[6]

Robust Transmission of Packet-Based H.264/AVC Video With Data Partitioning Over DS-CDMA Wireless Channels

Anand V.S. Mantravadi, Manu Bansal, Lisimachos P. Kondi

University at Buffalo, The State University of New York
Dept. of Electrical Engineering
Buffalo, NY 14260.

ABSTRACT

In this paper, we address the problem of robust transmission of packet based H.264/AVC video over direct sequence-code division multiple access (DS-CDMA) channels. H.264 based data partitioning is used to produce video packets of unequal importance with regards to their need in terms of the decoded video quality. In the proposed transmission system, the data partitioned video packets are packetized as per IP/UDP/RTP protocol stack and are sorted into different levels for giving unequal error protection (UEP) using Rate Compatible Punctured Convolutional (RCPC) codes. Constant size framing is done at the link layer and Cyclic Redundancy Check header (CRC) is attached for error detection. Link layer buffering and packet interleaving schemes are proposed to improve the efficiency of the system. A multipath Rayleigh fading channel with Additive White Gaussian Noise (AWGN) and interference from other users is considered at the physical layer. The link layer frames are channel encoded, spread and transmitted over the channel. The received data is despread/demodulated using the Auxiliary Vector (AV) filter or RAKE matched filter (RAKE-MF) receiver and subsequently channel and source decoded. Our experimental results show the effectiveness of using data partitioning for wireless transmissions when compared to the system not using data partitioning. Also the superior interference mitigation capabilities of AV receiver is shown in comparison to the RAKE-MF receiver.

Keywords: H.264/AVC, Data partitioning, DS-CDMA, Auxiliary Vector (AV) receiver, Unequal error protection, Wireless video transmission

1. INTRODUCTION

Efficient transmission of video through different networks not only needs encoding efficiency but also seamless and easy integration of the coded video into all present and future transport protocol architectures. In Ref. 1 and 2, the performance of two previous video coding standards namely H.263 and MPEG-4 were discussed. These video codecs were designed around a bit oriented format and had certain limited error resilience features. Thus in packet oriented transmission systems, such codecs introduced dependency between packets and hence the loss of certain packets would render other packets undecodable. Also, such codecs need source based re-encoding in heterogeneous network environments. This led to a system-specific decoder design since the encoder had to adapt to the network environment and so the decoder had to be designed accordingly.

The concept of distinguishing between video encoding and video transmission using H.264/AVC was introduced in Ref. 3 and 4. H.264/AVC introduced two conceptual layers called *Video Coding Layer* (VCL) and *Network Abstraction Layer* (NAL). Figures 1 and 2 show the H.264 encoder block diagram and conceptual layered structure in H.264. The VCL employs techniques like integer transforms, multiple block size motion estimation, multi-frame motion prediction, quarter pixel prediction, different intra encoding modes, Context Adaptive Binary Arithmetic Coding (CABAC), deblocking filter etc. to achieve high compression efficiency, using up to half the bit rate for the same video quality compared to previous standards.[3] The NAL is responsible for the transport of the video encoded by the VCL over a wide range of transport media and provides a wide range of error resilience

Further author information: (Send correspondence to M.B.)
A.V.S.M.: E-mail: sm236@eng.buffalo.edu
M.B.: E-mail: mbansal@eng.buffalo.edu
L.P.K.: E-mail: lkondi@eng.buffalo.edu

Visual Communications and Image Processing 2006, edited by John G. Apostolopoulos, Amir Said,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6077, 607725, © 2006 SPIE-IS&T · 0277-786X/06/$15

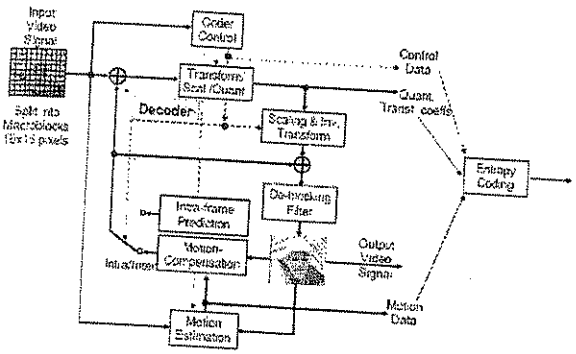

Figure 1. H.264 Encoder Block Diagram.

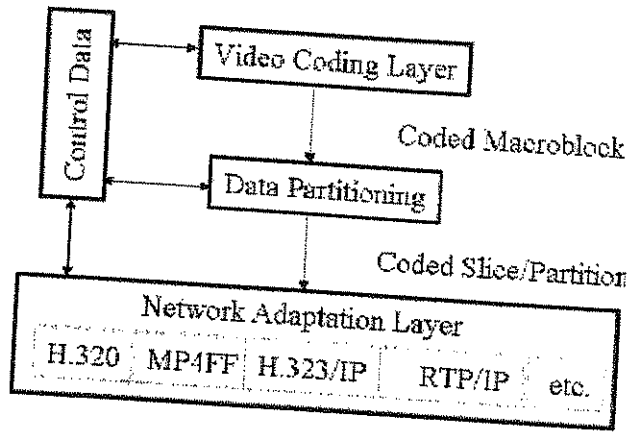

Figure 2. Conceptual Layered Structure in H.264.

features compared to previous standards. The interface between VCL and NAL is obtained by the *slice layer*. Slice is a group of macroblocks which forms the fundamental data structure of the VCL and can be decoded independently. A NAL unit (NALU)[5] represents a slice encoded data along with additional header and hence the NAL is responsible for encapsulation of VCL packets for transparent delivery over heterogeneous networks, framing, timing issues, and synchronization. Currently the most well defined NAL is for the IP/UDP/RTP based networks[6,7] and the same has been adopted in this paper.

The problem of video transmission over CDMA has been addressed in Ref. 8 and 9 and the performance of such systems for superior transmission of video over wireless has been established. Joint source-channel coding and optimal rate allocation for MPEG4 and H.263 based video has also been extensively addressed in the literature. In Ref. 6 and 10, the performance of H.264 over wired and wireless networks was studied. However, the area of packet based wireless transmissions of H.264/AVC video over DS-CDMA channels is still not addressed properly. In this work, we propose a robust video transmission system employing error resilient tools such as data partitioning and unequal error protection. Also, we discuss the channel model in detail and propose the use of Auxiliary Vector (AV) filter as an efficient CDMA receiver with superior interference mitigation capabilities.

