Chapter 1

Joint Source-Channel Coding for Video Communications

Fan Zhai[1], Yiftach Eisenberg[2], and Aggelos K. Katsaggelos[2]

[1] Department of Digital Video, DAV, HPA
Texas Instruments, Dallas, TX 75243, USA
[2] Department Electrical and Computer Engineering
Northwestern University, Evanston, IL 60208, USA

1.1 Introduction

The compression or coding of a signal (e.g., speech, text, image, video) has been a topic of great interest for a number of years. Numerous results and successful compression standards exist. Source compression is the enabling technology behind the multimedia revolution we are experiencing. The two primary applications for data compressing are storage and transmission. Video transmission is the topic of this chapter. Video transmission applications, such as on-demand video streaming, videotelephony, and videoconferencing, have gained increased popularity.

In a video communication system, the video is first compressed and then segmented into fixed or variable length packets and multiplexed with other types of data, such as audio. Unless a dedicated link that can provide a guaranteed quality of service (QoS) is available between the source and the destination, data bits or packets may be lost or corrupted, due to either traffic congestion or bit errors due to impairments of the physical channels. Such is the case, for example, with the current Internet and wireless networks. Due to its best effort design, the current Internet makes it difficult to provide the QoS, such as bandwidth, packet loss probability, and delay needed by video communication applications.

Compared to wired links, wireless channels are much noisier because of fading, multi-path, and shadowing effects, which results in a much higher bit error rate (BER) and consequently an even lower throughput. Figure 1.1 illustrates the effect of channel errors to a typical compressed video sequence in the presence of packet loss.

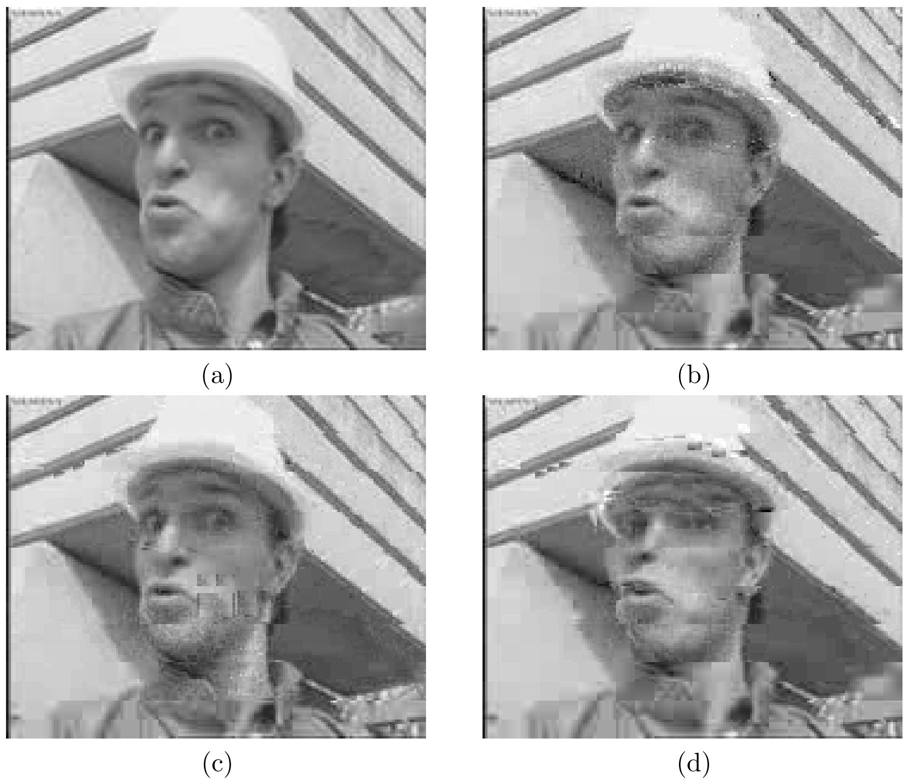


Figure 1.1: Illustration of the effect of channel errors to a video stream compressed using the H.263 standard: (a) Original Frame; Reconstructed frame at (b) 3% packet loss (c) 5% packet loss (d) 10% packet loss (QCIF Foreman sequence, frame 90, coded at 96 kbps and frame rate 15 fps).

Due to the "unfriendliness" of the channel to the incoming video packets, they have to be protected so that the best possible quality of the received video is achieved at the receiver. A number of techniques, which are collectively called error resilient techniques have been devised to combat transmission errors. They can be grouped into [1]: (i) those introduced at the source and channel coder to make the bitstream more resilient to potential errors; (ii) those invoked at the decoder upon detection of errors to conceal the effects of errors, and (iii) those which require interactions between the source encoder and decoder so that the encoder can adapt its operations based on the loss conditions detected at the decoder.

A number of reasons make the error resiliency problem a challenging one. First, compressed video streams are sensitive to transmission errors because of the use of predictive coding and variable-length coding (VLC) by the source encoder. Due to the use of spatio-temporal prediction, a single bit error can propagate in space and time. Similarly, because of the use of VLCs, a single bit error can cause the decoder to loose synchronization, so that even successfully received subsequent bits become unusable. Second, both the video source and the channel conditions are time-varying, and therefore it is not possible to derive an optimal solution for a specific transmission of a given video signal. Finally, severe computational constraints are imposed for real-time video communication applications.

The development of error resilient approaches or approaches for increasing the robustness of the multimedia data to transmission errors is a topic of the utmost importance and interest. To make the compressed bitstream resilient to transmission errors, redundancy must be added into the stream. Such redundancy can be added either by the source or the channel coder. Shannon said it fifty years ago [2], that source coding and channel coding can be separated for optimal performance communication systems. The source coder should compress a source to a rate below the channel capacity while achieving the smallest possible distortion, and the channel coder can add redundancy through Forward Error Correction (FEC) to the compressed bitstream to enable the correction of transmission errors. Following Shannon's separation theory resulted in major advances on source coding (e.g., rate-distortion optimal coders and advanced entropy coding algorithms) and channel coding (e.g., Reed-Solomon codes, Turbo codes, and Tornado codes). The separation theory not only promises that the separate design of source and channel coding does not introduce any performance sacrifice, but it also greatly reduces the complexities of a practical system design. However, the assumptions on which separation theory is based (infinite length codes, delay, and complexity) may not hold in a practical system. This leads to the development of the approach of joint consideration of source coding and channel coding, referred to as joint source-channel coding (JSCC). JSCC can greatly improve the system performance when there are, for example, stringent end-to-end delay constraints or implementation complexity concerns.

Our purpose in this chapter is to review the basic elements of some of the more recent approaches towards JSCC for wired and wireless systems alike. The rest of the chapter is organized as follows. We first provide basic rate-distortion definitions in addressing the need for JSCC in Sect. 1.2. We then describe the basic components in a video compression and transmission systems in Sect. 1.3, followed by a discussion on channel coding techniques that are widely used for video communications in Sect. 1.4. In Sect. 1.5 the JSCC problem formulation is presented, followed by examples of several practical implementations. Finally, Sect. 1.6 contains concluding remarks.

1.2 On Joint Source-Channel Coding

Due to the high bandwidth of the raw video data, compression is necessary for reducing the source redundancy. With ideal lossless compression, source redundancy can be entirely removed without any quality loss. Since the minimum bit rate achieved using lossless compression is usually much higher than the available channel capacity, lossy compression is generally required for video transmission applications. The same way entropy determines the lowest possible rate for lossless compression, rate-distortion (R-D) theory [2, 3] addresses the same question for lossy compression.

1.2.1 Rate-distortion theory

A high level block diagram of a video transmission system is shown in Fig. 1.2. In it, X and $\hat{X}$ represent respectively the source and reconstructed video, X_s and $\hat{X}_s$ the source encoder output and the source decoder input, and X_c and $\hat{X}_c$ the channel encoder output and the channel decoder input.

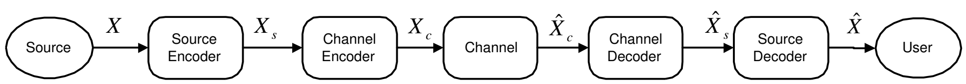


Figure 1.2: Block diagram of a communication system.

The central entity of R-D theory is the R-D function $R(D)$, which provides the theoretical information bound on the rate necessary to represent a certain source with a given average distortion. It is given by [2]

$$R(D) = \min_{\{P(\hat{x}_j|x_i)\}\in\Gamma} I(X;\hat{X}) \tag{1.1}$$

where R is the source rate, D the average source distortion, $I(X;\hat{X})$ the average mutual information between X and $\hat{X}$, $x_i \in X$, $\hat{x}_j \in \hat{X}$, $P(\hat{x}_j|x_i)$ the conditional probability, and the set Γ is defined as

$$\Gamma = \{\{P(\hat{x}_j|x_i)\} \quad \text{such that} \quad D(\{P(\hat{x}_j|x_i)\}) \leq D^*\}\,, \tag{1.2}$$

where D^* is the distortion constraint. The distortion D can be written as

$$D = \sum_{i=0}^{N-1}\sum_{j=0}^{M-1} d(x_i,\hat{x}_j)P(x_i)P(\hat{x}_j|x_i),$$

where d is the distortion metric and N and M are the sizes of the source and reconstruction alphabets, respectively.

The R-D function can be used to find the minimum channel bandwidth for a certain source with a given distortion constraint, as expressed by (1.1) and (1.2). Conversely, it can also be used to determine the information theoretical bounds on the average distortion subject to a channel capacity constraint, via the distortion-rate function, $D(R)$, which is the dual of (1.1), defined as,

$$D(R) = \min_{\{P(\hat{x}_j|x_i)\} \in \Phi} D(X; \hat{X}) \tag{1.3}$$

where Φ is defined as

$$\Phi = \{\{P(\hat{x}_j|x_i)\} \quad \text{such that} \quad R(\{P(\hat{x}_j|x_i)\}) \leq R^*\},$$

with R^* the rate constraint.

