

# Modeling and Optimization of a Systematic Lossy Error Protection System based on H.264/AVC Redundant Slices\*

Shantanu Rane, Pierpaolo Baccichet\*\* and Bernd Girod

Information Systems Laboratory, Department of Electrical Engineering  
Stanford University, Stanford, CA 94305  
{srane,bacci,bgirod}@stanford.edu

**Abstract.** We model the decoded picture quality at the output of a Systematic Lossy Error Protection (SLEP) scheme for error-resilient transmission of H.264/AVC compressed video signals. In this scheme, a video signal transmitted without channel coding constitutes the systematic portion of the transmission. Error-resilience is provided by transmitting a supplementary bit stream generated by Wyner-Ziv encoding. Our Wyner-Ziv codec uses H.264/AVC redundant slices in conjunction with Reed Solomon coding. When channel errors occur, the Wyner-Ziv bit stream allows the decoding of a coarsely quantized description of the video signal, which is used for lossy error protection. We study the error-resilience properties of SLEP using the model equations and utilize them for optimization of the SLEP system.

**Index Terms :** Wyner-Ziv coding, distributed video coding, side information, systematic source-channel coding, H.264/AVC, redundant slices.

## 1 Introduction

In our recent work [1–3], we have proposed an error-resilient scheme called Systematic Lossy Error Protection (SLEP), which uses Wyner-Ziv coding ideas to achieve error resilience. This scheme achieves a graceful trade-off between the decoded video quality and the resilience to channel errors, effectively mitigating the cliff effect of FEC. In this paper, we derive a model for the picture quality obtained using SLEP. Using the model, we study the error-resilience properties of SLEP and quantify the trade-offs involved. We cast the model equations into an optimization problem to find the source/channel rate allocation which maximizes the received picture quality. The present work was conducted in parallel with the actual implementation

and experimental evaluation of SLEP using redundant slices and flexible macroblock ordering [4].

The SLEP system is based on the systematic lossy source-channel coding framework in which, a source  $X$  is transmitted over an analog channel  $A$  without coding. A second encoded version of  $X$  is sent over a digital channel  $D$  as enhancement information. The noisy output  $Y$  of the analog channel serves as side information to decode the output of channel  $D$  and produce the enhanced version  $Z$ . Thus, source coding with decoder side information, i.e., Wyner-Ziv coding [5], is an integral part of the lossy source-channel coding configuration. The term systematic coding is used as an extension of systematic error-correcting channel codes, and refers to a partially uncoded transmission. Shamai, Verdú, and Zamir established information theoretic bounds and conditions for optimality of this configuration in [6].

The remainder of this paper is organized as follows. Section 2 outlines the principle of Systematic Lossy Error Protection. In Section 3, we describe an implementation of the SLEP system using redundant slices, a feature supported in the “Baseline Profile” of the H.264/AVC standard [7]. In Section 4, we model the average end-to-end distortion experienced by a decoded video packet, and use it to explain some properties of the SLEP scheme. Section 5 describes a method to choose encoding rates for the primary and redundant slices in the SLEP scheme with the aim of maximizing the average received picture quality.

## 2 SLEP Concept

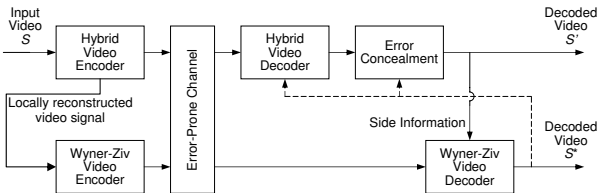
As shown in Fig. 1, consider a compressed video bit stream which is transmitted across an error-prone channel without channel coding, comprising the systematic part of the transmission. For error resilience, a second independent bit stream is generated using Wyner-Ziv encoding of the video signal. The Wyner-Ziv bit stream allows the decoding of a coarsely quantized description of the original video