2. CONCEPT OF DATA PARTITIONING

The IP/UDP/RTP NAL structure operates in two modes: single slice mode where data for particular slice is encapsulated in one packet while in data partitioning mode the encoded slice data is separated as per the video syntactical elements and their importance for decoding the slice. There are about 20 syntactical elements defined in Ref. 4. In the current H.264 standard, the three-partition approach is used for data partitioning[3,6,10] and Ref. 11 discuss data partitioning in more detail.

- Partition A contains 7 elements of slice header information: HEADER (I), PVPE (2), MBTYPE (3), REFFRAME (4), MVD (5), BFRAME (6), EOS (7). It requires highest channel protection.
- Partition B contains 7 of Intra-coded elements: MTRAPREDMODE (8), CBP-INTRA (9), LUM-DC-INTRA (10), CHF-DC-INTRA (11), LUM-AC-INTRA (12), CHF-AC-INTRA (13), DELTA-QUANT-INTRA (14).
- Partition C contains 6 Inter-coded elements: CBP-INTER(15), LUM-DC-INTER (16), CHF-DC-INTER (17), LUM-AC-INTER (18) CHF-AC-INTER (19), DELTA-QUANT-INTER (20).

Clearly, partition A contains the most important information for decoding the slice and needs the highest level of protection. Partition C requires the less protection when compared to partition B since the latter carries the INTRA elements which prevent the propagation of error at the decoder. Hence it's logical to use unequal error protection (UEP) for these partitions. Detailed procedures and performance gains obtained with UEP are also discussed in detail in Ref. 12 and 13.

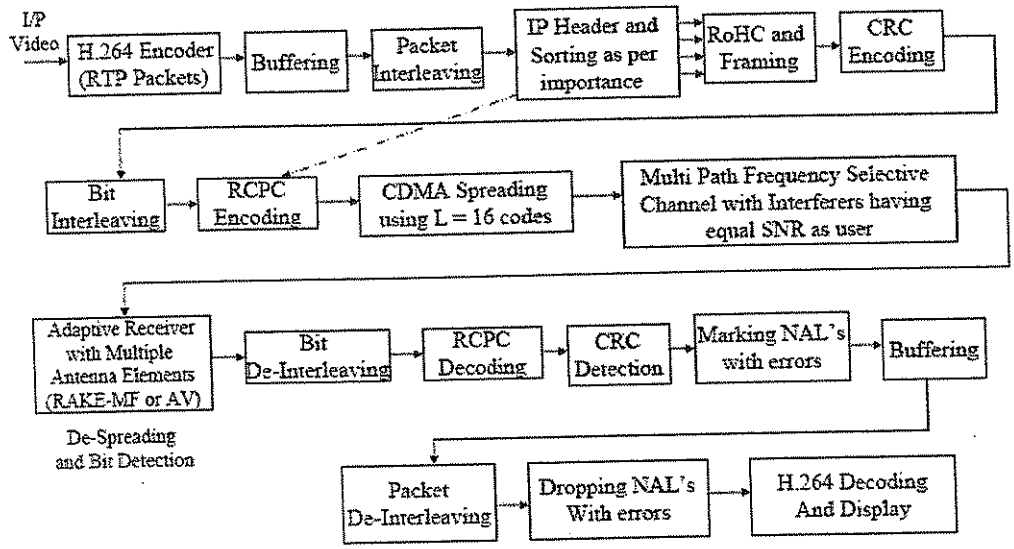

Figure 3. Block Diagram of the Transmission System.

3. VIDEO TRANSMISSION SYSTEM

In this section, the end to end system setup for transmission of the H.264 NAL units over DS-CDMA channels is discussed in detail. Link layer buffering and packet interleaving techniques are introduced and their need to improve system robustness is emphasized. Detailed channel modeling, encoding and receiver types are also discussed. Figure 3 gives a block diagram representation of the entire end-to-end system model.

For packet-switched wireless services, 3GPP/3GPP2 agreed on an IP-based protocol stack with RTP at the application layer and UDP at the transport layer. Thus each H.264 NALU is encapsulated in a RTP/UDP/IP protocol header. Detailed structure of the headers for RTP, UDP and IP is discussed in Ref. 7, 14 and 15. Figure 4 shows the typical packetization of a NALU encapsulated in RTP/UDP/IP through the 3GPP2 user plane protocol stack. The length of the RTP/UDP/IP header is 40 bytes (12+8+20) and so this is adds to a header overhead. Robust header compression techniques (RoHC)[16] are used to compress this header to three bytes. After RoHC, this RTP/UDP/IP packet is encapsulated into one packet data convergence protocol/point-to-point protocol (PDCP/PPP) packet that becomes a radio link control (RLC) service data unit. The RLC protocol can operate is three modes: 1) Transparent; 2) Unacknowledged; and 3) Acknowledged mode. We have used the unacknowledged mode for the RLC protocol where no retransmissions or data delivery is guaranteed. CRC error detection is used for all the modes on the physical layer. CRC algorithm is discussed in detail in Ref. 17 and we have used the standard 32-bit CRC polynomial given in the IEEE 802.XX standard.

3.1. Link Layer Buffering - Efficiency vs Error Resilience Tradeoff

NAL units are of varying lengths and their lengths depend upon the encoded video content. This also makes RLC-service data units (SDU's) of varying lengths and it is up to the RLC protocol to enclose these packets in constant size frames. In our simulations, a link layer frame size of 100 bytes was considered which is a typical value used for such wireless systems.[10] The most straightforward way is to pad each packet with unnecessary bits so that the loss of a single link layer frame will effect only one NAL unit. However, it was found in our simulations that such packetization will lead to excessive padding bits and huge overhead. Another way is to keep the flow of these SDU's continuous by buffering the NAL units and making frames (of constant size) from this continuous data stream. Intuitively, for a large buffer size, the padding bits will be reduced, but the decoder has to wait for all the frames in the buffer before it can decode the video. This could hamper real-time video applications. However, by proper selection of the buffer size, we can minimize the padding bits overhead and

