for all the frames ( $P_n, P_{nE1}$  and  $P_{nE2}$ ). Also, as explained in Section III, each of these layers is divided into packets of constant size  $\gamma$ . The packet error rate for these packets depends on the channel model, channel SNR, packet size, RCPT coding rate and constellation selected. Let us define the packet error rate for these constant size packets as  $\text{PER}(R_{c,\mu_l}, M_{\mu_l})$ , which, for given and fixed channel model (i.e., Rayleigh flat-fading), channel SNR and packet size  $\gamma$ , depends on the channel parameters: RCPT coding rate and constellation selected for the layer  $\mu_l$ . These packet error rate values for both MIMO systems are calculated using simulations and establish a reference as to the performance of the transmission over the wireless channel with the given parameters. Furthermore, this channel performance analysis needs to be done only once for a set of channel conditions of interest. Now, the probabilities  $P_n, P_{nE1}$  and  $P_{nE2}$  are obtained as

$$P_n = 1 - (1 - \text{PER}(R_{c,\mu_l}, M_{\mu_l}))^{\lceil \frac{N_{n,\mu_l}}{\gamma} \rceil} \quad l \in \{1, 2, \dots, L-2\} \quad (15)$$

$$P_{nE1} = 1 - (1 - \text{PER}(R_{c,\mu_{L-1}}, M_{\mu_{L-1}}))^{\lceil \frac{N_{n,\mu_{L-1}}}{\gamma} \rceil} \quad (16)$$

$$P_{nE2} = 1 - (1 - \text{PER}(R_{c,\mu_L}, M_{\mu_L}))^{\lceil \frac{N_{n,\mu_L}}{\gamma} \rceil} \quad (17)$$

where  $N_{n,\mu_l}$  is the size of FGS0 of the frame  $n$  which belongs to the layer  $\mu_l$ ;  $N_{n,\mu_{L-1}}$  and  $N_{n,\mu_L}$  are the size of the layers FGS1 and FGS2 of frame  $n$ , respectively.

Next, we define the bandwidth (symbol rate)  $B_{s+c}$  as

$$B_{s+c} = \sum_{l=1}^L B_{s+c,\mu_l} \quad (18)$$

where  $B_{s+c,\mu_l}$  is the bandwidth allocated for layer  $\mu_l$  and is defined by

$$B_{s+c,\mu_l} = \frac{R_{s,\mu_l}}{R_{c,\mu_l} \times \log_2(M_{\mu_l})} \times \frac{T}{K} \quad (19)$$

where  $R_{s,\mu_l}$  is the source rate for layer  $\mu_l$ , it is in bits/s and depends on the quantization parameter value used for that layer ( $R_{s,\mu_{L-1}}$  and  $R_{s,\mu_L}$  depend on the quantization parameter values used for FGS1 and FGS2, respectively);  $R_{c,\mu_l}$  is the Turbo channel coding rate for layer  $\mu_l$  and is dimensionless;  $M_{\mu_l}$  is the constellation used by layer  $\mu_l$ ,  $\log_2(M_{\mu_l})$  is the number of bits per symbol and  $T$  is the number of time slots required to transmit  $K$  symbols in each codeword over the MIMO system.

The problem in (14) is a constrained optimization problem and is solved as an unconstrained one by using the Lagrangian method as follows:

$$\mathcal{L}(\lambda) = D_{s+c} + \lambda B_{s+c} \quad (20)$$

where  $\lambda$  is the Lagrangian multiplier [8], [10]. The solution to this problem,  $B_{s+c}^*$  and, hence,  $D_{s+c}^*$  is also the solution to the constrained problem of (14) if and only if  $B_{s+c}^* = B_{\text{budget}}$ . In practice, since there is only a finite set of choices for source coding rate, RCPT coding rates and constellation choices, it is not always possible to exactly meet  $B_{\text{budget}}$ . In this case, the solution is the bandwidth that is closest to  $B_{\text{budget}}$  while being lower than  $B_{\text{budget}}$ . By varying  $\lambda$  from zero to  $\infty$ , the result of (20) will trace out the operational bandwidth-distortion curve for the system. For each combination of GOP sizes, source coding rates, RCPT channel coding rates and constellations selected for all layers, a set of operating points are obtained. The convex hull of these operating points is denoted by  $D_{s+c}^*$ , that is

$$D_{s+c}^* = C(\text{GOP}, \text{QP}, R_c, M) \quad (21)$$

where  $C$  represents the operational rate-distortion function (ORDF).

