Analysis of Video Transmission over Lossy Channels

Klaus Stuhlmüller, Niko Färber, *Member, IEEE*, Michael Link, and Bernd Girod, *Fellow, IEEE*

***Abstract*—A theoretical analysis of the overall mean squared error (MSE) in hybrid video coding is presented for the case of error prone transmission. Our model covers the complete transmission system including the rate-distortion performance of the video encoder, forward error correction, interleaving, and the effect of error concealment and interframe error propagation at the video decoder. The channel model used is a 2-state Markov model describing burst errors on the symbol level. Reed–Solomon codes are used for forward error correction. Extensive simulation results using an H.263 video codec are provided for verification. Using the model, the optimal tradeoff between INTRA and INTER coding as well as the optimal channel code rate can be determined for given channel parameters by minimizing the expected MSE at the decoder. The main focus of this paper is to show the accuracy of the derived analytical model and its applicability to the analysis and optimization of an entire video transmission system.**

***Index Terms*—Error resilience, intra-update, joint source-channel coding, robust video transmission, tradeoff source-channel coding, video transmission system model.**

I. Introduction

TO TRANSMIT video over noisy channels, one uses both source and channel coding. According to Shannon's *Separation Principle*, these components can be designed independently without loss in performance [1]. However, this important information-theoretic result is based on several assumptions that might break down in practice. In particular, it is based on 1) the assumption of an infinite block length for both source and channel coding, and 2) an exact and complete knowledge of the statistics of the (ergodic) transmission channel. As a result of the first assumption, the Separation Principle cannot be applied without performance loss to applications with real-time constraints. This holds especially for bursty channels which are characteristic for mobile radio transmission or the Internet. As a consequence of the second assumption, it applies only to point-to-point communications. Therefore, *Joint Source-Channel Coding* and *Error Resilient Coding* can be advantageous, in practice, and have become an important research topic. Recent reviews and special issues in the context of video coding include [2]–[5].

Despite increased research activity, joint source-channel coding schemes for video are still in their infancy today. A pragmatic approach for today's state of the art is to keep the source coder and the channel coder separate, but to optimize their parameters jointly. A key problem of this optimization is the bit allocation between source and channel coding that is also discussed in this paper. Interestingly, even this less ambitious problem is not well investigated in the literature. Often, the underlying transmission system is regarded as a "black box," and the video codec has to cope with whatever bit error rate or packet error rate is offered. This approach is indeed justified if video is added as another application on top of a fixed transmission system. However, current and future transmission systems provide increasing flexibility at the interface to the transport level. For example, the enhanced air interface of the GSM system (EDGE [6]) will include a flexible link adaptation where either 1/1, 3/4, 2/3, or 1/2 of the total bit rate can be allocated to the source while the rest is used for channel coding. In fact, the advantage of this flexibility for speech transmission is already exploited in the next generation speech codec of the GSM system, called Adaptive Multi Rate (AMR, [7]). In the future, *software radios* may even allow configuration of the modulation scheme [8]. This trend toward increased flexibility allows inclusion of channel coding (and modulation) into the optimization.

More flexibility, on the other hand, also increases the complexity of the system and makes parameter optimization more difficult. The overall performance depends on many interrelated issues, such as the distortion-rate performance and error resilience of the source codec, the error correction capability of the channel codec, and the characteristic of the channel. Because of this interaction of system components, the influence of individual parameters is difficult to understand, and the design of the overall system might become a formidable task. Often, simulations are used to study overall system performance (e.g., [9]). However, measurements can rarely be generalized, and provide only limited insight in the underlying problem. Furthermore, simulations can become very complex for a large parameter space. It is therefore desirable to develop appropriate models to study and understand the interaction and tradeoffs between system parameters.

The scope of this paper is to provide such a model for a complete video transmission system. We use this model to analyze the overall performance as a function of the most important system parameters. In particular, the optimum bit allocation between source and channel coding is found analytically while also considering the optimal tradeoff between INTER and INTRA coding. Similar investigations have been performed for vector quantization [10] and Lempel–Ziv compression [11]. However, no analysis has been presented for motion-compensated video coding that forms the basis of all common video coding standards, including H.261, H.263(+),

Manuscript received May 5, 1999; revised November 11, 1999. This work was supported in part by the German DFN-Verein.

K. Stuhlmüller and N. Färber are with the Telecommunications Laboratory, University of Erlangen-Nuremberg, Cauerstrasse 7/NT, 91058 Erlangen, Germany (e-mail: stuhl@LNT.de; faerber@LNT.de).

M. Link is with Lucent Technologies, Nuremberg, Germany (e-mail: mlink@lucent.com).

B. Girod is with the Information Systems Laboratory, Stanford University, Stanford, CA USA (e-mail: girod@ee.stanford.edu).

Publisher Item Identifier S 0733-8716(00)04338-9.

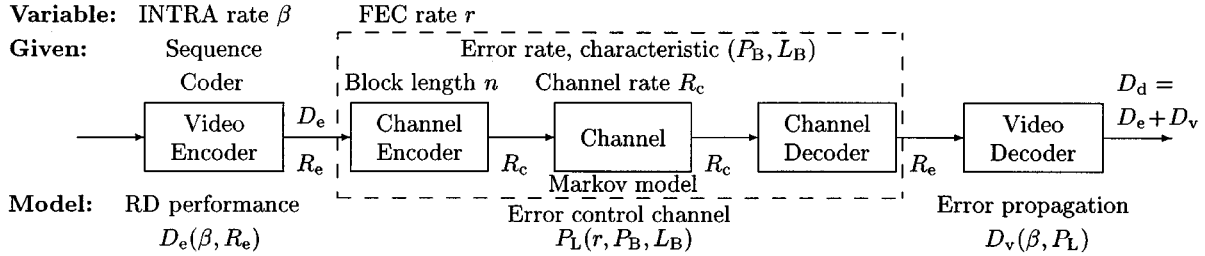

Fig. 1. Video transmission scheme. The video encoder is described by its distortion-rate function $D_e(\beta, R_e)$ depending on the INTRA rate β. The influence of the transmission using FEC is described by the residual word error rate $P_L(r, P_B, L_B)$ depending on the channel code rate r and the channel characteristics P_B (error probability) and L_B (average burst length). At the video decoder, the effect of error propagation is given by $D_v(\beta, P_L)$. The overall decoded video quality is denoted D_d.

MPEG-1, MPEG-2, and MPEG-4 [12]–[17]. In previous work, we addressed the related problem of optimal transmission of a given scalable video bit stream over a packet network by optimizing the unequal error protection for the layers of the video stream [18], [19]. In this paper we consider only a single layer codec but include the effects of transmission errors and INTRA coding as well as the distortion-rate behavior of the video encoder.

This paper is organized as follows. We first outline the transmission system in Section II. In Section III we model a hybrid motion compensated video codec. The distortion-rate performance of the video encoder is analyzed in Section III-A, while a theoretical framework for interframe error propagation is presented in Section III-B. The influence of channel coding and channel parameters are discussed in Section IV. Then, we combine the models to describe the overall system performance, and show in Section V that our model can approximate the decoded picture quality very accurately. The impact of INTRA coding and FEC is studied in Sections V-A and B, respectively. Finally, joint optimization of source and channel coding parameters is investigated in Section V-C.

TABLE I
SUMMARY OF MODEL PARAMETERS

Free Parameters		
β	INTRA rate (as percentage of macroblocks)	
r	code rate of channel code	
n	channel coding block length	influences delay, $n = 88; 222$
Sequence Depended Parameters		**also depended on**
R_0	rate offset	encoder, β
D_0	distortion offset	encoder, β
θ	RD factor	encoder, β
σ_{w0}^2	error energy per lost packet	packetizing, concealment
γ	leakage	encoder, packetizing
Channel Parameters		**default settings**
R_c	channel rate	$R_c = 80; 200$kbps
P_B	symbol error rate	
L_B	average burst length	$L_B = 8$
Dependent Parameters		**dependent on**
t_c	error correction capability of channel code	n, k
k	number of information symbols per block	r, n
R_e	video source rate	R_c, r
D_e	video source coding distortion	$\beta, R_e, R_0, D_0, \theta$
α	power transfer factor	γ, β
σ_w^2	error energy	σ_{w0}^2, P_L
D_v	distortion at decoder due to error propagation	$\sigma_w^2, \gamma, \beta$
P_D	block error density function	n, P_B, L_B
P_L	residual word error rate	P_D, n, r
D_d	overall distortion at decoder	D_e, D_v

II. VIDEO TRANSMISSION SYSTEM

A. Overview

In this section we provide an overview of the video transmission system under consideration, and introduce the most important model parameters. As can be seen from Fig. 1, the system consists of three parts: the video encoder, the video decoder, and the *error control channel*, which is defined as the combination of the channel codec and the channel [20]. These components are described briefly in the following paragraphs and are discussed in more detail in Sections III and IV. All model parameters are summarized in Table I for quick reference.

We assume that a space-time discrete video signal is used as input to the video encoder which is characterized by its operational distortion-rate (DR) function $D_e(\beta, R_e)$; i.e., the average distortion D_e is expressed as a function of the average bit rate R_e and INTRA rate β. The common DR relationship is extended by the INTRA rate because of its significant influence on error resilience. In fact, it is used as the first important parameter for system optimization in this paper.

After source coding, the compressed video bitstream is prepared for transmission by the channel codec. Often, this involves packetization and some form of *error control*. In this paper we focus on *forward error correction* (FEC) that can be combined with *interleaving* to reduce the effect of burst errors. More specifically, we assume an (n, k) Reed–Solomon (RS) block code with a block size of n symbols including $k < n$ information symbols. The second important parameter that is used for system optimization is the code rate $r = k/n$. By reducing the code rate, more channel coding redundancy is added to each codeword which improves the error correction capability of the code while reducing the throughput at the same time.

After channel encoding, the RS codewords are transmitted over the channel. We use a two-state Markov model to describe errors on the symbol level. As intuitive channel parameters, we use the average symbol error rate P_B and the average burst length L_B. Together with the total bit rate R_c, these two parameters completely describe the channel and can be used to, e.g., study the influence of burst errors versus independent symbol errors. Furthermore, the selected channel model allows calculation of the residual word error rate $P_L(r, P_B, L_B)$ after channel decoding from the parameters of the Markov model and the code rate. Thus, the overall performance of the error control channel,

including a burst channel and an RS channel codec, can be described analytically.

Finally, the influence of residual errors on the decoded video quality has to be considered. Depending on the error resilience capabilities of the video decoder, a single lost codeword may cause severe image distortion. Fast resynchronization of the bitstream and error concealment are two important issues that can help to mitigate the effect of residual errors. Another important issue is interframe error propagation because errors may be visible over many consecutive frames. Therefore, a model for interframe error propagation is derived in this paper that describes the additional distortion at the decoder $D_v(\beta, P_L)$ as a function of the INTRA rate β and the residual word error rate P_L.

After this brief description of each system component, it is interesting to discuss the interactions and tradeoffs that influence the overall distortion $D_d = D_e + D_v$. First consider a variation of the code rate r (see Section V-A). Note that for a given channel bit rate R_c, the code rate controls the bit allocation between source and channel coding. This has two effects on the picture quality of the video signal at the decoder output. First, a reduction of r reduces the bit rate available to the video encoder and thus increases the distortion at the encoder regardless of transmission errors. The actual D_e increase is determined by the operational DR function $D_e(\beta, R_e)$ of the video encoder. On the other hand, the residual word error rate is reduced when reducing r, determined by the properties of the error control channel according to $P_L(r, P_B, L_B)$. Finally, a reduction in P_L leads to a reduction in $D_v(\beta, P_L)$ depending on several implementation issues as discussed above. Considering the total distortion D_d at the video decoder output, these interactions of the various components make it difficult to select the optimum code rate. Basically, the characteristic of each component may have significant influence.

Now consider a variation of the INTRA rate β which is used as the second important optimization parameter in this paper (see Section V-B). Since INTRA coded macroblocks do not depend on the previous frame, error propagation can be reduced by increasing the number of INTRA coded macroblocks, thus reducing D_v. However, INTRA coding also reduces the coding efficiency compared to motion compensated prediction. Hence, the distortion at the encoder D_e is increased for a fixed bit rate R_e. Whether or not an increase in β is advantageous for the overall distortion $D_d = D_e + D_v$ depends on the actual amount of increase/decrease in each component. This illustrates that each component needs to be modeled accurately before system optimization can be attempted. This is particularly true for a joint optimization of r and β (see Section V-C).

B. Simulation Environment

The simulation environment we use in this paper to verify the derived model is described as follows. As source signals, we use the QCIF test sequences *Mother&Daughter* and *Foreman* which are encoded at 12.5 fps using 150 and 125 frames, respectively. The sequences are selected because of their different characteristic in motion and spatial detail. Although the model can also be applied to other test sequences (see [21]), we do not provide additional results because the selected sequences are sufficient to discuss the effect of different source statistics.

For source coding, we use an H.263 compliant video encoder. No H.263 options are used, however, each Group Of Blocks (GOB) is encoded with a header to improve resynchronization. The encoder operates at a constant bit rate R_e which is enforced by a simple rate control that is described as follows. Each frame is encoded with a fixed quantizer step size, which is adapted frame by frame to obtain a given target bit budget. The adaptation of the quantizer step size is performed as follows. First, the mode decision is performed according to TMN5 [22] for the whole frame, and then the resulting prediction error is transformed and quantized with different quantizer step sizes. Finally, the value that minimizes the difference between the accumulated number of transmitted bits and target bits is selected. This rate control reduces buffer variations to an acceptable amount, and hence allows the transmission over a constant bit rate channel with limited delay. In practice, other rate control algorithms should be used that can further reduce buffer variations at improved performance. However, since rate control is not the focus of the paper, the above approach is sufficient.

