Motion-compensated Scalable Video Transmission over MIMO Wireless Channels under Imperfect Channel Estimation

Hobin Kim†, Sun Yong Kim‡, Pamela C. Cosman† and Laurence B. Milstein†
† ECE Dept., University of California at San Diego, La Jolla, CA 92093
‡ Dept. of Electronic Engineering, Konkuk University, Seoul, Korea

***Abstract*—We study motion compensated fine granular scalable (MC-FGS) video transmission over multi-input multi-output (MIMO) wireless channels, where leaky and partial prediction schemes are applied in the enhancement layer of MC-FGS to exploit the tradeoff between error propagation and coding efficiency. For reliable transmission, we propose unequal error protection (UEP) by considering a tradeoff between reliability and data rates, which are controlled by forward error correction (FEC) and MIMO mode selection to minimize the average distortion. In a high Doppler environment where it is hard to get an accurate channel estimate, we investigate the performance of the proposed MC-FGS video transmission scheme with joint control of the leaky and partial prediction parameters and the UEP.**

I. Introduction

Recently, the transmission of multimedia over wireless channels has been in high demand. However, due to the high error probability and fluctuating channel bandwidth in a high Doppler environment, it becomes challenging to maintain the quality of service when a multimedia stream is transmitted over a wireless channel. Fine granularity scalable (FGS) video coding is suitable for mobile users with variable channel bandwidth, since it makes decoding possible even in the case of partial loss of the bitstream, where the FGS bitstream is encoded in a progressive manner. For example, in the MPEG-4 FGS video coding, a scalable enhancement layer (EL) is generated from the quantization errors of the DCT coefficients compressed in the base layer (BL) with bit-plane coding. When the base layer is transmitted reliably, the scalable enhancement layer bitstream can be decoded, even though it is truncated at any point.

In motion compensated prediction (MCP) of conventional MPEG-4 FGS coding, current base and enhancement layers are only predicted from the base layer of the previous frame. By excluding the enhancement layer from the MCP loop, MPEG-4 FGS coding can avoid error propagation which can be caused by the corruption of the enhancement layer. However, this can decrease the coding efficiency due to the use of a low quality reference frame. To enhance the compression efficiency, motion compensated fine granularity scalable (MC-FGS) coding was proposed in [1]. In this video coding scheme, a high quality reference is generated using the enhancement layer as well as the base layer, which allows the system to achieve a high coding efficiency. However, the loss of the enhancement layer can result in severe error propagation, since there can be a mismatch between the reconstructed references at the encoder and the decoder.

In [2], progressive FGS (PFGS) was introduced to improve the coding efficiency and alleviate error propagation simultaneously. For higher coding efficiency, PFGS uses a separate prediction loop that contains a high quality reference frame in the encoding of the enhancement layer video. In order to address the drift problem, PFGS keeps a prediction path from the base layer to the highest bitplanes at the enhancement layer across several frames to make sure that the coding schemes can gracefully recover from errors over a few frames. Robust FGS (RFGS) [3] uses different approach to control the tradeoff between the coding efficiency and the error propagation, where the two distinct parameters of leaky and partial prediction are used jointly. UEP can be also used to enhance the error resilience of MC-FGS by reducing the loss probability of the enhancement layer to be used for the reconstruction of the reference. As in [4], different numbers of parity bits are allocated to the packets of the enhancement layer according to their impact on average distortion.

In this paper, we study the transmission of an MC-FGS bitstream over a MIMO channel with joint control of the UEP and the prediction parameters. Specifically, we propose a UEP policy consisting of FEC and MIMO mode selection to exploit a fundamental tradeoff between spatial diversity and multiplexing. Originally, the idea of combining the cooperative diversity gain with UEP in a progressive image bitstream was proposed by Kwasinski in [5], where additional diversity was applied to high priority packets. We extend this to a MIMO system which can provide higher diversity orders. That is, for each packet, we choose an appropriate MIMO mode and FEC code rate to minimize average distortion.

The rest of this paper is organized as follows. In Section II, the source and channel models are described. The UEP scheme, based on a tradeoff between reliability and data rate, is proposed in Section III. In Section IV, we provide simulation results and a discussion regarding the selection of the prediction parameters. Finally, in Section V, conclusions

This work was supported in part by the Center for Wireless Communications of UCSD, LG Electronics, and the UC Discovery Grant Program.