### V. EXPERIMENTAL RESULTS

For our simulations, we implemented the SDDE algorithm by modifying the SVC. The source is taken as 160 frames of two sequences “Foreman” and “Akiyo” at 30 fps with a constant Intrapdate at every 16 or 32 frames. After source encoding with temporal and quality scalability, each frame is associated with a particular temporal level and a base and two FGS layers are obtained. These layers are then divided into packets of constant size  $\gamma = 100$  bytes. We consider the admissible set for  $\{\text{QP}\}$  as  $\text{QP} = \{20, 25, 30, 35, 40, 45, 50\}$  and GOP sizes as  $\text{GOP} = \{4, 8, 16\}$ . Each packet is channel encoded using the RCPT coding rates of  $R_c = \{1/3, 1/2\}$  which are obtained by puncturing a mother code of rate  $1/3$  with constraint length  $K = 3$  and a code generator  $g = [07; 05]_{\text{octal}}$ . The turbo encoder interleaver is of size  $L_{\text{inter}} = 1600$  bits and is randomly generated. At the RCPT decoder the log-MAP algorithm [27] is used for three decoding iterations. The data are modulated using quadrature amplitude modulation (QAM) with the possible constellations chosen from  $M = \{4\text{QAM}, 8\text{QAM}, 16\text{QAM}\}$ . The wireless MIMO channel between each transmitter and each receiver is modeled as a Rayleigh flat-fading channel with AWGN and an SNR of 12 dB. At the decoder, scheme 2 is used for error concealment.

We first demonstrate the performance of the proposed system for the optimal selection of constellation in comparison with a fixed constellation across all the layers transmitted over the MIMO channel. Fig. 7(a) and (b) illustrates this comparison for the transmission of the “Foreman” sequence encoded with optimal selection of QP,  $R_c$  and  $M$  for  $\text{GOP} = 4$  and  $\text{Intra update} = 32$  for both the MIMO systems used here. It is clear from these figures that using optimal but different modulations across the layers (multiple QAM) outperforms the scenario when the modulation for all the layers is optimally selected but kept fixed (constant QAM) for a wide range of bandwidth.

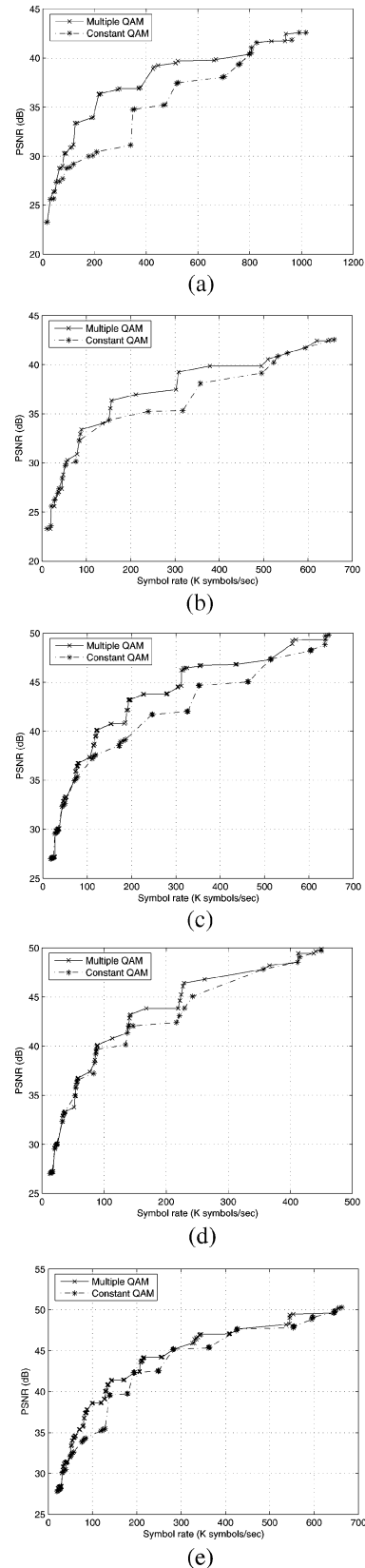


Fig. 7. Comparison of the system performance between multiple QAM versus constant QAM. (a), (b) “Foreman” sequence encoded at  $\text{GOP} = 4$  and  $I = 32$  for MIMO system 1 and MIMO system 2, respectively. (c), (d) “Akiyo” sequence encoded at  $\text{GOP} = 8$  and  $I = 16$  for MIMO system 1 and MIMO system 2, respectively. (e) “Akiyo” sequence encoded at  $\text{GOP} = 16$  and  $I = 16$  for MIMO system 2.