Another issue that is related to the coding control of the video encoder is the INTRA update scheme employed. Several schemes have been proposed in the literature that either consider the activity of image regions [23], [24], vary the shape of INTRA update patterns [9], or include the INTRA mode decision in a rate-distortion optimized encoding framework [25]–[27]. In a very common scheme, which is also recommended in H.263, each macroblock is assigned a counter that is incremented if the macroblock is encoded in interframe mode. If the counter reaches a threshold T (= update interval), the macroblock is encoded in INTRA mode and the counter is reset to zero. By assigning a different initial offset to each macroblock, the updates of individual macroblocks can be spread out in time. In our simulations, we use a very similar update scheme, however, with a variable threshold T instead of the fixed value of $T = 132$ that is recommended in H.263. The only difference is that we also increment the counter for skipped (i.e., UNCODED) macroblocks to guarantee a regular update of all image regions.

The channel parameters are selected as follows. Considering the different complexity of the sequences, we chose a total channel bit rate of $R_c = 80$ kbps for *Mother&Daughter* and $R_c = 200$ kbps for *Foreman*. This allows variation of the INTRA rate and code rate over a wide range without suffering too high distortions or buffer overflows. Unless otherwise noted, the average burst length is set to $L_B = 8$ while the symbol error rate is selected from the set $P_B \in \{10, 7.5, 5.0, 2.5, 1.0, 0.5\}$[%].

The parameters of the RS code are considered next. We use the very common choice of 8 bit per symbol, i.e., one symbol corresponds to one byte. The block size is set to the average GOB size which results in $n = 88$ bytes (80 000/12.5/9 $= 712$ bit) for the *Mother&Daughter* and $n = 222$ bytes (200 000/12.5/9 $= 1778$ bit) for the *Foreman* sequence. Note that this limits the delay introduced by channel coding to one GOB, and therefore also allows for conversational services

with their strict delay constraints. The amount of information symbols k is varied in increments of 8 bytes to achieve different code rates.

Finally, we need to consider the operation of the video decoder in the case of errors. If the RS decoder fails to correct the transmission errors in a block, the video decoder receives an error indication or detects that there has been an error due to bit stream syntax violations. In either case, error concealment is done for any GOB that overlaps with the lost packet. No special packetization is used, i.e., new GOB's are not necessarily aligned with the beginning of a packet. For error concealment, the previous-frame GOB is simply copied to the current frame buffer.

C. Distortion Measure

For the evaluation of the video transmission system, it is necessary to average the distortion over the whole sequence in order to provide a single figure of merit. Even though the time averaged squared error is somewhat questionable as a measure of subjective quality, this approach is still very useful, e.g., to provide an overview for a large set of simulations. Therefore, the video quality is measured as the Mean-Squared-Error (MSE) averaged over all frames of the video sequence throughout this paper. Since PSNR is a measure more common in the video coding community, we use $\text{PSNR} = 10\log_{10}(255^2/\text{MSE})$ to illustrate simulation results. Note that the average PSNR is often computed by *first* computing the PSNR for each frame and averaging in time afterwards. The definition used in this paper allows a better theoretical analysis (see Section III-B) and is more consistent with subjective quality for strong quality variations. In practice, however, there is no significant difference between the two definitions.

Note that we need to distinguish between the picture quality at the encoder and the picture quality at the decoder. Using D_e to describe the overall MSE for a whole sequence after encoding, we obtain

$$\text{PSNR}_e = 10\log_{10}\left(255^2/D_e\right) \tag{1}$$

for the corresponding PSNR value. At the decoder side we need to recall that the result depends on the probabilistic nature of the channel. Hence, the averaged distortion over many channel realizations has to be considered. For the simulation results in this paper, we use 30 random channel realizations for each particular setting of the video transmission system and average the MSE over all frames and realizations. The resulting MSE and PSNR are denoted D_d and

$$\text{PSNR}_d = 10\log_{10}\left(255^2/D_d\right) \tag{2}$$

respectively. In order to ensure that the distortion at the decoder is measured in a steady state, only the last 50 encoded frames are used to calculate D_e and D_d.

As mentioned above, the overall MSE D_d is actually a superposition of two distortion types. The distortion caused by signal compression D_e and the distortion D_v which is caused by residual errors and interframe error propagation. Assuming that D_e and D_v are uncorrelated, we can calculate the overall MSE as

$$D_d = D_e + D_v. \tag{3}$$

Our experiments indicate that this assumption is valid. Even though transmission errors may be clustered around active regions, and thus their magnitude may be correlated with the coding errors, usually their sign is not correlated to the sign of the coding errors.

However, it should be noted that (3) combines two distortion types that are likely to be perceived differently. The distortion D_e is caused by signal compression and consists of blocking artifacts, mosquito noise, ringing, blurring, etc. The distortion introduced by transmission errors D_v consists of severe destruction of image content and may be large and infrequent. Subjective tests are needed to determine how D_e and D_v shall be combined to give the best possible approximation of subjective quality.

If subjective tests show that, e.g., the distortion D_v caused by transmission errors is more annoying than the distortion caused by the video encoder D_e, (3) and (2) can be changed to a weighted sum or some other function of D_e and D_v. The determination of such a *subjective quality function* is beyond the scope of this paper and is left to future research.

III. Analysis of the Video Codec

In this section, we analyze the performance of the video encoder and decoder. Although we use the ITU-T H.263 [13] video compression standard throughout this paper, the model derived can be used for other codecs that are based on hybrid motion compensation.

In the following section we model the distortion-rate performance of the video encoder. Then we introduce an analytical model for the error propagation at the video decoder which can explain the cumulative effect of transmission errors. We focus on the main results and refer to the Appendixes for most derivations.

A. Video Encoder

In this section we model the Distortion-Rate (DR) performance of a hybrid motion compensated video encoder. The proposed model is an empirical model that is not derived analytically. Instead, we focus on the *input–output behavior* of the video encoder and emphasize simplicity and usability over a complete theoretical description. On the one hand, this approach is taken because we want to describe a complete transmission system, which requires the complexity of individual components to be kept at a reasonable level. On the other hand, we found that theoretically founded models often cannot describe experimental results very accurately due to simplistic assumptions. For example, such a theoretically founded model for the performance of motion compensated prediction is described in [28] and [29], where the DR performance is analyzed by deriving the power spectral density of the prediction error with respect to the probability density function of the displacement error. Although this model provides very interesting insights, it cannot describe the measured

DR performance of an H.263 encoder with sufficient accuracy. Similar problems can be observed for the description of the DR performance in transform coding [30] and DCT coding in particular. Although several empirical distortion-rate models have been published (e.g., [31]–[34], they are usually used for rate control and cannot be used to model the distortion of an entire video encoder for a given rate.

To avoid these limitations without an increase in model complexity, we use a simple equation that relates the distortion at the encoder D_e to the relevant parameters. In the simulation scenario that we consider, there are two parameters with a significant impact on D_e, namely the source rate R_e that is allocated to the video encoder, and second, the percentage of INTRA coded macroblocks (INTRA rate) β that is enforced by the coding control to improve error robustness. The general idea to use empirical models to describe DR performance has also been used for rate control as, for example, in [32], however, our focus is on the description of the overall performance, i.e., the average distortion for a whole sequence given R_e and β.

One drawback of this approach is that the necessary model parameters cannot be derived from commonly used signal statistics, like variance, correlation, or the power spectral density. Instead, the parameters need to be estimated by fitting the model to a subset of measured data points from the DR curve. Since the proposed model uses only six parameters (see below), the necessary subset is relatively small and can be obtained with reasonable complexity. However, the obtained parameters are specific for a given video sequence and video codec. Furthermore, the interpretation of these parameters is not always obvious. This makes it difficult to, e.g., extend results from a sequence with "complex motion" to a sequence with "moderate motion." However, we found that the model can describe the DR performance of a wide range of test sequences with very good accuracy, once the parameters are selected correctly. Furthermore, the simplicity of the model significantly increases its usability and thus, in practice, outweighs the described drawbacks. Nevertheless, it should be noted that a model of similar simplicity that is founded on theoretical analysis would be highly desirable.

We use the DR model

$$D_e = \frac{\theta}{R_e - R_0} + D_0 \tag{4}$$

where D_e is the distortion of the encoded sequence, measured as the MSE, and R_e is the output rate of the video encoder. The remaining variables (θ, R_0, and D_0) are the parameters of the DR model which depend on the encoded sequence as well as on the percentage of INTRA coded macroblocks β. We have found that the relationship with β is approximately linear, i.e.,

$$\begin{aligned} \theta &= \theta_P + \Delta\theta_{IP}\beta \\ R_0 &= R_{0P} + \Delta R_{0IP}\beta \\ D_0 &= D_{0P} + \Delta D_{0IP}\beta \end{aligned} \tag{5}$$

such that the total number of model parameters is six. According to (5), it is sufficient to measure the DR curves for only two different INTRA rates. Intermediate values can then be obtained by linear interpolation. This is also the approach used in the following to obtain the model parameters.

Fig. 2 shows that (4) and (5) approximate the DR performance of the video encoder very accurately. Although the experimental results are obtained with an H.263 encoder, the DR curves for other hybrid motion compensated video encoders, e.g., H.261 [12], MPEG-1 [15], or MPEG-2 [16], exhibit very similar behavior.

The model (4) was fitted to the measured points for $\beta = 1\%$ and for $\beta = 33\%$ ($\beta = 55\%$ for *Foreman*) INTRA coded macroblocks. The fitting was done by minimizing the sum of squared MSE differences between the model and the measured points. This resulted in two sets of parameters $\{\theta, R_0, D_0\}$ for each sequence. These two parameter sets together consist of six values, thus allowing us to determine $\theta_P, \Delta\theta_{IP}, R_{0P}, \Delta R_{0IP}, D_{0P}$, and ΔD_{0IP} from (5).

The model parameters $\theta_P, \Delta\theta_{IP}, R_{0P}, \Delta R_{0IP}, D_{0P}$, and ΔD_{0IP} are used to interpolate the DR curves for other INTRA rates β. The intermediate curves in Fig. 2 for 3%, 6%, 11%, and 22% (and 33%, 44% for *Foreman*) were generated by using (4) and interpolating the parameters according to (5). The maximal PSNR deviation between the model fitted that way and the measured DR points is 0.22 dB for the *Mother&Daughter* sequence and 0.3 dB for the *Foreman* sequence (Fig. 2).

Note that the parameters $\theta_P, \Delta\theta_{IP}, R_{0P}, \Delta R_{0IP}, D_{0P}$, and ΔD_{0IP} characterize the coding of the input video sequence with the given hybrid motion compensated encoder, in this example *Mother&Daughter* or *Foreman* coded with H.263 in baseline mode. The parameters depend very much on the spatial detail and the amount of motion in the sequence; e.g., for a sequence with high motion and little spatial detail $\Delta\theta_{IP}$ is low, whereas for a sequence with moderate motion and high spatial detail $\Delta\theta_{IP}$ is high.

B. Video Decoder

While motion compensated prediction yields significant gains in coding efficiency, it also introduces interframe error propagation in the case of transmission errors. Since these errors decay slowly, they are very annoying. To optimize the overall performance of video transmission systems in noisy environments, it is therefore important to consider the effect of error propagation. While several heuristic approaches have been investigated in the literature to reduce the influence of error propagation (e.g., [23], [24], and [35]), up until now no theoretical framework has been proposed to model the influence of transmission errors on the decoded picture quality. The model proposed in the following includes the effects of INTRA coding and spatial loop filtering and corresponds to simulation results very accurately.

Note that two different types of errors contribute to the overall distortion at the decoder. First, the errors that are caused by signal compression at the encoder D_e and, second, errors that are caused by residual errors which cannot be corrected by the channel decoder. Since the first type of error is sufficiently described by (4), we now focus on the second type of error and use the variable D_v to refer to it.

