

SYSTEMATIC LOSSY ERROR PROTECTION  
OF VIDEO SIGNALS

A DISSERTATION  
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING  
AND  
THE COMMITTEE ON GRADUATE STUDIES  
OF STANFORD UNIVERSITY  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

Shantanu Rane  
September 2007

© Copyright by Shantanu Rane 2007  
All Rights Reserved

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

---

(Bernd Girod) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

---

(Andrea Goldsmith)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

---

(John Apostolopoulos)

Approved for the University Committee on Graduate Studies.

# Abstract

This thesis addresses the problem of error-resilient video transmission. In most video transmission applications, a video signal is compressed, packetized and transmitted over an error-prone channel. Owing to multipath fading on wireless channels and/or congestion in the Internet, some video packets are lost or arrive in error. When feedback is unavailable or limited, this problem is traditionally solved by applying a Forward Error Correction (FEC) code and transmitting a suitable amount of redundant (parity) information along with the video packets. However, the number of parity symbols are insufficient for error correction at high error probabilities, and this results in a precipitous drop in video quality, which is commonly referred to as the “cliff” effect. This thesis describes and analyzes a scheme which mitigates the FEC cliff and obtains graceful degradation of video quality at the receiver by leveraging distributed source coding ideas for error resilience.

The Systematic Lossy Error Protection (SLEP) scheme is based on the principle of systematic lossy source/channel coding. The systematic portion of the transmission consists of a compressed video signal which is sent to the decoder without channel coding. For error resilience, a supplementary bit stream generated by Wyner-Ziv encoding of the video signal is also transmitted to the receiver. The Wyner-Ziv bit stream allows the decoding of a coarsely quantized redundant video description, which can be used in lieu of the lost or error-prone portions of the systematic signal. The resulting error protection scheme is based on a flexible tradeoff between the coarseness (quality) of the redundant video description and the error robustness provided by that description.

We first perform a theoretical analysis of a simplified transmission system in which SLEP is used for robust transmission of samples generated by a first-order Markov source. The source is compressed using a DPCM-style encoder, and a Wyner-Ziv encoded version of the prediction error is transmitted to provide robustness to symbol erasures. Using high-rate quantization theory, we derive a closed-form expression for the overall rate-distortion tradeoff and study the error resilience properties of this simplified SLEP system.

Next, a practical SLEP scheme is presented, in which the Wyner-Ziv bit stream is generated by applying Reed-Solomon codes to redundantly encoded video descriptions. Under the H.264/AVC specification, this is accomplished using a standardized feature known as “redundant slices”. The error resilience of this SLEP scheme exceeds that of FEC, both, in the sense of providing graceful degradation of average video quality, and in the sense of reducing the instantaneous fluctuation in video quality owing to channel errors. Additionally, using “Flexible Macroblock Ordering” (FMO), it is possible to provide preferential Wyner-Ziv protection to a region of interest within a video frame. Further, by allocating the Wyner-Ziv bit rate among several embedded redundant video descriptions, it is possible to exploit the resilience-quality tradeoff, even when there is a large variation in the channel error probabilities with time.

Irrespective of the implementation of the SLEP system, the picture quality at the receiver of a SLEP system is determined by the rate-distortion tradeoff of the systematic transmission, the rate-distortion tradeoff of the redundant video description, and the Wyner-Ziv bit rate. We derive a model which describes the average received video quality as a function of these three quantities, and use it to study the properties of SLEP. We demonstrate that the model closely predicts the results obtained by experimental simulation. Finally, using the model, we obtain the source coding bit rates of the primary (systematic) and redundant video descriptions, and the Wyner-Ziv bit rate such that the received picture quality is maximized for a realistic Internet video transmission experiment.

# Acknowledgments

I would like to express my gratitude to Prof. Bernd Girod who advised me with patience and acute insight during my doctoral tenure at Stanford. I am thankful to Prof. Andrea Goldsmith and Dr. John Apostolopoulos for their helpful comments before, during and after the thesis defense.

