Transactions Letters

Layered Video Transmission Over Wireless Multirate DS-CDMA Links

Lisimachos P. Kondi, *Member, IEEE*, Deepika Srinivasan, Dimitris A. Pados, *Member, IEEE*, and Stella N. Batalama, *Member, IEEE*

***Abstract*—In this paper, we consider the transmission of video over wireless direct-sequence code-division multiple access (DS-CDMA) channels. A layered (scalable) video source codec is used. The layers may be time-multiplexed and transmitted over a single CDMA channel or each layer can be transmitted over a different CDMA channel. For the latter case, spreading codes of different lengths are allowed for each CDMA channel (multirate CDMA). Thus, a different number of chips per bit can be used for the transmission of each scalable layer. For a given fixed energy value per chip and chip rate, the selection of a spreading code length affects the transmitted energy per bit and bit rate for each scalable layer. An MPEG-4 source encoder is used to provide a two-layer signal-to-noise-ratio scalable bit stream. Each of the two layers is channel-coded using rate-compatible punctured convolutional codes. Then, the data are interleaved, spread, carrier-modulated, and transmitted over the wireless channel. A multipath Rayleigh fading channel model is assumed. At the other end, the signal is collected by an antenna array front. After carrier demodulation, multiple-access-interference suppressing despreading is performed using adaptive space–time auxiliary-vector (AV) filters. The choice of the AV space–time receiver is dictated by realistic channel fading rates that limit the data record available for receiver adaptation and redesign. Our experimental results demonstrate the effectiveness of such a multirate DS-CDMA system for wireless video transmission.**

***Index Terms*—Adaptive filtering, antenna arrays, direct-sequence code-division multiple access (DS-CDMA), rate distortion optimization, small sample support, wireless video transmission.**

I. Introduction

During the past several years, there has been an increasing interest in multimedia communications over different types of channels. A significant amount of research has been focused on multimedia transmission over wireless channels. In this paper, we consider the problem of operational rate-distortion optimal scalable video transmission over wireless multipath fading multirate direct-sequence code-division multiple access (DS-CDMA) channels [1].

In a compressed video bit stream, the various parts of the bit stream are not equally important to the quality of the decoded video sequence. Thus, instead of protecting them equally, it would be advantageous to protect the most important parts of the bit stream more than the less important parts. This is the idea of data partitioning and unequal error protection (UEP). In this work, we apply UEP to the layers of a scalable bit stream. We accomplish UEP through the use of channel codes of variable rates and spreading codes of variable lengths.

The break-up of the bit stream into subsets of varying importance using a scalable codec lends itself naturally to employing a UEP scheme. The base layer is typically better protected than the enhancement layers. This allows for added degrees of freedom in selecting the rates that will minimize the overall distortion. In [2], the benefits of using scalability in an error-prone environment are demonstrated by examining all scalability modes supported by MPEG-2 in an ATM network.

In [3] and [4], we assumed video transmission from one transmitter to one receiver using binary phase-shift-keying (BPSK) modulation. The channel behavior was modeled as nonfrequency-selective Rayleigh fading. In [5], we assumed video transmission over a DS-CDMA system. The channel model that we considered was frequency-selective (multipath) Rayleigh fading. At the receiver, we employed an adaptive antenna array auxiliary-vector (AV) linear filter [6], [7] that provides space–time RAKE-type processing (thus, taking advantage of the multipath characteristics of the channel) and multiple-access-interference (MAI) suppression [8]. The choice of the AV receiver was dictated by realistic channel fading rates that limit the data record available for receiver adaptation and redesign. Under small-sample support adaptation, AV filter short-data-record estimators have been shown [6] to exhibit superior bit-error rate (BER) performance in comparison with least mean squares (LMS), recursive least squares (RLS), sample matrix inversion (SMI) [9], diagonally loaded SMI, or multistage Wiener filter implementations [10].

In this paper, we generalize and extend our previous results on DS-CDMA video transmission by allowing a single video user to utilize more than one CDMA channel and different processing gain values. The proposed system is shown in Fig. 1. Each video scalable layer is individually channel-coded and transmitted over a separate CDMA channel. As shown in the figure, other users are simultaneously transmitting data (video, audio, or other). In this paper, we compare scalable video transmission over a single DS-CDMA channel with transmission over two DS-CDMA channels. In both cases, an operational rate-distortion problem is defined and solved in order to optimally select the source coding rates, channel

Manuscript received July 11, 2003; revised April 18, 2005. This work was supported in part by the National Science Foundation under Grant CCF-0219903 and in part by the Air Force Office of Scientific Research under Grant FA9550-04-1-0256. This paper was recommended by Associate Editor O. Al-Shaykh.

The authors are with the Department of Electrical Engineering, State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: lkondi@eng.buffalo.edu;ds37@eng.buffalo.edu; pados@eng.buffalo.edu; batalama@eng.buffalo.edu).

Digital Object Identifier 10.1109/TCSVT.2005.856922

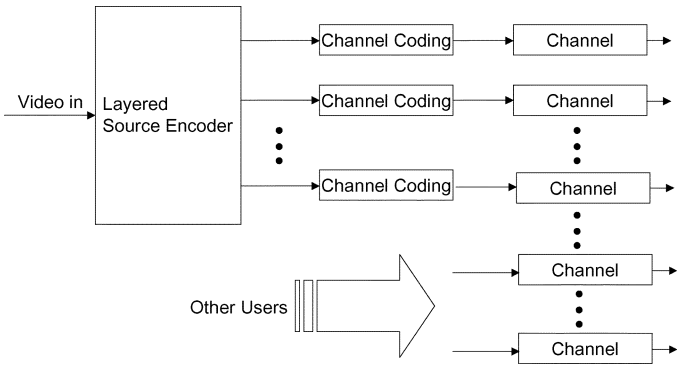

Fig. 1. Multilayer wireless video transmission system.

coding rates, and spreading code lengths (processing gains) used for the transmission. Some preliminary results of this work were reported earlier in [11].