A simplified block diagram of a hybrid motion compensated video codec is illustrated in Fig. 3, together with the relevant parameters that are introduced in the following. We describe errors

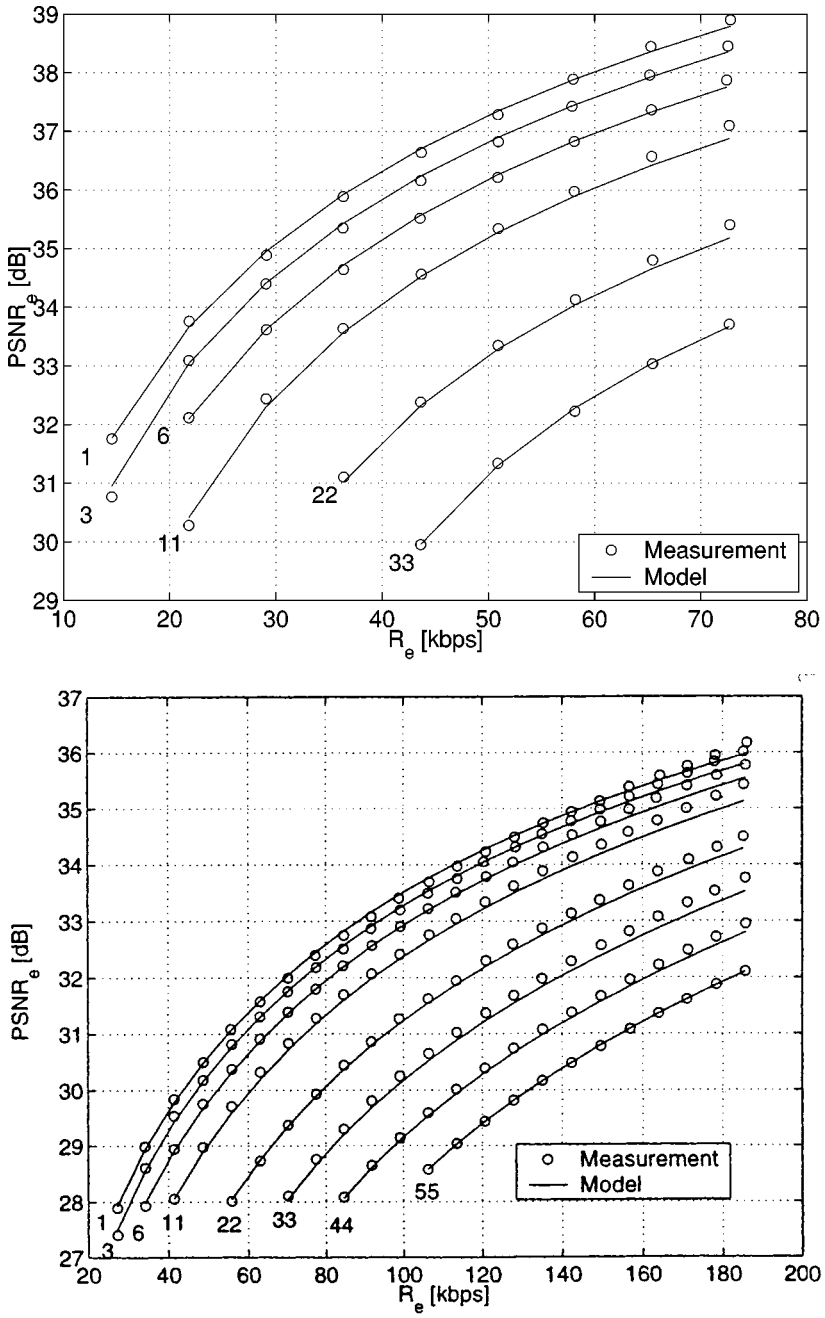

Fig. 2. Distortion-rate curves at the encoder for the test sequences *Mother&Daughter* (top) and *Foreman* (bottom) for 0%, 1%, 3%, 11%, 22%, and 33% (44%, 55%) of INTRA coded macroblocks β.

that are introduced by residual transmission errors using a stationary random process U which generates the zero-mean error signal $u[x, y]$. In other words, we assume that, on average, the same error variance σ_u^2 is introduced in each frame. Obviously, the parameter σ_u^2 is directly related to the residual error rate P_L, since an increased number of lost packets will also increase the variance of introduced errors. However, it also depends on several implementation issues, like packetization, resynchronization, and error concealment, as well as on the encoded video sequence. For a given sequence, fixed packet size, and given decoder implementation, it can be shown that the error variance that is introduced can be expressed as

$$\sigma_u^2 = \sigma_{u_0}^2 P_L. \tag{6}$$

This linear relation is only valid for low residual error rates, i.e., $P_L < 0.1$. Since reasonable picture quality is very difficult to obtain for higher error rates, even when advanced error resilience techniques are employed, the given linear relation is sufficient for relevant operation conditions. Note that $\sigma_{u_0}^2$ can be treated as a constant value that does not depend on other model parameters. It describes the sensitivity of the video decoder to an increase in error rate. If the decoder can cope well with residual errors, the value is low. For example, $\sigma_{u_0}^2$ can be reduced by an advanced error concealment technique.

Errors that are introduced at a given point in time propagate due to the recursive structure of the decoder. This temporal error propagation is typical for hybrid video coding that relies on mo-

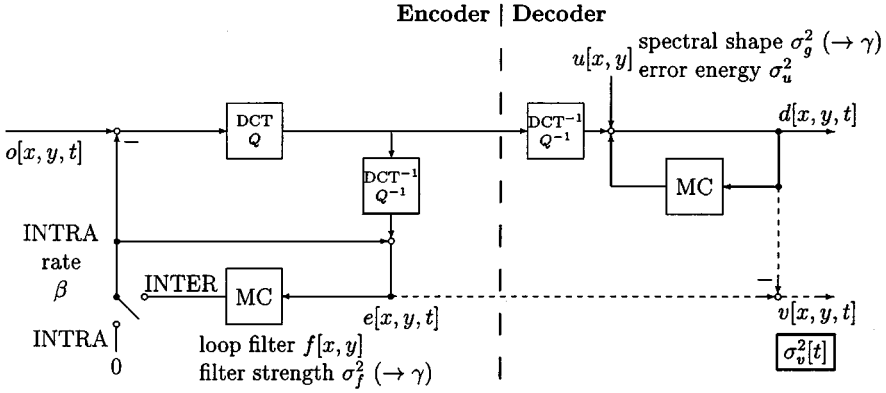

Fig. 3. Block diagram of a hybrid motion compensated video codec with transmission errors.

tion compensated prediction in the interframe mode. It is very important to consider this effect for the design of the overall system since it has a significant influence on the sensitivity of the video decoder to residual errors. For example, even small values of σ_u^2 may result in unacceptable picture quality if errors are accumulated in the decoder loop without being attenuated in some way.

Referring to Fig. 3, we are therefore interested in the accumulated error signal $v[x,y,t]$ which is the difference between the reconstructed frames at encoder and decoder. In Appendix I, we show that the energy of this error signal decays over time due to (explicit and/or implicit) spatial filtering in the prediction loop and due to INTRA coding of macroblocks. More precisely, we derive that if the error signal $u[x,y]$ is introduced at $t=0$, such that $v[x,y,0] = u[x,y]$, the variance of the propagated error signal is given by

$$\sigma_v^2[t] = \sigma_u^2 \frac{1-\beta t}{1+\gamma t} \tag{7}$$

for $0 \le t < T$, where T is the *INTRA update interval*. For $t \ge T$ we assume that the introduced error energy is removed completely by INTRA coded macroblocks, and hence obtain $\sigma_v^2[t] = 0$. In other words, we assume that the implemented INTRA update scheme encodes each macroblock once in INTRA mode within an interval of T encoded frames. In a practical system, some error energy might remain for $t \ge T$ due to migration by motion compensation. However, simulations show that this effect is negligible when the average error is considered. The relationship between T and the INTRA rate β is given by

$$\beta = \frac{1}{T}. \tag{8}$$

The *leakage* γ describes the efficiency of explicit and/or implicit (e.g., due to sub-pel motion compensation, overlapped block motion compensation) loop filtering to remove the introduced error (see Appendix I). Its value depends on the strength of the loop filtering as well as on the shape of the power spectral density of the introduced error $u[x,y]$. If no spatial filtering is applied in the predictor (i.e., full-pel motion compensation, etc.), $\gamma = 0$ and the decay in error energy is only influenced by INTRA coding. The value of γ usually increases when more spatial filtering is applied in the predictor or when the introduced error includes high spatial frequencies that can easily be removed by the loop filter. The range of typical values is given by $0 < \gamma < 1$.

In the following, we are interested in the time averaged distortion D_v that is introduced by transmission errors. Since each individual error propagates over at most T successive frames and the decoder is linear, we can derive the average distortion D_v as the superposition of T error signals that are shifted in time. If we further assume that the superimposed error signals are uncorrelated from frame to frame, we can calculate D_v directly from (7), yielding

$$D_v = \sigma_{u_0}^2 P_L \sum_{t=0}^{T-1} \frac{1-\beta t}{1+\gamma t}. \tag{9}$$

In practice, the above assumptions are less restrictive than they may seem. For example, the assumption of uncorrelatedness is automatically met when individual error signals are spatially and/or temporally separated in the decoded video sequence. This is very common for low residual error rates but may become a problem otherwise. Therefore, we expect that the accuracy of the model will decrease at high error rates. The assumption of stationarity, on the other hand, means that the effect of lost packets is approximately constant for each transmitted packet. However, we found that (9) can also be used for typical variations of errors introduced by packet loss. Only for extreme variations, e.g., for an almost static video scene with a short sequence of heavy motion, can the dominant effect of a few packets cause problems for the proposed model. Scalability is not considered in the model at the moment but can be incorporated by modeling each layer separately using (9). This extension is currently under investigation, but is beyond the scope of this paper.

Parameter Estimation and Experimental Comparison: The calculation of D_v according to (9) requires the knowledge of several model parameters. In the following we will discuss how to obtain these parameters for a given simulation scenario and compare the model calculation with experimental data. The INTRA rate β can be regarded as a control parameter that is enforced by the coding control of the video encoder, and hence is known *a priori*. Note that the effective average number of INTRA coded macroblocks per frame might be a little bit higher since the mode decision prefers the INTRA mode sometimes even when it is not enforced. However, we have observed that in the simulations presented in this paper, there is no significant difference between the enforced and the measured INTRA rate. The parameter P_L is the residual

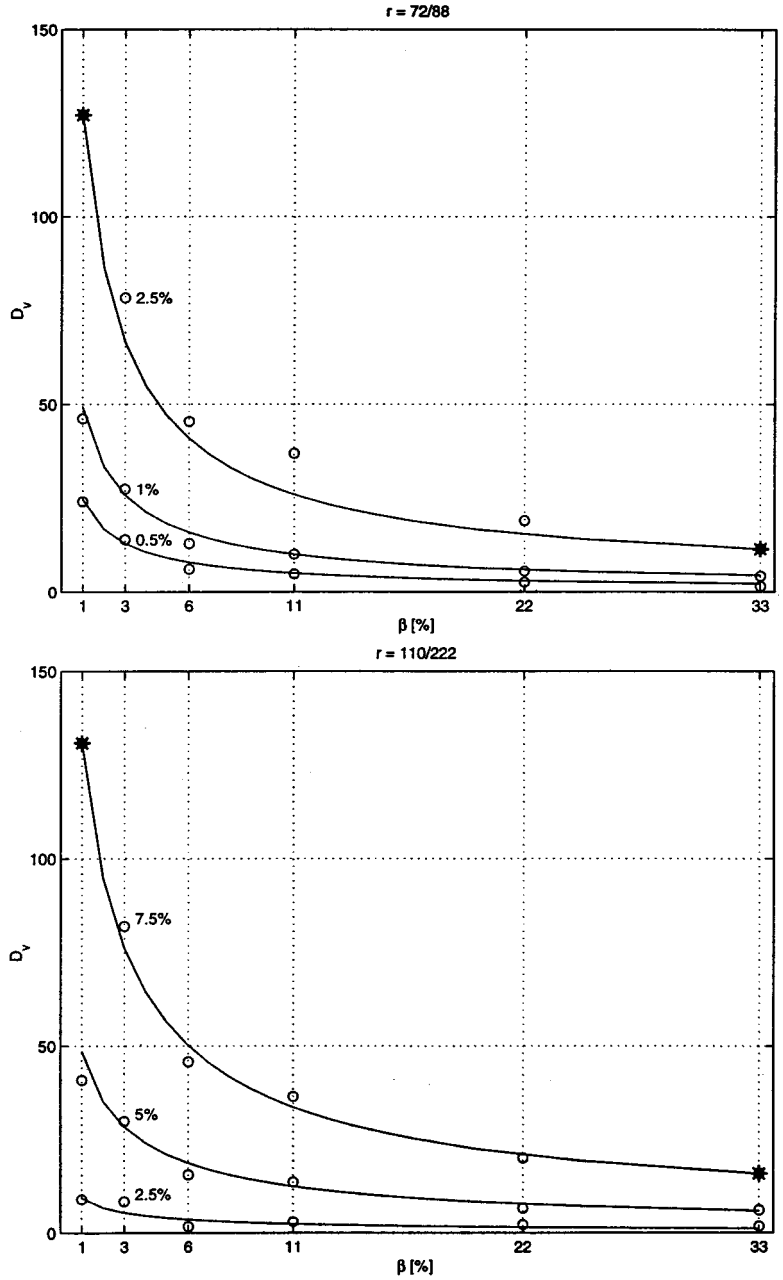

Fig. 4. Distortion caused by transmission errors (D_v) over the INTRA rate (β). The measurements (o) are compared to the model calculation (—) for three different symbol error rates P_B with fixed code rate r. The two parameters of the model (γ and $\sigma^2_{u_0}$) are obtained by matching the two indicated measurement points ($*$). The test sequences are *Mother&Daughter* (top) and *Foreman* (bottom).

word error rate and depends on the channel characteristic as well as on the channel codec used. In Section IV we show how P_L can be derived for a particular channel codec and channel model.

The remaining two parameters, γ and $\sigma^2_{u_0}$, have to be estimated for a given video codec, packetization, and video sequence. Although it is possible to derive their values theoretically under simplifying assumptions, we use a subset of measurement points to match the model to experimental data. This approach is very similar to the approach taken in the previous section to fit the empirical DR model. In theory, any two measurement points are sufficient to match γ and $\sigma^2_{u_0}$. More formally, assume that two distortion values D_v are given for known values of β and P_L, i.e., $D_v^{(1)}(\beta^{(1)}, {P_L}^{(1)})$, and $D_v^{(2)}(\beta^{(2)}, {P_L}^{(2)})$. Then we can obtain the corresponding values for γ and $\sigma^2_{u_0}$ from (9) using numerical minimization. Although fitting to more measurement points can increase the robustness, we found that two points are sufficient if selected carefully. Basically, the range of interesting INTRA rates and distortion values should be covered. Hence, we use $\beta^{(1)} = 0.01$ and $\beta^{(2)} = 0.33$ and select the parameters of the error control channel (code rate, symbol error rate) such that the range $140 < D_v < 10$ is covered (starting from an initial quality of 30 dB this range corresponds to a loss in PSNR of approximately 5–0.5 dB).

The experimental data presented in Fig. 4 show that (9) approximates the cumulative effect of transmission errors very accurately. The measured and calculated value for D_v is plotted as

a function of β, whereas γ and $\sigma_{u_0}^2$ are kept constant. The two data points which are used to calibrate the model are indicated by an asterisk. The environment that is described in Section II is used for the simulation.