---

\* This work has been supported in part by NSF Grant No:CCR-0310376

\*\* On leave from I.E.I.I.T. - Consiglio Nazionale delle Ricerche, Italy, partially supported by MIUR (Italian Ministry of Education and Research) under Research Project PRIMO - “Reconfigurable platforms for wideband wireless communications”.

signal. Since the video bit stream is generated without consideration of the error resilience provided by the Wyner-Ziv coder, we refer to the overall scheme as systematic source-channel coding. The receiver



**Fig. 1.** A Wyner-Ziv decoder uses the decoded error-concealed video waveform as side information in a systematic lossy source-channel setup.

decodes the compressed video signal, and conceals erroneous or lost portions. Even after concealment, some portions of the recovered video signal may contain unacceptably large errors. These errors are corrected up to a certain residual distortion by the Wyner-Ziv decoder.

The Wyner-Ziv bits allow error-free reconstruction of the coarser second description, employing the decoded video signal  $S'$  as side information. The coarser second description and side information  $S'$  are then combined to yield an improved decoded video signal  $S^*$ . In portions where the waveform  $S'$  is not affected by transmission errors,  $S^*$  is essentially identical to  $S'$ . However, in portions of the waveform where  $S'$  is substantially degraded by transmission errors, the second coarser representation transmitted at very low bit rate in the Wyner-Ziv bit stream limits the maximum degradation that can occur. Instead of the error-concealed decoded signal  $S'$ , the signal  $S^*$  at the output of the Wyner-Ziv encoder is fed back to the video decoder to serve as a more accurate reference frame for decoding of the future frames.

### 3 SLEP using H.264/AVC Redundant Slices

#### 3.1 Wyner-Ziv encoding

The implementation of the SLEP scheme is shown in Fig. 2. The following operations are performed on the encoder side:

1. *Generation of Redundant Slices:* Each macroblock belonging to the redundant description

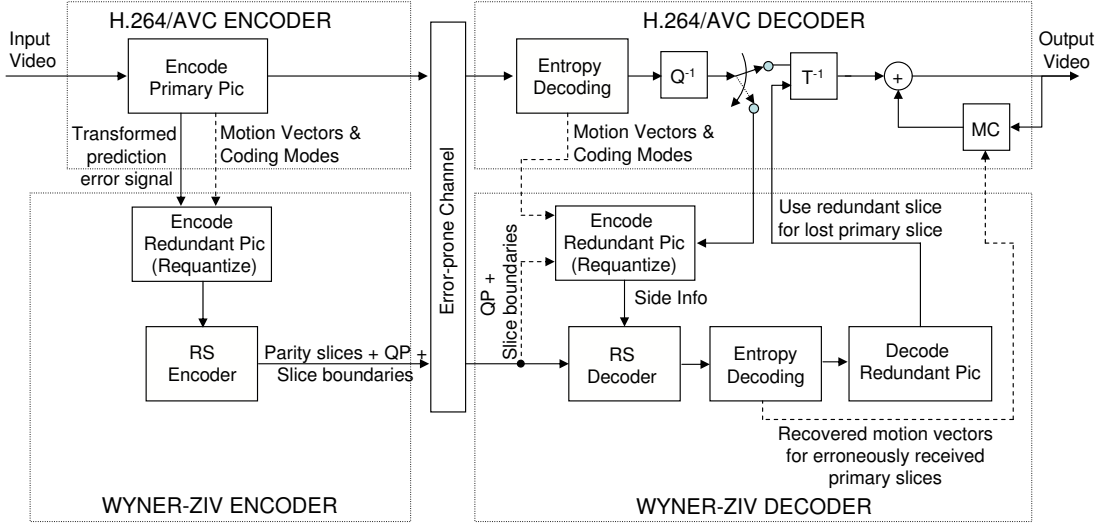
is encoded with the same coding mode, motion vectors and reference pictures of the corresponding primary coded macroblock.