The rest of this paper is organized as follows. In Section II, we describe the elements of the proposed video transmission system, i.e., scalable video coding (Section II-A), channel encoding (Section II-B), received signal (Section II-C), and AV filtering (Section II-D). In Section III, the joint source-coding optimization algorithm is described. Experimental results are presented in Section IV, and final conclusions are drawn in Section V.

II. Video Transmission System

A. Scalable Video Coding

A scalable video codec produces a bit stream which can be divided into embedded subsets. The subsets can be independently decoded to provide video sequences of increasing quality. Thus, a single compression operation can produce bit streams with different rates and reconstructed quality. A subset of the original bit stream can be initially transmitted to provide a base-layer quality with extra layers subsequently transmitted as enhancement layers. The base layer is more important to the quality of the reconstructed video sequence than the enhancement layers, since it can be decoded by itself.

There are three main types of scalability: signal-to-noise ratio (SNR), spatial, and temporal. In SNR scalability, the enhancement in quality translates in an increase in the SNR of the reconstructed video sequence; in spatial and temporal scalability, the spatial and temporal resolution, respectively, is increased.

An important application of scalability is in video transmission from a server to multiple users over a heterogeneous network such as the Internet. Users are connected to the network at different speeds, thus, the server needs to transmit the video data at bit rates commensurate to these connection speeds. Scalability allows the server to compress the data only once and serve each user at an appropriate bit rate by transmitting a subset of the original bitstream.

Yet another important application of scalability is in error-resilient video transmission. It has been shown [2] that it is advantageous to use scalability and apply stronger error protection to the base layer than to the enhancement layers (UEP). This way, the base layer will be successfully decoded with high probability even during adverse channel conditions. Had we not used scalability but instead protected the whole bit stream equally, there would be a much higher probability of catastrophic errors that would result in a poor quality reconstructed video sequence.

This study is concerned with video transmission over wireless DS-CDMA channels. Since such channels can exhibit relatively high error probabilities, error resilience is critical. For this reason, we propose to use an MPEG-4 video codec in SNR scalability mode and allow for different error protection for each scalable layer. UEP is accomplished jointly through different channel coding rates and spreading code lengths (processing gains).

B. Channel Coding

Rate-compatible punctured convolutional (RCPC) codes for channel coding are used in this work. Convolutional coding is accomplished by convolving the source data with a convolutional matrix $\mathbf{G}$. 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. Convolution is the process of modulo-2 addition of the current source bit with previously delayed source bits; the generator matrix specifies which delayed inputs to add with the current input. The operation is equivalent to passing the input data through a linear finite-state register where the tap connections are defined by the generator matrix. 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.

Decoding convolutional codes is most commonly done using the Viterbi algorithm [12], which is a maximum-likelihood (ML) sequence estimation procedure. There are two types of Viterbi decoding: soft and hard decoding. In soft decoding, channel output decision statistics are passed to the decoder for final source-string decision making based on the appropriate distance metric [for example, Euclidean distance for additive white Gaussian noise (AWGN) channels]. In hard decoding, the channel output is used to produce an ML decision on the channel input which is then used for final source-string decision making under the proper distance metric (Hamming distance, for example, for binary symmetric channels).

If a variable rate code is desired, then a higher rate code can be obtained by puncturing the output of the code [13]. 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, thus leading to a higher coding rate. The idea of puncturing was extended to include the concept of rate compatibility [14]. 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.

RCPC codes are very well suited to UEP since they provide a family of codes with different rates. Furthermore, all of these codes can be decoded by the same Viterbi decoder.

C. Received Signal

The time-varying multipath propagation characteristics of (mobile) CDMA radio communications suggest a channel

model that undergoes a process known as 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 low-pass 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 $\mathbf{r}_m \in \mathbb{C}^{L+P}$ given by

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

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 bits, 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 zero-padded by P, p-cyclic-shifted version of the signature $\mathbf{s}_k$ of the kth SS signal, $\mathbf{s}_{k,p}^-$ is the zero-filled $(L-p)$-left-shifted version of $\mathbf{s}_{k,0}$, and $\mathbf{s}_{k,p}^+$ is the zero-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 signal, pth path, array response vector

$$\mathbf{a}_{k,p}[\mathrm{m}] = e^{j2\pi(m-1)\frac{\sin\theta_{k,p}d}{\lambda}}, \quad m = 1,\ldots,M \tag{2}$$

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 two-dimensional (2-D) data notation and filtering operations, we decide at this point to "vectorize" the $(L+P)\times M$ space–time data matrix $[\mathbf{r}_1\ \mathbf{r}_2\ \cdots\ \mathbf{r}_M]$ by sequencing all matrix columns in the form of a single $(L+P)M$-long column vector

$$\mathbf{r}_{(\mathrm{L+P})\mathrm{M}\times 1} = \mathrm{Vec}\left\{[\mathbf{r}_1,\ \mathbf{r}_2,\ \cdots,\ \mathbf{r}_M]_{(\mathrm{L+P})\mathrm{M}}\right\}. \tag{3}$$

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

For conceptual and notational simplicity, we may rewrite the vectorized space–time data equation as follows:

$$\mathbf{r} = \sqrt{E_0}b_0\mathbf{w}_{\mathrm{R-MF}} + \mathbf{i} + \mathbf{n} \tag{4}$$

where $\mathbf{w}_{\mathrm{R-MF}} = E_{b_0}\{\mathbf{r}b_0\} = \mathrm{Vec}\{[\sum_{p=0}^{P} c_{0,p}\mathbf{s}_{0,p}, \mathbf{a}_{0,p[1]}, \cdots, \cdots, \sum_{p=0}^{P} c_{0,p}\mathbf{s}_{0,p}\mathbf{a}_{0,p[\mathrm{M}]}]\}$ is the effective space–time signature of the CDMA signal of interest (signal 0) and $\mathbf{i}$ identifies comprehensively both the intersymbol and the CDMA interference present in $\mathbf{r}$ ($E_{b_0}\{\cdot\}$ denotes statistical expectation with respect to b_0). We use the subscript $\mathrm{R-MF}$ in our effective S-T signature notation 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 our notation, the generalized S-T RAKE operation corresponds to linear filtering of the form $\mathbf{w}_{\mathrm{R-MF}}^H\mathbf{r}$ where H denotes the Hermitian operation.