Note that the $D(R)$ function may be more widely applicable in practical image/video communication systems, since, as a practical matter, the aim is usually to deliver the best quality image/video subject to certain channel bandwidth, rather than the opposite. R-D theory is of fundamental importance in that it conceptually provides the information bounds for lossy data compression. However, it is usually difficult to find closed-form expressions for the $R(D)$ or $D(R)$ functions. In this case, one resorts to numerical algorithm for specifying the operational R-D function [4].

1.2.2 Practical constraints in video communications

We can see from (1.1) that the process of finding the optimal compression scheme requires searching over the entire set of conditional probabilities that satisfy the distortion constraint shown in (1.2). Under the assumption that the source encoder output is identical to the source decoder input, i.e., $X_s = \hat{X}_s$, the problem becomes a pure source coding problem, since the conditional probabilities $P(\hat{x}_j|x_i)$ have nothing to do with the channel. However, such an assumption generally requires an ideal channel coding scheme, such that error free transmission of the source output over a noisy channel with source bit rate $R(D)$ less than the channel capacity can be guaranteed.

However, such ideal channel coding generally requires infinite length code words, which can only be realized without complexity and delay constraints, both of which are important factors in practical real-time systems. Due to these constraints, channel coding cannot achieve channel capacity. Hence, most practical channel coding schemes do not provide an idealized error free communication path between source and destination, and thus the overall distortion consists of both source distortion and channel distortion. For this reason, minimizing the total distortion usually requires jointly designing the source and channel encoders, which is referred to as joint source-channel coding.

At the receiver side, gains may be obtained by jointly designing the channel and source decoders, which is referred to as joint source-channel decoding. In using joint source-channel decoding, the channel decoder does not make *hard* decisions on the output $\hat{X}_s$. Instead, the decoder makes *soft* decisions in order to allow the source decoder to make use of information such as the signal-to-noise ratio (SNR) of the corrupted code. Alternatively, such soft decisions can be regarded as hard decisions plus a confidence measure. Soft-decision processing used in joint source channel decoding can usually help improve the coding gain by about 2 dB compared to hard-decision processing [5]. In this chapter, we focus on what is performed at the sender side.

1.2.3 Illustration

The basic idea of JSCC is illustrated in Fig. 1.3. When the channel is error free, increased bit rate leads to decreased distortion, as in standard rate-distortion (R-D) theory. This is illustrated by the lowest curve in Fig. 1.3, in which the lowest distortion is obtained by utilizing the largest available source bit rate, represented by the point (R1, D1). However, when channel errors are present, this trend may not hold, since the overall distortion consists of both source and channel distortions. For a given channel rate, as more bits are allocated to source coding, fewer will be left for channel coding, which leads to less protection and higher channel distortion. As shown in Fig. 1.3, an optimal point exists for given channel distortions in allocating bits between source and channel coding. Note that different channel error rates result in different optimal allocations. This is indicated by the points (R2, D2) and (R3, D3) on the two curves with different channel error rates.

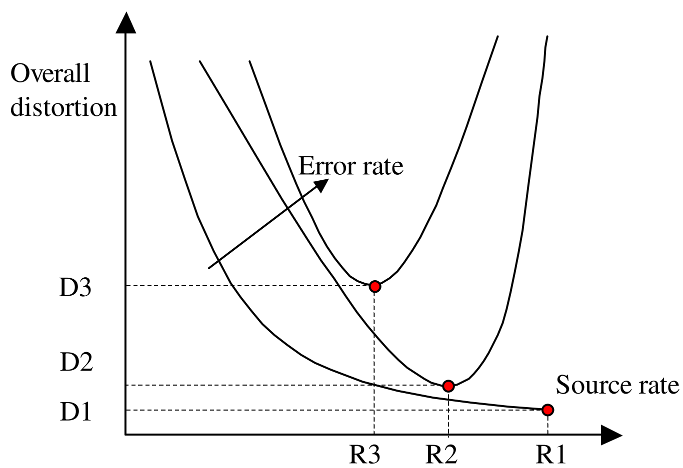


Figure 1.3: Illustration of joint source-channel coding.

The tradeoff between source and channel coding has been studied from a theoretical standpoint based on the use of vector quantizers [6, 7]. In general, JSCC is accomplished by designing the quantizer and entropy coder jointly for given channel characteristics, as in [6, 8]. There is a substantial number of research results in this area. Interested readers can refer to [9] for a comprehensive review on this topic. In this chapter, we focus on the specific application of JSCC in image and video communications, where JSCC usually faces three tasks: finding an optimal bit allocation between source coding and channel coding for given channel loss characteristics; designing the source coding to achieve the target source rate; and designing the channel coding to achieve the required robustness [10, 11]. These tasks, although stated separately, are inter-related, forming the backbone of the integrated nature of JSCC.

We have discussed the basic concept underlying JSCC and its significance in image/video communications. Next, we will first provide an overview of the video compression and transmission systems. Then we will highlight the key components of JSCC for video applications, such as the different forms of error resilient source coding, channel codes used to deal with different types of channel errors, the general problem formulation and the general solution approach. In addition to the commonly used video compression standards, such as MPEGx and H.26x, we also briefly discuss wavelet and subband-based video compression schemes, since they are also widely used.

1.3 Video Compression and Transmission

1.3.1 Video transmission system

We begin by providing a brief high-level overview of a packet-based video transmission system. Some of the major conceptual components found in such a system are shown in Fig. 1.4.

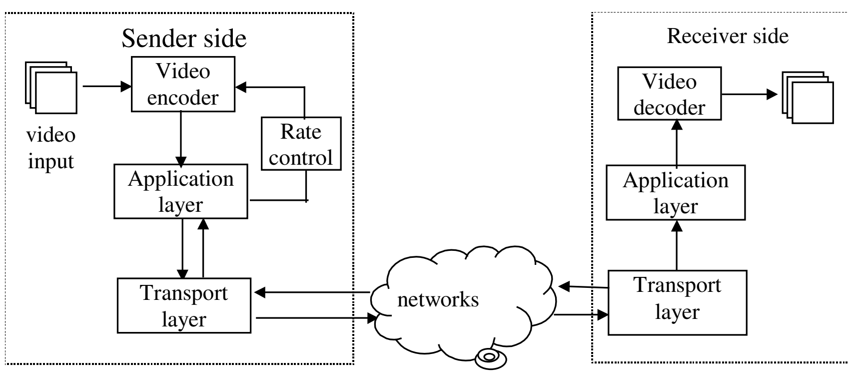


Figure 1.4: Video transmission system architecture.

Most practical communication networks have limited bandwidth and are lossy by nature. Facing these challenges, the *video encoder* has two main objectives: to compress the original video sequence and to make the encoded sequence resilient to errors. Compression reduces the number of bits used to represent the video sequence by exploiting both temporal and spatial redundancy. On the other hand, to minimize the effects of losses on the decoded video quality, the sequence must be encoded in an error resilient way. A recent review of resilient video coding techniques can be found in [1]. The source bit rate is shaped or constrained by a rate controller that is responsible for allocating bits to each video frame or packet. This bit rate constraint is set based on the estimated channel state information (CSI) reported by the lower layers, such as the application and transport layers. It is mentioned here that the system in Fig. 1.4 is a simplified version of a seven or a five layer Open Systems Interconnection (OSI) reference model. For example, in both OSI models the network, data link, and physical layers are below the transport layer. In the subsequent sections we will be referring to the various layers, since allowing the various layers to exchange information leads to cross-layer design of a video communication system, which is a central theme in this chapter.

In Fig. 1.4, the *network* block represents the communication path between the sender and the receiver. This path may include routers, subnets, wireless links, etc. The network may have multiple channels (e.g., a wireless network) or paths (e.g., a network with path diversity), or support QoS (e.g., integrated services or differentiated services networks). Packets may be dropped in the network due to congestion, or at the receiver due to excessive delay or unrecoverable bit errors in a wireless network. To combat packet losses, parity check packets used for FEC may be generated at the application/transport layer. In addition, lost packets may be retransmitted if the application allows.

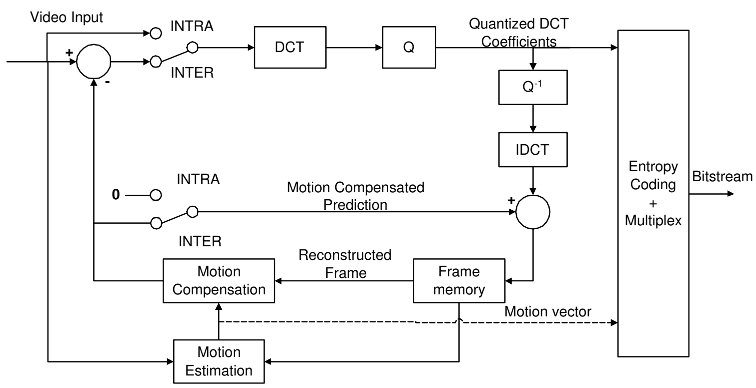
For many source-channel coding applications, the exact details of the network infrastructure may not be available to the sender and they may not always be necessary. Instead, what is important in JSCC is that the sender has access to or can estimate certain network characteristics, such as the probability of packet loss, the transmission rate and the round-trip-time (RTT). In most communication systems, some form of CSI is available at the sender, such as an estimate of the fading level in a wireless channel or the congestion over a route in the Internet. Such information may be fed back from the receiver and can be used to aid in the efficient allocation of resources.

On the receiver side, the transport and application layers are responsible for de-packetizing the received transport packets, channel decoding (if FEC is used), and forwarding the intact and recovered video packets to the video decoder. The video decoder then decompresses the video packets and displays the resulting video frames in real-time (i.e., the video is displayed continuously without interruption at the decoder). The video decoder typically employs error detection and concealment techniques to mitigate the effects of packet loss. The commonality among all error concealment strategies is that they exploit correlations in the received video sequence to conceal lost information.