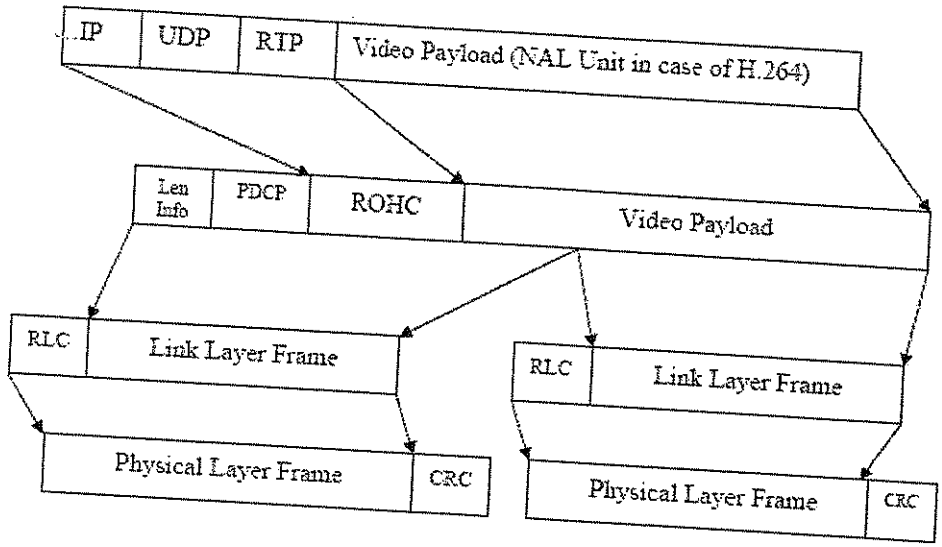

Figure 4. Packetization through 3GPP2 user plane protocol stack.

also meet the delay constraints in the network. Also the amount of padding bits for UEP cases will be more that EEP. However proper buffering of the NAL units can help us to implement an optimal system under this setup. In this work, the buffer size is defined as the number of frames whose packets are buffered before transmission, i.e., number of buffered NAL units is always a multiple of number of NAL units per frame.

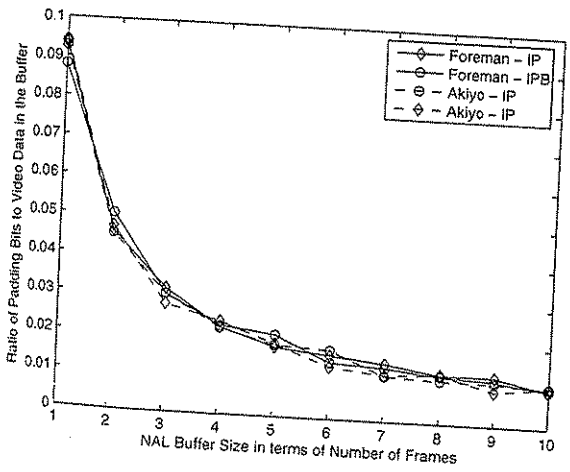

Figure 5. Ratio of Padding Bits to Video Data in the Buffer VS Buffer Size in Number of Frames for EEP schemes.

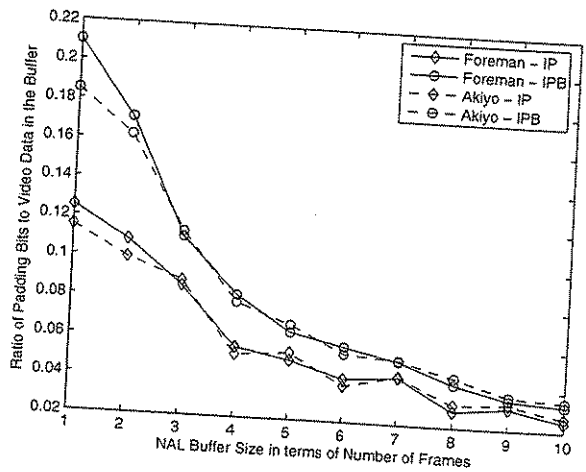

Figure 6. Ratio of Padding Bits to Video Data in the Buffer VS Buffer Size in Number of Frames for UEP schemes with Data Partitioning.

Figure 5 and 6 show the behavior of ratio of padding bits to actual video data in the buffer with increase in buffer size (in frames) for EEP and UEP cases, respectively. Both "Foreman" and "Akiyo" sequences have been investigated (for both IP and IPB encoding modes). We can see that the behavior is same irrespective of the video sequence and the encoding mode. From figure 5 we can see that there is not much reduction in the padding bits ratio after a buffer size of five frames. The same was also observed for other frame size like 200, 300 and 400 bytes. Also for a five frame buffer size, the difference between the UEP and EEP cases is not significant.

Thus we can conclude that a buffer size of five frames is suitable for our proposed system for all video sequences and encoding modes.

3.2. Packet Interleaving

As discussed in the previous section, if we are using a buffer to have a continuous flow of packets and reduce the padding bit overhead, then the loss of a single link layer frame can result in the loss of two successive NAL units. Thus, there is a possibilty of a burst packet loss which could leave the video undecodable. To avoid this, we introduce a *packet interleaver*, which scrambles the NAL units in the buffer in an predefined manner before the link layer framing is done. This separates successive NAL units and hence reduces the possibility of a burst loss of packets.

Depth of the interleaver is defined as the number of packets that come between two successively numbered packets. Clearly, the greater is the depth of the interleaver, the more is protection against burst packet loss, but it introduces more delay. In our simulations, we have encoded the video to produce nine NAL units per frame and since we buffer five frames, the number of packets in the buffer is 45. And since the decoder has to wait for these packet to arrive, we have chosen an interleaver depth of 15 ($15 \times 3 = 45$) which is the maximum possible for our buffer size.

It has to be observed here that the idea behind packet interleaving is different from bit interleaving. While the former is used at the IP layer level to prevent burst packet losses, the latter is used at the physical layer level to spread out any burst bit errors so that the channel coding is able to correct them.

3.3. Physical Layer Transmission

Cyclic Redundancy Check (CRC) code is attached to the fixed size link layer frames which make the physical layer frames and are the basic unit of transmission over the physical channel. Each frame unit is channel encoded unequally using the Rate Compatible Punctured Convolutional (RCPC) codes. The channel code rate for each frame depends on the partition it belongs to in the H.264 encoded bitstream. This channel encoded frame is then spread/modulated and is transmitted over a multipath Rayleigh fading channel with interferers. The received data is despread/demodulated using the Auxiliary Vector (AV) receiver or conventional RAKE matched filter (RAKE-MF) receiver and channel decoded using Viterbi decoding scheme. The received frame is then checked for errors using the CRC detection scheme. If the frame is erroneous, then all the IP layer packets to which this frame belongs are marked to be having errors. Before feeding to the H.264 decoder, all the NALU's that are marked with errors are dropped and the remaining NALU's are decoded to get output video.

4. CHANNEL CODING AND RECEIVED SIGNAL MODEL

In this section, we will discuss the details of the channel coding, DS-CDMA channel model and receivers that has been used in this work.