In this paper, we consider a combination of *multicode multirate CDMA* and *variable-sequence-length multirate CDMA*. In multicode CDMA, more than one CDMA channel (code) can be allocated to a user. In variable-sequence-length CDMA, the spreading sequences of different CDMA channels can have different lengths. Consider as an example two CDMA channels: channel 1 with a spreading code of length $L = 16$ and channel 2 with a spreading code of length $L = 32$. Assuming equal energy per chip for the two channels, channel 2 transmits twice as many chips per bit (and energy per bit) than channel 1. However, channel 1 transmits at twice the bit rate of channel 2. Thus, channel 2 exhibits a lower raw BER than channel 1, but channel 1 allows for transmission at a higher data rate.

D. AV Adaptive Filtering

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

$$\mathbf{w}_0 = \frac{\mathbf{w}_{\mathrm{R-MF}}}{\|\mathbf{w}_{\mathrm{R-MF}}\|^2} \tag{5}$$

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

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu_{k+1}\mathbf{g}_{k+1}. \tag{6}$$

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

$$\mathbf{g}_{k+1} = \mathbf{R}\mathbf{w}_k - \frac{\mathbf{w}_{\mathrm{R-MF}}^H\mathbf{R}\mathbf{w}_k}{\|\mathbf{w}_{\mathrm{R-MF}}\|^2}\mathbf{w}_{\mathrm{R-MF}} \tag{7}$$

where $\mathbf{R}$ is the input autocorrelation matrix, $\mathbf{R} \triangleq E\{\mathbf{r}\mathbf{r}^H\}$. 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

$$\mu_{k+1} = \frac{\mathbf{g}_{k+1}^H\mathbf{R}\mathbf{w}_k}{\mathbf{g}_{k+1}^H\mathbf{R}\mathbf{g}_{k+1}}. \tag{8}$$

The AV filter recursion is completely defined by (5)–(8). Theoretical analysis of the AV algorithm was pursued in [6]. 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 (5)–(8).

1) Successive AVs generated through (6)–(8) 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 $|k-j| \neq 1$).
2) The generated sequence of AV weights $\{\mu_k\}$, $k = 1, 2, \ldots$, is real-valued, positive, and bounded: $0 < (1/\lambda_{\max}) \leq \mu_k \leq (1/\lambda_{\min})$, $k = 1, 2, \ldots$, where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and minimum, correspondingly, eigenvalues of $\mathbf{R}$.
3) The sequence of AVs $\{\mathbf{g}_k\}$, $k = 1, 2, \ldots$, converges to the $\mathbf{0}$ vector: $\lim_{n\to\infty} \mathbf{g}_n = \mathbf{0}$.
4) The sequence of AV filters $\{\mathbf{w}_k\}$, $k = 1, 2, \ldots$, converges to the minimum-variance-distortionless-response (MVDR) filter: $\lim_{k\to\infty} \mathbf{w}_k = (\mathbf{R}^{-1}\mathbf{w}_{\mathrm{R-MF}} / \mathbf{w}_{\mathrm{R-MF}}^H \mathbf{R}^{-1} \mathbf{w}_{\mathrm{R-MF}})$. □

If $\mathbf{R}$ is unknown and sample-average estimated from a packet data record of D points, $\hat{\mathbf{R}}(D) = (1/D)\sum_{d=1}^{D} \mathbf{r}_d \mathbf{r}_d^H$, then Theorem 1 shows that

$$\hat{\mathbf{w}}_k(D) \underset{k\to\infty}{\longrightarrow} \hat{\mathbf{w}}_\infty(D) = \frac{\left[\hat{\mathbf{R}}(D)\right]^{-1} \mathbf{w}_{\mathrm{R-MF}}}{\mathbf{w}_{\mathrm{R-MF}}^H \left[\hat{\mathbf{R}}(D)\right]^{-1} \mathbf{w}_{\mathrm{R-MF}}} \quad (9)$$

where $\hat{\mathbf{w}}_\infty(D)$ is the widely used MVDR filter estimator known as the sample-matrix-inversion (SMI) filter [9]. The output sequence begins from $\hat{\mathbf{w}}_0(D) = (\mathbf{w}_{\mathrm{R-MF}}/\|\mathbf{w}_{\mathrm{R-MF}}\|^2)$, which is a zero-variance fixed-valued estimator that may be severely biased ($\hat{\mathbf{w}}_0(D) = (\mathbf{w}_{\mathrm{R-MF}}/\|\mathbf{w}_{\mathrm{R-MF}}\|^2) \neq \mathbf{w}_{\mathrm{MVDR}}$) unless $\mathbf{R} = \sigma^2\mathbf{I}$ for some $\sigma > 0$. In the latter trivial case, $\hat{\mathbf{w}}_0(D)$ is already the perfect MVDR filter. Otherwise, the next filter estimator in the sequence, $\hat{\mathbf{w}}_1(D)$, has a significantly reduced bias due to the optimization procedure employed at the expense of nonzero estimator (co)variance. As we move up in the sequence of filter estimators $\hat{\mathbf{w}}_k(D)$, $k = 0, 1, 2, \ldots$, the bias decreases rapidly to zero while the variance rises slowly to the SMI ($\hat{\mathbf{w}}_\infty(D)$) levels [see (9)].