IV. Error Control Channel

The reliability of a transmission can be improved by Forward Error Correction (FEC), thus lowering the residual word error rate P_L and the additional distortion D_v as described by (9). However, in order to maintain a constant channel data rate R_c, the available data rate for the source encoder must be reduced to

$$R_e = rR_c \tag{10}$$

where $r \in [0,1]$ is the channel code rate. This implies that the distortion D_e (4) introduced by the source encoder increases. Hence, a tradeoff between source coding distortion D_e and channel induced distortion D_v results. For the optimization of the total distortion $D_d = D_e + D_v$, it is therefore important to understand how much reliability can actually be gained by a certain reduction in code rate.

In this paper we use Reed–Solomon (RS) codes. For symbols composed of m bits, the encoder for an (n,k) RS code groups the incoming data stream into blocks of k information symbols (km bits) and appends $n-k$ parity symbols to each block. Hence, the code rate is

$$r = k/n. \tag{11}$$

For RS codes operating on m-bit symbols, the maximum block length is $n_{\max} = 2^m - 1$. By using *shortened* RS codes, any smaller value for n can be selected, which provides a great flexibility in system design. In this paper we chose $m = 8$, and therefore a symbol corresponds to a byte and also (n,k) are measured in bytes. For an (n,k) RS code, any error pattern resulting in less than

$$t_c = \left\lfloor \frac{n-k}{2} \right\rfloor \tag{12}$$

symbol errors can be corrected. Other error patterns containing more than t_c symbol errors may also be corrected with a certain probability. However, the decoder usually reports an uncorrected error instead (*bounded minimum distance decoding* [20]). The probability of undetected errors is very unlikely, especially for large n. If it still happens, it is likely to be detected by the video decoder due to syntax violations (which is a result of left redundancy in the source coding, e.g., sync words). In the following we assume that errors can always be detected.

Note that the block length n determines the delay introduced by the FEC scheme, because a buffer at the receiver is necessary which can hold n symbols. On the other hand, the error correction capability of the code is usually improved by increasing the block length. In fact, one result of Shannon's classical information theory is that reliable communication can always be achieved for $n \to \infty$ as long as the code rate is selected to be less than or equal to the channel capacity. However, in practical systems, the block length is limited due to delay constraints. Moreover, the complexity of the channel code does increase with the block length n and thus becomes prohibitive for practical systems at some point. Hence, an error free transmission cannot be guaranteed, i.e., there are always residual errors.

The *residual word error rate* P_L is the probability that a block cannot be corrected. Based on (12), it can be calculated as

$$P_L = \sum_{\kappa=t_c+1}^{n} P_D(n,\kappa) \tag{13}$$

where P_D is the *block error density function*. $P_D(n,\kappa)$ denotes the probability of κ symbol errors within a block of n successively transmitted symbols; e.g., for the Binary Symmetric Channel (BSC) with symbol error probability P_B, P_D is given by the binomial distribution

$$P_{D,\mathrm{BSC}}(n,\kappa) = \binom{n}{\kappa} P_B^{\kappa}(1-P_B)^{n-\kappa}.$$

For channels with memory, it is more complicated to calculate P_D. We use a simple and analytically tractable 2-state Markov model with only two parameters to describe the errors on the symbol level. The two states of the model are denoted G (good) and B (bad). In state G symbols are received correctly, whereas in state B symbols are erroneous. The model is fully described by the transition probabilities p_{GB} between states G and B and p_{GB} between states B and G. Since these parameters are not intuitive, we prefer to use the error probability

$$P_B = \Pr(\mathrm{B}) = \frac{p_{\mathrm{GB}}}{p_{\mathrm{GB}} + p_{\mathrm{BG}}}$$

and the average burst length

$$L_B = \frac{1}{p_{\mathrm{BG}}}$$

which is the average number of consecutive symbol errors. The derivation of P_D for this model can be found in [36] or [19]. For completeness it is repeated in Appendix II. An overview of burst error models is given in [37]. Finally, note that the model includes the memoryless BSC as a special case by setting $L_B = 1/(1-P_B)$.

The behavior of the investigated error control channel is illustrated in Fig. 5 for a block size of $n = 88$ byte. Note that a variation of the code rate can reduce the residual word error rate P_L by several orders of magnitude. This is particularly true for small average burst lengths where high reliability can be achieved while maintaining a reasonable throughput. For example, assume that one GOB corresponds to one FEC block, and that $R_e = 32$ kbps are used for video coding given a total bit rate of $R_c = 80$ kbps $(r = 0.4)$. Then, for a channel characterized by $P_B = 0.1$ and $L_B = 3/2$, only one out of 10 000 blocks will have to be discarded, i.e., less than one GOB within 1000 frames. In this situation, advanced error resilience techniques in the video decoder are hardly necessary due to the powerful error correction capability of RS codes. However, Fig. 5 also illustrates that the efficiency of FEC is reduced significantly for bursty channels. For an increased average burst length of $L_B = 8$ and $r > 0.5$, the residual word error rate is higher than 10%, i.e., on average at least about one GOB is lost in each frame. The classic approach to combat this problem is to use *interleaving*. In this paper we assume a simple block interleaver, where i blocks

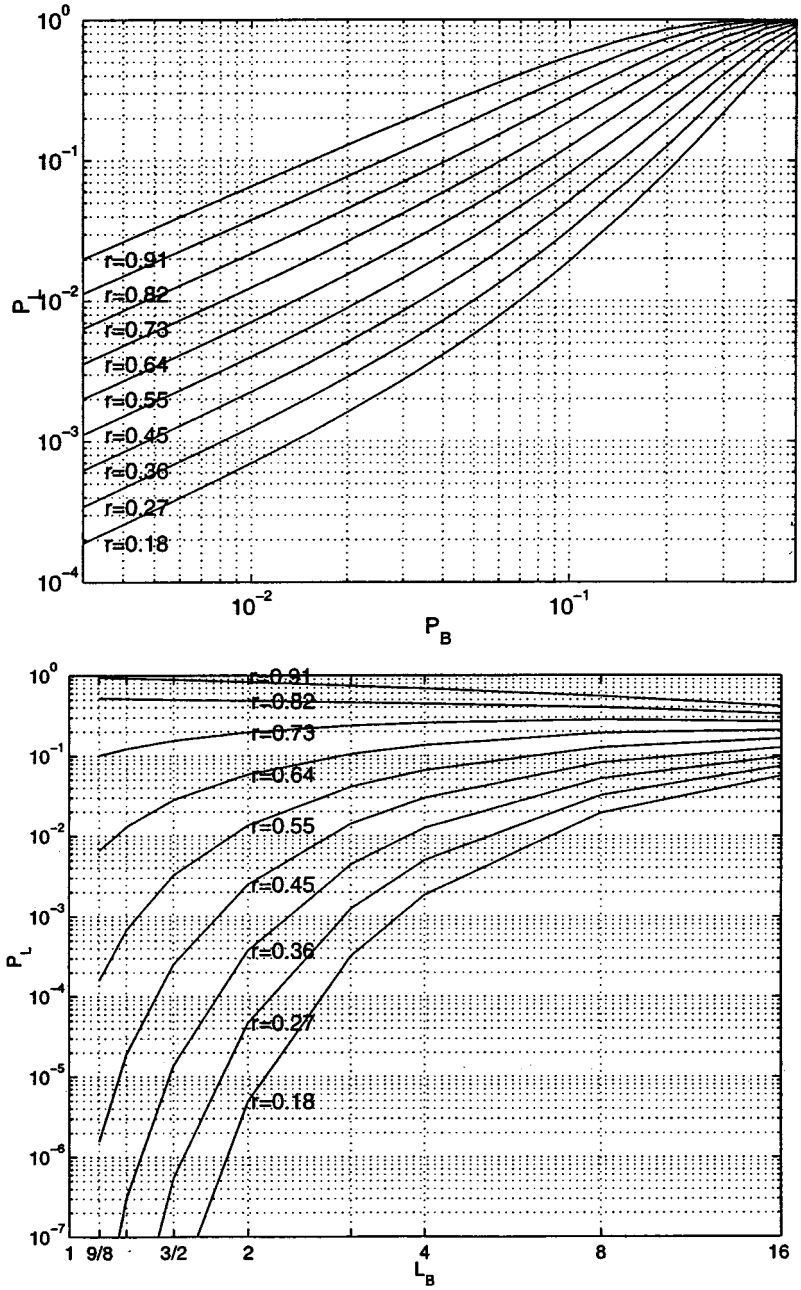

Fig. 5. Residual word error rate P_L for the investigated error control channel. *Top*: Variation of the symbol error rate P_B for a fixed average burst length of $L_B = 8$. *Bottom*: Variation of the average burst length L_B for a fixed symbol error rate of $P_B = 0.1$. The block size of the (n, k) RS code is set to $n = 88$ byte. Each curve corresponds to a fixed code rate $r = k/n$.

are read into a rectangular matrix—each block corresponding to one row. By transmitting the filled matrix column by column, the error burst is spread over i FEC blocks. However, note that interleaving introduces additional delay. Interleaving by a factor of i results in an i times higher delay. Since the FEC delay in our simulations corresponds to one GOB, interleaving by a factor of $i = 9$ would correspond to the delay of one frame interval.

We can easily incorporate interleaving in our model. As shown in Appendix II, the effective channel after interleaving can still be described by the same Markov model but with changed transition probabilities. In other words, the combination of a channel model with parameters P_B and L_B with an interleaver covering i blocks can be described by an equivalent channel model with parameters $P_{B,i}$ and $L_{B,i}$. Since the symbols are just transmitted in a different order, the symbol error rate remains constant, i.e., $P_{B,i} = P_B$. The average burst length, however, is effectively reduced and can asymptotically approach the memoryless case. This is illustrated in Fig. 6, which shows how the burstiness of a channel with average burst length $L_B = 8$ is reduced by interleaving. Note that for $P_B = 0.5$, the channel has no memory if $L_B = 1/(1 - 0.5) = 2$. Therefore, interleaving by a factor of 10 is already very close to the BSC. Considering the results from Fig. 5, it is obvious that interleaving is a very effective tool if the additional delay is acceptable.

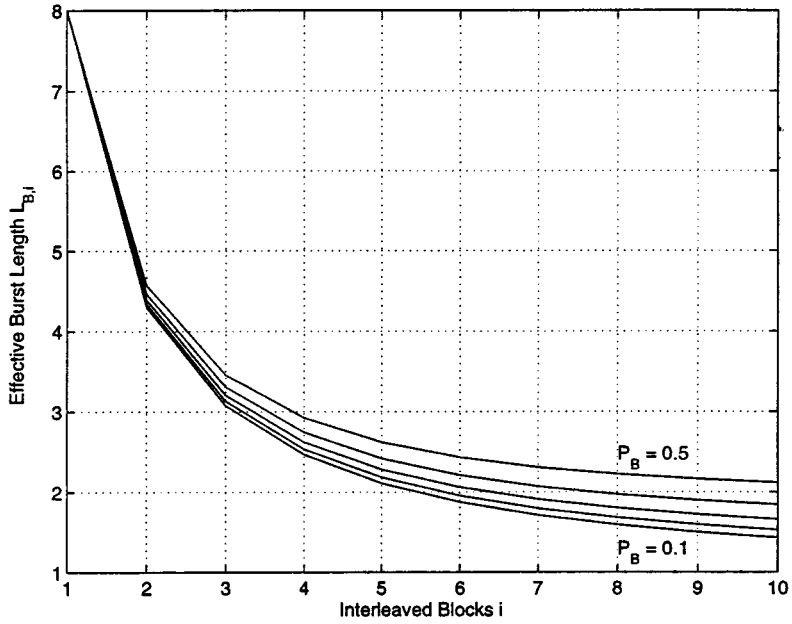

Fig. 6. Effective average burst length $L_{B,i}$ over the number of interleaved blocks i for $P_B = 0.1, 0.2, 0.3, 0.4, 0.5$. The channel burst length without interleaving is $L_B = 8$.

V. Decoded Video Quality

In the following we verify the derived system model by a comparison to simulation results for a practical problem. The simulation results are obtained with an H.263 video codec, and the problem at hand is to minimize the overall MSE D_d for a lossy channel by adjusting the INTRA rate (percentage of INTRA coded macroblocks) β and the FEC code rate r. The simulation environment used for the experiments is described in Section II. We first focus on the influence of each parameter separately and then consider the joined optimization of β and r.

A. Optimal INTRA Rate

In this subsection, the influence of the INTRA rate β on the decoded picture distortion D_d is studied for a fixed channel code rate r. Obviously there is a tradeoff to be considered for the selection of the INTRA rate β. On the one hand, an increased percentage of INTRA coded macroblocks helps to reduce interframe error propagation, and therefore reduces D_v as described by (8) and (9). On the other hand, a high INTRA rate increases the distortion D_e that is caused by compression at a given target bit rate. The influence of β on D_e is given by (4) and (5) in our model.

In Fig. 7 the video quality at the decoder PSNR$_d$ is plotted over the INTRA rate β for four symbol error rates. It can be seen that the model gives a very good approximation of the PSNR$_d$ at the decoder.