4.1. Channel Coding: RCPC Codes

RCPC codes form a class of convolutional codes that are obtained by *puncturing* the output of a "mother" convolutional code. Convolutional coding is accomplished by convolving the source data with a convolutional matrix G. In essence, rather than having a number of channel code symbols for a corresponding block of source symbols as in linear block codes, convolutional coding generates one codeword for the entire source data. The rate of a convolutional code is defined as k/n where k is the number of input bits and n is the number of output bits. *Puncturing* is the process of deleting bits from the output sequence in a predefined manner so that fewer bits are transmitted than in the original coder leading to a higher coding rate. The idea of puncturing was extended to include the concept of rate compatibility.[18] Rate compatibility requires that a higher rate code be a subset of a lower rate code, or that lower protection codes be embedded into higher protection codes. This is accomplished by puncturing a "mother" code of rate $1/n$ to achieve higher rates. At the receiver, RCPC codes are decoded using the Viterbi algorithm which is a maximum-likelihood sequence estimation technique.

4.2. Channel Model and the Received Signal

In order to design good communication systems, it is essential to have knowledge about the transmission characteristics of the channel or medium involved. This can be obtained by developing models of propagation channels that take into account the following radio channel parameters.

4.2.1. Radio Channel Parameters

The parameters used to characterize multipath wireless channels are the time delay spread, the coherence bandwidth and the number of resolvable multipaths. The time delay (T_{delay}) spread is a measure of the length of the impulse response for the multipath wireless channel. Time delay spread causes Inter-symbol interference (ISI) and complicates the receiver design. For an exponentially distributed delay power profile, the coherence bandwidth can be written as

$$B_{coherence} = 1/(2\pi T_{delay}). \tag{1}$$

If two frequencies lie within the same coherence bandwidth they may experience correlated fading. If the bandwidth of the symbol (B_{symbol}) is such that

$$B_{symbol} << B_{coherence}, \tag{2}$$

the channel model is a frequency non-selective fading narrowband channel. If

$$B_{symbol} >> B_{coherence}, \tag{3}$$

the channel is a frequency selective fading wideband channel.

The number of resolvable paths in a multipath wireless channel is given by

$$\rho = \lfloor T_{delay}/T_{coherence} \rfloor + 1, \tag{4}$$

where is the largest integer that is less than or equal to x. In narrowband systems, there is usually one resolvable path. For wideband systems, a number of resolvable paths exist. For each resolvable path there exists an independent and random attenuation, phase shift and time delay.

4.2.2. Received Signal

The time-varying multipath propagation characteristics of (mobile) CDMA radio communications suggest a channel model that undergoes frequency-selective fading. Under uniform linear antenna-array reception, the baseband received signal at each antenna element m, $m = 1, \ldots, M$, is viewed as the aggregate of the multipath received CDMA signal of interest with signature code $\mathbf{s}_0 \in \{\pm 1\}^L$ (if T is the symbol period and T_c is the chip period, then $L = T/T_c$), $K-1$ multipath received CDMA interferers with unknown signatures $\mathbf{s}_k \in \{\pm 1\}^L$, $k = 1, \ldots K-1$, and white Gaussian noise. For notational simplicity and without loss of generality, we adopt a chip-synchronous signal formulation. We assume that the multipath spread is of the order of a few chip intervals, P, and since the signal is bandlimited to $B = 1/2T_c$ the lowpass channel can be represented as a tapped delay line with $P+1$ taps spaced at chip intervals T_c. After conventional chip-matched filtering and sampling at the chip rate over a multipath extended symbol interval of $L+P$ chips, the $L+P$ data samples from the mth antenna element, $m = 1, \ldots, M$, are organized in the form of a vector $\mathbf{r}_m$ given by

$$\mathbf{r}_m = \sum_{k=0}^{K-1} \sum_{p=0}^{P} c_{k,p} \sqrt{E_k} (b_k \mathbf{s}_{k,p} + b_k^- \mathbf{s}_{k,p}^- + b_k^+ \mathbf{s}_{k,p}^+) \mathbf{a}_{k,p}[\mathrm{m}] + \mathbf{n} \ , \ m = 1 \ldots, M, \tag{5}$$

where, with respect to the kth CDMA signal, E_k is the transmitted energy per chip, b_k, b_k^-, and b_k^+ are the present, the previous, and the following transmitted bit, respectively, and $\{c_{k,p}\}$ are the coefficients of the frequency-selective slowly fading (quasi-static) channel modeled as independent zero-mean complex Gaussian random variables that are assumed to remain constant over a few symbol intervals. $\mathbf{s}_{k,p}$ represents the 0-padded by P, p-cyclic-shifted version of the signature of the kth SS signal $\mathbf{s}_k$, $\mathbf{s}_{k,p}^-$ is the 0-filled $(L-p)$-left-shifted version of $\mathbf{s}_{k,0}$, and $\mathbf{s}_{k,p}^+$ is the 0-filled $(L-p)$-right-shifted version of $\mathbf{s}_{k,0}$. Finally, $\mathbf{n}$ represents additive complex

Gaussian noise with mean $\mathbf{0}$ and autocorrelation matrix $\sigma^2\mathbf{I}_M$, and $\mathbf{a}_{k,p}[\mathrm{m}]$ is the mth coordinate of the kth CDMA signal, pth path, array response vector:

$$\mathbf{a}_{k,p}[\mathrm{m}] = e^{j2\pi(m-1)\frac{\sin\theta_{k,p}d}{\lambda}}\ , m = 1,\ldots,M, \tag{6}$$

where $\theta_{k,p}$ identifies the angle of arrival of the pth path of the kth CDMA signal, λ is the carrier wavelength, and d is the element spacing (usually $d = \lambda/2$).

To avoid in the sequel cumbersome 2-D data notation and filtering operations, the $(L+P)\times M$ space-time data matrix $[\mathbf{r}_1\ \mathbf{r}_2\ \cdots\ \mathbf{r}_M]$ are "vectorized" at this point by sequencing all matrix columns in the form of a single $(L+P)M$-long column vector:

$$\mathbf{r}_{(\mathrm{L+P})\mathrm{M}\times 1} = Vec\{[\mathbf{r}_1,\ \mathbf{r}_2,\ \cdots,\ \mathbf{r}_M]_{(\mathrm{L+P})\times\mathrm{M}}\}\ . \tag{7}$$

From now on, $\mathbf{r}$ denotes the joint space-time data in the $C^{(L+P)M}$ complex vector domain.