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

are presented.

II. System Model

A. Source Model

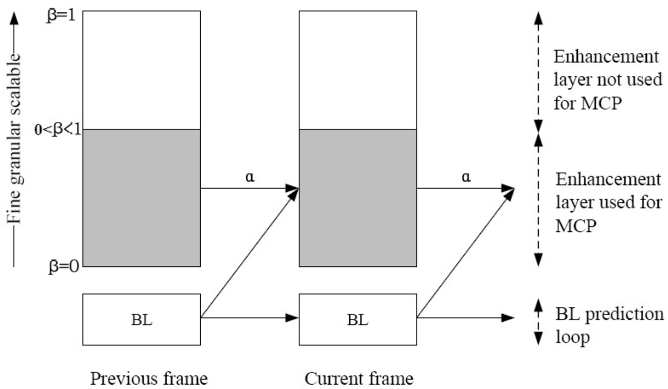


Fig. 1: A motion compensated FGS coder with leaky and partial prediction [4].

We consider an MC-FGS video coder employing leaky and partial prediction, as introduced in [3] and [4]. In conventional MC-FGS video coding, both the base and the enhancement layers are used to reconstruct a high-quality reference. However, this can result in error propagation. To compensate for the error propagation, RFGS introduced leaky prediction as presented in Fig. 1, where the enhancement layer is scaled by the leaky prediction parameter, $\alpha \in [0,1]$, before it is incorporated into the MCP loop. That is, the reference for the prediction of the current enhancement layer at time n, $\hat{F}_n^{EL}$, is a weighted sum of the previous base layer, F_{n-1}^{BL} and the partial enhancement layer, F_{n-1}^{EL}, i.e., [4]

$$\hat{F}_n^{EL} = (1-\alpha)F_{n-1}^{BL} + \alpha F_{n-1}^{EL}. \quad (1)$$

Therefore, if the leaky prediction parameter is set to 0, the scheme becomes the MPEG-4 FGS video coding, where the enhancement layer is entirely excluded from the MCP loop. In contrast, if the leaky prediction parameter is fixed at 1, then it works as the conventional MC-FGS video coding. However, by choosing the parameter less than 1, the effect of error propagation can be reduced at the price of coding efficiency.

In partial prediction, the encoder designates the number of bitplanes in the enhancement layer to be used for the reconstruction of the reference frame in the MCP loop. By including more bitplanes of the enhancement layer into the MCP loop, better coding efficiency can be achieved. However, if the instantaneous channel bandwidth cannot support the number of bitplanes used in the MCP loop, then it can result in error propagation. Therefore, the partial prediction parameter needs to be chosen based on knowledge of the channel bandwidth. In this paper, we allow any arbitrary number of symbols in the enhancement layer bitstream to be used in the MCP loop. This is also a reasonable approach because the enhancement layer of MC-FGS coding can be truncated at

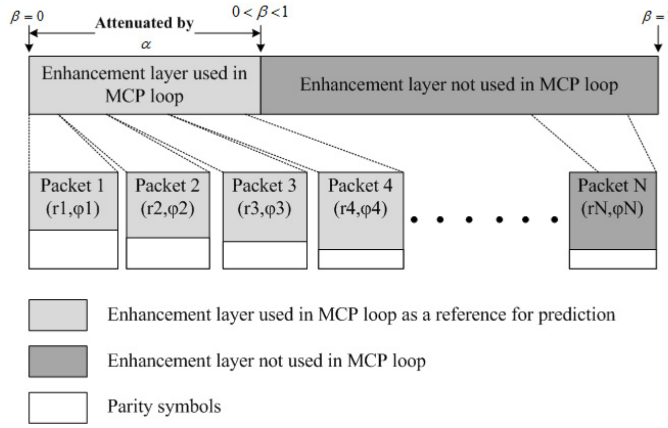


Fig. 2: Packetization of the MC-FGS enhancement layer.

any point. We define the partial prediction parameter, β, to be the ratio of the number of enhancement layer symbols used for MCP and the maximum number of enhancement layer symbols for that frame.