First, consider the error-free case (see $P_B = 0\%$). Obviously, increasing β has quite a large influence on the PSNR at the encoder. This is particularly true for the *Mother&Daughter* sequence, because of the static background and little motion. Hence, the additional cost by coding macroblocks in INTRA mode, instead of using motion compensated prediction, is large. For sequences with more complex motion, e.g., *Foreman*, the same increase in INTRA rate has less effect. Therefore, the INTRA mode can be used more generously, and higher optimal INTRA rates β^* result. Note that for $\beta \to 0$, the PSNR falls rapidly since $T \to \infty$ and therefore $D_v \sim \sum_{t=0}^{\infty} 1/(1+\gamma t) \to \infty$ according to (9). Therefore, at least a small amount of INTRA coding should always be used if transmission errors may occur. On the other hand, the exact selection of β^* is less critical, since the optimum is rather flat. As expected, the optimal INTRA rate increases with increasing symbol error rates. However, the optimum value is also sequence specific.

B. Optimal FEC Code Rate

Analogous to the previous subsection, we now study the influence of the channel code rate r on the decoded video quality PSNR$_d$ for a fixed INTRA rate β. Fig. 8 shows that our model approximates the PSNR$_d$ at the video decoder for different channel code rates very well. Only for severe channel induced distortion D_v is the accuracy of the model slightly lower. As explained in Section III-B, this is due to the fact that the introduced errors are not independent any more. In this case, however, the overall quality is usually far from acceptable anyway such that the achieved accuracy there is without practical relevance.

Note that the variation of PSNR$_d$ as a function of r is more severe for the *Foreman* sequence than for the *Mother&Daughter* sequence. This is partly because errors in the *Foreman* sequence are more difficult to conceal and, hence, the sensitivity to errors $(\sigma_{u_0}^2)$ is increased. More importantly, the same reduction in code rate is more effective for the *Foreman* sequence because of the increased block size. Although the average burst length was set to $L_B = 8$ in both simulations, the block size is $n = 88$ and $n = 222$ for the *Mother&Daughter* and *Foreman* sequence, respectively. Because an increased block size has an effect similar to interleaving, FEC is more effective for larger blocks. In this case the selection of the code rate becomes very important.

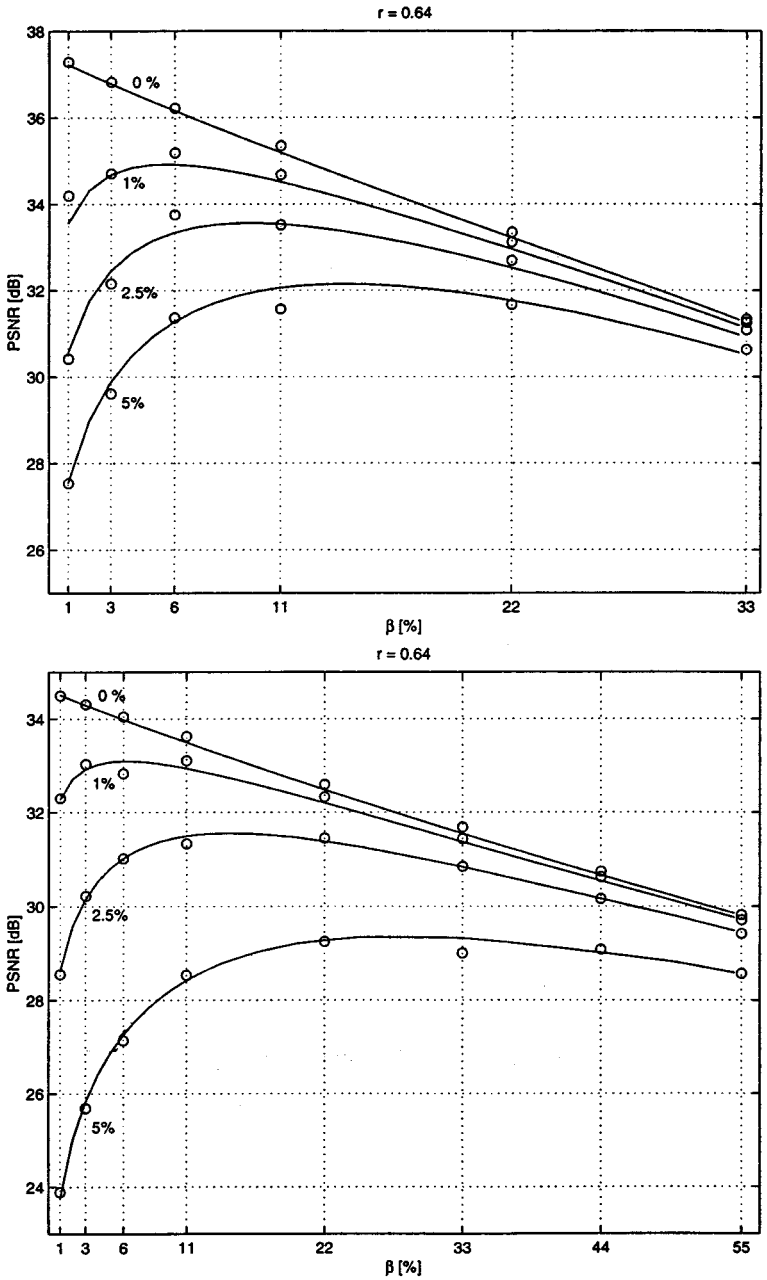

Fig. 7. Measured (o) and modeled (—) PSNR_d at the decoder over the INTRA rate β. The channel is characterized by the average burst length $L_B = 8$ and the symbol error rates $P_B = 0\%, 1\%, 2.5\%, 5\%$. The channel code rate is $r = 0.64$. The test sequences are *Mother&Daughter* (top) and *Foreman* (bottom).

C. Optimal Parameter Selection for the Transmission System

In this subsection we optimize the rate of INTRA coded macroblocks β and the channel code rate r jointly. In Fig. 9 the maximum achievable quality at the decoder PSNR_d^* is depicted over the INTRA rate β. As a reference, the corresponding PSNR_e^* at the encoder is also included. For each β, the channel code rate r is optimized such that PSNR_d is maximized. The circles mark measurements with different r for the given β. Note that PSNR_d^* denotes the upper limit for the given channel, i.e., it is the convex hull of all PSNR_d achievable for the channel with the given INTRA rate β. The gap between PSNR_e^* and PSNR_d^* corresponds to the distortion which is introduced by transmission errors D_v.

Fig. 10 shows the same data as Fig. 9 from a different perspective, i.e., for a given channel code rate r, the INTRA rate β is optimized such that PSNR_d gets maximal. The flatness of PSNR_d^* in Figs. 9 and 10 indicates that INTRA coding and FEC can be exchanged to some extent without losing too much in performance. Note that this involves an exchange between encoding distortion D_e and distortion caused by transmission errors D_v with approximately constant overall distortion D_d at the decoder.

All the following results are obtained by using only the model. Fig. 11 shows the optimal INTRA rate β^* and the optimal channel code rate r^* for a transmission over burst channels with different average burst lengths $L_B = 2, 4, 8, 16, 32$ and symbol error rates in the range $P_B = 0\%, \cdots, 25\%$. The

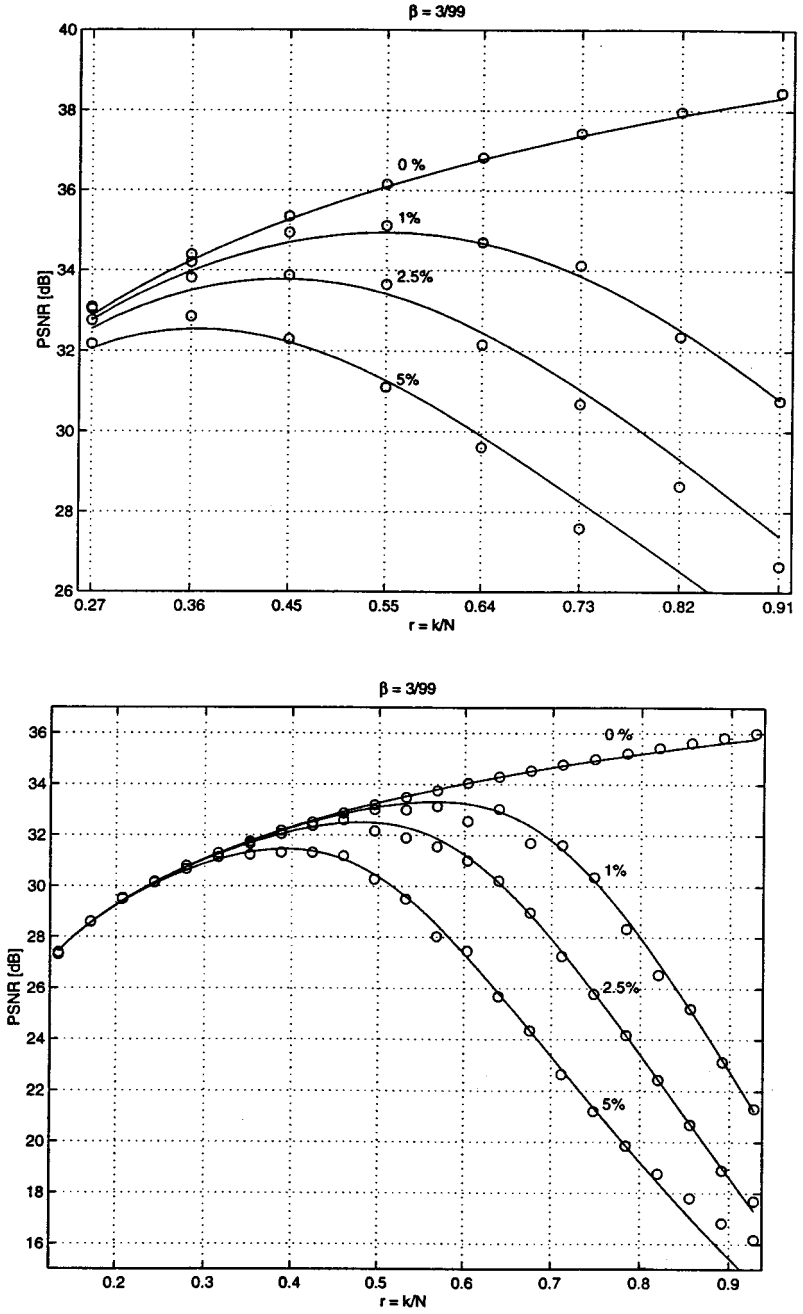

Fig. 8. Measured (o) and modeled (—) PSNR$_d$ at the decoder over the channel code rate r. The channel is characterized by the average burst length $L_B = 8$ and the symbol error rates $P_B = 0\%, 1\%, 2.5\%, 5\%$. The INTRA rate is $\beta = 3\%$. The test sequences are *Mother&Daughter* (top) and *Foreman* (bottom).

shown optimal values were found by numerical minimization of (3). Obviously, the optimal parameters β^* and r^* are very much dependent on both the symbol error rate P_B and the average burst length L_B. For increasing error rates, the optimal INTRA rate β^* increases monotonically and the optimal code rate r^* generally decreases corresponding to a stronger FEC. For large average burst lengths L_B, the optimal INTRA rate β^* goes up to large values since in this case FEC is not very efficient. In contrast, for small L_B or independent errors (BSC), only a small amount of INTRA coding is needed since the FEC is very reliable.

Interestingly the optimal code rate r^* decreases for bursty channels only to a certain point and goes up again beyond that point. This is due to the fact that in this case a high INTRA rate β^* is chosen which results in a large R_0 [(4) and (5)], thus making it very expensive in terms of D_e to further reduce the rate of the video encoder R_e by reducing the code rate r.

In Fig. 12 the optimal INTRA rate β^* and the optimal code rate r^* are plotted over the burst length L_B for various symbol error rates P_B. It can be seen that the optimal INTRA rate β^* increases monotonically with the burst length L_B. In contrast, the optimal code rate r^* first decreases very steeply with the burst length L_B and then increases again after reaching a minimum value. This behavior is caused (ultimately) by the fact that FEC becomes ineffective for large average burst lengths. If the burst length L_B is long enough, the channel behaves like an *on–off*

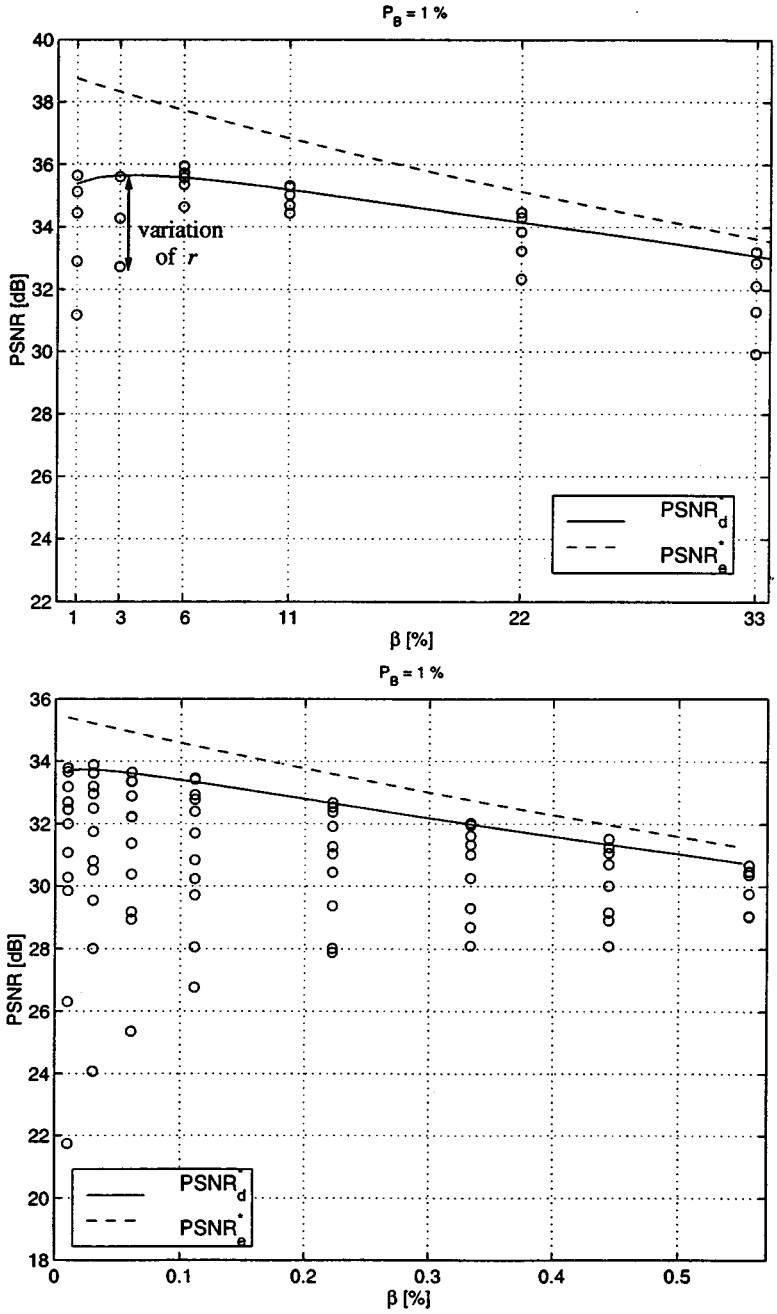

Fig. 9. Optimal PSNR_d^* at decoder and corresponding encoder PSNR_e^* over the INTRA rate β. The channel code rate r is used as a free parameter for optimization. The channel is characterized by the average burst length $L_B = 8$ and the symbol error rate $P_B = 1\%$. The test sequences are *Mother&Daughter* (top) and *Foreman* (bottom). The measurements (o) correspond very well with the performance bound predicted by the model (—).

channel, i.e., a block is either completely correct or completely wrong. Hence, FEC is not sufficient for some packets and not needed at all for other packets.