For conceptual and notational simplicity the vectorized space-time data equation may be rewritten as follows:

$$\mathbf{r} = \sqrt{E_0}b_0\mathbf{w}_{\text{R-MF}} + \mathbf{i} + \mathbf{n} \tag{8}$$

where $\mathbf{w}_{\text{R-MF}} = E_{b_0}\{\mathbf{r}\, b_0\} = Vec\{[\ \sum_{p=0}^{P} c_{0,p}\mathbf{s}_{0,p}, \mathbf{a}_{0,p}[1],\ \cdots,\ \cdots,\ \sum_{p=0}^{P} c_{0,p}\mathbf{s}_{0,p}\mathbf{a}_{0,p}[\mathrm{M}]\ ]\}$ is the effective space-time signature of the CDMA signal of interest (signal *0*) and $\mathbf{i}$ identifies comprehensively both the inter-symbol and the CDMA interference present in $\mathbf{r}$ ($E_{b_0}\{\cdot\}$ denotes statistical expectation with respect to b_0). The subscript R-MF in the effective S-T signature notation is used to make a direct association with the RAKE matched-filter time-domain receiver that is known to correlate the signature $\mathbf{s_0}$ with $P+1$ size-L shifted windows of the received signal (that correspond to the $P+1$ paths of the channel), appropriately weighted by the conjugated channel coefficients $c_{0,p}, p = 0,\ldots,P$. In this notation, the generalized S-T RAKE operation corresponds to linear filtering of the form $\mathbf{w}_{\text{R-MF}}^{\mathrm{H}}\mathbf{r}$, where H denotes the Hermitian operation.

5. DS-CDMA RECEIVERS - AUXILIARY VECTOR (AV) FILTER AND RAKE-MF

5.1. Auxiliary-Vector Adaptive Filtering

The AV receiver is chosen based on the realistic channel fading rates that limit the data record available for receiver adaptation. Under small sample support adaptation, AV filter short-data-record estimators have been shown to exhibit superior bit error rate (BER) performance in comparison to least mean squares (LMS), recursive least squares (RLS), sample matrix inversion (SMI), diagonally-loaded SMI, or multistage nested Wiener filter implementations. The AV algorithm generates a sequence of AV filters making use of two basic principles: (i) The maximum magnitude cross-correlation criterion for the evaluation of the auxiliary vectors (ii) The conditional mean-squared optimization criterion for the evaluation of the scalar AV weights. This algorithm is more clearly explained below.

The AV algorithm generates an infinite sequence of filters $\{\mathbf{w}_k\}_{k=0}^{\infty}$. The sequence is initialized at the S-T RAKE filter

$$\mathbf{w}_0 = \frac{\mathbf{w}_{\text{R-MF}}}{\|\mathbf{w}_{\text{R-MF}}\|^2}, \tag{9}$$

which is here scaled to satisfy $\mathbf{w}_0^H\mathbf{w}_{\text{R-MF}} = 1$. At each step $k+1$ of the algorithm, $k = 0,1,2,\ldots$, an "auxiliary" vector component $\mathbf{g}_{k+1}$ that is orthogonal to $\mathbf{w}_{\text{R-MF}}$ is incorporated in $\mathbf{w}_k$ and weighted by a scalar μ_{k+1} to form the next filter in the sequence,

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu_{k+1}\mathbf{g}_{k+1}. \tag{10}$$

The auxiliary vector $\mathbf{g}_{k+1}$ is chosen to maximize, under fixed norm, the magnitude of the cross-correlation between its output, $\mathbf{g}_{k+1}^H\mathbf{r}$, and the previous filter output, $\mathbf{w}_k^H\mathbf{r}$, and is given by

$$\mathbf{g}_{k+1} = \mathbf{R}\mathbf{w}_k - \frac{\mathbf{w}_{\text{R-MF}}^H\mathbf{R}\mathbf{w}_k}{\|\mathbf{w}_{\text{R-MF}}\|^2}\mathbf{w}_{\text{R-MF}} \tag{11}$$

where $\mathbf{R}$ is the input autocorrelation matrix, $\mathbf{R} = E\{\mathbf{r}\mathbf{r}^H\}$. The scalar μ_{k+1} is selected such that it minimizes the output variance of the filter $\mathbf{w}_{k+1}$ or equivalently minimizes the mean-square (MS) error between $\mathbf{w}_k^H\mathbf{r}$ and $\mu_{k+1}^*\mathbf{g}_{k+1}^H\mathbf{r}$. The MS-optimum μ_{k+1} is

$$\mu_{k+1} = \frac{\mathbf{g}_{k+1}^H \mathbf{R}\mathbf{w}_k}{\mathbf{g}_{k+1}^H \mathbf{R}\mathbf{g}_{k+1}}. \tag{12}$$

The AV filter recursion is completely defined by (9)-(12). More detailed discussion is given in Ref. 19 and 20. Theoretical analysis of the AV algorithm was pursued in Ref. 21. The results are summarized below in the form of a theorem.

THEOREM 1. *Let* $\mathbf{R}$ *be a Hermitian positive definite matrix. Consider the iterative algorithm of eqs. (9)-(12).*
(i) Successive *auxiliary vectors generated through (10)-(12) are orthogonal:* $\mathbf{g}_k^H\mathbf{g}_{k+1} = 0$, $k = 1, 2, 3, \ldots$ *(however, in general* $\mathbf{g}_k^H\mathbf{g}_j \neq 0$ *for* $|k-j| \neq 1$*).*
(ii) The generated sequence of auxiliary-vector weights $\{\mu_k\}$, $k = 1, 2, \ldots$, *is real-valued, positive, and bounded:* $0 < \frac{1}{\lambda_{\max}} \leq \mu_k \leq \frac{1}{\lambda_{\min}}$, $k = 1, 2, \ldots$, *where* $\lambda_{\max}$ *and* $\lambda_{\min}$ *are the maximum and minimum, correspondingly, eigenvalues of* $\mathbf{R}$*.*
(iii) The sequence of auxiliary vectors $\{\mathbf{g}_k\}$, $k = 1, 2, \ldots$, *converges to the* $\mathbf{0}$ *vector:* $\lim_{n\to\infty} \mathbf{g}_n = \mathbf{0}$*.*
(iv) The sequence of auxiliary-vector filters $\{\mathbf{w}_k\}$, $k = 1, 2, \ldots$, *converges to the minimum-variance-distortionless-response (MVDR) filter:* $\lim_{k\to\infty} \mathbf{w}_k = \frac{\mathbf{R}^{-1}\mathbf{w}_{R\text{-}MF}}{\mathbf{w}_{R\text{-}MF}^H\mathbf{R}^{-1}\mathbf{w}_{R\text{-}MF}}$.

An adaptive data-dependent procedure for the selection of the most appropriate member of the AV filter estimator sequence $\{\hat{\mathbf{w}}_k(D)\}$ for a given data record of size D is presented in Ref. 22. The following criterion summarizes the corresponding AV filter estimator selection rule.