B. Channel Model

In this paper, we assume an $M_r \times M_t$ wireless MIMO channel, where M_r and M_t represent the number of receive and transmit antennas, respectively. Then, the generalized baseband signal model can be expressed as

$$r[t] = H[t]s[t] + n[t], \quad (2)$$

where $r[t]$ represents the $M_r \times 1$ received signal vector, $s[t]$ is the $M_t \times 1$ transmitted signal vector, and $n[t]$ is the $M_r \times 1$ independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) vector at time t. $H[t]$ is the $M_r \times M_t$ MIMO channel coefficients matrix whose elements are i.i.d. zero-mean complex Gaussian random variables with variance σ_h^2. We consider a high Doppler environment, which results in rapidly time-varying channels. Given the Doppler spread, to model the channel estimation error, we consider the following system :

- Pilot symbol assisted modulation (PSAM) [6]
- Orthogonal pilot symbols for each transmit antenna
- Channel estimation by using K nearest pilot samples in conjunction with a Wiener filter

If we denote the estimation error of the channel from the jth transmit antenna to the ith receive antenna by $\varepsilon_{ij}[t]$, then it can be modelled as a complex Gaussian random variable, and its variance, $\sigma_\varepsilon^2[t]$, can be expressed as [6]

$$\sigma_\varepsilon^2[t] = 1 - w^\dagger[t]R^{-1}w[t], \quad (3)$$

where R represents the $K \times K$ autocorrelation matrix of the nearest K received pilot samples, $w[t]$ is the cross-covariance vectors between the received pilot samples and $h_{ij}[t]$, and we assume $\sigma_h^2 = 1$. Note that $w[t]$ and R are dependent on the pilot signal-to-noise ratio (SNR), Doppler spread, and pilot spacing. In this paper, we assume that the pilot SNR and its

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

spacing are selected to be equal to the data symbol SNR and the channel coherence time, respectively.

By the assumption that all MIMO channels are independent, channel estimation errors in the MIMO system can be modelled as a matrix consisting of i.i.d. complex Gaussian random variables with the variance of (3). To maximize the signal-to-interference-and-noise ratio (SINR) under the condition of imperfect channel estimation, we consider the modified minimum-mean-squared-error (MMSE) detection scheme proposed in [7]. We assume BPSK modulation, and that N packets, whose size is fixed at m symbols, are allocated to a frame of the FGS enhancement layer bitstream, as presented in Fig. 2. Note that each packet is protected by FEC. For the transmission of the packet, we choose either spatial diversity, spatial multiplexing or a hybrid of the two. We have a total transmit power constraint to make the system consume a constant power for any selection of MIMO modes. The specific MIMO mode (multiplexing, diversity, hybrid) is chosen at the transmitter on a packet by packet basis in order to minimize distortion. The selection is at the transmitter because the source statistics and encoder characteristics are known. Due to the lack of instantaneous channel information at the transmitter, the long-term average channel gain is used to choose the UEP policy and the appropriate MIMO mode.

1) Spatial Diversity: Spatial diversity is achieved by transmitting and receiving symbol streams with the same information content through multiple transmit and receive antennas. There are various ways to implement spatial diversity, but quasi-orthogonal space time block coding (QOSTBC) [8] is considered in this paper.

2) Spatial Multiplexing: Spatial multiplexing makes it possible to increase the transmission rate proportional to $min(M_t, M_r)$ without allocating additional bandwidth or transmit power [9]. Spatial multiplexing can be achieved by transmitting different data streams through multiple transmit antennas over independently fading channels. However, under the constraint of the fixed total transmit power, spatial multiplexing assigns less energy per bit compared to spatial diversity.

3) Hybrid Mode: We consider the hybrid mode to have twice the data rate and half the diversity order, as compared to the spatial diversity mode. For the purpose of simulation, we implement Double Space Time Transmit Diversity (D-STTD), as proposed in [10].

III. Unequal Error Protection with FEC and MIMO Mode Selection

In this section, we investigate how to find a UEP policy for minimizing the average distortion by choosing the FEC code rate and the appropriate MIMO mode. We assume that N fixed size packets are available for the transmission of the FGS enhancement layer in a frame, where the packet length is equal to the duration of m symbols. For the ith packet, the channel code rate and MIMO mode are denoted r_i and ϕ_i, respectively, where r_i is chosen from the possible channel code rate set and ϕ_i represents one of the MIMO modes. If we