2. *Reed-Solomon encoding:* Reed-Solomon (RS) codes perform the role of Slepian-Wolf coding in this system. A Reed-Solomon code over  $GF(2^8)$  is applied across the redundant slices, to generate parity slices, as shown in Fig. 2. The number of parity slices generated per frame depends upon the allowable error resilience bit rate, and can vary slightly from frame to frame. The redundant slices are then *discarded* and *only the parity slices* are included for transmission in the Wyner-Ziv bit stream.
3. *Wyner-Ziv bit stream generation:* In addition to the parity slices resulting from the previous step, we encode the slice boundaries, i.e., the number of macroblocks in each redundant slice, and the quantization parameter (QP).

#### 3.2 Wyner-Ziv decoding

Wyner-Ziv decoding is activated only when transmission errors result in the loss of one or more slices from the bit stream of the primary coded picture. It consists of the following operations:

1. *Requantization to obtain Redundant Slices:* This step involves the requantization of the prediction residual signal of the primary coded picture, followed by entropy coding. This generates the redundant slices used as side information for the Wyner-Ziv decoder. Redundant slices can be generated only for those portions of the frame, where the primary bit stream has not experienced channel errors. Redundant slices containing errors are treated as erasures. This simplification is a slight departure from the SLEP concept of Section 1, which requires a full re-encoding of the error-concealed primary video signal. This would incur a large complexity cost because motion estimation would have to be re-performed at the decoder. The requantization process sacrifices a small amount of coding efficiency, but requires very low complexity. In addition, since Wyner-Ziv encoding is applied to the prediction residual of the current frame, this method is robust to Wyner-Ziv decoder failure that would otherwise occur due to error propagation from previous frames.
2. *Reed-Solomon Slepian-Wolf decoding:* The parity slices received in the Wyner-Ziv bit stream are combined with the redundant slices, and erasure decoding is performed to recover the



**Fig. 2.** Implementation of a SLEP system using H.264/AVC redundant slices. Reed Solomon codes applied across the redundant slices play the role of Slepian-Wolf codes in distributed source coding. At the receiver, the Wyner-Ziv decoder obtains the correct redundant slices using the error-prone primary coded slices as side information. The redundant description is used in lieu of the lost portions of the primary (systematic) signal.

slices which were erased from the redundant bit stream, as shown in Fig. 2. In the language of distributed video coding, the RS decoder functions as a Slepian-Wolf decoder, and recovers the correct redundant bit stream using the error-prone redundant bit stream as side information.

3. *Concealment of lost primary slices:* If Wyner-Ziv decoding succeeds, the lost portions of the prediction residual from the primary (systematic) signal are replaced by the quantized redundant prediction error signal. The H.264/AVC decoder then performs motion compensation in the conventional manner, using the motion vectors recovered from Wyner-Ziv decoding. This operation results in a quantization mismatch which propagates to the future frames, but avoids drastic reduction in picture quality.

## 4 Video Distortion Model

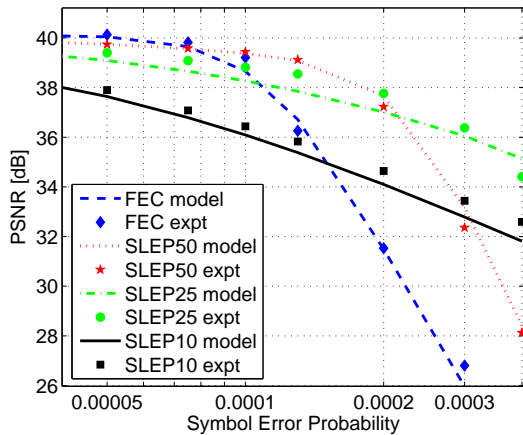
We now derive a model for the average end-to-end distortion incurred by a SLEP system implemented using H.264/AVC redundant slices, extending the model derived in [3] where we considered MPEG-2 video transmission. Let the distortion-rate pairs for encoding the primary and redundant pictures be denoted by  $(D_p, R_p)$  and  $(D_r, R_r)$  re-

spectively. As the redundant description is coarser,  $R_r \leq R_p$ ,  $D_r \geq D_p$ .