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 [7]. The procedure selects the estimator $\hat{\mathbf{w}}_k$ from the generated sequence of AV filter estimators that exhibits maximum J-divergence between the filter output conditional distributions given that $+1$ or -1 is transmitted. Under a Gaussian approximation on the conditional filter output distribution, it was shown in [7] that the J-divergence of the filter estimator with k AVs is

$$J(k) = \frac{4E^2\left\{b_0 \mathrm{Re}\left[\hat{\mathbf{w}}_k^H(D)\mathbf{r}\right]\right\}}{Var\left\{b_0 \mathrm{Re}\left[\hat{\mathbf{w}}_k^H(D)\mathbf{r}\right]\right\}}. \quad (10)$$

To estimate the J-divergence $J(k)$ from the data packet of size D, the transmitted information bits b_0 are required to be known. We can obtain a blind *approximate* version of $J(k)$ by replacing the information bit b_0 in (10) by the detected bit $\hat{b}_0 = \mathrm{sgn}(\mathrm{Re}\{\hat{\mathbf{w}}_k^H(D)\mathbf{r}\})$ (output of the sign detector that follows the linear filter). In particular, using $\hat{b}_0$ in place of b_0 in (10), we obtain the following J-divergence expression:

$$J_B(k) = \frac{4E^2\left\{\hat{b}_0 \mathrm{Re}\left[\hat{\mathbf{w}}_k^H(D)\mathbf{r}\right]\right\}}{Var\left\{\hat{b}_0 \mathrm{Re}\left[\hat{\mathbf{w}}_k^H(D)\mathbf{r}\right]\right\}} = \frac{4E^2\left\{\left|\mathrm{Re}\left[\hat{\mathbf{w}}_k^H(D)\mathbf{r}\right]\right|\right\}}{Var\left\{\left|\mathrm{Re}\left[\hat{\mathbf{w}}_k^H(D)\mathbf{r}\right]\right|\right\}} \quad (11)$$

where the subscript "B" identifies the blind version of the J-divergence function. To estimate $J_B(k)$ from the data packet of size D, we substitute the statistical expectations in (11) by sample averages. 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 AVs with k shown at the bottom of the page. □

Criterion 1 completes the design of the joint S-T AV filter estimator.

III. Optimal Resource Allocation

We next describe the optimal resource allocation procedure for the cases of video transmission over one or two CDMA channels. The optimization constraint in both cases is the available chip rate, $R_{\mathrm{budget}}^{\mathrm{chip}}$. Fixed energy per chip is assumed.

A. Single-CDMA-Channel Case

If a single CDMA channel (code) is used for the transmission of all scalable layers of a video user, the layers need to be time-multiplexed. The available transmission bit rate is

$$R_{\mathrm{budget}} = \frac{R_{\mathrm{budget}}^{\mathrm{chip}}}{L} \quad (13)$$

where L is the system processing gain (spreading code length). The available bit rate R_{budget} has to be allocated between scalable layers and, within each layer, between source and channel coding.

The formal statement of the problem that we are solving is as follows. Given an overall bit rate R_{budget}, we want to optimally allocate bits between source and channel coding such that the overall MS distortion $D_{\mathrm{s+c}}$ is minimized, that is,

$$\min D_{\mathrm{s+c}} \text{ subject to } R_{\mathrm{s+c}} \leq R_{\mathrm{budget}} \quad (14)$$

where $R_{\mathrm{s+c}}$ is the total bit rate used for source and channel coding for all layers and $D_{\mathrm{s+c}}$ is the resulting expected squared error distortion which is due to both source coding errors and channel errors. The distortion that is caused by source coding

$$k = \arg\max_k \left\{\hat{J}_B(k)\right\} = \arg\max_k \left\{ \frac{4\left[\frac{1}{D}\sum_{d=1}^{D}\left|\mathrm{Re}\left[\hat{\mathbf{w}}_k^H(D)\mathbf{r}_d\right]\right|\right]^2}{\frac{1}{D}\sum_{d=1}^{D}\left|\mathrm{Re}\left[\hat{\mathbf{w}}_k^H(D)\mathbf{r}_d\right]\right|^2 - \left[\frac{1}{D}\sum_{d=1}^{D}\left|\mathrm{Re}\left[\hat{\mathbf{w}}_k^H(D)\mathbf{r}_d\right]\right|\right]^2} \right\} \quad (12)$$

is due to quantization and is deterministic. The distortion due to channel errors is stochastic. Thus, the total distortion is also stochastic and we use its expected value.

Utilizing Lagrangian optimization, the constrained problem in (14) is transformed into the unconstrained problem of minimizing

$$\mathcal{L}(\lambda) = D_{\mathrm{s+c}} + \lambda R_{\mathrm{s+c}} \tag{15}$$

where λ is the Lagrange multiplier.

For T scalable layers, $R_{\mathrm{s+c}}$ is equal to

$$R_{\mathrm{s+c}} = \sum_{l=1}^{T} R_{\mathrm{s+c,l}} \tag{16}$$

where $R_{\mathrm{s+c,l}}$ is the bit rate used for source and channel coding for the scalable layer l. $R_{\mathrm{s+c,l}}$ is equal to

$$R_{\mathrm{s+c,l}} = \frac{R_{\mathrm{s,l}}}{R_{\mathrm{c,l}}} \tag{17}$$

where $R_{\mathrm{s,l}}$ and $R_{\mathrm{c,l}}$ are the source and channel rates, respectively, for the scalable layer l. It should be emphasized that $R_{\mathrm{s,l}}$ is in b/s and $R_{\mathrm{c,l}}$ is a dimensionless number.

The task described in (14) is a discrete optimization problem: $R_{\mathrm{s,l}}$ and $R_{\mathrm{c,l}}$ can only take values from discrete sets that are predefined as part of the problem. Optimization can be carried out by a Lagrange procedure. To reduce the computational complexity of the procedure, it is useful to write the overall distortion $D_{\mathrm{s+c}}$ as the sum of distortions per scalable layer.

$$D_{\mathrm{s+c}} = \sum_{l=1}^{T} D_{\mathrm{s+c,l}}. \tag{18}$$

In a subband-based scalable codec, it is straightforward to express the distortion as the sum of distortions per layer since each layer corresponds to different transform coefficients. However, in our scalable codec, we need to redefine distortion per layer as the *differential improvement* of including the layer in the reconstruction [3]. Therefore, in the absence of channel errors, only the distortion for Layer 1 (base layer) would be positive and the distortions for all other layers would be negative since inclusion of these layers reduces the overall mean squared error (MSE).