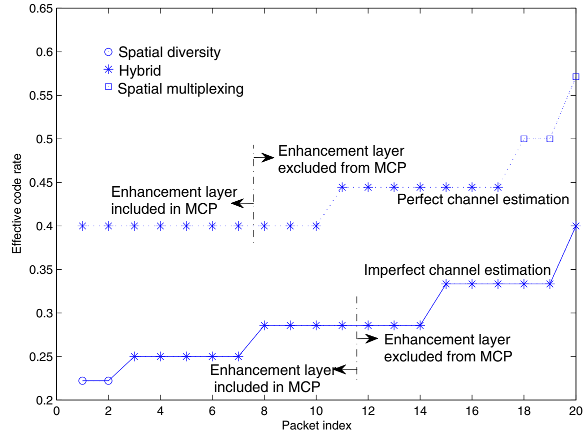


Fig. 3: UEP policy and MIMO mode selection for *"Foreman"* sequence where $\alpha = 1.0$ and $\beta = .15$. Total transmit signal to noise power ratio is fixed at 8 dB.

denote the packet error rate of the ith packet by $p_i(r_i, \phi_i)$, then the probability that the first k packets are received successfully and the first packet error happens at the $(k+1)$th packet is

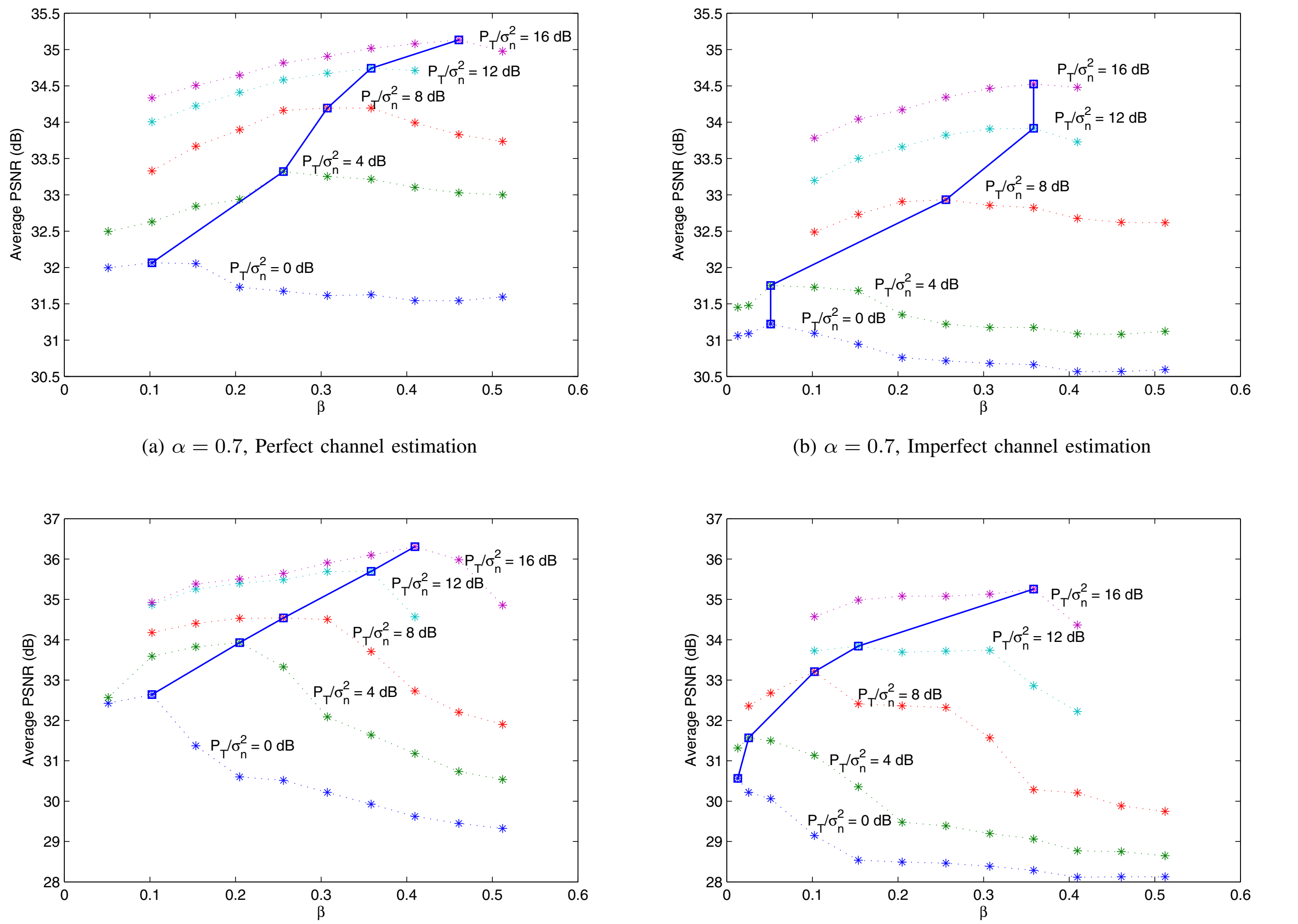
$$\pi_k \triangleq \Pr(\text{first packet error occurs at packet } k{+}1) = \begin{cases} p_1(r_1, \phi_1), & k{=}0 \\ \prod_{j=1}^{k}(1 - p_j(r_j, \phi_j))p_{k+1}(r_{k+1}, \phi_{k+1}), & 0 < k < N \\ \prod_{j=1}^{N}(1 - p_j(r_j, \phi_j)), & k{=}N. \end{cases}$$