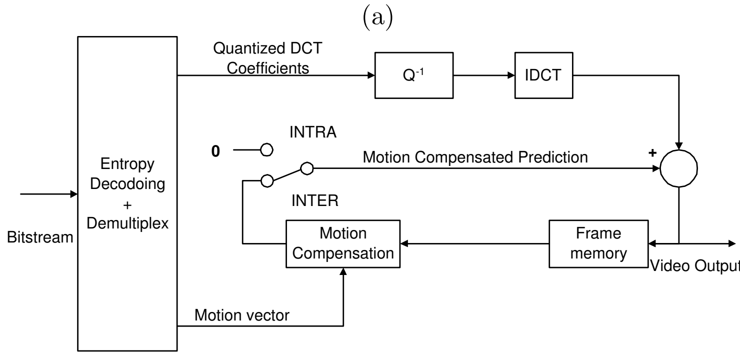
1.3.2 Video compression basics

In this section, we focus on one of the most widely used video coding techniques, that of Hybrid Block-based Motion-Compensated (HBMC) video coding, as utilized in the H.26x and MPEGx standards. In this type of video codecs, each video frame is presented in block-shaped units of associated luma and chroma samples (16×16 region) called MBs (macroblocks).

As shown in Fig. 1.5(a), the core of the encoder is motion compensated prediction (MCP). The first step in MCP is motion estimation (ME), aiming



(a)



(b)

Figure 1.5: Hybrid block-based motion-compensated video (a) encoder and (b) decoder.

to find the region from the previous frame that best matches each MB in the current frame. The offset between the MB and the prediction region is known as a motion vector. The motion vectors form a motion field, which is differentially entropy encoded. The second step is motion compensation (MC), where the reference frame is produced by applying the motion field to the previously reconstructed frame. The prediction error, known as the displaced frame difference (DFD), is obtained by subtracting the reference frame from the current frame.

Following MCP, the DFD is processed by three major blocks, namely, transform, quantization, and entropy coding. The key reason in using a transform coding is to de-correlate the data so that the associated energy in the transform domain is more compactly represented and thus the resulting transform coefficients are easier to encode. The Discrete Cosine Transform (DCT) is one of the most widely used transforms in image and video coding due to its high transform coding gain and low computational complexity. Quantization introduces loss of information, and is the primary source of the compression gain. Quantized coefficients are entropy encoded, e.g., using Huffman or arithmetic coding. The DFD is first divided into 8×8 blocks, and the DCT is then applied to each block, with the resulting coefficients quantized. In most Block-based Motion-Compensated (BMC) standards, a given MB can be intra-frame coded, inter-frame coded using motion compensated prediction, or simply replicated from the previously decoded frame. These prediction modes are denoted as INTRA, INTER, and SKIP modes, respectively. Quantization and coding are performed differently for each MB according to its mode. Thus, the coding parameters for each MB are typically represented by its prediction mode and quantization parameter.

At the decoder, as shown in Fig. 1.5(b), the inverse DCT (IDCT) is applied to the quantized DCT coefficients to obtain a reconstructed version of the DFD; the reconstructed version of the current frame is obtained by adding the reconstructed DFD to the previously reconstructed frame.

Besides DCT-based video compression, the wavelet representation provides a multi-resolution/multi-scale expression of a signal with localization in both time and frequency. One of the advantages of wavelet coders in both still image and video compression is that they are free of blocking artifacts. In addition, they usually offer continuous data rate scalability.

During the last decade, the discrete wavelet transform (DWT) and subband decomposition have gained increased popularity in image coding due to the substantial contributions in [12], [13], JPEG2000 [14], and others. Recently, there has also been active research applying the DWT to video coding [15–18]. Among the above studies, 3D wavelet or subband video codecs have received special attention due to their inherent feature of full scalability [17, 18]. Until recently, the disadvantage of these approaches has been their poor coding efficiency caused by inefficient temporal filtering. A major breakthrough which has greatly improved the coding efficiency and led to renewed efforts toward the standardization of wavelet-based scalable video coders has come from the contributions of combining lifting techniques with 3D wavelet or subband coding [19, 20].

1.3.3 Channel Models

The development of mathematical models which accurately capture the properties of a transmission channel is a very challenging but extremely important problem. For video applications, two fundamental properties of the communication channel are the probability of packet loss and the delay needed for each packet to reach the destination. In wired networks, channel errors usually appear in the form of packet loss and packet truncation. In wireless networks, besides packet loss and packet truncation, bit error is another common source of error. Packet loss and truncation are usually due to network traffic and clock drift, while bit corruption is due to the noisy air channel [21].

Internet

In the Internet, queuing delays experienced in the network can be a significant delay component. The Internet, therefore, can be modeled as an independent time-invariant packet erasure channel with random delays, as in [22]. In real-time video applications, a packet is typically considered lost and discarded if it does not arrive at the decoder before its intended playback time. Thus the packet loss probability is made up of two components: the packet loss probability in the network and the probability that the packet experiences excessive delay. Combining these two factors, the overall probability of loss for packet k is given by

$$\rho_k = \epsilon_k + (1 - \epsilon_k) P\{\Delta T_n(k) > \tau\},$$

where ϵ_k is the probability of packet loss in the network, $\Delta T_n(k)$ is the network delay for packet k, and τ is the maximum allowable network delay for this packet.

Wireless Channel

Compared to their wire-line counterparts, wireless channels exhibit higher bit error rates, typically have a smaller bandwidth, and experience multi-path fading and shadowing effects. At the IP level, the wireless channel can also be treated as a packet erasure channel, as it is "seen" by the application. In this setting, the probability of packet loss can be modeled by a function of transmission power used in sending each packet and the CSI. Specifically, for a fixed transmission rate, increasing the transmission power will increase the received SNR and result in a smaller probability of packet loss. This relationship could be determined empirically or modeled analytically. For example, in [23], an analytical model based on the notion of outage capacity is used. In this model, a packet is lost whenever the fading realization results in the channel having a capacity less than the transmission rate. Assuming a Rayleigh fading channel, the resulting probability of packet loss is given by

$$\rho_k = 1 - exp\left(\frac{1}{P_k S(\theta_k)}(2^{R/W} - 1)\right),$$

where R is the transmission rate (in source bits per sec), W the bandwidth, P_k the transmission power allocated to the k-th packet, and $S(\theta_k)$ the normalized expected SNR given the fading level, θ_k. Another way to characterize channel state is to use bounds for the bit error rate with regard to a given modulation and coding scheme; for example, in [24, 25], a model based on the error probability of BPSK (Binary Phase Shift Keying) in a Rayleigh fading channel is used.

1.3.4 End-to-end distortion

In a error prone channel, the reconstructed images at the decoder usually differ from those at the encoder due to random packet losses, as shown in Fig. 1.6. Even with the same channel characteristics, the reconstruction quality at the decoder may vary greatly based on the specific channel realization, as indicated in Figs. 1.6 (c) and (d). In this case, the most common metric used to evaluate video quality in communication systems is the expected end-to-end distortion, where the expectation is with respect to the probability of packet loss. The expected distortion for the k-th packet can be written as

$$E[D_k] = (1 - \rho_k)E[D_{R,k}] + \rho_k E[D_{L,k}], \tag{1.4}$$

where $E[D_{R,k}]$ and $E[D_{L,k}]$ are the expected distortion when the k-th source packet is either received correctly or lost, respectively, and ρ_k is its loss probability. $E[D_{R,k}]$ accounts for the distortion due to source coding as well as error propagation caused by Inter frame coding, while $E[D_{L,k}]$ accounts for the distortion due to concealment. Predictive coding and error concealment both introduce dependencies between packets. Because of these dependencies, the distortion for a given packet is a function of how other packets are encoded as well as their probability of loss. Accounting for these complex dependencies is what makes the calculation of the expected distortion a challenging problem.

Methods for accurately calculating the expected distortion have recently been proposed [10, 26]. With such approaches, it is possible, under certain conditions, to accurately compute the expected distortion with finite storage and computational complexity by using per-pixel accurate recursive calculations. For example, in [26], a powerful algorithm called ROPE is developed, which efficiently calculates the expected mean squared error by recursively computing only the first and second moments of each pixel in a frame. Model-based distortion estimation methods have also been proposed (for example, [27–29]), which are useful when the computational complexity and storage capacity are limited.

1.3.5 Error resilient source coding

If source coding removes all the redundancy in the source symbols and achieves entropy, a single error occurring at the source will introduce a great amount

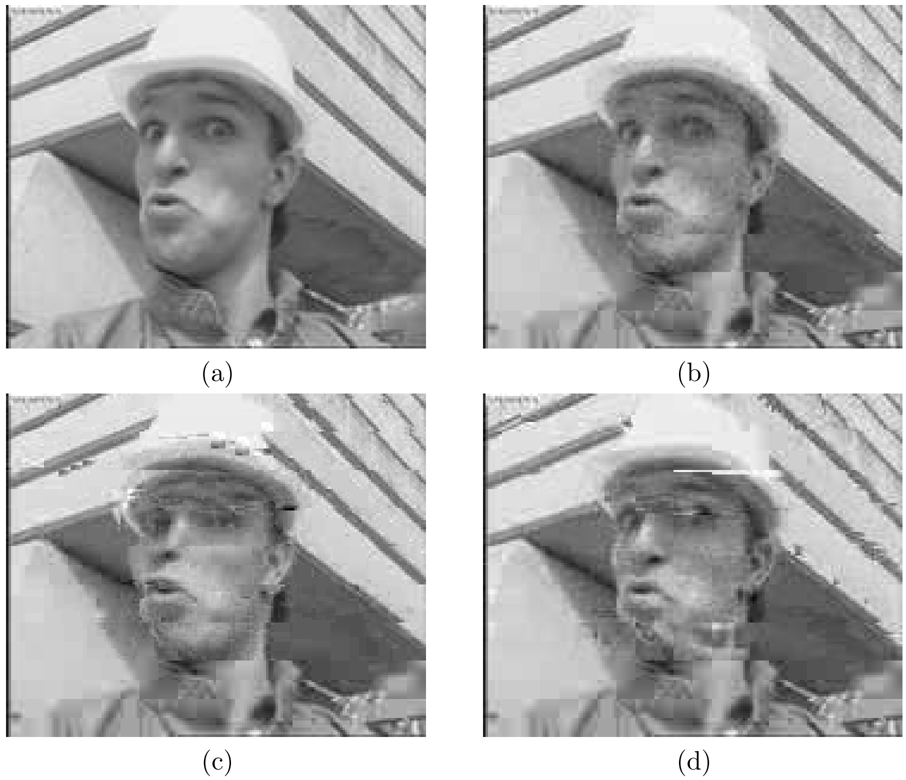


Figure 1.6: (a) Illustration of effect of random channel errors to a video stream compressed using the H.263 standard. (a) Original Frame; (b) Reconstructed frame at the encoder; and Reconstructed frame at the decoder (c) simulation 1 (d) simulation 2 (QCIF Foreman sequence, frame 92, coded with 96 kbps, frame rate 15 fps, and packet loss probabilty 10%).

of distortion. In other words, an ideal source coding is not robust to channel errors. In addition, designing an ideal or near-ideal source coder is complicated, especially for video signals, which are usually not stationary, have memory, and their stochastic distribution may not be available during encoding (especially for live video applications). Thus, redundancy certainly remains after source coding. Joint source-channel coding should not aim to remove the source redundancy completely, but should make use of it and regard it as an implicit form of channel coding [9].