Another observation that should be made is that the differential improvement in MSE due to a given layer depends on the rates of the previous layers. For example, for a two layer case an enhancement layer of 28 kb/s will cause a different improvement in MSE depending on the rate used for the base layer. The differential improvement depends on the picture quality before the inclusion of the scalable layer in question. Therefore, the distortion per layer is better written as

$$D_{\mathrm{s+c}} = \sum_{l=1}^{T} D_{\mathrm{s+c,l}}(R_{\mathrm{s+c,1}}, \ldots, R_{\mathrm{s+c,l}}). \tag{19}$$

Our optimization problem reduces to finding the *operational rate-distortion functions* (ORDF) $D_{\mathrm{s+c,l}}(\cdot, \ldots, \cdot)$ for each scalable layer. One way to proceed is to experimentally obtain the expected distortion for each layer for all possible combinations of source and channel rates and "all possible" channel conditions. However, this becomes prohibitively complex for even a small number of admissible source and channel rates and channel conditions. Thus, we choose instead to utilize the *universal rate-distortion characteristics* (URDC) at the expense of a small performance penalty. URDC characteristics show the expected distortion per layer as a function of the BER (after channel coding). Their use is discussed in Section III-C.

B. Multiple-CDMA-Channel Case

We next discuss the case where each scalable video layer is transmitted over a separate DS-CDMA channel (code). In that case, the available bit rate for layer i is

$$R_{\mathrm{s+c,i}} = \frac{R_{\mathrm{budget}}^{\mathrm{chip}}}{L_i} \tag{20}$$

where L_i is the spreading length for layer i. Thus, if two layers are assumed, the ratio $R_{\mathrm{s+c,1}}/R_{\mathrm{s+c,2}}$ is fixed and equal to L_2/L_1. This is in contrast to the single CDMA channel case where the allocation of the available bit rate to each individual scalable layer is part of the optimization.

Our optimization problem now is as follows (for the case of T layers):

$$\min D_{\mathrm{s+c}} \text{ subject to } L_i R_{\mathrm{s+c,i}} \le R_{\mathrm{budget}}^{\mathrm{chip}}, \text{ for } i = 1, \ldots, T. \tag{21}$$

For each layer, we need to determine the source coding rate $R_{s,i}$, the channel coding rate $R_{c,i}$, and the spreading length L_i. As mentioned previously, the total bit rate allocated to a scalable layer depends only on L_i and not on any decisions made for another layer. Since $D_{\mathrm{s+c}} = \sum_{i=1}^{T} D_{\mathrm{s+c,i}}(R_{\mathrm{s+c,1}}, \ldots, R_{\mathrm{s+c,i}})$, our problem can be broken into separate problems for each layer, thus simplifying the optimization when compared to the single-CDMA-channel case. For the case of two layers, the two problems to be solved are

$$\left\{R_{s,1}^*, R_{c,1}^*, L_1^*\right\} = \arg\min D_{\mathrm{s+c,1}}(R_{\mathrm{s+c,1}}) \quad \text{subject to } L_1 R_{\mathrm{s+c,1}} \le R_{\mathrm{budget}}^{\mathrm{chip}} \tag{22}$$

and

$$\left\{R_{s,2}^*, R_{c,2}^*, L_2^*\right\} = \arg\min D_{\mathrm{s+c,2}}\left(R_{\mathrm{s+c,1}}^*, R_{\mathrm{s+c,2}}\right) \quad \text{subject to } L_2 R_{\mathrm{s+c,2}} \le R_{\mathrm{budget}}^{\mathrm{chip}}. \tag{23}$$

Thus, optimization for the two-channel video transmission case is algebraically simpler than optimization in the single-channel case. Since there is no dependency between layers, the above two problems can be solved independently using Lagrangian optimization. As in the single-channel case, we first need to obtain the rate-distortion functions $D_{\mathrm{s+c,i}}(.)$, $i = 1, 2$. The rate-distortion functions can be obtained using universal rate-distortion characteristics, as explained next.

C. Universal Rate-Distortion Characteristics

So far, we have formulated the problem and outlined the solution for video layer rate design assuming that the optimal rate-distortion characteristics of the individual layers, $D_{\mathrm{s+c,i}}(.)$, are given. We now describe the technique used to obtain $D_{\mathrm{s+c,i}}(.)$, $i = 1, \ldots, T$.

While it is possible to simulate transmission of the actual source coded data over a channel, gather statistics, and develop an optimal strategy based on these results, in practice, this leads to extremely high computational complexity and makes

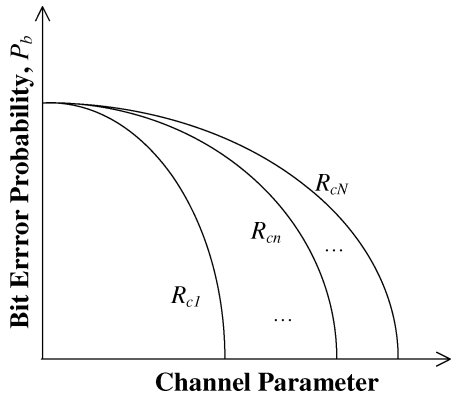

Fig. 2. Typical channel characteristic plot, i.e., BER versus channel parameters for a set of channel coding rates $R_{cn}, n = 1, \ldots, N$.

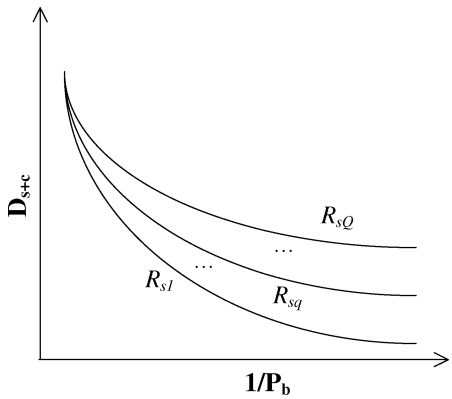

Fig. 3. URDCs of the source; typical end-to-end distortion versus BER response characterizing the source and channel for a set of source coding rates, R_{sq} for $q = 1, \ldots, Q$.

the process impractical in many ways. For every bit stream, we will have to simulate transmission of the data over the channel using all combinations of source and channel coding rates per scalable layer and channel conditions of interest. Clearly, the computational complexity of this approach is prohibitive. To circumvent this problem, *universal* rate-distortion characteristics of the source coding scheme are utilized [3], [15]. This approach is described next.