Let  $D[i]$  be the *average* end-to-end distortion experienced by a packet in the  $i^{\text{th}}$  frame (assume it is a P frame). We consider three distinct scenarios: (1) There are no errors, and error energy in frame  $i$  is contributed only by the distortion propagating from the previous frame, denoted as  $D[i-1]$ , (2) Wyner-Ziv decoding is successful and the total distortion contribution from error propagation and Wyner-Ziv decoding is  $D[i-1] + D_r - D_p$ , with  $D_r - D_p$  representing the error energy corresponding to the quantization mismatch between the primary and redundant descriptions, (3) Wyner-Ziv decoding fails and the resulting distortion from error propagation and previous frame error concealment is modeled as  $D[i-1] + \text{MSE}[i, i-1]$ , where  $\text{MSE}[i, i-1]$  is the mean squared error between frames  $i$  and  $i-1$ . The derivation of these three distortions by averaging per-pixel squared errors is explained in detail in [3]. Combining the three distortions and weighting each by its probability of occurrence,

$$\begin{aligned}
 D[i] = & (1-p)D[i-1] \\
 & + pp_{wz} (D[i-1] + D_r - D_p) \\
 & + pp_{ec} (D[i-1] + \text{MSE}[i, i-1]) \quad (1)
 \end{aligned}$$

where  $p$  is the packet erasure probability experienced by the Wyner-Ziv decoder,  $p_{wz}$  is the prob-



**Fig. 3.** SLEP results for the *Foreman* CIF sequence, using  $R_p = 1$  Mbps,  $R_{WZ} = 200$  kbps. The error resilience increases when the bit rate of the redundant description is decreased from  $R_r = R_p = 1$  Mbps (FEC) to  $R_r = R_p/10 = 100$  kbps (SLEP10).

ability that Wyner-Ziv decoding is successful,  $p_{EC}$  is the probability that Wyner-Ziv decoding fails, forcing the decoder to use error concealment.

For the implementation of Fig. 2, the success or failure of Wyner-Ziv decoding is completely determined by the number of erasures seen by the Reed-Solomon decoder located inside the Wyner-Ziv decoder. Thus,

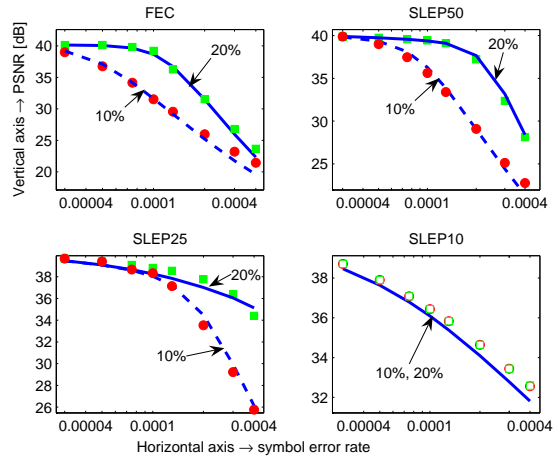
$$p_{WZ} = \sum_{m=k}^{n-1} \binom{n-1}{m} (1-p)^m p^{n-1-m} \quad (2)$$

$$p_{EC} = 1 - p_{WZ} \quad (3)$$

where  $(n, k)$  are the parameters of the Reed-Solomon code used in the Wyner-Ziv encoder. Since  $n$  and  $k$  differ slightly from frame to frame, the model uses values of  $n$  and  $k$  averaged over a GOP of  $N$  frames.