For wireless video, error resilient source coding may include data partitioning, resynchronization, and reversible variable-length coding (RVLC) [1,27]. For packet-switched networks, error resilience may include the selection of the encoding mode for each packet, the use of scalable video coding, and multiple description coding (MDC) [1, 10]. In addition, packet dependency control has been recognized as a powerful tool to increase error robustness. The common methods of packet dependency control are long-term memory (LTM) prediction for MBs, reference picture selection (RPS), intra-MB insertion, and video redundancy coding (VRC) [30].

Layered video coding produces a hierarchy of bitstreams, where the different parts of an encoded stream have unequal contributions to the overall quality. Layered coding has inherent error resilience benefits, especially if the layered property can be exploited in transmission, where, for example, available bandwidth is partitioned to provide unequal error protection (UEP) for different layers with different importance. This approach is commonly referred to as *layered coding with transport prioritization* [31].

Mode selection refers to the choice between temporal prediction (Inter coding) to encode a macro-block and encoding it independently of previous frames (Intra coding). Inter coding has higher compression efficiency and thus results in lower source coding distortion than Intra coding for the same bit budget. Intra coding, on the other hand, is more resilient to error propagation and thus results in lower channel distortion. The gist of optimal mode selection method is to find the trade-off between coding efficiency and error robustness.

Mode selection algorithms have traditionally focused on Single Frame BMC coding (SF-BMC), i.e., they consider the case where the previous frame is defined as the reference for the current frame. Recently, there has been significant work on mode selection using Multiple Frame BMC (MF-BMC). Unlike SF-BMC, these approaches choose the reference frame from a group of previous frames. MF-BMC techniques capitalize on the correlation between multiple frames to improve compression efficiency and increase error resilience.

1.4 Channel Coding

In this section, we discuss the channel coding techniques that are widely used for the transmission of images and video. Two basic techniques used for video transmission are FEC and Automatic Repeat reQuest (ARQ). Each has its own benefits with regard to error robustness and network traffic load [32, 33].

As the name indicates, FEC refers to techniques in which the sender adds extra information known as check or parity information to the source information in order to make the transmission more robust to channel errors; the receiver analyzes the parity information to locate and correct errors. FEC techniques have become an important channel coding tool used in modern communication systems. One advantage of FEC techniques is that they do not require a feedback channel. In addition, these techniques improve system performance at significantly lower cost than other techniques aiming to improve channel SNR, such as increased transmitter power or antenna gain [9].

Of the two error correction techniques, FEC is usually preferred in real-time video applications, because of the delay requirements of these applications. Also, ARQ may not be appropriate for multicast scenarios due to their inherent scalability problems [1, 11]. This is because retransmission typically benefits only a small portion of receivers while all others wait unproductively, resulting in poor throughput. For these reasons, FEC-based techniques are currently under consideration by the Internet Engineering Task Force (IETF) as a proposed standard in supporting error resilience [34].

The error detection/correction capability of FEC is limited due to the restrictions on the block-size dictated by the application's delay constraints. In addition, the appropriate level of FEC usually depends heavily on the accurate estimation of the channel's behavior. ARQ, on the other hand, can automatically adapt to the channel loss characteristics by transmitting only as many redundant packets as are lost. Compared to FEC, ARQ can usually achieve a level closer to channel capacity. Of course, the tradeoff is that larger delays are introduced by ARQ. Thus, if the application has a relatively loose end-to-end delay constraint (e.g., on-demand video streaming), ARQ may be better suited. Even for real-time applications, delay constrained application-layer ARQ has been shown to be useful in some situations [22, 32, 35].

1.4.1 Forward error correction

The choice of the FEC method depends on the requirements of the system and the nature of the channel. For video communications, FEC can usually be applied across packets (at the application or transport layer) and within packets (at the link layer) [36]. In inter-packet FEC, parity packets are usually generated in addition to source packets to perform cross-packet FEC, which is usually achieved by erasure codes. At the link layer, redundant bits are added within a packet to perform intra-packet prediction from bit errors.

The Internet can usually be modeled as a packet erasure channel [11, 21, 22]. For Internet applications, many researchers have considered using erasure codes to recover packet losses [37]. With such approaches, a video stream is first partitioned into segments and each segment is packetized into a group of m packets. A block code is then applied to the m packets to generate additional l redundant packets (also called parity packets) resulting in a n-packet block, where $n = m+l$. With such a code, the receiver can recover the original m packets if a sufficient number of packets in the block are received. The most commonly studied erasure codes are Reed-Solomon (RS) codes [38]. They have good erasure correcting properties and are widely used in practice, as for example in storage devices (VCD, DVD), mobile communications, satellite communications, digital television, and high speed modems (ADSL) [37]. Another class of erasure codes that have recently been considered for network applications are Tornado codes, which have slightly worse erasure protecting properties, but can be encoded and decoded much more efficiently than RS codes [31].

RS codes are a subset of BCH codes and are linear block codes. An RS code is represented as RS(n, m) with s-bit symbols, where m is the number of source symbols and $l = n - m$ is the number of parity symbols. Figure 1.7 shows a typical RS codeword. RS codes are based on Galois fields (GF) or finite fields. RS codes with codewords from GF(q) have length equal to $q - 1$. Given a symbol size s, the maximum codeword length for an RS code is $n = 2^s - 1$. A popular RS code is chosen from the field GF($2^8 - 1$), since each

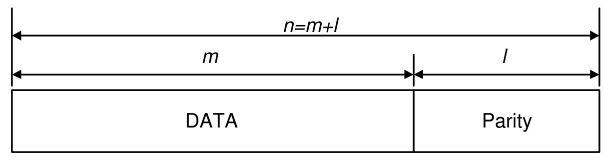


Figure 1.7: Illustrate of RS(n,m) codeword

symbol can be represented as a byte. For the detailed encoding and decoding operation rules and implementations in hardware or software, refer to [39, 40] for a comprehensive tutorial.

An RS code can be used to correct both errors and erasures (an erasure occurs when the position of an error symbol is known). An RS(n, m) decoder can correct up to $(n-m)/2$ errors or up to $(n-m)$ erasures, regardless of which symbols are lost. The code rate of an RS(n, m) code is defined as m/n. The protection capability of an RS code depends on the block size n and the code rate m/n. These are limited by the extra delay introduced by FEC. The block size can be determined based on the end-to-end system delay constraints.

The channel errors in wired links are typically in the form of packet erasures, so an RS(n, m) code applied across packets can recover up to $(n-m)$ lost packets. Thus, the block failure probability (i.e., the probability that at least one of the original m packets is in error) is $P_b(n,m) = 1 - \sum_{j=0}^{n-m} P(n,j)$, where $P(n,j)$ represents the probability of j errors out of n transmissions. As for wireless channels, channel coding is applied within each packet to provide protection. Source bits in a packet are first partitioned into m symbols, and then $(n-m)$ parity symbols are generated and added to the source bits to form a block. In this case, the noisy wireless channel causes symbol errors within packets (but not erasures). As a result, the block error probability for an RS (n, m) code can be expressed as $P_b(n,m) = 1 - \sum_{j=0}^{(n-m)/2} P(n,j)$.

Another popular type of codes used to perform intra-packet FEC is Rate-Compatible Punctured Convolutional (RCPC) codes [36], first introduced in [41]. These codes are easy to implement, and have the property of being rate compatible, i.e., a lower rate channel code is a prefix of a higher rate channel code. A family of RCPC codes is described by the mother code of rate $1/N$ and memory M with generator tap matrix of dimension $N \times (M+1)$. Together with N, the puncturing period G determines the range of code rates as $R = G/(G+l)$, where l can vary between 1 and $(N-1)G$. RCPC codes are punctured codes of the mother code with puncturing matrices $\boldsymbol{a}(l) = (a_{ij}(l))$ (of dimension $N \times G$), with $a_{ij}(l) \in (0,1)$ and 0 denoting puncturing.

The decoding of convolutional codes is most commonly achieved through the Viterbi algorithm, which is a maximum-likelihood sequence estimation algorithm. The Viterbi upper bound for the bit error probability is given by

$$p_b \leq \frac{1}{G} \sum_{d=d_{free}}^{\infty} c_d p_d$$

where d_{free} is the free distance of the convolutional code, which is defined as the minimum Hamming distance between two distinct codewords, p_d the probability that the wrong path at distance d is selected, and c_d the number of paths at Hamming distance d from the all-zero path. d_{free} and c_d are parameters of the convolutional code, while p_d depends on the type of decoding (soft or hard) and the channel. Both the theoretical bounds of BER and the simulation methods to calculate BER for RCPC codes can be found in [40, 41].