For given channel SNRs, spreading lengths, and choice of channel codes, the probability of bit error P_b is calculated for the set of channel coding rates of interest. P_b establishes a reference as to the performance of channel coding over the particular channel with the given parameters. This channel performance analysis needs to be done only once. An illustration plot showing the performance of channel coding as a function of a given channel parameter is shown in Fig. 2 for a set of channel coding rates $R_{cn}, n = 1, \ldots N$. We will call these plots *channel characteristic plots*.

Toward calculating the impact of the errors due to both lossy source coding and channel disturbance on a set of data, it is realized that, for a given set of preceding layer source rates, the distortion for a particular layer $D_{\mathrm{s+c,i}}$, given a particular source coding rate $R_{s,i}$, is a function of the BER. Thus, the rate-distortion function of the layer for a fixed source rate $R_{s,i}$ (given the preceding layer source rates) is a function of the BER (after channel decoding) P_b. It is then possible to plot a family of $D_{\mathrm{s+c,i}}$ versus $1/P_b$ curves given a set of source coding rates of interest, as shown in Fig. 3. These are defined as the universal rate-distortion characteristics (URDCs) of the source. Due to the use of variable length codes in the video coding standards, it would be a formidable task to analytically obtain the URDCs. Thus, the URDCs are obtained experimentally using simulations. To obtain the URDC for $D_{\mathrm{s+c,i}}$, the ith layer of the bit stream is corrupted with independent errors with BER P_b. Layers $1, \ldots, i-1$ are not corrupted. The bit stream is then decoded and the MSE is calculated. The experiment is repeated many times (in our studies, it was 30 times). If $i > 1$, i.e., we are calculating the URDC for an enhancement layer, we need to subtract the distortion of the first $i-1$ uncorrupted layers, since $D_{\mathrm{s+c,i}}$ in this case is the differential improvement of including layer i, as mentioned earlier.

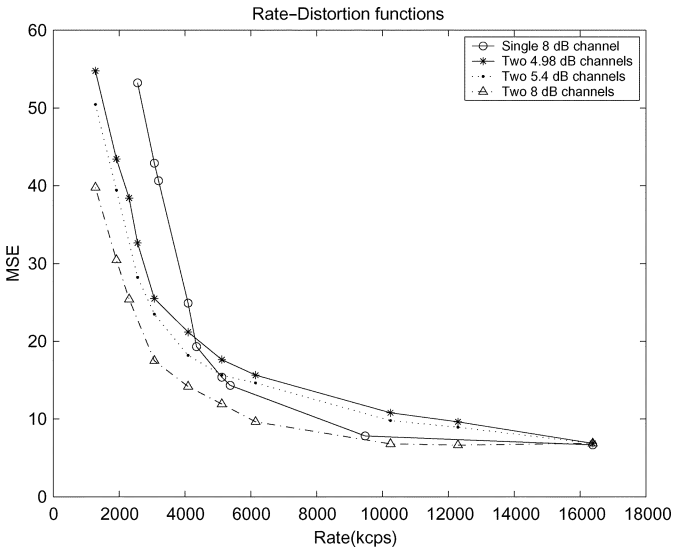

Fig. 4. Rate-distortion performance of scalable video transmission of the Foreman sequence over a DS-CDMA wireless system.

Using the channel characteristic plots and the URDCs, operational rate-distortion functions for each scalable layer are constructed as follows. First, for the given channel parameters, we use the channel characteristic plot to determine the resulting BERs for each of the available channel coding rates. Then, for each of these probability values, we use the URDC to obtain the resulting distortion for each available source coding rate. By obtaining also the total rate $R_{\mathrm{s+c}}$ for each combination of source and channel codes, we have the rate-distortion operating points for the given channel conditions.

IV. Experimental Results

We next present experimental results for the signal model in (1) that compare scalable video transmission over one and two DS-CDMA channels. The physical-layer receiver is equipped with a uniform linear antenna array of $M = 4$ elements. All received CDMA signals, $k = 0, 1, \ldots, K-1$, experience $P = 2$ resolvable multipaths with independent fading per path and equal mean power $E\{|c_{k,p}|^2\}$, $p = 0,1,2$. All paths of all signals have independent angles of arrival $\theta_{k,p}$ drawn uniformly in

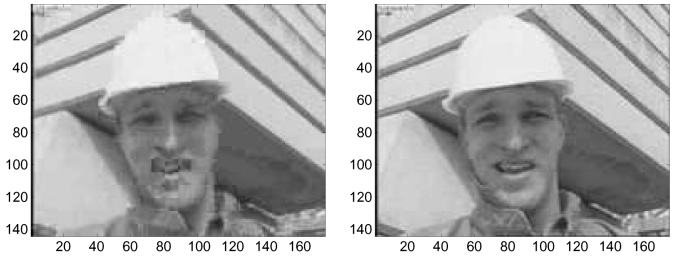

Fig. 5. Frame 14 of Foreman sequence with base plus enhancement decoding; total chip rate of 2560 kc/s. (a) Single CDMA channel. (b) Two CDMA channels.

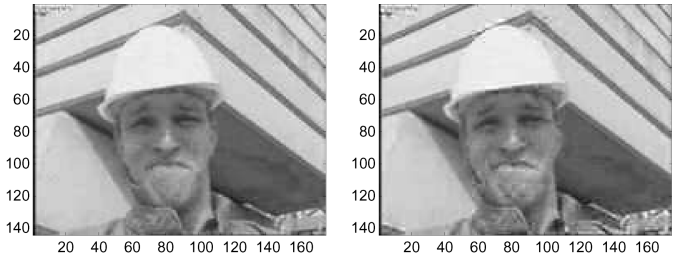

Fig. 6. Frame 12 of the Foreman sequence with base plus enhancement decoding; total chip rate of 5120 kc/s. (a) Single CDMA channel. (b) Two CDMA channels.

$(-90°, 90°)$. The fading realization $c_{k,p}$ of each path of each signal is the same across the antenna elements (antenna diversity effects are not considered/exploited herein). All SNR values identified in the experimental study at this section refer to the total SNR per chip defined as $(\sum_{p=0}^{P} E\{|c_{k,p}|^2\}E_k/\sigma^2)$, $k = 0, 1, \ldots, K-1$.