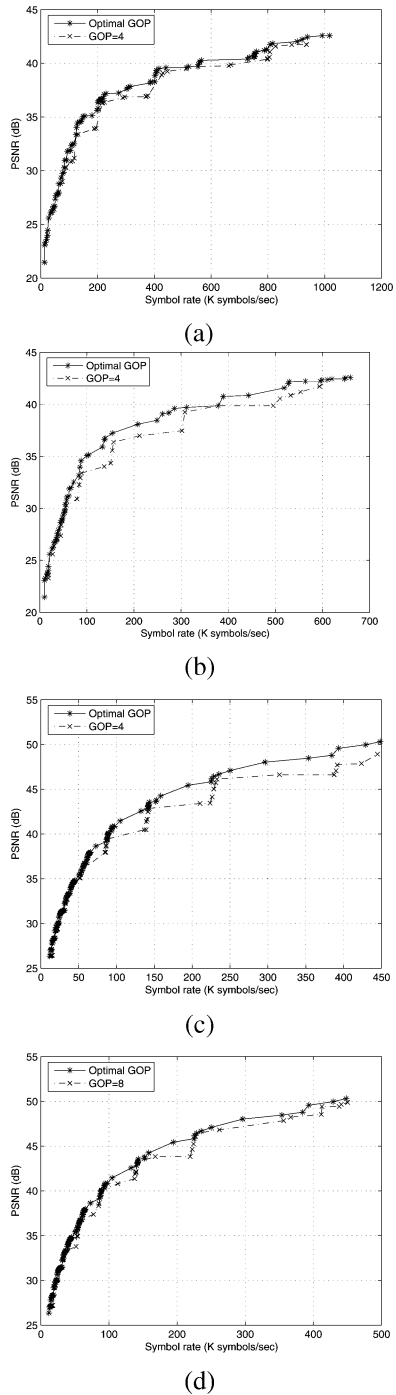


Fig. 8. Comparison of the system performance between optimal GOP size versus fixed GOP size. (a), (b) “Foreman” sequence encoded at  $I = 32$  compared against fixed  $GOP = 4$  for MIMO system 1 and MIMO system 2, respectively. (c), (d) “Akiyo” sequence encoded at  $I = 16$  for MIMO system 1 compared against fixed  $GOP = 4$  and fixed  $GOP = 8$ .

It is only because of the structure of O-STBC used here that we can use different constellations independently over multiple antennae, which results in unequal error protection (without any added redundancy) in addition to what is provided by RCPT codes. It is also observed that the PSNR improvement gained by using multiple QAM is higher in the MIMO system 2 ( $4 \times 1$ ) as compared to the other MIMO system ( $2 \times 2$ ). This can be explained by the fact that MIMO system 2 allows the transmission

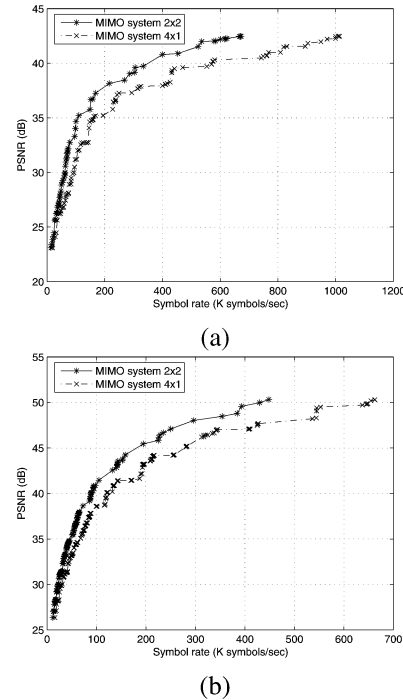


Fig. 9. Comparison of the performance of the two MIMO systems with optimal parameters  $QP$ ,  $GOP$ ,  $R_c$  and  $M$ . (a) “Foreman” sequence encoded at  $I = 16$ . (b) “Akiyo” sequence encoded at  $I = 16$ .

using three different constellations as compared to a maximum of two different constellations in MIMO system 1. Similar results can also be observed in Fig. 7(c)–(e) for the “Akiyo” sequence for Intra update = 16 and various GOP sizes.