1.4.2 Retransmission

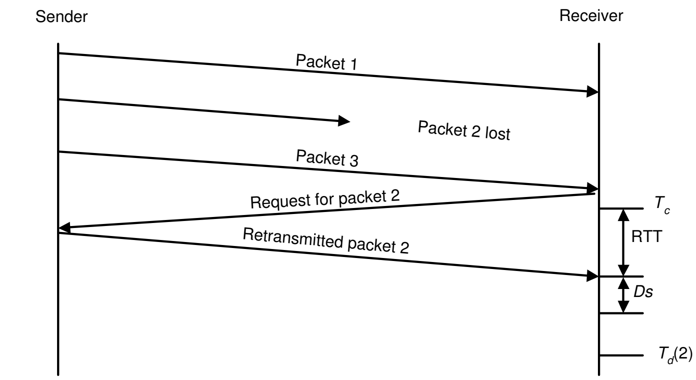
Due to the end-to-end delay constraint of real-time applications, retransmissions used for error control should be delay-constrained. Various delay-constrained retransmission schemes for unicast and multicast video are discussed in [11]. In this chapter, we focus on the unicast case, where the delay-constrained retransmissions can be classified into sender-based, receiver-based, and hybrid control, according to [11].

We illustrate the basic idea of receiver-based retransmission control in Fig. 1.8(a), where T_c is the current time, D_s is a slack term, and $T_d(n)$ is the scheduled playback time for packet n. The slack term D_s is introduced to take into account the error in estimating the RTT and other processing time, such as error correction and decoding. For a detected loss of packet n, if $T_c + RTT + D_s < T_d(n)$, which means if the retransmitted packet n can arrive at the receiver before its playback time, the receiver sends a retransmission request of packet n to the sender. This is the case depicted in Fig. 1.8(a) for packet 2.

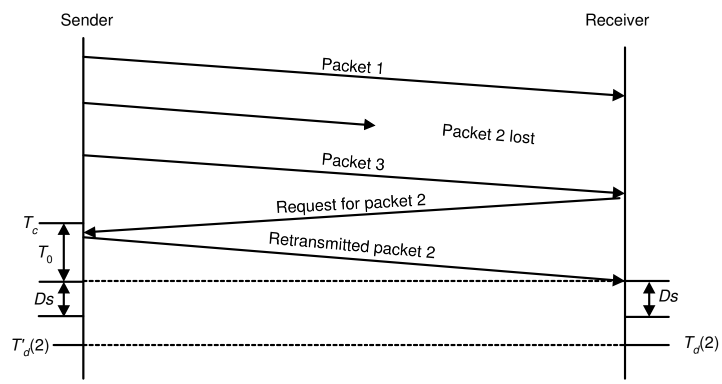
Different from the receiver-based control, in sender-based control, decisions are made at the sender end. The basic idea is illustrated in Fig. 1.8(b), where T_0 is the estimated forward-trip-time, and $T'_d(n)$ is an estimate of $T_d(n)$. If $T_c + T_0 + D_s < T'_d(n)$ holds, it can be expected that the retransmitted packet n will arrive at the receiver in time for playback. The hybrid control is a direct combination of the receiver-based and sender-based control, so that better performance can be achieved at the cost of higher complexity. After laying out the basic concept of delay-constraint retransmission, we next discuss how retransmission techniques are implemented in a network.

Delay-constrained retransmission can be implemented in multiple network layers. First, it is well known that Transport Control Protocol (TCP) is a reliable end-to-end transport protocol that provides reliability by means of a window-based positive acknowledgement (ACK) with a go-back-N retransmission scheme in the IP suite [42].

In an IP-based wireless network for the emerging 3G and 4G systems, such as CDMA2000, transport packets are transferred to the radio link control (RLC) frames and further to the Medium Access Control (MAC) frames in the link layer. 3G and 4G systems allow both RLC frame retransmissions and MAC frame retransmissions [43]. The current Wireless Local-Area Network (WLAN) standard IEEE 802.11 also allows MAC frame retransmission [44]. Compared to transport-layer retransmission TCP, link-layer and MAC-layer retransmission techniques introduce smaller delays, because the lower layers react to the



(a)



(b)

Figure 1.8: Timing diagram for delay-constrained retransmission (a) receiver-based control (b) sender-based control (adapted from [11]).

network faster than the upper layers [45]. Due to the strict end-to-end delay constraint, TCP is usually not preferred for real-time video communications. However, because delay-constrained retransmission at the link and MAC layers introduce much shorter delays, they are widely used in real-time video communications [44]. For example, researchers have been studying how many retransmissions in the MAC layer are appropriate for multimedia transmission applications in order to achieve the best tradeoff between error correction and delay [44–46].

1.5 Joint Source-Channel Coding

Error resilient source coding and channel coding are both error control mechanisms available at the sender. In this chapter, we focus on techniques that aim to allocate the available resources to these two components of the transmission chain in order to provide the best end-to-end video quality.

As a preliminary matter of a formal approach to problem solving, several factors need to be clarified. An appropriate system performance evaluation metric should first be selected. Second, the constraints need to be specified. Third, a model of the relationship between the system performance metric and the set of adaptation parameters needs to be established. The final step is to find the best combination of adaptation parameters that maximizes the system performance while meeting the required constraints. Keeping those four steps in mind, we next present a formal approach to formulate and provide solutions to the joint source-channel coding problem.

1.5.1 Problem formulation

A commonly used criterion for the evaluation of the system performance is the expected distortion. The expectation is required due to the stochastic nature of the channel. As shown in (1.4), in calculating the expected distortion for each source packet, the two distortion terms, $E[D_{R,k}]$ and $E[D_{L,k}]$, and the loss probability for the source packet ρ_k need to be determined. The two distortion terms depend on the source coding parameters such as quantization stepsize and prediction mode, as well as error concealment schemes used at the decoder. The relationship between the source packet loss probability and channel characteristics depends on the specific packetization scheme, the channel model, and the adaptation parameters chosen.

Let $\mathcal{S}$ be the set of source coding parameters, and $\mathcal{C}$ the channel coding parameter. Let $\boldsymbol{s} = \{s_1, ..., s_M\} \in \mathcal{S}^M$ and $\boldsymbol{c} = \{c_1, ..., c_M\} \in \mathcal{C}^M$ denote, respectively, the vector of source coding parameters and channel coding parameters for the M packets in one video frame or a group of frames. The general formulation then is to minimize the total expected distortion for the frame(s), given the corresponding bit rate constraint [47], i.e.,

$$\begin{aligned} &\min_{\{\boldsymbol{s}\in\mathcal{S}^M, \boldsymbol{c}\in\mathcal{C}^M\}} E[D(\boldsymbol{s},\boldsymbol{c})] \\ &\text{s.t.} \quad R(\boldsymbol{s},\boldsymbol{c}) \le R_0, \end{aligned} \tag{1.5}$$

where $R(\boldsymbol{s},\boldsymbol{c})$ represents the total number of bits used for both source and channel coding, and R_0 the bit rate constraint for the frame(s). The bit rate constraint is usually obtained based on the estimated channel throughput. Note that since video packets are usually of different importance, the optimal solution will result in an UEP cross video packets.

As shown in Fig. 1.6, with the same channel characteristics, different simulations may diverge to a large extent with regard to reconstruction quality. A novel approach called VAPOR (Variance-Aware per-Pixel Optimal Resource allocation) is proposed in [48] to deal with this. Besides the widely used expected distortion matric, the VAPOR approach aims to limit error propagation from random channel errors by accounting for both the expected value and the variance of the end-to-end distortion when allocating source and channel resources. By accounting for the variance of the distortion, this approach increases the reliability of the system by making it more likely that what the end-user sees, closely resembles the mean end-to-end distortion calculated at the transmitter.

This type of constrained problem can be solved in general using the Lagrangian relaxation method; that is, instead of the original problem, the following problem is solved

$$\min_{\{\boldsymbol{s}\in\mathcal{S}^M,\boldsymbol{c}\in\mathcal{C}^M\}} J(\boldsymbol{s},\boldsymbol{c},\lambda) = \min_{\{\boldsymbol{s}\in\mathcal{S}^M,\boldsymbol{c}\in\mathcal{C}^M\}} \{E[D(\boldsymbol{s},\boldsymbol{c})] + \lambda R(\boldsymbol{s},\boldsymbol{c})\}, \tag{1.6}$$

The solution of (1.5) can be obtained, within a convex hull approximation, by solving (1.6) with the appropriate choice of the Lagrange multiplier, $\lambda \geq 0$, so that the bit rate constraint is satisfied. The difficulty in solving the resulting relaxed problem depends on the complexity of the inter-packet dependencies. Depending on the nature of such dependencies, an iterative descent algorithm based on the method of alternating variables for multivariate minimization [49] or a deterministic dynamic programming algorithm [50] can be employed to efficiently solve the minimization problem.

The JSCC problem formulation (1.5) is general for the fact that both the source coding and channel coding can take a variety of forms, depending on the specific application. For example, when FEC is utilized, the packet loss probability becomes a function of the FEC choice. The details of this model will depend on how transport packets are formed from the available video packets [51]. In addition to FEC, retransmission-based error control may be used in the form of ARQ protocols. In this case, the decision whether to retransmit a packet or to send a new one forms another channel coding parameter, which also affects the probability of loss as well as the transmission delay. When considering the transmission of video over a network, a more general joint source-channel coding scheme may cover modulation and demodulation [52], power adaptation [23], packet scheduling [53], and data rate adaptation [53]. These adaptation components can all be regarded as channel coding parameters. Source coding parameters, on the other hand, can be in the form of mode selection [10, 23, 51], packetization [44], intra-MB refreshment rate [28], and entropy coding mechanism [6]. By solving problem (1.5) and selecting the source coding and channel coding parameters within their sets, we can obtain the optimal tradeoff among all those adaptation components. We next provide examples of the applications of JSCC to video transmission in different network infrastructures.

1.5.2 Internet video transmission

For video transmission over the Internet, channel coding usually takes the form of FEC and/or ARQ at the transport layer. FEC is usually preferred for applications that impose a relatively short end-to-end delay constraint. Joint source coding and FEC has been extensively studied in the literature [24, 37, 54–56]. Such studies focus on the determination of the optimal bit allocation between source coding and FEC. In [57], the authors introduced the integrated joint source-channel coding (IJSCC) framework, where error resilient source coding, channel coding, and error concealment are jointly considered in a tractable optimization setting. In using the IJSCC framework, an R-D optimized hybrid error control scheme has been presented in [58], which results in the optimal allocation of bits among source, FEC, and ARQ.