CRITERION 1. *For a given data record of size* D*, the unsupervised (blind) J-divergence AV filter estimator selection rule chooses the estimator* $\hat{\mathbf{w}}_k(D)$ *with* k *auxiliary vectors where*

$$k = \arg\max_k \left\{\hat{J}_B(k)\right\} = \arg\max_k \left\{ \frac{4\left[\frac{1}{D}\sum_{d=1}^{D}\left|\mathrm{Re}\left[\hat{\mathbf{w}}_k^H(D)\mathbf{r}_d\right]\right|\right]^2}{\frac{1}{D}\sum_{d=1}^{D}\left|\mathrm{Re}\left[\hat{\mathbf{w}}_k^H(D)\mathbf{r}_d\right]\right|^2 - \left[\frac{1}{D}\sum_{d=1}^{D}\left|\mathrm{Re}\left[\hat{\mathbf{w}}_k^H(D)\mathbf{r}_d\right]\right|\right]^2} \right\}. \tag{13}$$

5.2. RAKE Matched Filter

We have also considered a space-time RAKE-MF receiver. A multiple antenna front end was used to receive the space-time signal. We also assumed that the receiver has the complete knowledge of the channel characteristics before the RAKE-MF filter is designed. We observed that the RAKE-MF performance is not good in the low SNR region. Though RAKE-MF is optimal for a Gaussian disturbance, the performance is bad in case of our channel where a practical interference situation is simulated.

6. SIMULATION PARAMETERS AND RESULTS

In this section we give an overview of the parameters that we considered in our simulations.

- The "Foreman" and "Akiyo" QCIF sequences were considered for simulation. Both IP and IPB encoding modes were used for encoding the sequences.
- Video sequences were encoded at a target source bit rate of 60 kbps. Each frame was divided into nine slices and hence nine NAL units. If data partitioning was employed, each NAL was further divided into three NAL units and hence 27 NAL units were produced for each frame.
- The allowed channel coding rates were 1/3, 1/2, 2/3, 4/5 and 8/9. Walsh Hadamard codes of length L = 16 were used as spreading sequences. The standard 32-bit CRC[17] polynomial used in IEEE 802.5 standard was used for CRC error detection scheme.

- A Rayleigh fading channel with three multipaths was considered as the channel model. Nine interferer having equal SNR as the user of interest were simulated. The number of antenna elements at the receiver were taken to be four.

For the unequal error protection case, the channel coding rates used are as below. Two different UEP schemes UEP-1 and UEP-2 have been implemented and tested against two equal error protection schemes EEP-1 and EEP-2, respectively.

- Rates allocated for UEP schemes for data partitioned video transmission. The target chip rates for UEP-1 and UEP-2 were close enough to EEP-1 and EEP-2, respectively, making this comparison of performance meaningful.
  - Level-1 which had the I-NALU's and partition-A was given a channel code rate of 1/3.
  - Level-2 which had the partition-B was given a channel code rate of 1/2.
  - Level-3 which had partition-C was given a channel code rate of 1/2 for UEP-1 and a rate of 8/9 for UEP-2.
- Under equal error protection scheme, constant channel code rates of 1/3 for EEP-1 and 1/2 for EEP-2 were considered. No data partitioning was used for equal error protection case.

Figures 7 and 8 give the packet loss rates for both RAKE-MF and AV receiver, respectively, over the simulated DS-CDMA channel. These are the loss rates generated for a specific link layer packet size of 100 bytes and they are not unique for all values. Clearly, these are dependent on the link layer packet size as well. We can clearly see that AV receiver has lower packet loss rates compared to RAKE-MF especially at low SNRs.

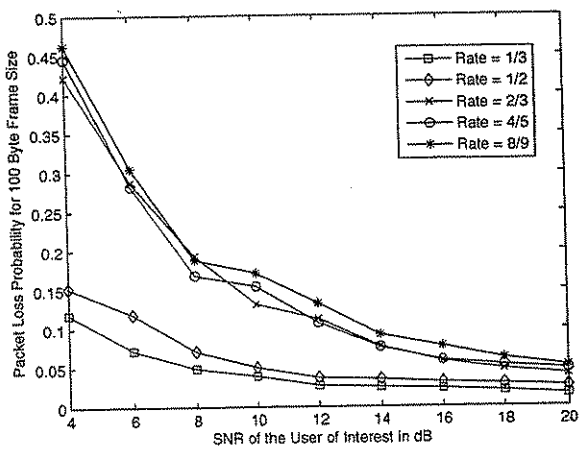

Figure 7. Packet Loss Rate for RAKE-MF Receiver for frame size of 100 bytes.

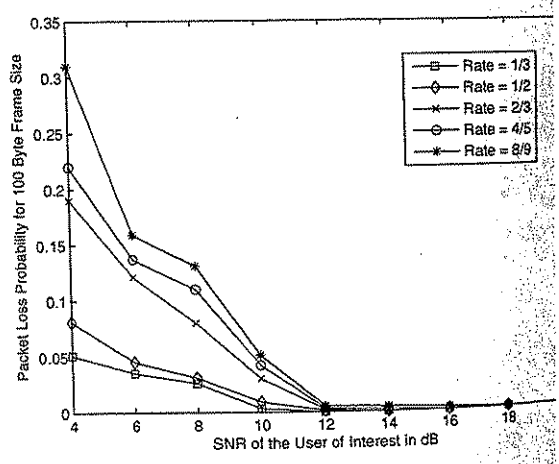

Figure 8. Packet Loss Rate for AV Receiver for frame size of 100 bytes.

6.1. Performance Comparison between UEP and EEP Schemes

From figures 9, 10, 11 and 12 we can clearly see that the UEP-1 and UEP-2 schemes with data partitioning outperform the EEP-1 and EEP-2 schemes, respectively, at low SNR's. At high SNR's of above 12 dB, the packet loss rates are small and the error concealment tools concealed the errors successfully. Hence the PSNR for UEP and EEP cases are almost the same. However, at low SNR's where the packet loss rates are really high the concealment tools fail causing a sharp drop in PSNR for the EEP cases. The UEP schemes are more robust under such conditions and hence result in higher decoded PSNR values when compared to EEP cases at low SNR's.

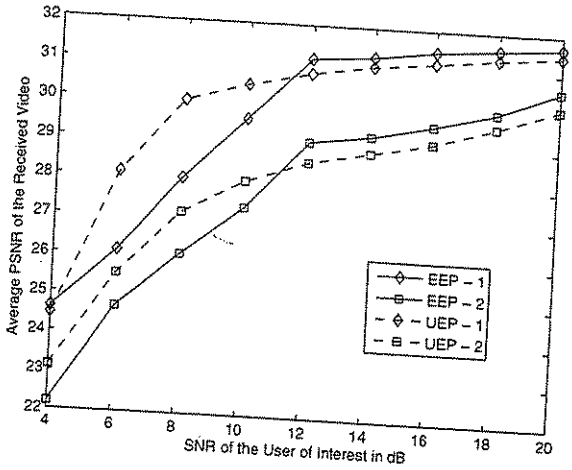

Figure 9. Comparison of Unequal and Equal Error Protection schemes for Foreman Sequence in **IP** Encoding Mode for RAKE-MF Receiver.