The results of Figs. 11 and 12 can be illustrated in a compact way by plotting the trajectory of optimal parameters in the $\beta^* - r^*$-plane as shown in Fig. 13. Each trajectory corresponds to a given burst length L_B, while the symbol error rate P_B is used as the free parameter that is varied along the trajectory. It can be seen clearly that only for very bursty channels is a high INTRA rate β needed. Note that for a large average burst length L_B, the optimal code rate is $r^* = 100\%$ for low symbol error rates $P_B \leq 0.7\%$. In this case, errors occur very infrequently, but if one occurs it is followed by a whole burst of errors. A lot of FEC would be needed to correct this error burst, thus lowering the available rate for the video bitstream in many blocks which are not affected by channels errors at all. This becomes too expensive at some point and it is better to increase the INTRA rate β.

In Fig. 14 it is shown how the PSNR_d^* drops with increasing symbol error rate P_B for different average burst lengths $L_B = 2, 4, 8, 16, 32$ and for the memoryless channel (BSC). As expected, the quality is generally lower for more bursty channels, i.e., if L_B is higher, since the FEC is less efficient in this case. For very large burst lengths, PSNR_d^* increases again slightly. This is due to the fact that the errors are being clustered together, and thus long periods remain without any transmission error.

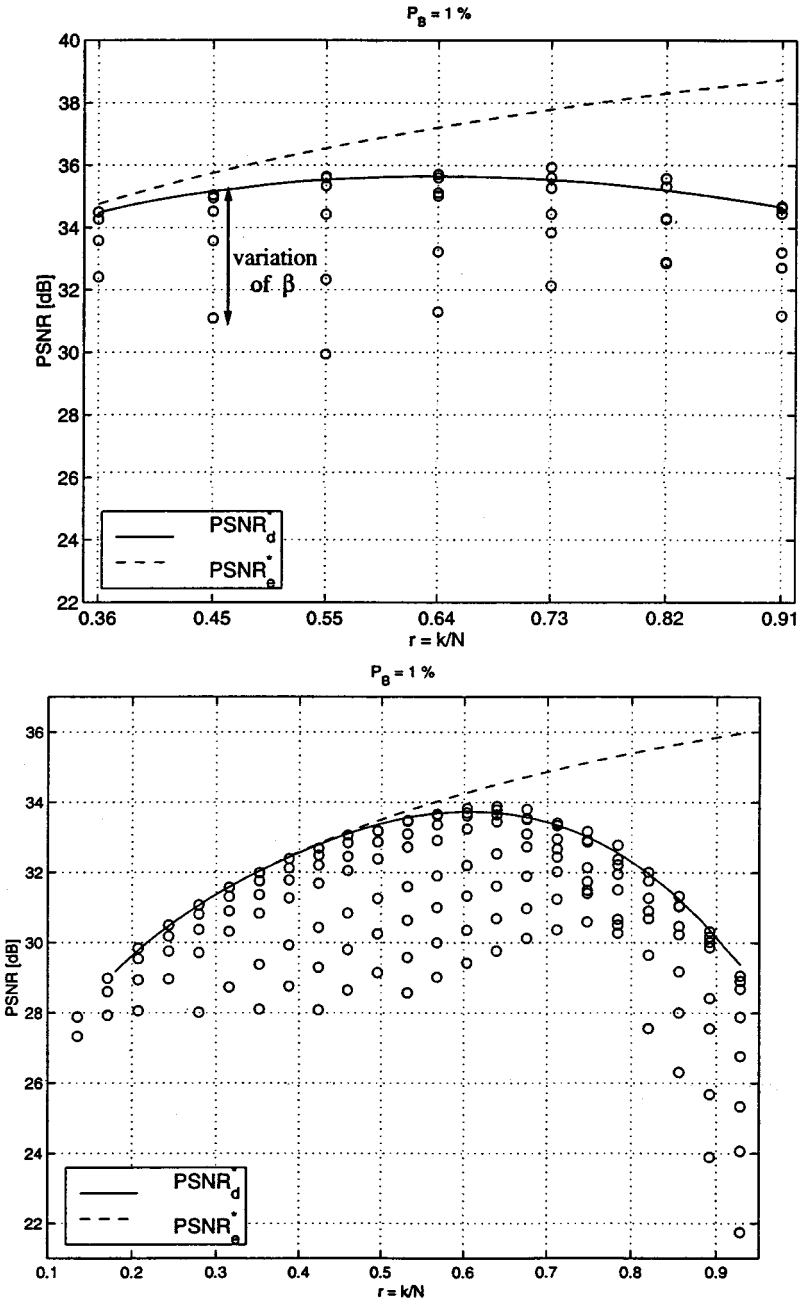

Fig. 10. Optimal PSNR$^*_d$ at decoder and corresponding encoder PSNRe^* over the code rate r. The INTRA rate β is used as a free parameter for optimization. The channel is characterized by the average burst length $L_B = 8$ and the symbol error rate $P_B = 1\%$. The test sequences are *Mother&Daughter* (top) and *Foreman* (bottom). The measurements (o) correspond very well with the performance bound predicted by the model (—).

Hence, the code rate is reduced (see Fig. 13) and therefore the encoded video quality is increased.

VI. Conclusion and Future Work

We have derived a theoretical framework for the decoded picture quality after video transmission over lossy channels. Models for the video encoder, a bursty transmission channel, and error propagation at the video decoder have been combined into a complete model of the entire video transmission system. The proposed model for error propagation includes the effects of INTRA coding and spatial loop filtering. It has been shown that the models correspond to simulation results very accurately while only using a small set of parameters. The model has been used to determine the optimal percentage of INTRA coded macroblocks and the optimal channel code rate for a given channel characteristic. We have also studied the impact of the channel error rate and burstiness on the optimal parameter settings.

We have used a low latency scenario for the simulations. However, if more latency is acceptable, interleaving should be used to reduce the burstiness of the channel. Interleaving can easily be included in the channel model.

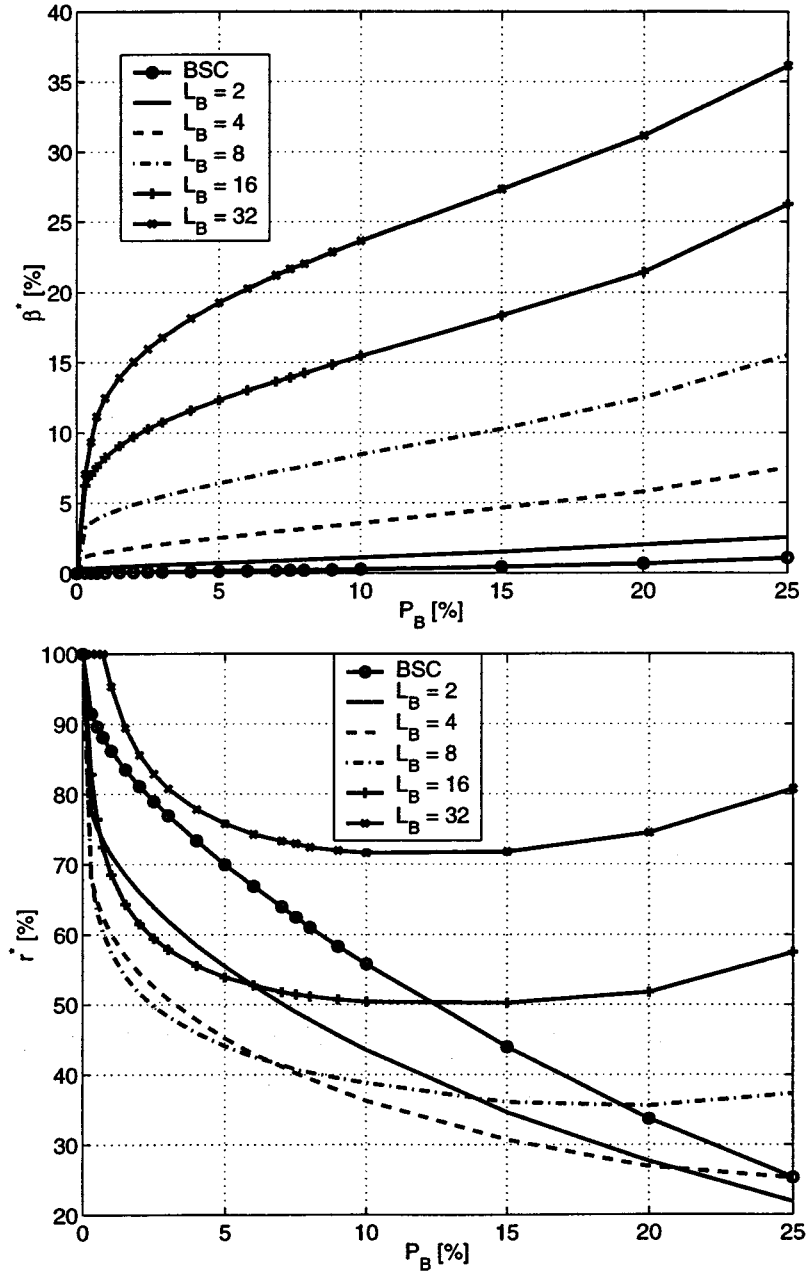

Fig. 11. Optimal parameters β^* and r^* for a transmission of the test sequence *Mother&Daughter* over channels with different symbol error rates P_B and average burst lengths L_B. The optimal parameters for a channel with independent errors (BSC) are also shown.

The results show that for a memoryless channel, FEC is very important, whereas error resilience, i.e., INTRA coding, is not so important in this case. In contrast, for bursty channels the use of FEC is limited and the INTRA update is essential.

More work is left to refine the modeling of a complete video transmission system. For the video encoder, a more theoretically based model should be used. Another important topic is the weighting between the coding errors and the transmission errors in (3). Subjective tests are needed to find a function which provides a better approximation of the decoded video quality. At the moment we are working on the online estimation of the model parameters and its incorporation in an optimal mode selection. The extension of the model to a scalable video transmission system with unequal error protection is also in progress.