#### 4.1 Changing the redundant description at constant Wyner-Ziv bit rate

Consider the effect of changing the redundant description, i.e., varying  $R_r$ , with no change in the systematic bit rate  $R_p$  or the Wyner-Ziv bit rate  $R_{WZ}$ . As shown in Fig. 3, increasing the coarseness of the Wyner-Ziv description increases the error resilience for a constant  $R_{WZ}$ , in exchange for a small loss in video quality at low symbol error



**Fig. 4.** The error resilience of SLEP increases when the Wyner-Ziv bit rate  $R_{WZ}$  is increased from 10% to 20% of the primary description bit rate  $R_p$ . For SLEP10, the schemes with 10% and 20% Wyner-Ziv bit rate have identical performance over the depicted range of symbol error rates.

probabilities. This is due to the quantization mismatch from Wyner-Ziv decoding, and is expressed in the difference  $D_r - D_p$  in (1). This trade off is calculated explicitly in the appendix. The results shown in Fig. 3 are for a wireless scenario, where symbol errors result in packet erasures at the input of the Wyner-Ziv decoder. The packet erasure probability  $p$  is obtained from the symbol error rate  $s$  using  $p = 1 - (1-s)^l$ , where  $l$  is the packet length.

#### 4.2 Increasing the Wyner-Ziv bit rate

As expected SLEP provides superior error-resilience when the Wyner-Ziv bit rate  $R_{WZ}$  is increased, for the same redundant description, i.e., for constant  $R_r$ . In Fig. 4, we observe that the range of error probabilities over which acceptable decoded picture quality can be obtained, increases when the Wyner-Ziv bit rate,  $R_{WZ}$  is increased from 10% to 20% of  $R_p$ , the bit rate of the primary description. This is shown for 4 Wyner-Ziv descriptions, encoded at  $R_r = R_p = 1$  Mbps (which is the same as FEC),  $R_r = R_p/2$  (designated SLEP50 in Fig. 4),  $R_r = R_p/4$  (SLEP25),  $R_r = R_p/10$  (SLEP10). The experimental results and the model are in close agreement as to the above behavior.

## 5 Optimizing a SLEP system

We use the model to find  $R_p^*$ , the optimum bit rate for encoding the primary pictures,  $R_r^*$ , the optimum bit rate for encoding the redundant description and  $R_{WZ}^*$ , the optimum Wyner-Ziv bit rate, such that average distortion in the decoded video sequence is minimized for a given worst case packet erasure probability  $p_e$ . Recall that, irrespective of whether video packets are lost or corrupted, the Wyner-Ziv decoder sees erasures in the recovered redundant description. In the following, we assume that the *average* Wyner-Ziv bit rate is just large enough to ensure that Wyner-Ziv decoding is successful, at the maximum erasure probability  $p_e$  encountered by the system. This is reasonable because error concealment results in degradation of video quality, and we wish to avoid it for all  $p \leq p_e$ . With this assumption, we can set  $p = p_e$ ,  $p_{wz} = 1$  and  $p_{ec} = 0$  in (1) and calculate the average distortion for a GOP of length  $N$  frames as follows,

$$\mathcal{D} = \frac{1}{N} \sum_{i=1}^N D[i] = D_p + \frac{N+1}{2} p_e (D_r - D_p) \quad (4)$$

The right hand side of (4) shows that Wyner-Ziv decoding contributes some excess error energy which depends upon the quality of the redundant description  $D_r$ , the probability of packet erasure  $p_e$ , and the number of frames  $N$  across which this energy can propagate. The above assumption means that the average Wyner-Ziv bit rate is given by

$$R_{WZ} = \frac{n-k}{k} R_r = \frac{p_e}{1-p_e} R_r \quad (5)$$

The total transmitted bit rate is  $R_p + R_{WZ}$ . It now remains to find the rates  $R_p^*$  and  $R_r^*$ , which maximize the average MSE  $\mathcal{D}$ , following which  $R_{WZ}^*$  can be obtained from (5).