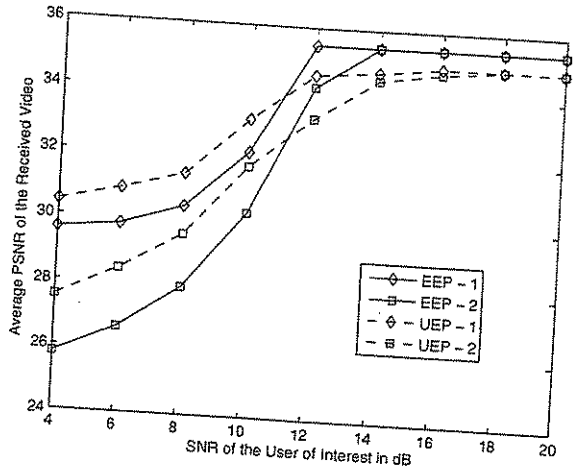

Figure 10. Comparison of Unequal and Equal Error Protection schemes for Foreman Sequence in **IP** Encoding Mode for Auxiliary Vector Filtering Receiver.

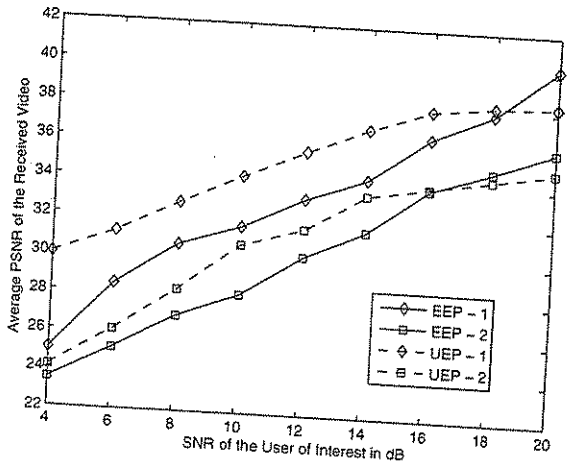

Figure 11. Comparison of Unequal and Equal Error Protection schemes for Akiyo Sequence in **IP** Encoding Mode for RAKE-MF Receiver.

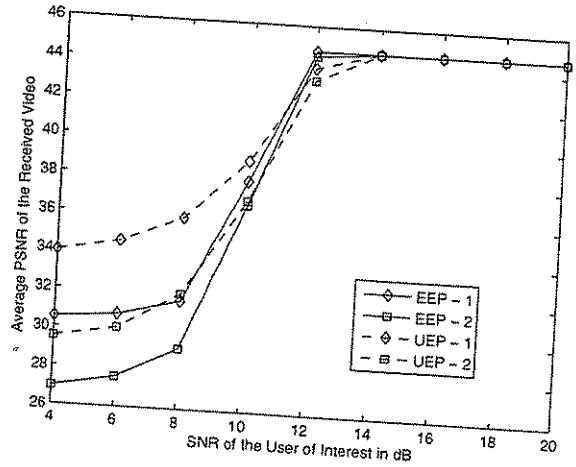

Figure 12. Comparison of Unequal and Equal Error Protection schemes for Akiyo Sequence in **IP** Encoding Mode for Auxiliary Vector Filtering Receiver.

6.2. Performance Comparison between RAKE-MF and AV Receivers

From figures 13, 14. 15 and 16 we can clearly see that AV receiver outperforms RAKE-MF receiver for all channel rates, SNR's and video encoding modes for both the video sequences. This clearly emphasizes the powerful interference suppression capabilities of AV and also its superior performance for short data records.

7. CONCLUSIONS AND FUTURE SCOPE

In this work, we proposed a robust transmission system for H.264/AVC encoded video and analyze it's performance over a IP/UDP/RTP based DS-CDMA wireless environment. We showed the effectiveness of using data partitioning and unequal error protection for low SNR values. We also established the advantage of using AV receiver as compared to conventional RAKE-MF receiver for such systems under different video sequences and channel conditions.

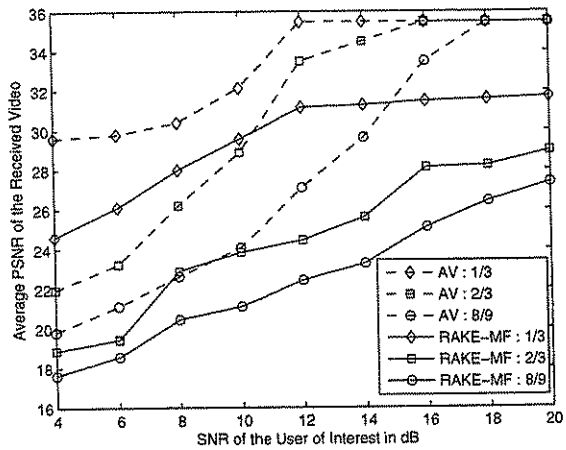

Figure 13. Comparison of Auxiliary Vector Filtering and RAKE-MF Receiver for Foreman Sequence in **IP** Encoding Mode and Different Channel Rates.

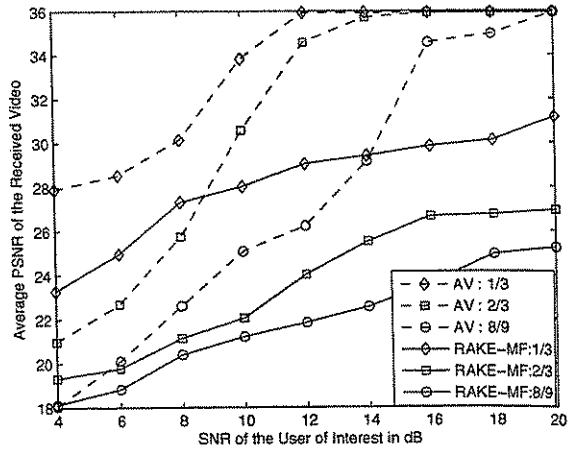

Figure 14. Comparison of Auxiliary Vector Filtering and RAKE-MF Receiver for Foreman Sequence in **IPB** Encoding Mode and Different Channel Rates.