APPENDIX I
ANALYSIS OF INTERFRAME ERROR PROPAGATION

In this appendix we derive an analytical model for the distortion that is caused by transmission errors. In particular, we investigate how the energy of introduced errors propagates due to the recursive DPCM structure in the decoder. We are interested in the signal $v[x, y, t]$, which is the difference between the reconstructed frames at encoder and decoder (see Fig. 3 in Section III-B). We assume that the residual error $u[x, y]$ is introduced at $t = 0$ (after resynchronization and error concealment)

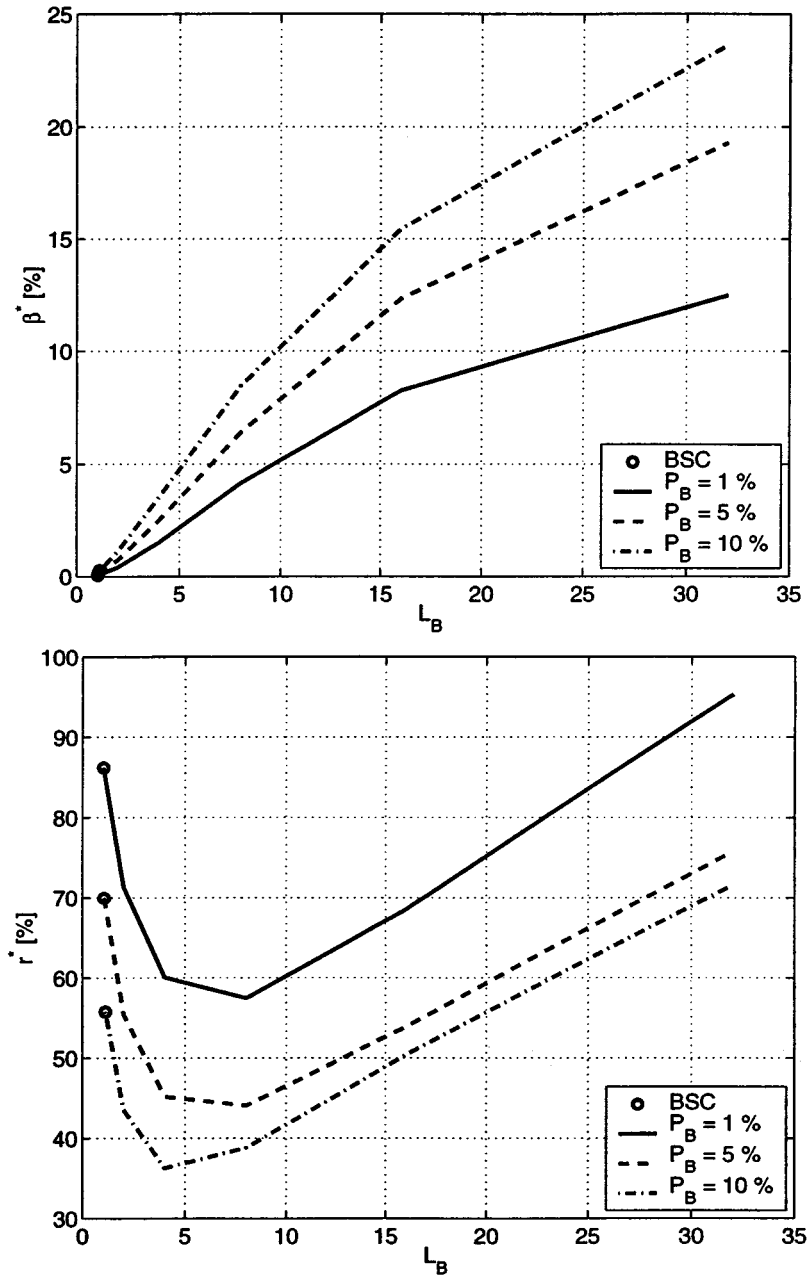

Fig. 12. Optimal parameters β^* and r^* for a transmission of the test sequence *Mother&Daughter* over channels with different average burst lengths L_B for symbol error rates $P_B = 1\%, 5\%$, and 10%.

such that $v[x, y, 0] = u[x, y]$. We are mainly interested in the variance $\sigma_v^2[t]$ of the propagated error signal and in its average over time. The analysis is an extension of previous work [38], [39] that first appeared in [21].

We assume that a separable loop filter is applied to the reconstructed frames after each time step. This loop filter shall describe the overall effect of various spatial filter operations that are performed during encoding. Spatial filtering can either be introduced by an explicit loop filter, as e.g., in H.261, or implicitly as a side effect of half-pel motion compensation with linear interpolation, as in H.263 or MPEG-2. Other prediction techniques like overlapped block motion compensation (OBMC) or deblocking filters inside the DPCM loop may also contribute to the overall loop filter. Although the exact effect is difficult to derive theoretically for the individual prediction techniques, we found that the overall effect can be described by a separable *average loop filter* $F(\omega)$ with impulse response $f[k]$. We will first analyze the effect of this loop filter on the propagated error energy and then add the effect of INTRA coding to the derived model.

When the decoder is regarded as a linear system $H_t(\omega_x, \omega_y)$ with parameter t, the variance of $v[x, y, t]$ can be obtained as

$$\sigma_v^2[t] = \frac{1}{4\pi^2} \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} |H_t(\omega_x, \omega_y)|^2 \cdot \Phi_{uu}(\omega_x, \omega_y)\, d\omega_x\, d\omega_y \tag{14}$$

where Φ_{uu} is the power spectral density (PSD) of the signal $u[x, y]$. As mentioned above, the spatial loop filter $F(\omega)$ is ap-

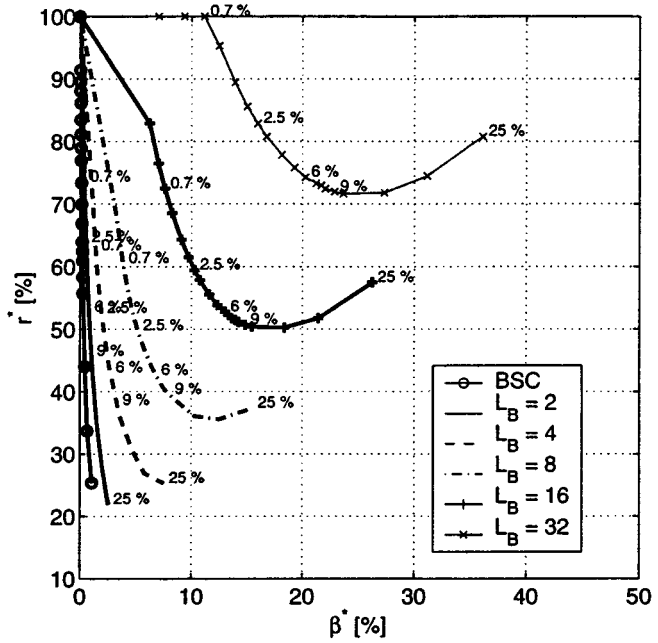

Fig. 13. Optimal parameters β^* and r^* for transmission of the test sequence *Mother&Daughter* dependent on the symbol error rate P_B of the channel for various average burst lengths L_B and for the memoryless channel (BSC).

plied to the reconstructed frame in horizontal and vertical direction in each time step. The resulting two-dimensional filter shall be denoted $F(\omega_x, \omega_y)$ with impulse response $f[x, y]$. Then, the impulse response of the decoder $h_t[x, y]$ can be defined recursively as $h_t[x, y] = h_{t-1}[x, y] * f[x, y]$, where $*$ denotes discrete two-dimensional convolution. Based on the central limit theorem, we expect $h_t[x, y]$ to be Gaussian for large t. Therefore, the squared magnitude of the transfer function of the decoder can be approximated in the base band $|\omega_x|, |\omega_y| < \pi$, by

$$|\hat{H}_t(\omega_x, \omega_y)|^2 = \exp\left[-\left(\omega_x^2 + \omega_y^2\right) t\sigma_f^2\right] \tag{15}$$

where the *filter strength* σ_f^2 is defined as

$$\sigma_f^2 = \sum_k k^2 f[k] - \left(\sum_k k f[k]\right)^2. \tag{16}$$

For example, in the case of bilinear interpolation with $f[k] = [1/2\ \ 1/2]$, we obtain $\sigma_f^2 = 1/4$. In addition to the Gaussian approximation for $H_t(\omega_x, \omega_y)$, we also approximate the PSD of the introduced error signal $u[x, y]$ by

$$\hat{\Phi}_{uu}(\omega_x, \omega_y) = \sigma_u^2 4\pi\sigma_g^2 \exp\left[-\left(\omega_x^2 + \omega_y^2\right)\sigma_g^2\right] \tag{17}$$

i.e., a separable Gaussian PSD with the energy σ_u^2. The parameter σ_g^2 describes the *spectral shape* of the PSD and can be used to match (17) with the true PSD. Note that the same shape parameter is assumed in horizontal and vertical direction. Although this is not necessarily a very accurate assumption, it greatly simplifies the following analysis and provides sufficient accuracy. With the approximations for $|H_t(\omega_x, \omega_y)|$ and $\Phi_{uu}(\omega_x, \omega_y)$, we can solve (14) analytically (after extending the integration interval to $[-\infty, +\infty]$), yielding

$$\hat{\sigma}_v^2[t] = \frac{\sigma_u^2}{1 + \gamma t} = \sigma_u^2 \alpha[t] \tag{18}$$

where $\gamma = \sigma_f^2/\sigma_g^2$ is the *leakage* and $\alpha[t]$ is the power transfer factor after t time steps. Thus, the introduced energy σ_u^2 decays proportional to $1/t$, and the decay is determined by the strength of the loop filter (σ_f^2) as well as the shape of the PSD of the introduced error u (σ_g^2).

So far we did not consider INTRA coded macroblocks, which cause a faster decay in error energy. If the INTRA mode is selected once every T frames for each macroblock, and the update time for a specific macroblock is selected randomly in this interval, the effect on the variance can be modeled as a linear decay. With $\beta = 1/T$ being the percentage of INTRA coded macroblocks, the final equation for the power transfer factor becomes

$$\alpha[t] = \frac{1 - \beta t}{1 + \gamma t} \tag{19}$$

for $0 \le t < T$. For $t \ge T$, the error energy is removed completely and, thus, $\alpha[t] = 0$. Note that motion compensated prediction may cause spatial error propagation, such that errors may actually "survive" one INTRA update period T. However, our simulation results show that this effect is neglectable when considering the average transmission error energy.

APPENDIX II
CHANNEL MODEL FOR BURST CHANNELS

A. Derivation of Block Error Density

We show how the block error density function $P_D(n, \kappa)$ can be calculated from the parameters P_B and L_B of the 2-state Markov model described in Section IV. The derivation can also be found in [19] and [18] and follows the one given in [36]. It is repeated here for completeness.

The Markov model possesses a characteristic distribution of error-free intervals (gaps). Let a gap of length ν be the event that after an error $\nu - 1$ symbols are received correctly and then again an error occurs. The gap density function $g(\nu)$ gives the probability of a gap length ν, i.e., $g(\nu) = \Pr(0^{\nu-1}1 \mid 1)$, where "1" denotes an error, and "$0^{\nu-1}$" denotes $\nu - 1$ consecutively correctly received symbols. The gap distribution function $G(\nu)$ gives the probability of a gap length greater than $\nu - 1$, i.e., $G(\nu) = \Pr(0^{\nu-1} \mid 1)$. In state B, all symbols are lost ("1"), while in state G all symbols are received ("0"), yielding

$$g(\nu) = \begin{cases} 1 - p_{\mathrm{GB}}, & \nu = 1 \\ p_{\mathrm{GB}}(1 - p_{\mathrm{GB}})^{\nu-2} p_{\mathrm{GB}}, & \nu > 1 \end{cases}$$

$$G(\nu) = \begin{cases} 1, & \nu = 1 \\ p_{\mathrm{GB}}(1 - p_{\mathrm{GB}})^{\nu-2}, & \nu > 1. \end{cases}$$

Let $R(n, \kappa)$ be the probability of $\kappa - 1$ erroneous symbols within the next $n - 1$ symbols following an erroneous symbol. It can be calculated using the recurrence

$$R(n, \kappa) = \begin{cases} G(n), & \kappa = 1 \\ \displaystyle\sum_{\nu=1}^{n-\kappa+1} g(\nu) R(n - \nu, \kappa - 1), & 2 \le \kappa \le n. \end{cases}$$

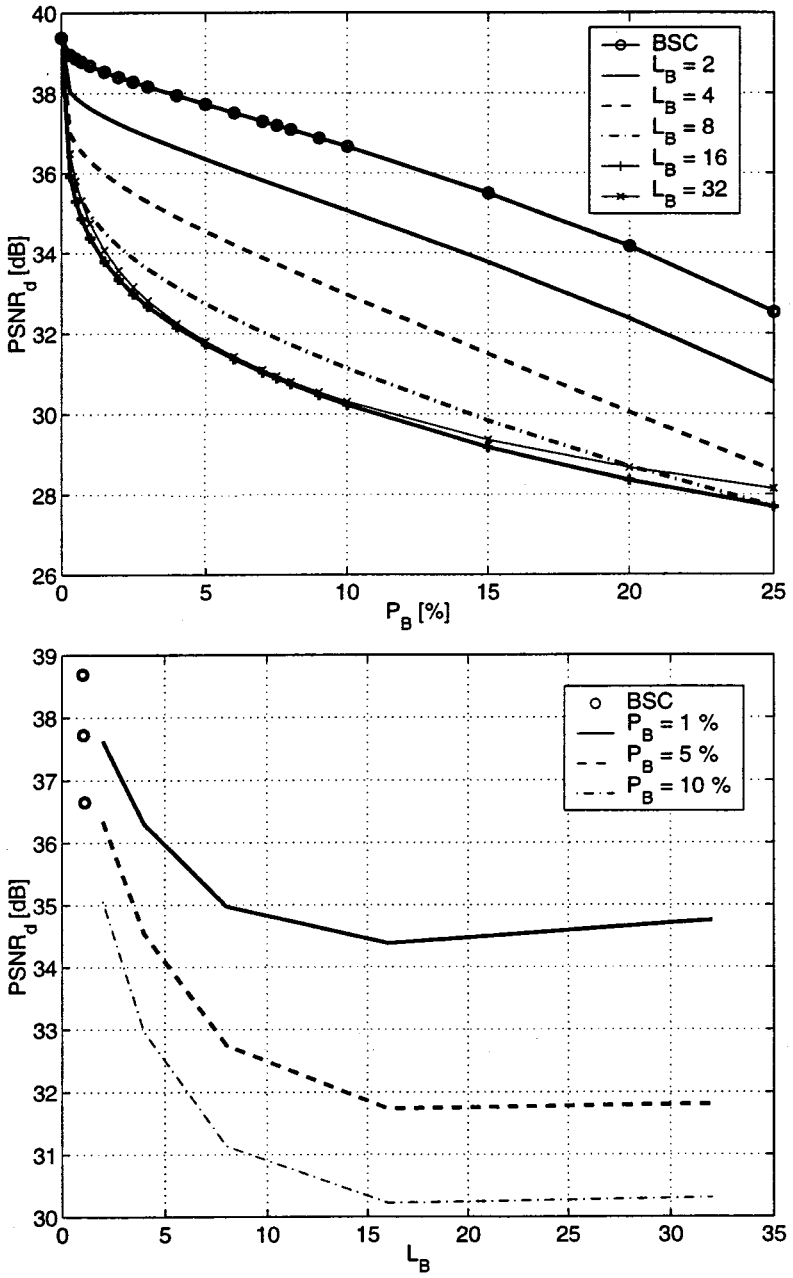

Fig. 14. PSNR_d^* at the decoder for optimally chosen parameters β^* and r^* is plotted over the symbol error rates P_B (top) and over the average burst length L_B (bottom). BSC denotes the memoryless channel (independent symbol errors). The test sequence is *Mother&Daughter*.