To model the relationship between the distortions  $D_p$  and  $D_r$ , and their respective H.264/AVC source encoding rates  $R_p$  and  $R_r$ , we use the encoder model of [8]. According to this model,

$$D_p = D_{0p} + \frac{\theta_p}{R_p - R_{0p}} \quad (6)$$

$$D_r = D_{0r} + \frac{\theta_r}{R_r - R_{0r}} \quad (7)$$

where  $(D_{0p}, R_{0p}, \theta_p)$  and  $(D_{0r}, R_{0r}, \theta_r)$  are parameters which can be determined from trial encodings at the encoder. When the total allowable

bit rate is  $C$ , and video packets are erased with probability  $p_e$ , then the optimal FEC scheme has

$$R_p^\dagger + \frac{p_e}{1-p_e} R_p^\dagger = C \Rightarrow R_p^\dagger = (1-p_e)C. \quad (8)$$

$$D_{FEC}^\dagger = D_{0p} + \frac{\theta_p}{R_p^\dagger - R_{0p}} \quad (9)$$

For  $p \leq p_e$ , this gives a constant low distortion  $D_{FEC}^\dagger$ . The optimal SLEP scheme is now obtained by solving the following optimization problem:

$$\begin{aligned} & \text{Maximize } R_p \\ & \text{subject to} \\ & R_p + \frac{p_e}{1-p_e} R_r \leq C \\ & \mathcal{D} = D_{FEC}^\dagger \\ & 0 \leq R_r \leq (1-p_e)C \leq R_p \leq C \end{aligned} \quad (10)$$

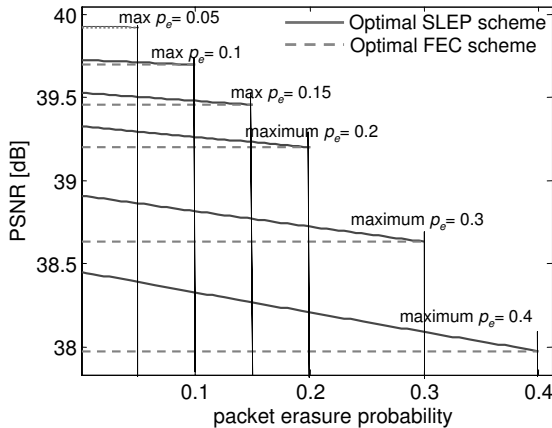
From the maximum  $R_p^*$ , we obtain  $R_r^*$  from (10) and  $R_{WZ}^*$  from (5). Fig. 6 shows a plot of the optimal FEC and SLEP schemes for several maximum packet erasure probabilities. It is clear the video quality delivered by FEC has a flat profile for  $0 \leq p \leq p_e$  while SLEP degrades gracefully in the same range. Furthermore, when  $p < p_e$ , the video quality is higher for SLEP as compared to FEC.

## 6 Conclusions

A model is derived for the end-to-end average video quality delivered by a SLEP system, implemented using H.264/AVC redundant slices in conjunction with Reed-Solomon coding. This scheme involves transmission of a Wyner-Ziv bit stream to add error robustness to a compressed video signal. The model accounts for the small distortion introduced due to the quantization mismatch from Wyner-Ziv decoding, the large distortion due to error concealment, and the effect of error propagation. The model closely approximates the observed performance of the SLEP system, which mitigates the FEC cliff effect and ensures graceful degradation of video quality. The model has been used to find the combination of the primary description bit rate, the redundant description bit rate and the Wyner-Ziv bit rate which maximizes the average received video quality at the decoder.

## Appendix: Loss from WZ decoding

We will now calculate the minimum increase in video distortion that must be tolerated by the



**Fig. 5.** When SLEP is optimized for a certain maximum probability of packet loss  $p_e$ , it ensures that the received average video quality is better than that with FEC for  $0 \leq p \leq p_e$ . The graceful degradation in SLEP is due to the quantization mismatch from Wyner-Ziv decoding. The modeling and optimization is carried out for the *Foreman CIF* sequence, with capacity  $C = 1$  Mbps.