When the first k packets are successfully received, the number of information bits available at the receiver is $\sum_{j=1}^{k} r_j M_j$, where M_j represents the total number of symbols in the jth packet, which is decided by the MIMO mode. Assume that the rate-distortion function of the source input is available at the transmitter. Then, the average distortion, $E[D(\alpha, \beta)]$, can be computed as

$$E[D(\alpha, \beta)] = D(0, \alpha, \beta)\pi_0 + \sum_{k=1}^{N} D(\sum_{j=1}^{k} r_j M_j, \alpha, \beta)\pi_k, \quad (4)$$

where α and β are the leaky and partial prediction parameters, respectively, and $D(R, \alpha, \beta)$ is the distortion of the decoded MC-FGS video with leaky and partial prediction parameters α and β, when the entire base layer and R bits of enhancement layer are received successfully.

At this point, we focus on the choice of $\underline{r} = [r_1, r_2, \cdots, r_N]$ and $\underline{\phi} = [\phi_1, \phi_2, \cdots, \phi_N]$ for minimizing $E[D(\alpha, \beta)]$, given α and β. In [11], the authors proposed a local search algorithm to get a sub-optimal distortion-minimizing UEP policy when the length of the channel codeword is fixed. By using this algorithm, we can get a sub-optimal rate allocation and MIMO mode selection to transmit the FGS enhancement layer



(a) $\alpha = 0.7$, Perfect channel estimation

(b) $\alpha = 0.7$, Imperfect channel estimation

(c) $\alpha = 1.0$, Perfect channel estimation

(d) $\alpha = 1.0$, Imperfect channel estimation

Fig. 4: Average PSNR performance versus partial prediction parameter, β. Leaky parameter is fixed at 0.7 and 1.0, and both perfect and imperfect channel estimations are considered.

bitstream. To do this, we define the effective code rate, s_i, as the ratio of the number of information symbols and the maximum number of transmitted symbols in a packet, where the maximum number of transmitted symbols is defined as that in a packet where spatial multiplexing is used. For example, in a 4×4 MIMO system, $4 \times m$ symbols can be transmitted with spatial multiplexing. Then, the effective code rate of the ith packet, s_i, is defined as

$$s_i = \begin{cases} r_i/4, & \text{if spatial diversity is used} \\ r_i/2, & \text{if hybrid mode is used} \\ r_i, & \text{if spatial multiplexing is used.} \end{cases}$$