Heartfelt thanks go out to all members of the Image, Video and Multimedia Systems group; they have helped me in more ways than they know. In particular, I want to thank David Rebollo-Monedero for sharing his mathematical brilliance, Anne Aaron for many intuitive insights into distributed video coding, and Pierpaolo Baccichet for generously sharing his expertise with video standards while teaching by his own example the proper way to write and debug large code.

Patrick Burke and Keith Gaul provided excellent technical support to the Stanford Center for Image Systems Engineering, and my work as a teaching assistant was greatly simplified because of them. Kelly Yilmaz was always at hand to alleviate concerns over funding documentation, travel reimbursements, software registration and equipment orders.

For four years, Tzen Ong shared an apartment and a tennis court with me, helping me negotiate the crests and troughs of graduate life. Rutu and Rucha Tamhankar offered an unlikely gift of a home away from Pune, and Pratibha Gupta kept me honest by reminding me, gently but persistently, that my dissertation must take priority over everything else.

Through the long and sometimes difficult time at Stanford, my parents and sister remained fiercely concerned about my well-being and endured much vicarious distress on my account. This thesis is dedicated to them.

*For Aai, Papa & Swati*

# Contents

<b>Abstract</b>	<b>iv</b>
<b>Acknowledgments</b>	<b>vi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Research Contributions . . . . .	3
1.2 Organization . . . . .	4
<b>2 Background</b>	<b>7</b>
2.1 Hybrid Video Compression . . . . .	9
2.2 Robust Video Transmission . . . . .	15
2.2.1 Forward Error Correction of Video Signals . . . . .	15
2.2.2 Layered Video Coding and Priority Encoding Transmission . .	17
2.2.3 Multiple Description Coding and Path Diversity . . . . .	18
2.2.4 Feedback-Based Error Control . . . . .	19
2.3 Foundations of Distributed Source Coding . . . . .	21
2.3.1 Slepian-Wolf Theorem for Lossless Distributed Coding . . . . .	22
2.3.2 Practical Slepian-Wolf Coding . . . . .	23
2.3.3 Rate-Distortion Theory for Lossy Compression with Receiver Side Information . . . . .	26
2.3.4 Practical Wyner-Ziv Coding . . . . .	27
2.3.5 Low Complexity Distributed Video Encoding . . . . .	28
2.3.6 Error-Resilient Video Compression using Distributed Source Coding . . . . .	30



2.3.7	Systematic Lossy Source/Channel Coding . . . . .	31
2.4	Summary . . . . .	33
<b>3</b>	<b>Systematic Lossy Error Protection</b>	<b>34</b>
3.1	Concept of SLEP . . . . .	34
3.2	SLEP of a First-Order Markov source . . . . .	36
3.2.1	DPCM Source Coding Scheme . . . . .	37
3.2.2	Wyner-Ziv Coding of the Prediction Residual . . . . .	38
3.3	Rate-Distortion Tradeoffs in SLEP . . . . .	40
3.4	Observations on Lossy Versus Lossless Protection . . . . .	47
3.5	Summary . . . . .	49
<b>4</b>	<b>SLEP based on H.264/AVC Redundant Slices</b>	<b>51</b>
4.1	H.264/AVC Tools . . . . .	52
4.1.1	Standard Support for Redundant Slices . . . . .	52
4.1.2	Standard Support for FMO . . . . .	53
4.2	SLEP Implementation in H.264/AVC . . . . .	54
4.2.1	Wyner-Ziv Video Encoding . . . . .	54
4.2.2	Wyner-Ziv Video Decoding . . . . .	57
4.3	Applying SLEP to a Region-of-Interest . . . . .	62
4.3.1	Rationale for ROI-based SLEP . . . . .	62
4.3.2	Determination of ROI . . . . .	64
4.3.3	Specification of ROI in Wyner-Ziv Bit Stream . . . . .	65
4.4	SLEP Experimental Results . . . . .	66
4.4.1	Video Codec Settings . . . . .	66
4.4.2	Channel Simulations . . . . .	69
4.4.3	Designation of SLEP Schemes . . . . .	71
4.4.4	Comparison of SLEP and FEC . . . . .	71
4.4.5	Effect of increasing Wyner-Ziv Bit Rate . . . . .	78
4.4.6	Benefit of ROI-Based SLEP . . . . .	80
4.4.7	SLEP with Multiple Redundant Descriptions . . . . .	81
4.5	Summary . . . . .	82