The comparison of the performance of the system for optimal selection of GOP size (for the whole sequence) versus a fixed GOP size for the range of transmission bandwidth is shown in Fig. 8(a)–(d). In these figures, it is evident that selecting the GOP size optimally along with the other parameters  $QP$ ,  $R_c$  and  $M$  results in better PSNR performance for the reconstructed “Foreman” and “Akiyo” sequences.

In Fig. 9(a) and (b), we show the comparison of the two MIMO systems with all the parameters  $QP$ ,  $GOP$ ,  $R_c$  and  $M$  selected after the optimization. Although MIMO system 2 allows more flexibility compared to MIMO system 1 in selecting the number of constellations, MIMO system 1 shows better PSNR performance because of the rate of the O-STBC associated with each of the MIMO systems (rate of  $G_2(x_1, x_2)$  equals 1, whereas rate of  $G_4(x_1, x_2, x_3)$  equals  $3/4$ ). MIMO system 2 may be considered more practical to use in scenarios where hardware restrictions limit the number of antennas at the receiver to be only one. The optimization results for different target bandwidth values are also shown in Table II. Table II shows some of the optimal choices of the parameters under consideration ( $QP$ ,  $GOP$ ,  $R_c$  and  $M$ ) for all the layers for the MIMO system 1.

## VI. CONCLUSION

We have proposed a new wireless video transmission system that uses the SVC codec and exploited video scalability along with spatial diversity using O-STBC over broadband

TABLE II  
OPTIMAL PARAMETER VALUES FOR "AKIYO" SEQUENCE TRANSMITTED OVER MIMO SYSTEM I

BW (Ksps)	PSNR (dB)	GOP	QP	TL0		TL1		TL2		TL3		TL4		FGS1		FGS2	
				$M$	$R_c$	$M$	$R_c$	$M$	$R_c$	$M$	$R_c$	$M$	$R_c$	$M$	$R_c$	$M$	$R_c$
15.58	27.83	16	50	4	1/3	8	1/2	–	–	–	–	–	–	–	–	–	–
20.00	29.18	4	45	4	1/3	8	1/2	–	–	–	–	–	–	–	–	–	–
24.13	30.06	8	45	4	1/3	4	1/2	16	1/3	16	1/3	–	–	–	–	–	–
28.65	31.41	16	45	4	1/3	4	1/2	4	1/2	16	1/3	16	1/3	–	–	–	–
32.92	32.48	4	40	4	1/3	16	1/3	–	–	–	–	–	–	–	–	–	–
36.76	33.32	8	40	4	1/3	4	1/2	4	1/2	4	1/2	–	–	–	–	–	–
42.22	34.5	16	40	4	1/3	4	1/2	16	1/3	16	1/3	16	1/3	–	–	–	–
53.96	35.89	4	35	4	1/3	4	1/2	16	1/3	–	–	–	–	–	–	–	–
57.00	36.54	8	35	4	1/3	16	1/3	16	1/3	16	1/3	–	–	–	–	–	–
64.95	37.89	16	40	4	1/3	4	1/2	4	1/2	4	1/2	16	1/3	16	1/3	–	–
88.41	39.95	8	30	4	1/3	4	1/2	16	1/3	16	1/3	–	–	–	–	–	–
281.48	42.84	8	25	4	1/3	4	1/2	16	1/3	16	1/3	–	–	–	–	–	–
140.68	43.35	16	30	4	1/3	4	1/2	4	1/2	4	1/2	16	1/3	16	1/3	–	–
225.76	46.18	16	25	4	1/3	4	1/2	4	1/2	4	1/2	4	1/2	16	1/3	–	–
384.43	48.79	16	20	4	1/3	4	1/2	4	1/2	4	1/2	4	1/2	4	1/2	16	1/3
448.17	50.31	16	20	4	1/3	4	1/2	4	1/2	4	1/2	4	1/2	8	1/3	8	1/3