An MPEG-4-compatible video source codec is used along with RCPC channel codes from [14]. AV filtering followed by soft-decision Viterbi decoding is used at the receiver. Eight active CDMA interferers occupy eight distinct channels (spreading codes), all at an SNR of 8 dB. For the single-CDMA-channel video transmission case, the video user of interest occupies one channel (code) at the same SNR of 8 dB. For the dual-CDMA-channel case, two channels (codes) are used by the video user of interest and have an SNR of 4.98 dB each. Thus, the video user has the same total power in the one- and two-channel case.

For both the single- and dual-CDMA-channel cases, the admissible source coding rates are 64 000, 96 000, 128 000, and 256 000 b/s for both the base and enhancement layers. The admissible channel coding rates for both layers are 1/2, 2/3, and 4/5, while the admissible spreading codes are Walsh–Hadamard of length 16 and 32. Future generalization of this present work may consider the minimum total-squared-correlation optimal Karystinos–Pados spreading codes [16]–[19] that are available for all lengths L and number of signatures K except $K = L = 4n + 1$, $n = 1, 2, \ldots$.

Fig. 4 shows a comparison of the performance of scalable video transmission of the Foreman sequence over a DS-CDMA system when the video user of interest occupies one or two CDMA channels. The MSE is plotted against the total chip rate. It can be seen that, for the SNRs under consideration, when the video user has the same total power, the two-channel case outperforms the one-channel case at chip rates lower than 5.120 Mc/s[1]

Sample frames from the Foreman sequence are shown in Figs. 5 and 6. By inspecting frame 14 decoded at a total chip rate of 2560 kc/s (Fig. 5), it is indeed seen that the two-channel video quality with an MSE of 32.64 is superior to the single-channel case that has an MSE of 53.24. However, at 5120 kc/s, the single CDMA channel with an MSE of 15.39 starts performing slightly better than the two CDMA channels with an MSE of 17.64 (frame 12 in Fig. 6 has a few more motion artifacts in the two-channel case).

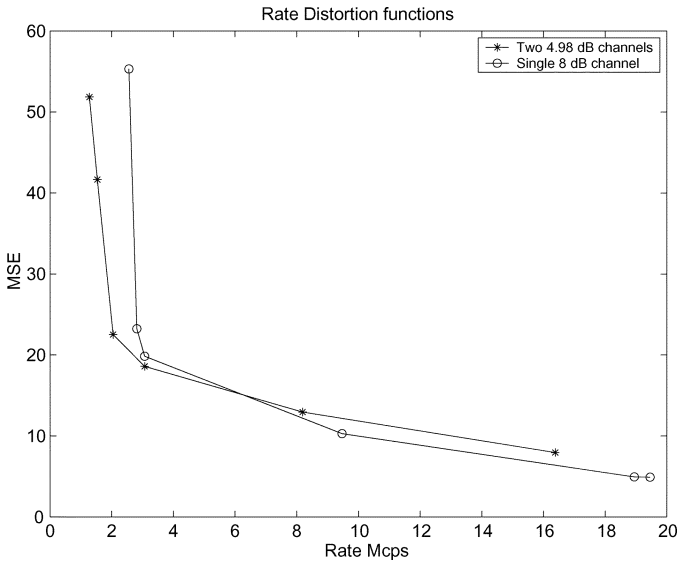

Fig. 7. Rate-distortion performance of scalable video transmission of the News sequence over a DS-CDMA wireless system.

Fig. 7 shows the rate-distortion functions for the News sequence when the video user of interest occupies one or two CDMA channels. Again, it can be seen that, for the SNRs under consideration, when the video server transmits at the same total power, the two-channel case outperforms the one-channel case at chip rates lower than 6.4 Mc/s.

From Tables I and II, it can be seen that, for the base layer, higher source coding rates or lower channel coding rates can be achieved with multirate CDMA, leading to better video quality for devices whose decoder complexity allows decoding of the base layer only. For example, at $R_{\text{budget}}^{\text{chip}}$ of 3.072 Mc/s, in the two-channel system $R_{s,1} = 128$ kb/s, $R_{c,1} = 2/3$, $L = 16$, and the MSE for base-layer decoding is 26.71. In the single-channel system, $R_{s,1} = 64$ kb/s, $R_{c,1} = 2/3$, $L = 16$, and the MSE is 46.99 (see also Fig. 8 with Frame 24 of the Foreman sequence under base-only decoding).

[1]Third-generation mobile Communication systems based on W-CDMA standards typically employ a chip rate of 3.84 Mc/s. If a different selection of admissible source coding rates is used, it is possible that the two-channel case will outperform the one-channel case for all chip rates. However, it is also possible that single-channel video transmission can perform better under different channel conditions.

TABLE I
OPTIMAL RATE ALLOCATION FOR TWO-LAYER SNR SCALABLE VIDEO OVER A SINGLE 8-dB DS-CDMA CHANNEL

Total Rate	Total Distortion	Base Layer			Enhancement Layer			Base Distortion
R^{chip*}_{budget}	D^*_{s+c}	$R^*_{s,1}$	$R^*_{c,1}$	L^*_1	$R^*_{s,2}$	$R^*_{c,2}$	L^*_2	$D^*_{s+c,1}$
2560	53.24	64	4/5	16	64	4/5	16	54.34
3072	42.89	64	2/3	16	64	2/3	16	46.99
3200	40.64	96	4/5	16	64	4/5	16	44.74
4096	24.92	128	4/5	16	64	2/3	16	28.92
4352	19.32	128	2/3	16	64	4/5	16	21.06
5120	15.39	128	2/3	16	64	1/2	16	21.06
5376	14.33	128	1/2	16	64	4/5	16	17.95
9472	7.83	256	1/2	16	64	4/5	16	8.97
16384	6.68	256	1/2	16	256	1/2	16	8.97