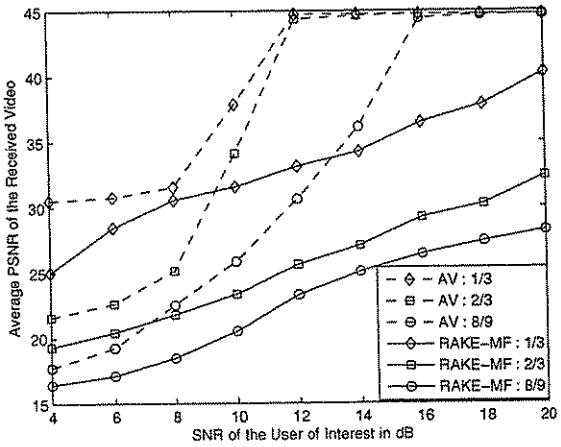

Figure 15. Comparison of Auxiliary Vector Filtering and RAKE-MF Receiver for Akiyo Sequence in **IP** Encoding Mode and Different Channel Rates.

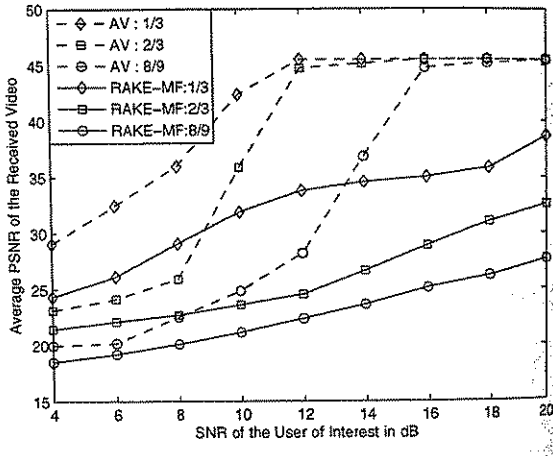

Figure 16. Comparison of Auxiliary Vector Filtering and RAKE-MF Receiver for Akiyo Sequence in **IPB** Encoding Mode and Different Channel Rates.

It is important to mention that the structure of the proposed system is transparent to other types of channel models and hence allow us to replace DS-CDMA with any other channel models of interest like Orthogonal Frequency Division Multiplexing (OFDM) system, Multiple-Input Multiple-Output (MIMO) system etc., and test its overall performance. Also, the current H.264 encoder used here does not include the concept of scalability, i.e., the encoded video does not have any scalable layers. If provided, it will definitely improve the performance of this codec over wireless channels and provide optimal source-channel rate allocation.

REFERENCES

1. ITU-T, "Video Coding for Low Bitrate Communicaitons, Version-1," in *ITU-T Recommendation H.263*, 1995.
2. "Coding of Audio Visual Objects - Part-2 - Visual," *ISO/IEC JTC1, ISO/IEC 14496-2* , 1999.
3. T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding

standard," in *IEEE Transactions on Circuits and Systems for Video Technology*, **13**, pp. 560 – 576, July 2003.
4. T. Wiegand, "Final committee draft: Editors proposed revisions," tech. rep., Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, Feb 2003.
5. T. Stockhammer, M. M. Hannuksela, and S. Wenger, "H.26L/JVT Coding Network Abstraction Layer and IP-Based Transport," in *ICIP - International Conference on Image Processing*, 2002.
6. S. Wenger, "H.264/AVC over IP," in *IEEE Transactions on Circuits and Systems for Video Technology*, **13**, pp. 645 – 656, July 2003.
7. H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," *RFC 1889* , 1996.
8. L. P. Kondi, D. Srinivasan, D. A. Pados, and S. N. Batalama, "Layered Video Transmission over Multirate DS-CDMA Wireless Systems," *IEEE Transactions on Circuits and Systems for Video Technology* , to appear.
9. D. A. P. D. Srinivasan, L. P. Kondi, "Scalable Video Transmission Over Wireless DS-CDMA Channels Using Minimum TSC Spreading Codes," *IEEE Signal Processing Letters*, **11**, pp. 836–840., Oct 2004.
10. T. Stockhammer, M. H. Hannuksela, and T. Wiegand, "H.264/AVC in Wireless Environments," in *IEEE Transactions on Circuits and Systems for Video Technology*, **13**, pp. 657 – 673, July 2003.
11. N. D. Dao and W. A. C. Fernando, "Channel coding for H.264 video in constant bit rate transmission context over 3G mobile systems," in *Proceedings of the 2003 International Symposium on Circuits and Systems, 2003*, **2**, pp. 896–899, May 2003.
12. J. Kim, R. Mersereau, and Y. Altunbasak, "Error-resilient image and video transmission over the internet using unequal error protection," in *IEEE Transactions on Image Processing*, **12**, pp. 121–131, Feb 2003.
13. M. Gallant and F. Kossentini, "Rate-distortion optimized layered coding with unequal error protection for robust Internet video," in *IEEE Transactions on Circuits and Systems for Video Technology*, **11**, pp. 357–372, March 2001.
14. "User Datagram Protocol," *RFC 768* , 1980.
15. J. Postel, "Internet Protocol," *RFC 793* , 1981.
16. H. Hannu, L. E. Jonsson, R. Hakenberg, T. Koren, K. Le, Z. Liu, A. Martensson, A. Miyazaki, K. Svanbro, T. Wiebke, T. Yoshimura, and H. Zheng, "Robust header compression (ROHC): Framework and four profiles: RTP, UDP, ESP and Uncompressed," *RFC 3095* , July 2001.
17. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Cyclic Redundancy and Other Checksums*, ch. 20.3, pp. 888–895. Cambridge, England: Cambridge University Press, 1992.
18. J. Hagenauer, "Rate-Compatible Punctured Convolutional Codes (RCPC codes) and their Applications," in *IEEE Transactions on Communications*, **36**, pp. 389 – 399, April 1988.
19. A. Kansal, S. N. Batalama, and D. A. Pados, "Adaptive maximum SINR RAKE filtering for DS-CDMA multipath fading channels," *Special Issue on Signal Processing for Wireless Communications* **16**, pp. 1765 – 1773, Dec 1998.
20. J. S. Goldstein, I. S. Reed, and L. L. Sharf, "A multistage representation of the Wiener filter-based on orthogonal projections," *IEEE Transactions on Information Theory* **44**, pp. 2943 – 2959, Nov 1998.
21. D. A. Pados and G. N. Karystinos, "An iterative algorithm for the computation of the MVDR filter," *IEEE Transactions on Signal Processing* **49**, pp. 290–300, Feb 2001.
22. H. Qian and S. N. Batalama, "Data-record-based criteria for the selection of an auxiliary-vector estimator of the MMSE/MVDR filter," *IEEE Transactions on Communications* **51**, pp. 1700–1708, Oct 2003.