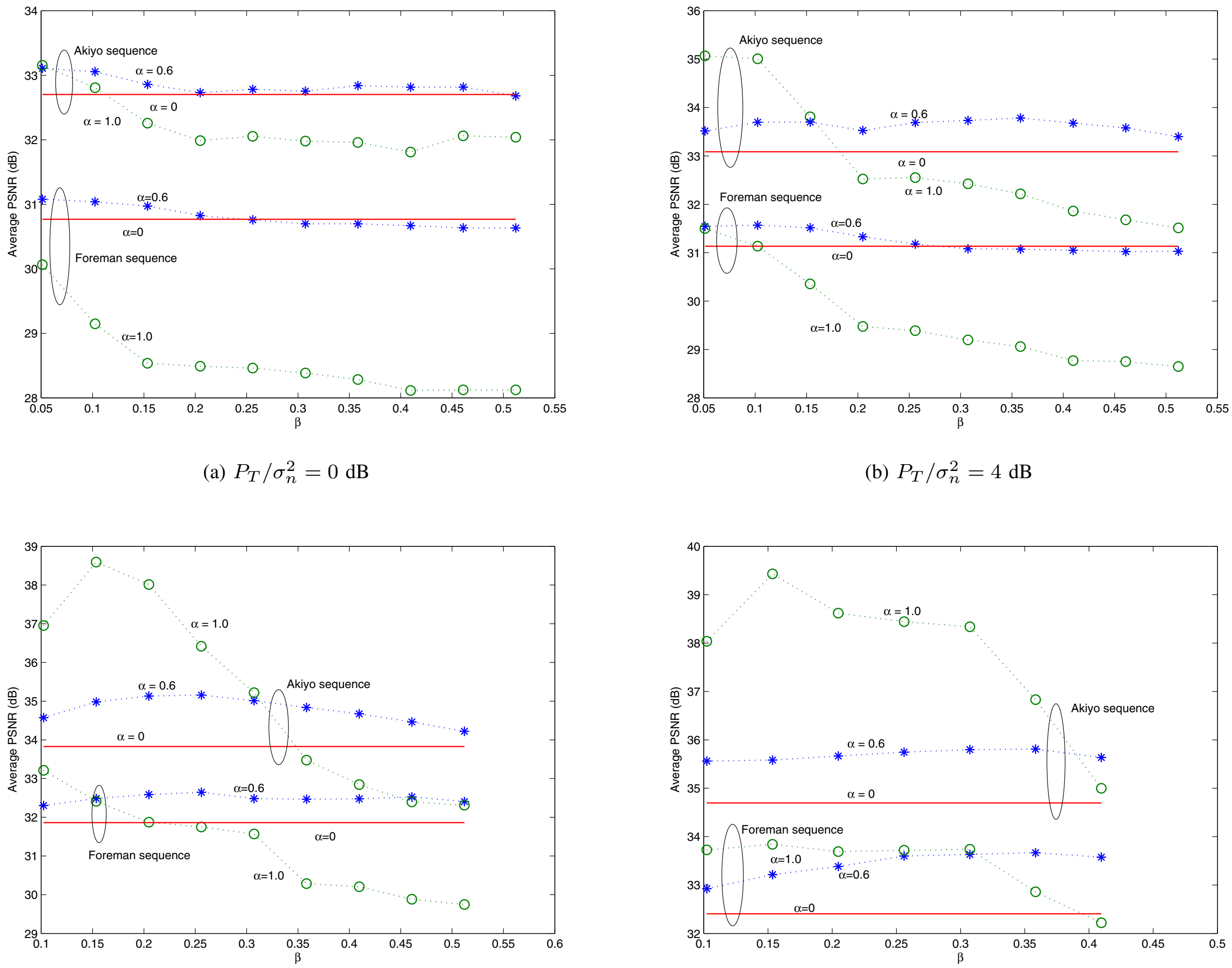
Therefore, instead of finding the sub-optimal $\underline{r}$ and $\underline{\phi}$ separately, the sub-optimal $\underline{s} = [s_1, s_2, s_3, \cdots s_N]$ can be found using the existing local search algorithm from [11].

In Fig. 3, the UEP policy and MIMO mode selection for the *"Foreman"*. We assume the leaky prediction parameter, α, is fixed at 1.0, and the partial prediction parameter, β, is set to 0.15. Average channel SNR, P_T/σ_n^2, is fixed at 8 dB, where P_T is the total transmit power. Then, for perfect channel estimation, the first 8 packets transmit the enhancement layer used for MCP. Since this part of the enhancement layer can result in error propagation if it is lost, additional protection is required. When there is imperfect channel estimation, more diversity gain and a lower channel code rate are required to protect the enhancement layer used for MCP, where the first 12 packets are involved. In particular, to reduce the loss probability for the beginning of the bitstream, spatial diversity is applied for the first 2 packets.

IV. Results and Discussion

In this section, we present simulation results in terms of average peak-signal-to-noise ratio (PSNR) for various leaky and partial prediction parameters, and a Rayleigh fading channel. For the simulation, we use MC-FGS video, which is implemented with an H.264 TML-9 codec for the base layer and an MPEG-4 FGS codec for the enhancement layer.