MIMO systems. We have proposed different error concealment schemes to handle packet losses during wireless transmission and, at the encoder, we have developed a method for the accurate estimation of the video distortion at the receiver for given channel conditions. This method takes into account loss of temporal and SNR scalable layers as well as error concealment at the decoder. The experimental results in comparison with simulated wireless video transmission have shown the accuracy of the distortion estimation algorithm. Using the decoder distortion estimation algorithm, the bandwidth-constrained optimization problem has been solved. We exploited the orthogonal structure of the O-STBC codes by allocating different layers over different codeword symbols modulated using different constellations. The results clearly indicate the advantage of this compared to using only single constellation for all the layers. Also, it has been shown that optimally selecting the GOP size for the whole video sequence for a given bandwidth and channel conditions results in better PSNR performance. The optimal selection for source coding parameters (QP and GOP) and channel coding parameters ( $R_c$  and  $M$ ) have been presented for two MIMO systems of interest for a wide range of bandwidth values.

## REFERENCES

- [1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
- [2] S. Wenger, "H.264/AVC over IP," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 645–656, Jul. 2003.
- [3] T. Stockhammer, M. H. Hannuksela, and T. Wiegand, "H.264/AVC in wireless environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 657–673, Jul. 2003.
- [4] H. Schwarz, T. Hinz, H. Kirchhoffer, D. Marpe, and T. Wiegand, "Technical description of the HHI proposal for SVC CE1," *ISO/IEC JTC1/SC29/WG11, M11244*, Oct. 2004.
- [5] H. Schwarz, D. Marpe, T. Schierl, and T. Wiegand, "Combined scalability support for the scalable extension of H.264/AVC," in *Proc. IEEE Int. Conf. Multimedia and Expo*, Jul. 2005, pp. 446–449.
- [6] J. Reichel, H. Schwarz, and M. Wien, "Scalable working draft—Working draft 1," in *Joint Video Team (JVT), Doc. JVT-N020*, Hong Kong, CN, Jan. 2005.
- [7] H. Schwarz, D. Marpe, and T. Wiegand, "MCTF and scalability extension of H.264/AVC," presented at the PCS, San Francisco, CA, Dec. 2004.
- [8] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint source-channel coding for motion-compensated DCT-based SNR scalable video," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 1043–1054, Sep. 2002.
- [9] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. 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.
- [10] L. P. Kondi, D. Srinivasan, D. A. Pados, and S. Batalama, "Layered video transmission over wireless multirate DS-CDMA links," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1629–1637, Dec. 2005.
- [11] Y. S. Chan and J. W. Modestino, "A joint source coding-power control approach for video transmission over CDMA networks," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1516–1525, Dec. 2003.
- [12] D. Srinivasan, L. P. Kondi, and D. A. Pados, "Scalable video transmission over wireless DS-CDMA channels using minimum TSC spreading codes," *IEEE Signal Process. Lett.*, vol. 11, no. 10, pp. 836–840, Oct. 2004.
- [13] S. Zhao, Z. Xiong, and X. Wang, "Optimal resource allocation for wireless video over CDMA networks," *IEEE Trans. Mobile Comput.*, vol. 4, no. 1, pp. 56–67, Jan. 2005.
- [14] Y. Shen, P. C. Cosman, and L. Milstein, "Error-resilient video communications over CDMA networks with a bandwidth constraint," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3241–3252, Nov. 2006.
- [15] R. Zhang, S. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 966–976, Jun. 2000.
- [16] S. Regunathan, R. Zhang, and K. Rose, "Scalable video coding with robust mode selection," *Signal Process.: Image Commun.*, vol. 16, pp. 725–732, 2001.
- [17] Y. Shen, P. C. Cosman, and L. B. Milstein, "Video coding with fixed-length packetization for a tandem channel," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 273–288, Feb. 2006.
- [18] Y. Eisenberg, F. Zhai, T. N. Pappas, R. Berry, and A. Katsaggelos, "VAPOR: Variance-aware per-pixel optimal resource allocation," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 289–299, Feb. 2006.
- [19] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1989.
- [20] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.
- [21] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
- [22] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.
- [23] C. Kuo, C. Kim, and C.-C. J. Kuo, "Robust video transmission over wideband wireless channel using space-time coded OFDM systems," in *Proc. IEEE Wireless Communications and Networking Conf.*, Mar. 2002, vol. 2, pp. 931–936.