Joint source coding and FEC

As mentioned above, the appropriate way of calculating the loss probability per packet depends on the chosen FEC as well as the way transport packets are formed from the available video packets. Next we show one example where the source packet is a video slice (a group of blocks).

Figure 1.9 illustrates a packetization scheme for a frame, where one row corresponds to one packet. In this packetization scheme, one video slice is directly packetized into one transport packet by the attachment of a transport packet header. Since the source packet sizes (shown by the shaded area in Fig. 1.9) are usually different, the maximum packet size of a block (a group of packets protected by one RS code) is determined first, and then all packets are padded with stuffing bits in the tail part to make their sizes equal. The stuffing bits are removed after the parity codes are generated and thus are not transmitted. The resulting parity packets are all of the same size (maximum packet size mentioned above). Each source packet in Fig. 1.9(a) is protected by an RS(N, M) code, where M is the number of video packets and N is the number of total packets including parity packets.

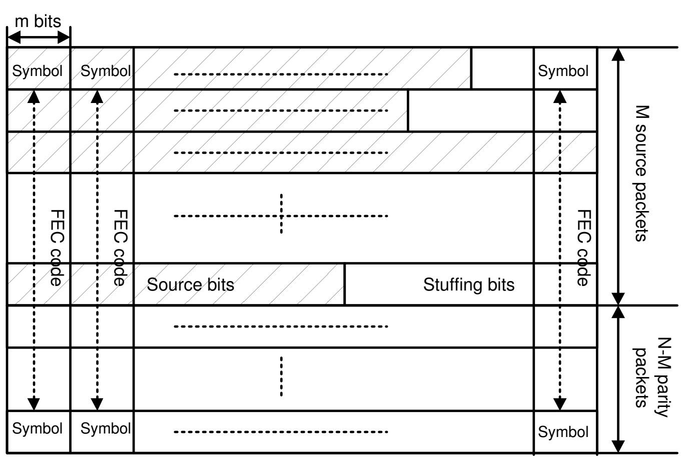


Figure 1.9: Illustration of a packetization scheme for inter-packet FEC.

In this case, the channel coding parameter $\boldsymbol{c}$ would be the choice of the RS code rate. If we take the source coding parameter $\boldsymbol{s}$ as the prediction mode and quantizer for each video packet, by solving (1.5), we can obtain the optimal JSCC solution, i.e., the optimal bit allocation as well as the optimal error resilient source coding and FEC.

To illustrate the advantage of the IJSCC approach, we compare two systems: i) system 1, which uses the proposed framework to jointly consider error resilient source coding and channel coding; ii) system 2, which performs error resilient source coding, but with fixed rate channel coding. Note that system 2 is also optimized, i.e., it performs optimal error resilient source coding to adapt to the channel errors (with fixed rate channel coding).

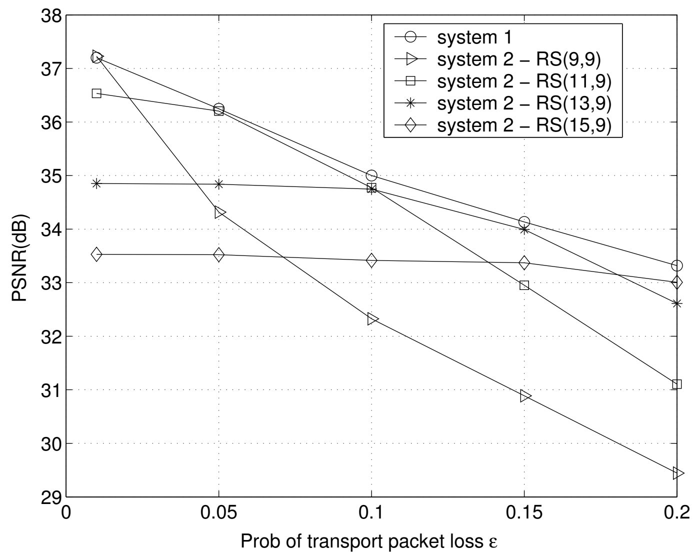


Figure 1.10: Average PSNR vs. transport packet loss probability (QCIF Foreman sequence, transmission rate 480 kbps, coded at 15 fps).

In Fig. 1.10, the performance of the two systems is compared, using the QCIF Foreman test sequence coded by an H.263+ codec at transmission rate 480 kbps and frame rate 15 fps. Here, we plot the average PSNR in dB versus different packet loss rates. It can be seen in Fig. 1.10 that system 1 outperforms system 2 at different pre-selected channel coding rates. In addition, system 1 is above the envelope of the four performance curves of system 2 by 0.1 to 0.4 dB. This is due to the flexibility of system 1, which is capable of adjusting the channel coding rate in response to the CSI as well as the varying video content.

Joint source coding and hybrid FEC/ARQ

When considering the use of both FEC and ARQ, the channel coding parameter $\boldsymbol{c}$ includes the FEC rate chosen to protect each packet and the retransmission policy for each lost packet. Hybrid FEC/retransmission has been considered in [22], where a general cost-distortion framework was proposed to study several scenarios such as DiffServ (Differentiated Services), sender-driven retransmission, and receiver-driven retransmission. In [58], optimal error control is performed by jointly considering source coding with hybrid FEC and sender-driven application-layer selective retransmission. This study is carried out with the use of (1.5), with a sliding window scheme in which lost packets are selectively retransmitted according to a rate-distortion optimized policy. Simulations in [58] show that the performance advantage in using either FEC or selective retransmission depends on the packet loss rate and the round-trip time. In that work, the proposed hybrid FEC and selective retransmission approach is able to derive the benefits of both approaches by adapting the type of error control based on the channel characteristics.

A receiver-driven hybrid FEC/Pseudo-ARQ mechanism is proposed for Internet multimedia multicast in [33]. In that work, the sender multicasts all the source layers and all the channel layers (parity packets obtained by using RS coding similar to what we have discussed in the previous section) to separate multicast groups. Each user computes the optimal allocation of the available bit rate between source and channel layers based on its estimated channel bandwidth and packet loss probability, and joins the corresponding multicast group. This is achieved through a pseudo-ARQ system, in which the sender continuously transmits delayed parity packets to additional multicast group, and the receivers can join or leave a multicast group to retrieve the lost information up to a given delay bound. Such a system looks like ordinary ARQ to the receiver and an ordinary multicast to the sender. This can be characterized as JSCC with receiver feedback. More specifically, the optimal JSCC is obtained by solving (1.5) at the receiver side, where the source coding parameter is the number of source layers, and the channel coding parameter is the number of channel layers.

1.5.3 Wireless video transmission

Wireless video communications is a broad, active, and well-studied field of research [59, 60]. Recently, several adaptation techniques have been proposed specifically for energy efficient wireless video communications. A trend in this field of research is the joint adaptation of source coding and transmission parameters based on the time-varying source content and channel conditions. The general JSCC framework (1.5) therefore encompasses these techniques with an additional constraint on the total energy consumed in delivering the video sequence to the end-user. Correspondingly, the channel coding parameters would cover more general channel adaptation parameters such as the transmission rate, physical-layer modulation modes, and the transmitter power.

Joint source coding and FEC

As with Internet video transmission, the problem of joint source coding and FEC for wireless video communications focuses on the optimal bit allocation between source and channel coding by solving (1.5). The difference is that due to the different type of channel errors (bit errors) in a wireless channel, FEC is achieved by adding redundant bits within packets to provide intra-packet protection. RCPC and RS codes are widely used in this case.

Optimal bit allocation has been studied in [55] based on a subband video codec. A Binary Symmetric Channel (BSC) with Additive White Gaussian Noise (AWGN) model have been considered for simulations. The source coding parameters are the bit rate of the source subband and the channel coding parameters are the FEC parameter for each subband. A similar problem has been studied for video transmission over a Rayleigh fading wireless channel in [24] based on an H.263+ SNR scalable video codec. In that work, Universal R-D Characteristics (URDC) of the source scheme are employed to make the optimization tractable. Both works used RCPC codes to achieve the intra-packet FEC.

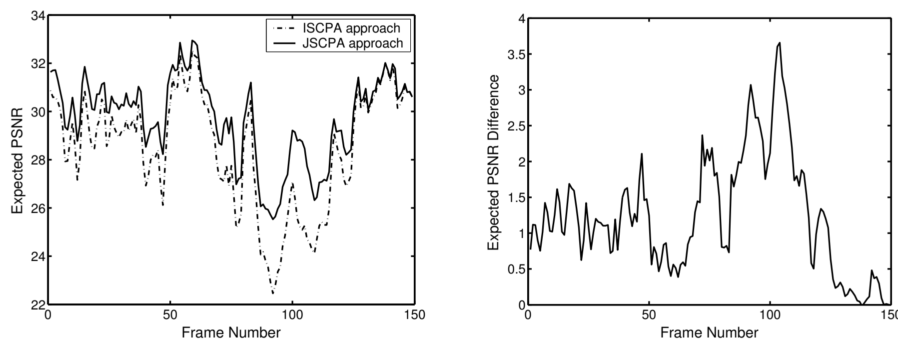
RS codes are used to perform channel coding in [28] for video transmission over a random BSC. Based on their proposed RD model, the source coding parameter is the intra-MB refreshment rate and the channel coding parameter is the channel rate.

Joint source coding and power adaptation

Joint source coding and power allocation techniques account for the varying error sensitivity of video packets by adapting the transmission power per packet based on the source content and the CSI. In other words, these techniques use transmission power as an UEP mechanism. In this case, the channel coding parameter is the power level for each video packet. Video transmission over CDMA networks using a scalable source coder (3-D SPIHT), along with error control and power allocation is considered in [61]. A scheme for allocating source rate and transmission power under bandwidth constraints is considered in [62]. In [23], optimal mode and quantizer selection is considered jointly with transmission power allocation.