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.



(a) $P_T/\sigma_n^2 = 0$ dB

(b) $P_T/\sigma_n^2 = 4$ dB

(c) $P_T/\sigma_n^2 = 8$ dB

(d) $P_T/\sigma_n^2 = 12$ dB

Fig. 5: Average PSNR performance versus partial prediction parameter, β, where P_T/σ_n^2 represents the ratio of the total transmit power and noise variance. The solid line represents the average PSNR without using MC-FGS.

Both partial and leaky prediction schemes are incorporated in the enhancement layer MCP loop. The first 150 frames of the *"Foreman"* video sequence, consisting of a single intra-frame (I-frame) and 149 predicted frames (P-frames), are encoded. Before transmitting the bitstream through the MIMO wireless channel, it is packetized and channel-encoded by rate compatible punctured convolutional (RCPC) codes. To ensure reliable delivery of the base layer, we use the lowest channel code rate and spatial diversity for the base layer bitstream. However, for the enhancement layer bitstream, its UEP policy and leaky/partial prediction parameters are jointly selected. Note that more protection is required for a packet which includes the enhancement layer bitstream involved in MCP or that has a larger leaky prediction parameter. We fix the size of a packet at 256 BPSK symbols, and 20 packets are available for each frame. A 4x4 MIMO antenna configuration is considered, and a minimum mean square error (MMSE) detection scheme is used for all the MIMO modes. We consider Jakes' model with a normalized Doppler frequency of 10^{-2} for the Rayleigh fading channel. At the transmitter, the power is evenly distributed among transmit antennas.

Fig. 4 presents the average PSNR versus various partial prediction parameters when the leaky prediction parameter is fixed at 0.7 and 1.0, under both perfect and imperfect channel estimation. Given the channel SNR, the optimal selection of the partial prediction parameter, β, is highlighted by the square($\square$). For all scenarios, it is shown that larger β can be chosen as the channel SNR increases. This can be explained by noting that more transmit power allows us to use more enhancement layers in the MCP loop. If we compare Fig. 4a with Fig. 4b, it can be observed that imperfect channel estimation results in decreasing partial prediction parameters. However, the choice of β is not sensitive since the propagated error can be forced to decay by the use of a leaky prediction

978-1-4244-4148-8/09/$25.00 ©2009
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2009 proceedings.

parameter, α, less than 1. In contrast, in Fig. 4c and 4d with α set to 1, it is seen that the choice of β is more sensitive with respect to the performance. Because the propagated error is not being forced to decay, a small value of β is preferred. Instead, if we choose a larger β than the optimal one, the corresponding performance is significantly degraded.

In Fig. 5, the average PSNR performance of the MC-FGS video with the partial and leaky prediction schemes is compared to that of a conventional FGS video under imperfect channel estimation for *"Foreman"* and *"Akiyo"* video sequences, where *"Akiyo"* has slower motion than *"Foreman"*. To implement conventional FGS video coding, α is set to 0. At low SNR, the MC-FGS video coding does not provide significant gain over conventional FGS video coding, since it suffers from error propagation even though a small value of β is chosen. However, by choosing a small α, comparable performance can be achieved. As the channel SNR increases, MC-FGS outperforms conventional FGS, even if a large value of the partial prediction parameter is used. Another observation is the sensitivity regarding the choice of prediction parameters. Especially, if α is close to 1, the performance seems to be more sensitive to the choice of β. For example, at an SNR of 8 dB, when $\alpha = 1$, the average PSNR corresponding to the partial prediction parameters of .1 and .35 are quite different for both video sequences. Therefore, for a larger leaky prediction parameter, the partial prediction parameter needs to be chosen more carefully. In addition, higher gain is achieved in a slow motion video sequence such as *"Akiyo"* since the leaky/partial prediction schemes are exploited when the frames are highly correlated. In a slow motion video, larger β is also preferred. This shows that the selection of leaky and partial prediction parameters is affected by the characteristics of the video sequence.

Finally, the PSNR performance of conventional FGS video coding is compared with that of MC-FGS video coding using leaky and partial prediction schemes with or without MIMO mode selection in Fig. 6, where the parameters are selected jointly. This figure shows that the joint selection of parameters can result in a significant gain over a conventional FGS over a wide range of channel SNR.