TABLE II
OPTIMAL RATE ALLOCATION FOR TWO LAYER SNR SCALABLE VIDEO OVER TWO SEPARATE 4.98-dB DS-CDMA CHANNELS

Total Rate	Total Distortion	Base Layer			Enhancement Layer			Base Distortion
R^{chip*}_{budget}	D^*_{s+c}	$R^*_{s,1}$	$R^*_{c,1}$	L^*_1	$R^*_{s,2}$	$R^*_{c,2}$	L^*_2	$D^*_{s+c,1}$
1280	54.78	64	4/5	16	64	4/5	16	59.97
1920	43.44	96	4/5	16	96	4/5	16	47.64
2304	38.41	96	2/3	16	96	2/3	16	41.83
2560	32.64	128	4/5	16	64	4/5	32	33.89
3072	25.49	128	2/3	16	96	1/2	16	26.71
4096	21.19	128	1/2	16	128	1/2	16	22.71
5120	17.64	256	4/5	16	128	4/5	32	19.04
6144	15.65	128	2/3	32	96	1/2	32	17.65
10240	10.8	256	4/5	32	256	4/5	32	11.00
12288	9.65	256	2/3	32	256	2/3	32	10.74
16384	6.89	256	1/2	32	256	1/2	32	8.79

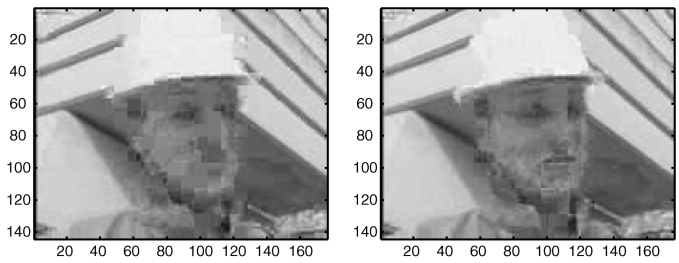

Fig. 8. Frame 24 of the Foreman sequence with base-only decoding; total chip rate 3072 kc/s. (a) Single CDMA channel. (b) Two CDMA channels.

V. CONCLUSION

We presented a wireless multiple-access scalable video communications scheme that is based on DS-CDMA technology with antenna-array reception and AV filtering. Each video user is allowed to occupy one or two CDMA channels (spreading codes). We developed resource optimization schemes for both cases. Our experimental results demonstrate that, for the chosen SNR values of practical interest, two-channel video transmission outperforms single-channel video transmission for chip rates used in third-generation wireless CDMA systems.

REFERENCES

[1] R. Wyrwas, M. J. Miller, R. Anjaria, and W. Zhang, "Multiple access options for multi-media wireless systems," in *Wireless Communications Future Directions*, J. M. Holtzman and D. J. Goodman, Eds. Boston, MA: Kluwer, Mar. 1993, ch. 8.

[2] R. Aravind, M. R. Civanlar, and A. R. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithms," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 6, no. 10, pp. 426–435, Oct. 1996.

[3] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint source-channel coding for SNR scalable video," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 1043–1054, Sep. 2002.

[4] ——, "Joint source-channel coding for scalable video," in *Proc. SPIE Conf. Image and Video Commun. and Processing*, San Jose, CA, Jan. 2000, pp. 324–335.

[5] L. P. Kondi, S. N. Batalama, D. A. Pados, and A. K. Katsaggelos, "Joint source-channel coding for scalable video over DS-CDMA multipath fading channels," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, Thessaloniki, Greece, Oct. 2001, pp. 994–997.

[6] D. A. Pados and G. N. Karystinos, "An iterative algorithm for the computation of the MVDR filter," *IEEE Trans. Signal Process.*, vol. 49, no. 2, pp. 290–300, Feb. 2001.

[7] H. Qian and S. N. Batalama, "Data-record-based criteria for the selection of an auxiliary-vector estimator of the MMSE/MVDR filter," *IEEE Trans. Commun.*, vol. 51, no. 10, pp. 1700–1708, Oct. 2003.

[8] A. Kansal, S. N. Batalama, and D. A. Pados, "Adaptive maximum SINR RAKE filtering for DS-CDMA multipath fading channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 12, pp. 1765–1773, Dec. 1998.

[9] I. S. Reed, J. M. Mallet, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, no. 6, pp. 853–863, Nov. 1974.

[10] J. S. Goldstein, I. S. Reed, and L. L. Sharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Trans. Inf. Theory*, vol. 44, no. 11, pp. 2943–2959, Nov. 1998.

[11] L. P. Kondi, D. Srinivasan, D. A. Pados, and S. N. Batalama, "Layered video transmission over multirate DS-CDMA wireless systems," *Proc. SPIE*, vol. 5022, pp. 289–300, Jan. 2003.

[12] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. PROC-61, no. 3, pp. 268–278, Mar. 1973.

[13] J. B. Cain, G. C. Clark, and J. M. Geist, "Punctured convolutional codes of rate $(n-1)/n$ and simplified maximum likelihood decoding," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 1, pp. 97–100, Jan. 1979.

[14] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, no. 4, pp. 389–400, Apr. 1988.

[15] M. Bystrom and J. W. Modestino, "Combined source-channel coding schemes for video transmission over an additive white gaussian noise channel," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 880–890, Jun. 2000.

[16] G. N. Karystinos and D. A. Pados, "New bounds on the total squared correlation and optimum design of DS-CDMA binary signature sets," *IEEE Trans. Commun.*, vol. 51, no. 1, pp. 48–51, Jan. 2003.

[17] C. Ding, M. Golin, and T. Kløve, "Meeting the Welch and Karystinos-Pados bounds on DS-CDMA binary signature sets," *Designs, Codes Cryptogr.*, vol. 30, pp. 73–84, Aug. 2003.

[18] P. Ipatov, "On the Karystinos-Pados bounds and optimal binary DS-CDMA ensembles," *IEEE Commun. Lett.*, vol. 8, no. 2, pp. 81–83, Feb. 2004.

[19] G. N. Karystinos and D. A. Pados, "The maximum squared correlation, sum capacity, and total asymptotic efficiency of minimum total-squared-correlation binary signature sets," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 348–355, Jan. 2005.