To illustrate some advantages of joint adaptation of the source coding and transmission parameters in wireless video transmission systems, we present experimental results which are discussed in detail in [23]. We compare a joint source coding and transmission power allocation (JSCPA) approach, i.e., the approach described by (1.5), with an independent source coding and power allocation (ISCPA) approach in which $\mathcal{S}$ and $\mathcal{C}$ are independently adapted. In Fig. 1.11, we plot the expected PSNR per frame of both approaches for the Foreman test sequence coded at 15 fps. It is important to note that both approaches use the same transmission energy and delay per frame.

As shown in Fig. 1.11, the JSCPA approach achieves significantly higher quality (expected PSNR) per frame than the ISCPA approach. Because the video encoder and the transmitter operate independently in the ISCPA approach, the relative importance of each packet, i.e., their contribution to the total distortion, is unknown to the transmitter. Therefore, the transmitter treats each packet equally and adapts the power in order to maintain a constant probability of packet loss. The JSCPA approach, on the other hand, is able to adapt the power per packet, and thus the probability of loss, based on the



(a) (b)

Figure 1.11: (a) Expected PSNR per frame for the ISCPA and JSCPA approaches; (b) Difference in expected PSNR between the two approaches (adopted from [23]).

relative importance of each packet. For example, more power can be allocated to packets that are difficult to conceal. As shown in Fig. 1.11, the PSNR improvement is the greatest during periods of high activity. For example, around frame 100 there is a scene change in which the camera pans from the foreman to the construction site. During this time, the JSCPA approach achieves PSNR improvements of up to 3.5 dB. This gain comes from the ability of the JSCPA approach to increase the power while decreasing the number of bits sent in order to improve the reliability of the transmission. The ISCPA scheme is unable to adapt the protection level and thus incurs large distortion during periods of high source activity.

We show the visual quality comparison of the two approaches in Fig. 1.12. An expected reconstructed frame is shown from the "Foreman" sequence when the same amount of energy is consumed in the two approaches. It can be clearly seen that the JSCPA approach achieves a much better video reconstruction quality than the ISCPA approach.

As mentioned above, the VAPOR approach is used to limit error propagation by accounting for not only the mean but also the variance of the end-to-end distortion [48]. In Fig. 1.13, we compare a series of reconstructed frames at the decoder for the Minimum Expected Distortion (MED) approach (1.5) and VAPOR approach using the same amount of transmission energy for the *Silent* sequence. These images are for a single channel loss realization when the same MBs are lost in both schemes. We can clearly see the advantage of using the VAPOR approach. For example, the error occurring at frame 109 persists utill frame 123 in the MED approach while it has been quickly removed by the VAPOR approach.

(a) (b) (c)

Figure 1.12: (a) Frame 184 in the Foreman sequence. (a) Original Frame; (b) expected frame at the decoder using the JSCPA approach; and (c) expected frame at the decoder using the ISCPA approach (adopted from [23]).

Joint source-channel coding and power adaptation

In an energy-efficient wireless video transmission system, transmission power needs to be balanced against delay to achieve the best video quality. Specifically, for a given transmission rate, increasing the transmission power will decrease BER, resulting in a smaller probability of packet loss. On the other hand, for a fixed transmission power, increasing the transmission rate will increase the BER but decrease the transmission delay needed for a given amount of data (or allow more data to be sent within a given time-period). Therefore, in order to efficiently utilize resources such as energy and bandwidth, those two adaptation components should be designed jointly.

The problem of joint source-channel coding and power adaptation was studied in [25, 63, 64]. In this case, the general channel coding parameters consist of both channel coding and power allocations. In [63, 64], the study is based on scalable video, and error resilient source coding is achieved through optimized transport prioritization for layered video. The study in [25] is based on

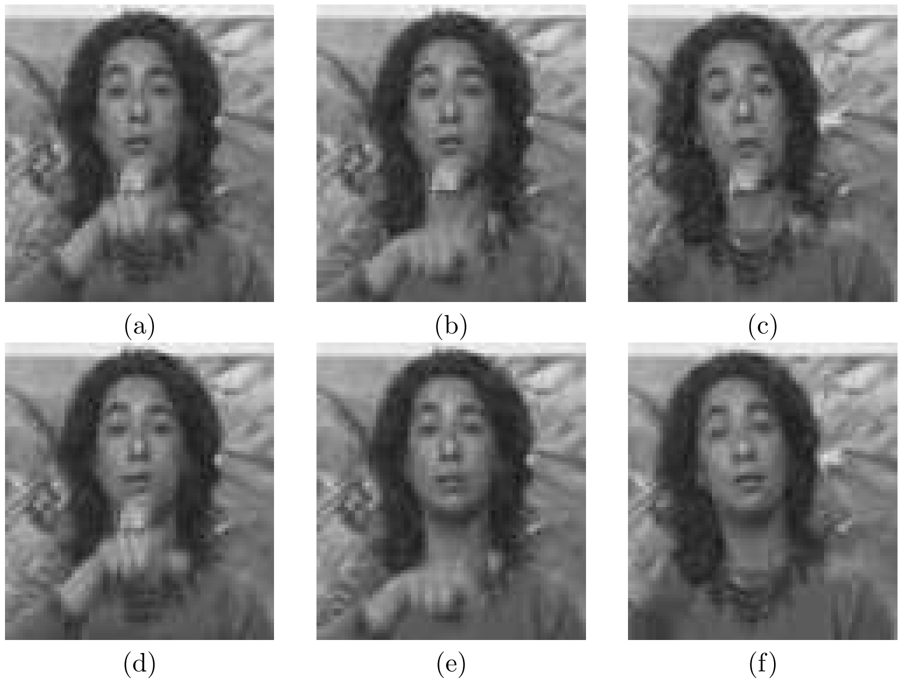


Figure 1.13: Illustration of error propagation effect using the QCIF Silent test sequence. MED approach: frame number (a) 109, (b) 110, (c) 123. VAPOR approach: frame number (d) 109, (e) 110, (f) 123. (adopted from [48])

an H.263+ codec, and the error resilient source coding is achieved by mode selection.

Joint source coding and data rate adaptation

Joint source coding and transmission rate adaptation has also been studied as a means of providing energy efficient video communications. In order to maintain a certain probability of loss, the energy consumption increases as the transmission rate increases [65]. Therefore, in order to reduce energy consumption, it is advantageous to transmit at the lowest rate possible [66]. In addition to affecting energy consumption, the transmission rate determines the number of bits that can be transmitted within a given period of time. Therefore, as the transmission rate decreases, the distortion due to source coding increases. Joint source coding and transmission rate adaptation techniques adapt the source coding parameters and the transmission rate in order to balance energy consumption against end-to-end video quality. In [53], the authors consider optimal source coding and transmission rate adaptation. Stochastic Dynamic Programming is used to find an optimal source coding and transmission policy based on a Markov state channel model. A key idea in this work is that the performance can be improved by allowing the transmitter to suspend or slow down transmissions during periods of poor channel conditions, as long as the delay constraints are not violated.

1.6 Discussion

While application of Shannon's separation theorem leads to the introduction of redundancy only during channel coding for achieving error-free transmission, this is not the case under real-time constraints. Redundancy needs to be introduced during both source and channel encoding in a judicious way. Furthermore, a well-designed decoder can recover some of the lost information utilizing error-concealment techniques. When a feedback channel is available, a retransmission protocol can be implemented, offering a different means for improving the error resiliency of the video communication system.

In this chapter, joint source-channel coding (JSCC) for video communications has been discussed. We have used the term "channel encoding" in a general way to include modulation and demodulation, power adaptation, packet scheduling, and data rate adaptation. We provided an overview of the state-of-the-art implementations of JSCC in various network infrastructures. Although the most recent video coding standards H.263, MPEG4, and H.264 provide a number of error resilient tools, there are a number of resource allocation problems which need to be resolved in order to efficiently utilize such tools. In addition, there is a plethora of issues that need to be addressed by considering new system structures.

As mentioned earlier, cross-layer design is a general term, which encompasses JSCC, and represents the current state of the art. In order to efficiently utilize limited network resources, the video transmission system needs to be adaptive to the changing network conditions. In the traditional layered protocol stack, each layer is optimized or adapted to the changing network conditions independently. The adaptation, however, is very limited due to the limited conversations between layers. Cross-layer design aims to improve the system's overall performance by jointly considering multiple protocol layers. The studies on this topic so far not only show the necessity to employ the joint design of multiple layers, but also point out the future direction of network protocol suite development to better support video communications over the current best effort networks.

Cross-layer design is a powerful approach to account for different types of channel errors in a hybrid wireless/wireline network that consists of both wired and wireless links. An initial investigation of this topic is described in [51], where lower layer adaptation includes inter-packet FEC at the transport layer and intra-packet FEC at the link layer, which are respectively used to combat packet losses in the wired line and bit errors in the wireless link. Such channel adaptations are jointly designed with mode selection in source coding to achieve optimal UEP, by solving (1.5).

Overall the topic addressed in this chapter does not represent mature technology yet. Although technologies providing higher bit rates and lower error rates are continuously being deployed, higher QoS will inevitably lead to higher user demands of service, which for video applications translates to higher resolution images of higher visual quality.

Bibliography

[1] Y. Wang, G. Wen, S. Wenger, and A. K. Katsaggelos, "Review of error resilience techniques for video communications," *IEEE Signal Processing Magazine*, vol. 17, pp. 61–82, July 2000.

[2] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE International Convention Records*, pp. 142–163, 1959, Part 4.

[3] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J. 27*, pp. 379–423 and 623–656, July 1948.

[4] G. M. Schuster and A. K. Katsaggelos, *Rate-Distortion Based Video Compression: Optimal Video Frame Compression and Object Boundary Encoding*, Kluwer Academic Publishers, 1997.

[5] B. Sklar, *Digital Communications: Fundamentals and Applications*, Englewood Cliffs, NJ: Prentice-Hall Inc., 2nd edition, 2001.

[6] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: an approach to combined source-channel coding," *IEEE Trans. Inform. Theory*, vol. IT-38, pp. 827–838, 1987.

[7] F. Farvardin, "A study of vector quantization for noisy channels," *IEEE Trans. Inform. Theory.*, vol. 36, pp. 799–809, July 1990.