- [24] S. Zhao, Z. Xiong, X. Wang, and J. Hua, "Progressive video delivery over wideband wireless channels using space-time differentially coded OFDM systems," *IEEE Trans. Mobile Comput.*, vol. 5, no. 4, pp. 303–316, Apr. 2006.
- [25] T. Schierl, H. Schwarz, D. Marpe, and T. Wiegand, "Wireless broadcasting using the scalable extension of H.264/AVC," in *Proc. IEEE Int. Conf. Multimedia and Expo*, Jul. 2005, pp. 884–887.
- [26] M. M. Ghandi and M. Ghanbari, "Layered H.264 video transmission with hierarchical QAM," *Elsevier J. Vis. Commun. and Image Represent.*, 2005.
- [27] D. N. Rowitch and L. B. Milstein, "On the performance of hybrid FEC/ARQ systems using rate compatible punctured turbo (RCPT) codes," *IEEE Trans. Commun.*, vol. 48, pp. 948–959, Jun. 2000.
- [28] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Cyclic Redundancy and Other Checksums*. Cambridge, U.K.: Cambridge Univ. Press, 1992.



**Mohammad K. Jubran** received the B.S. degree in electronic engineering from Al-Quds University, Abu-Dies, Palestine, in 2000, the M.S. and Ph.D. degrees, both in electrical engineering from the University at Buffalo, The State University of New York, in 2005 and 2007, respectively.

He is currently an Assistant Professor in the Department of Electrical Engineering at Birzeit University, Birzeit, Palestine. His research interests include video compression, joint source/channel coding, multimedia signal processing and transmission,

and wireless communications.



**Manu Bansal** received the B.Tech degree in electronics engineering from the YMCA Institute of Engineering, Haryana, India, in 2001, and the M.S. and Ph.D. degrees in electrical engineering from the University at Buffalo, The State University of New York, in 2004 and 2007, respectively.

During the summers of 2005–2007, he was an engineering intern at Qualcomm, Inc., San Diego, CA, where he worked on the development and implementation of MediaFLO and DVB-H standards of mobile TV technology. His research interests

include video processing and communications over wireless networks and joint source-channel coding. He is currently working as a Science Advisor in the intellectual property division of Goodwin Procter LLP, Boston, MA.



**Lisimachos P. Kondi** (S'92–M'99) received the diploma degree in electrical engineering from the Aristotle University of Thessaloniki, Greece, in 1994, and the M.S. and Ph.D. degrees in electrical and computer engineering from Northwestern University, Evanston, IL, in 1996 and 1999, respectively.

During the 1999–2000 academic year, he was a postdoctoral research associate at Northwestern University. He is currently an Assistant Professor in the Department of Computer Science at the University of Ioannina, Greece. He was previously with the faculty of the University at Buffalo, The State University of New York, and has also held summer appointments at the Naval Research Laboratory, Washington, DC, and the Air Force Research Laboratory, Rome, NY. His research interests are in the general areas of signal and image processing and communications, including image and video compression and transmission over wireless channels and the Internet, super-resolution of video sequences, and shape coding.

Dr. Kondi is an Associate Editor for the *EURASIP Journal of Advances in Signal Processing* and an Associate Editor for the *IEEE SIGNAL PROCESSING LETTERS*.



**Rohan Grover** received the B.Tech. degree in electronics engineering from the BMS College of Engineering, Karnataka, India, in 2001 and the M.S. and Ph.D. degrees in electrical engineering from University at Buffalo, The State University of New York, in 2004 and 2007, respectively.

From 2004–2007, he was a Research Assistant at the Communications and Systems laboratory at the University at Buffalo where he worked on multiuser CDMA communications, adaptive antenna arrays, and MIMO space-time block coding. His

research interests include OFDM, multiantenna systems, multiuser CDMA communications, and joint source channel coding. Since June 2007, he has been a Senior Communications Systems Engineer at Radiospire Networks, Hudson, MA, developing gigabit wireless PAN systems for transmission of high-def video signals.