V. Conclusion

In this paper, we studied the transmission of an MC-FGS video bitstream over a MIMO wireless channel under the condition of imperfect channel estimation. We proposed a UEP policy consisting of FEC and MIMO mode selection per packet for the enhancement layer of MC-FGS by exploiting the fundamental tradeoff between multiplexing and diversity. To compensate for reference mismatch and the resulting error propagation, leaky and partial prediction schemes were applied in the MCP loop. Then we investigated the average PSNR performance with various choices of the leaky and partial prediction parameters. Simulation results showed that the joint control of prediction parameters and UEP could enhance the system performance significantly. Moreover, the performance was more sensitive to the selection of partial prediction parameter rather than the leaky prediction parameter with respect to the given channel SNR and channel estimation accuracy. Therefore, knowledge regarding channels needs to be considered in the selection of the prediction parameters.

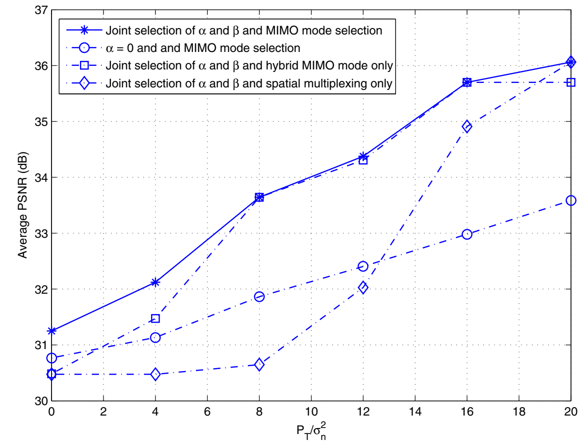


Fig. 6: Comparison of average PSNR for the conventional FGS and MC-FGS video coding with joint selection of leaky and partial prediction parameters under imperfect channel estimation.

References

[1] M. van der Schaar and H. Radha, "Adaptive Motion-compensation Fine Granular-scalability (AMC-FGS) for Wireless Video," *IEEE Trans. CSVT.,* vol. 12, no. 6, pp. 360-371, Jun. 2002.

[2] F. Wu, S. Li, and Y-Q. Zhang, "A Framework for Efficient Progressive Fine Granularity Scalable Video Coding," *IEEE Trans. CSVT.,* vol. 11, no. 3, pp. 332-344, Mar. 2001.

[3] H.-C. Huang, C.-N. Wang, and T. Chiang, "A Robust Fine Granularity Scalability Using Trellis-based Predictive Leak," *IEEE Trans. CSVT.,* vol. 12, no. 6, pp. 372-385, Jun. 2002.

[4] Y. S. Chan, P. C. Cosman, and L. B. Milstein "A Multiple Description Coding and Delivery Scheme for Motion-Compensated Fine Granularity Scalable Video," *IEEE Trans. Image Process.,* vol. 17, no. 8, pp. 1353-1366, Aug. 2008.

[5] A. Kwasinski and K. J. R. Liu, "Optimal Unequal Error Protection with User Cooperation for Transmission of Embedded Source-Coded Images," in *Proc. IEEE Int. Conf. Image Processing,* pp. 717-720, Oct. 2006.

[6] J. K. Cavers, "An Analysis of Pilot Symbol Assisted Modulation for Rayleigh Fading Channels," *IEEE Trans. Veh. Tech.,* vol. 40, no. 4, pp. 686-693, Nov. 1991.

[7] K. Lee and J. Chun, "Symbol Detection in V-BLAST Architectures under Channel Estimation Errors," *IEEE Trans. Wireless Comm.,* vol. 6, no. 2, pp. 593-597, Feb. 2007.

[8] H. Jafarkhani, "A Quasi-orthogonal Space-time Block Code," *IEEE Commun. Letters,* vol. 49, no. 1, pp. 1-4, Jan. 2001.

[9] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications,* U.K. : Cambridge Univ. Press, 2003

[10] 3GPP, "Technical Specification Group Radio Access Network : Multiple Input Multiple Output in UTRA," 3GPP, Technical Report, 25.876 V1.7.0, Aug. 2004.

[11] V. Stanković, R. Hamzaoui, and D. Saupe, "Fast Algorithm for Rate-Based Optimal Error Protection of Embedded Codes," *IEEE Trans. Commun.,* vol. 51, no. 11, pp. 1788-1795, Nov. 2003.