Then the probability of κ errors within a block of n symbols is

$$P_D(n,\kappa) = \begin{cases} \displaystyle\sum_{\nu=1}^{n-\kappa+1} P_B G(\nu) R(n-\nu+1,\kappa), & 1 \le \kappa \le n \\ \displaystyle 1 - \sum_{\nu=1}^{n} P_D(n,\nu), & \kappa = 0 \end{cases}$$

where P_B is the average error probability.

B. Interleaving

We show how interleaving can be incorporated in our channel model.

The one-step transition matrix of the Markov model is defined as

$$\mathbf{\Psi} = \begin{bmatrix} 1-p_{\text{GB}} & p_{\text{GB}} \\ p_{\text{GB}} & 1-p_{\text{GB}} \end{bmatrix}.$$

With i symbols interleaving, i.e., every ith transmitted symbol belongs to the same block, we can use the same Markov model in our analysis if we replace the one-step transition matrix $\mathbf{\Psi}$ by the i-step transition matrix [40], [41]

$$\mathbf{\Psi}_i = \mathbf{\Psi}^i$$

which is the ith power of $\mathbf{\Psi}$. From $\mathbf{\Psi}_i$ the probabilities $p_{GB,i} = \mathbf{\Psi}_i(1,2)$ and $p_{BG,i} = \mathbf{\Psi}_i(2,1)$ are extracted, and the symbol

error rate $P_{B,i}$ and the average burst length $L_{B,i}$ can be calculated as described in Section IV. Obviously the error probability P_B is independent of the interleaving but the burst length changes asymptotically to

$$L_B = \frac{1}{1 - P_B} \qquad \text{for } i \to \infty$$

which corresponds to a binary symmetric channel.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[2] Y. Wang and Q. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, pp. 974–997, May 1998.

[3] P. Bahl and B. Girod, Eds., *IEEE Commun. Mag.*, June 1998, vol. 36, Special Section on Wireless Video.

[4] J. C. Brailean, T. Sikora, and T. Miki, Eds., *Special Issue on Error Resilience*, May 1999, vol. 14, Signal Processing: Image Commun..

[5] H. Gharavi and L. Hanzo, Eds., *Special Issue on Video Transmission for Mobile Multimedia Applications*, Oct. 1999, vol. 87, Proc. IEEE.

[6] A. Furaskär, S. Mazur, F. Müller, and H. Olofsson, "Edge: Enhanced data rates for GSM and TDMA/136 evolution," *IEEE Personal Commun.*, vol. 6, pp. 56–66, June 1999.

[7] S. Bruh, P. Blöcher, K. Hellwig, and J. Sjöberg, "Concepts and solutions for link adaptation and inband signaling for the GSM AMR speech coding standard," in *Proc. Veh. Technol. Conf. (VTC)*, vol. 3, Houston, TX, May 1999, pp. 2451–2455.

[8] T. F. La Porta, Ed., *Special Issue on Software Radio*, Aug. 1999, vol. 6, IEEE Personal Commun..

[9] Q. F. Zhu and L. Kerofsky, "Joint source coding, transport processing, and error concealment for H.323-based packet video," in *Proc. Symp. Visual Commun. Image Processing*, vol. 3653, SPIE, San Jose, CA, Jan. 1999, pp. 52–62.

[10] B. Hochwald and K. Zeger, "Tradeoff between source and channel coding," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1412–1424, Sept. 1997.

[11] G. Buch, F. Burkert, J. Hagenauer, and B. Kukla, "To compress or not to compress?," in *Proc. 5th CMTC GLOBECOM'96*, London, Nov. 1996.

[12] ITU-T, "Rec. H.261 Vers. 3/1993, video codec for audiovisual services at $p \times 64$ kbit/s,", 1993.

[13] ITU-T, "Video coding for narrow telecommunication channels at <64 kbit/s, draft ITU-T recommendation H.263,", Apr. 1995.

[14] ITU-T, "Rec. H.263, video coding for low bitrate communication, version 2 (H.263+),".

[15] ISO/IEC 11172, "Information technology—Coding of moving pictures and associated audio for digital storage media at up to about 1.5 MBit/s,".

[16] ISO/IEC, "Document 13818-2 (MPEG-2), generic coding of moving pictures and associated audio, Part 2: Video, recommendation H.262, international standard,".

[17] ISE/IEC, "Document 14496-2, coding of audio-visual objects: Visual (MPEG-4 video),", 1998.

[18] B. Girod, K. W. Stuhlmüller, M. Link, and U. Horn, "Packet loss resilient Internet video streaming," in *Proc. Symp. Visual Commun. Image Processing*, SPIE, San Jose, CA, Jan. 1999.

[19] U. Horn, K. Stuhlmüller, M. Link, and B. Girod, "Robust internet video transmission based on scalable coding and unequal error protection," *Image Commun.*, vol. 15, no. 1–2, pp. 77–94, Sept. 1999.

[20] S. B. Wicker, *Error Control Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[21] N. Färber, K. W. Stuhlmüller, and B. Girod, "Analysis of error propagation in hybrid video coding with application to error resilience," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Kobe, Japan, Oct. 1999, pp. 550–554.

[22] Telenor Research. (1995) Video codec test model, TMN5. [Online]. Available: http://www.nta.no/brukere/DVC

[23] P. Haskell and D. Messerschmitt, "Resynchronization of motion compensated video affected by ATM cell loss," in *Proc. Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, vol. 3, San Francisco, CA, Mar. 1992, pp. 545–548.

[24] J. Y. Liao and J. D. Villasenor, "Adaptive intra update for video coding over noisy channels," in *Proc. Int. Conf. Image Processing*, vol. 3, Lausanne, Switzerland, Sept. 1996, pp. 763–766.

[25] T. Wiegand, N. Färber, K. Stuhlmüller, and B. Girod, "Error-resilient video transmission using long-term memory motion-compensation prediction," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1050–1061, June 2000.

[26] S. Wenger and G. Côté, "Using RFC2429 and H.263+ at low to medium bit-rates for low-latency applications," presented at the Packet Video Workshop, New York, Apr. 1999.

[27] G. Côté and F. Kossentini, "Optimal intra coding of blocks for robust video communication over the internet," *Signal Processing, Image Commun.*, vol. 15, no. 1–2, pp. 25–34, Sept. 1999.

[28] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 1140–1154, Aug. 1987.

[29] ——, "Motion-compensating prediction with fractional-pel accuracy," *IEEE Trans. Commun.*, vol. 41, pp. 604–612, Apr. 1993.

[30] H.-M. Hang and J.-J. Chen, "Source model for transform video coder and its application," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 287–311, Apr. 1997.

[31] W.C. Chung, F. Kossentini, and M. J. T. Smith, "An efficient motion estimation technique based on a rate-distortion criterion," in *Proc. Int. Conf. Accoust., Speech, Signal Processing (ICASSP)*, Atlanta, GA, May 1996, pp. 1927–1930.

[32] J. Ribas-Corbera and S. Lei, "Optimal quantizer control in DCT video coding for low-delay video communications," in *Proc. Pict. Cod. Symp.*, Berlin, Mar. 1997, pp. 749–754.

[33] J. L. H. Webb and K. Oehler, "A simple rate-distortion model, parameter estimation, and application to real-time rate control for DCT-based coders," in *Proc. Int. Conf. Image Processing*, vol. 2, Santa Barbara, CA, Oct. 1997, pp. 13–16.

[34] K.H. Yang, A. Jacquin, and N.S. Jayant, "A normalized rate-distortion model for H.263-compatible codecs and its application to quantizer selection," in *Proc. Int. Conf. Image Processing*, vol. 2, Santa Barbara, CA, Oct. 1997, pp. 41–44.

[35] E. Steinbach, N. Färber, and B. Girod, "Standard compatible extension of H.263 for robust video transmission in mobile environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 872–881, Dec. 1997.

[36] E. O. Elliott, "A model of the switched telephone network for data communications," *Bell. Syst. Tech. J.*, vol. 44, pp. 89–109, Jan. 1965.

[37] L. N. Kanal and A. R. K. Sastry, "Models for channels with memory and their applications to error control," *Proc. IEEE*, vol. 66, pp. 724–744, July 1978.

[38] B. Girod and N. Färber, "Feedback-based error control for mobile video transmission," *Proc. IEEE*, vol. 87, pp. 1703–1723, Oct. 1999.

[39] ——, "Error-resilient standard-compliant video coding," in *Recovery Techniques for Image and Video Compression and Transmission*, A. Katsaggelos and N. Galatsanos, Eds. Boston, MA: Kluwer Academic, Oct. 1998.

[40] R. Coleman, "Stochastic processes," in *Problem Solvers*. London: George Allen & Unwin Ltd., 1974.

[41] J. Huber, "Codierung für gedächtnisbehaftete Kanäle," Ph.D. dissertation, Hochschule der Bundeswehr, München, July 1982.

Klaus Stuhlmüller received the diploma in electrical engineering from the University Erlangen-Nürnberg in 1994.

From 1994 to 1996 he was at the Fraunhofer Institute for Applied Electronics. Since 1996 he has been working as a Research Assistant at the Telecommunications Laboratory, University Erlangen-Nürnberg. His research interests are in very low bit rate video coding, object based video coding, real-time video streaming, robust video transmission, and modeling of video transmission systems.

Mr. Stuhlmüller received the Young Investigator Award of the SPIE Visual Communication and Image Processing Conference in 1996.

Niko Färber (M'99) received the Diplom-Ingenieur degree (Dipl.-Ing.) in electrical engineering from the Technical University of Munich, Germany, in 1993.

He was with the research laboratory Mannesmann Pilotentwicklung, where he developed system components for satellite based vehicular navigation. In October 1993 he joined the Telecommunications Laboratory at the University of Erlangen-Nuremberg, Germany, and is now a member of the Image Communication Group. He started his research on robust video transmission as a Visiting Scientist at the Image Processing Laboratory of University of California, Los Angeles. Since then he has published several conference and journal papers on the subject, and has contributed successfully to the ITU-T Study Group 16 efforts for robust extensions of the H.263 standard. His doctoral thesis is entitled "Feedback-Based Error Control for Robust Video Transmission" and is supervised by Prof. B. Girod. He has also served as the Publicity Vice Chair for ICASSP-97 in Munich, Germany, and performed research for 8 × 8, Inc., Santa Clara, CA, and RealNetworks, Inc., Seattle, WA. In cooperation with RealNetworks, he holds a U.S. patent in the area of scalable video coding. His current research project is on scalable video streaming over UMTS, which is conducted with Ericsson Eurolab, Herzogenrath, Germany.

Michael Link received the diploma in electrical engineering from the University Erlangen-Nürnberg, Germany, in 1994.

From 1994 to 1999, he was a Research Assistant at the Telecommunications Laboratory, University Erlangen-Nürnberg. In September 1999 he joined the Network Technology Group at Lucent Technologies, Nürnberg. His research interests are in channel coding for the transmission of multimedia data in packet switched networks, channel modeling, and wireless multimedia protocols, especially for UMTS.

Bernd Girod (S'80–M'80–SM'97–F'98) received the M.S. degree in electrical engineering from Georgia Institute of Technology in 1980, and the Doctoral degree (with highest honors) from University of Hannover, Germany, in 1987.

He is a Professor of Electrical Engineering, Information Systems Laboratory, Stanford University, Stanford, CA. Until 1987 he was a member of the research staff at the Institut fur Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover, working on moving image coding, human visual perception, and information theory. In 1988, he joined Massachusetts Institute of Technology, Cambridge, first as a Visiting Scientist with the Research Laboratory of Electronics, then as an Assistant Professor of Media Technology at the Media Laboratory. From 1990 to 1993, he was a Professor of Computer Graphics and Technical Director of the Academy of Media Arts in Cologne, Germany, jointly appointed with the Computer Science Section of Cologne University. He was a Visiting Adjunct Professor with the Digital Signal Processing Group at Georgia Institute of Technology, Atlanta, GA, in 1993. From 1993 until 1999, he was a Chaired Professor of Electrical Engineering/Telecommunications at University of Erlangen-Nuremberg, Germany, and the Head of the Telecommunications Institute I, co-directing the Telecommunications Laboratory. He has served as the Chairman of the Electrical Engineering Department from 1995 to 1997, and as Director of the Center of Excellence "3-D Image Analysis and Synthesis" from 1995 to 1999. He has been a Visiting Professor with the Information Systems Laboratory of Stanford University, Stanford, CA, during the 1997/1998 academic year. His research interests include image communication, video signal compression, human and machine vision, computer graphics and animation, as well as interactive media. For several years, he has served as a consultant to government agencies and companies, with special emphasis on start-up ventures. He has been a founder and Chief Scientist of Vivo Software, Inc., Waltham, MA (1993–1998), Chief Scientist of RealNetworks, Inc., Seattle, WA (since 1998), and a board member of 8 × 8, Inc., Santa Clara, CA (since 1996). He has authored or coauthored one major textbook and over 200 book chapters, journal articles, and conference papers in his field, and he holds several international patents.

Dr. Girod has served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1991 to 1995, and as Reviewing Editor and Area Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS since 1995. He is also a member of the Editorial Board of the IEEE SIGNAL PROCESSING MAGAZINE. He served as General Chair of the 1998 IEEE Image and Multidimensional Signal Processing Workshop in Alpbach, Austria. He was a member of the IEEE Image and Multidimensional Signal Processing Committee from 1989 to 1997. He was elected Fellow of the IEEE in 1998 "for his contributions to the theory and practice of video communications."