[8] A. Kurtenbach and P. Wintz, "Quantizing for noisy channels," *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 291–302, Apr. 1969.

[9] R. E. Van Dyck and D. J. Miller, "Transport of wireless video using separate, concatenated, and joint source-channel coding," *Proceedings of the IEEE*, vol. 87, pp. 1734–1750, Oct. 1999.

[10] D. Wu, Y. T. Hou, B. Li, W. Zhu, Y.-Q. Zhang, and H. J. Chao, "An end-to-end approach for optimal mode selection in Internet video communication: theory and application," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 977–995, June 2000.

[11] D. Wu, Y. T. Hou, and Y.-Q. Zhang, "Transporting real-time video over the Internet: Challenges and approaches," *Proc. IEEE*, vol. 88, pp. 1855–1877, Dec. 2000.

[12] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3463, Dec. 1993.

[13] A. Said and W. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on Circ. and Syst. for Video Techn.*, vol. 6, pp. 243–250, June 1996.

[14] *JPEG-2000 VM3.1 A Software*, ISO/IECJTC1/SC29/WG1 N1142, Jan. 1999.

[15] K. Shen and E. J. Delp, "Wavelet based rate scalable video compression," *IEEE Trans. Circ. and Syst. for Video Techn.*, vol. 9, pp. 109–122, Feb. 1999.

[16] Y.-Q. Zhang and S. Zafar, "Motion-compensated wavelet transform coding for color video compression," *IEEE Trans. on Circ. and Syst. for Video Techn.*, vol. 2, pp. 285–296, Sept. 1992.

[17] J. R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Proc.*, vol. 3, pp. 559–571, Sept. 1994.

[18] S. Choi and J. W. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Trans. Image Proc.*, vol. 8, pp. 155–167, Feb. 1999.

[19] A. Secker and D. Taubman, "Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Thessaloniki, Greece, Oct. 2001.

[20] J.-R. Ohm, "Motion-compensated wavelet lifting filters with flexible adaptation," in *Proceedings of Tyrrhenian International Workshop on Digital Communications*, Capri, Italy, Sept. 2002.

[21] D. A. Eckhardt, *An Internet-style approach to managing wireless link errors*, Ph.D Thesis, Carnegie Mellon University, Pittsburgh, PA, May 2002.

[22] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Trans. on Multimedia*, 2001, Submitted.

[23] Y. Eisenberg, C. E. Luna, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source coding and transmission power management for energy efficient wireless video communications," *IEEE Trans. on Circuits and System for Video Technology*, vol. 12, no. 6, pp. 411–424, June 2002.

[24] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint source-channel coding for motion-compensated DCT-based SNR scalable video," *IEEE Trans on Image Processing*, vol. 11, pp. 1043–1052, Sept. 2002.

[25] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source-channel coding and power allocation for energy efficient wireless communications," in *Proc. 41st Allerton Conf. on Communication, Control and Computing*, Otc. 2003.

[26] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 966–976, June 2000.

[27] G. Côté, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error-prone networks," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 952–965, June 2000.

[28] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circ. and Syst. for Video Techn.*, vol. 12, pp. 511–523, June 2002.

[29] T. Wiegand, N. Farber, and B. Girod, "Error-resilient video transmission using long-term memory motion-compensated prediction," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 1050–1062, June 2000.

[30] B. Girod, J. Chakareski, M. Kalman, Y. J. Liang, E. Setton, and R. Zhang, "Advances in network-adaptive video streaming," in *Proc. 2002 Tyrrhenian International Workshop on Digital Communications*, 2002, pp. 1–8.

[31] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1737–1744, Nov. 1996.

[32] F. Hartanto and H. R. Sirisena, "Hybrid error control mechanism for video transmission in the wireless IP networks," in *Proc. of IEEE Tenth Workshop on Local and Metropolitan Area Networks (LANMAN'99)*, Sydney, Australia, Nov. 1999, pp. 126–132.

[33] P. A. Chou, A. E. Mohr, A. Wang, and S. Mehrotra, "Error control for receiver-driven layered multicast of audio and video," *IEEE Trans. Multimedia*, pp. 108–122, March 2001.

[34] J. Rosenberg and H. Schulzrinne, "An RTP payload format for generic forward error correction," Tech. Rep., Internet Engineering Task Force, Request for Comments (Proposed Standard) 2733, Dec. 1999.

[35] B. J. Dempsey, J. Liebeherr, and A. C. Weaver, "On retransmission-based error control for continuous media traffic in packet-switched networks," *Computer Networks and ISDN Syst.*, vol. 28, pp. 719–736, Mar. 1996.

[36] N. Celandroni and F. Pototì, "Maximizing single connection TCP goodput by trading bandwidth for BER," *Int. J. of Commun. Syst.*, vol. 16, pp. 63–79, Feb. 2003.

[37] M. Gallant and F. Kossentini, "Rate-distortion optimized layered coding with unequal error protection for robust Internet video," *IEEE Trans. on Circ. and Syst. for Video Techn.*, vol. 11, no. 3, pp. 357–372, March 2001.

[38] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *SIAM J. Applied Math.*, vol. 8, pp. 300–304, 1960.

[39] S. B. Wicker and V. K. Bhargava, *Reed-Solomon Codes and Their Applications*, John Wiley & Sons Inc., Sept. 1999.

[40] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, Aug. 2000.

[41] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, pp. 389–400, Apr. 1988.

[42] Douglas E. Comer, *Internetworking with TCP/IP*, vol. 1, Prentice-Hall, Upper Saddle River, NJ, 1995.

[43] P. Luukkanen, Z. Rong, and L. Ma, "Performance of 1XTREME system for mixed voice and data communications," in *Proceedings IEEE Inf. Conf. Communications*, Helsinki, Finland, June 2001, pp. 1411–1415.

[44] M. var der Schaar, S. Krishnamachari, S. Choi, and X. Xu, "Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 1752–1763, Dec. 2003.

[45] A. Chockalingam and G. Bao, "Performance of TCP/RLP protocol stack on correlated fading DS-CDMA wireless links," *IEEE Trans. Veh. Tech.*, vol. 49, pp. 28–33, Jan. 2000.

[46] H. Zheng, "Optimizing wireless multimedia transmissions through cross layer design," in *Proc. IEEE*, Baltimore, MD, July 2003, vol. 1, pp. 185–188.

[47] A. K. Katsaggelos, Y. Eisenberg, F. Zhai, R. Berry, and T. N. Pappas, "Advances in efficient resource allocation for packet-based real-time video transmission," *Proceedings of the IEEE*, Jan. 2005.

[48] Y. Eisenberg, F. Zhai, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "VAPOR: Variance-aware per-pixel optimal resource allocation," *IEEE Trans. Image Processing*, 2005, To appear.

[49] R. Fletcher, *Practical methods of optimization*, New York: Wiley, 2nd edition, 1987.

[50] D. P. Bertsekas, *Dynamic Programming: Determinstistic and Stochastic methods*, Prentice-Hall, 1987.

[51] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Rate-distortion optimized product code forward error correction for video transmission over IP-based wireless networks," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, Montreal, Canada, June 2004.

[52] Y. Pei ad J. W. Modestino, "Multi-layered video transmission over wireless channels using an adaptive modulation and coding scheme," in *Proc.2001 IEEE Int. Conf. Image Processing*, Thesaloniki, Greece, Oct. 2001.

[53] C. E. Luna, Y. Eisenberg, R. Berry, T. N. Pappas, and A. K. Katsaggelos, "Joint source coding and data rate adaption for energy efficient wireless video streaming," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 1710–1720, Dec. 2003.

[54] G. Davis and J. Danskin, "Joint source and channel coding for Internet image transmission," in *Proc. SPIE and Conf. Wavelet Applications of Digital Image Processing XIX*, Denver, CO, Aug. 1996.

[55] G. Cheung and A. Zakhor, "Bit allocation for joint source/channel coding of scalable video," *IEEE Trans. Image Processing*, vol. 9, pp. 340–356, March 2000.

[56] J. Kim, R. M. Mersereau, and Y. Altunbasak, "Error-resilient image and video transmission over the Internet using unequal error protection," *IEEE Trans. Image Processing*, vol. 12, pp. 121–131, Feb. 2003.

[57] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "An integrated joint source-channel coding framework for video communications," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 2004.

[58] F. Zhai, Y. Eisenberg, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Rate-distortion optimized hybrid error control for real-time packetized video transmission," in *Proc. IEEE Int. Conf. Communications (ICC'04)*, Paris, France, June 2004.

[59] B. Girod and N. Farber, "Wireless video," in *Compressed Video Over Networks*, New York: Marcel Dekker, 2000, pp. 124–133, A. Reibman and M.-T. Sun, Eds.

[60] D. Wu, Y. Hou, and Y.-Q. Zhang, "Scalable video coding and transport over broad-band wireless networks," *Proc. IEEE*, vol. 89, pp. 6–20, Jan. 2001.

[61] S. Zhao, Z. Xiong, and X. Wang, "Joint error control and power allocation for video transmission over CDMA networks with multiuser detection," *IEEE Trans. Circ. and Syst. for Video Techn.*, vol. 12, pp. 425–437, June 2002.

[62] Y. S. Chan and J. W. Modestino, "A joint source coding-power control approach for video transmission over CDMA networks," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 1516–1525, Dec. 2003.

[63] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "Network-adaptive scalable video streaming over 3G wireless network," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Thessaloniki, Greece, Oct. 2001, vol. 3, pp. 579–582.

[64] S. Appadwedula, D. L. Jones, K. Ramchandran, and L. Qian, "Joint source channel matching for wireless image transmission," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Chicago, IL, Oct. 1998, vol. 2, pp. 137–141.

[65] R. G. Gallager, "Energy limited channels: Coding, multi-access and spread spectrum," Tech. Rep., M.I.T. LIDS-P-1714, Nov. 1987.

[66] A. El Gamal, C. Nair, B. Prabhakar, E. Uysal-Biyikoglu, and S. Zahedi, "Energy-efficient scheduling of packet transmissions over wireless networks," in *Proceedings of INFOCOM'02*, 2002.