Wyner-Ziv decoder. Clearly, to minimize the quantization mismatch between the primary and redundant descriptions, the encoding bit rate  $R_r$  of the redundant description must be as close as possible to the primary description bit rate  $R_p$ . From (5) and (10) the maximum allowable bit rate for encoding the redundant description is given by:

$$R_r = \min \left( (C - R_p) \frac{1 - p_e}{p_e}, R_p \right)$$

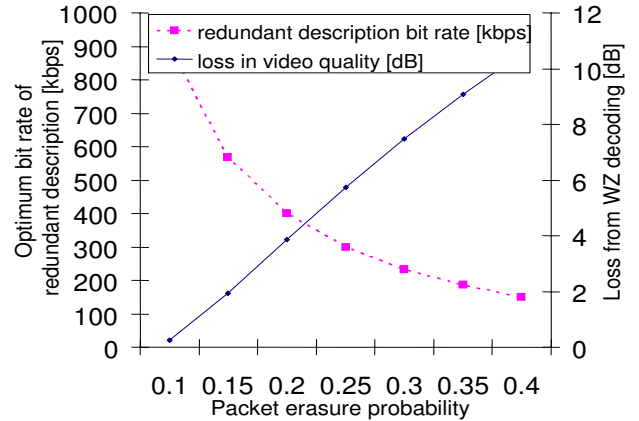
At packet erasure probability  $p_e$ , this value of  $R_r$  increases the video distortion by:

$$\Delta = p_e \frac{N + 1}{2} (D_r - D_p)$$

where  $D_r$  and  $D_p$  depend on  $R_r$  and  $R_p$  through (6) and (7). Fig. 6 plots this loss in dB, at various packet erasure rates, for  $C = 1.1$  Mbps, and  $R_p = 1$  Mbps for the *Foreman CIF* sequence. Thus, error resilience at high bit rates is achieved at the price of increased distortion from the quantization mismatch between the redundant and primary descriptions.

## References

1. Girod, B., Aaron, A., Rane, S., Rebollo-Monedero, D.: Distributed video coding. Proc. IEEE, Special Issue on Advances in Video Coding and Delivery **93** (2005) 71–83



**Fig. 6.** As the erasure probability increases, redundant descriptions encoded at a lower bit rate must be used to provide error robustness. The increased resilience is achieved at the cost of increased quantization mismatch after Wyner-Ziv decoding.

2. Rane, S., Aaron, A., Girod, B.: Systematic lossy forward error protection for error resilient digital video broadcasting - A Wyner-Ziv coding approach. In: Proc. IEEE International Conference on Image Processing, Singapore (2004)
3. Rane, S., Girod, B.: Analysis of Error-Resilient Video Transmission based on Systematic Source-Channel Coding. In: Picture Coding Symposium (PCS 2004), San Francisco, CA (2004)
4. Baccichet, P., Rane, S., Girod, B.: Systematic Lossy Error Protection using H.264/AVC Redundant Slices and Flexible Macroblock Ordering. In: Proc. IEEE Packet Video Workshop, Hangzhou, China (2006)
5. Wyner, A.D., Ziv, J.: The rate-distortion function for source coding with side information at the decoder. IEEE Transactions on Information Theory **IT-22** (1976) 1–10
6. Shamai, S., Verdú, S., Zamir, R.: Systematic lossy source/channel coding. IEEE Transactions on Information Theory **44** (1998) 564–579
7. ISO/IEC MPEG & ITU-T VCEG, J.V.T.J.: Draft ITU-T recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 — ISO/IEC 14496/10 AVC - JVT-G050r1.doc). ISO/IEC MPEG & ITU-T VCEG (2003)
8. Stuhlmüller, K., Färber, N., Link, M., Girod, B.: Analysis of video transmission over lossy channels. IEEE Journal on Selected Areas in Communications **18** (2000) 1012–1032