<b>5</b>	<b>SLEP Modeling and Optimization</b>	<b>83</b>
5.1	Distortion-Rate Modeling of SLEP . . . . .	84
5.1.1	Motion Compensated Encoding and Decoding . . . . .	84
5.1.2	Distortion in the Decoded Video Sequence . . . . .	87
5.1.3	Encoder Distortion-Rate Model . . . . .	91
5.2	Resilience-Quality Tradeoff in SLEP . . . . .	94
5.2.1	Overall Distortion for a Fixed Bit Rate Allocation . . . . .	94
5.2.2	Residual Distortion after Wyner-Ziv Decoding . . . . .	96
5.3	Optimization of a Practical SLEP System . . . . .	98
5.4	Summary . . . . .	102
<b>6</b>	<b>Conclusions</b>	<b>105</b>
6.1	Standardization Effort for SLEP . . . . .	107
6.2	Improvements and Extensions of SLEP . . . . .	108
<b>A</b>	<b>Stationarity Relations and Proofs</b>	<b>111</b>
<b>B</b>	<b>Embedded Redundant Descriptions</b>	<b>115</b>
B.1	Embedded Wyner-Ziv Codec . . . . .	116
B.1.1	Unequal Error Protection for I, P and B Slices . . . . .	117
B.1.2	Embedded Redundant Descriptions . . . . .	117
B.2	Experimental Results . . . . .	118
B.2.1	Unequal Protection for I, P, and B Slices . . . . .	119
B.2.2	Embedded Wyner-Ziv Coding . . . . .	119
	<b>Bibliography</b>	<b>122</b>

# List of Tables

4.1	Video sequences used in the SLEP simulations. CIF sequences have a size of $352 \times 288$ pixels, while SIF sequences have a size of $352 \times 240$ pixels. . . . .	68
-----	---	----

# List of Figures

2.1	A video encoder consists of a motion-compensated predictive coding loop and performs transform coding of the quantized prediction residual. The encoder contains a local copy of the decoder (shown in grey), and uses the locally decoded frames for motion estimation. . . . .	10
2.2	Distributed compression of two statistically dependent random processes, $X$ and $Y$ . The decoder jointly decodes $X$ and $Y$ and thus may exploit their mutual dependence. . . . .	21
2.3	Slepian-Wolf Theorem: Achievable rate region for distributed compression of two statistically dependent i.i.d. sources $X$ and $Y$ [158]. . . .	22
2.4	Compression of a sequence of random symbols $X$ using statistically related side information $Y$ . We are interested in the distributed case, where $Y$ is only available at the decoder, but not at the encoder. . . .	23
2.5	In the coset interpretation, the Slepian-Wolf decoder chooses that codeword in the coset of interest which is most likely given the value of the side information. . . . .	24
2.6	Lossy compression of a sequence $X$ using statistically related side information $Y$ . . . . .	26
2.7	A practical Wyner-Ziv coder is obtained by cascading a quantizer and a Slepian-Wolf encoder. . . . .	27
2.8	Digitally enhanced analog transmission. . . . .	32

3.1	A video transmission system in which the video waveform is protected by a Wyner-Ziv bit stream, in a systematic source/channel coding configuration. The receiver decodes the Wyner-Ziv bit stream using the error-prone received video signal as side information. . . . .	35
3.2	Systematic lossy error protection applied to the prediction residual signal of a DPCM coding scheme. . . . .	37
3.3	Residual distortion after Wyner-Ziv decoding increases when the erasure probability increases. . . . .	45
3.4	Residual distortion after Wyner-Ziv decoding increases as the prediction coefficient $\rho$ approaches unity. . . . .	46
3.5	Residual distortion after Wyner-Ziv decoding increases when the step-size of the Wyner-Ziv quantizer is increased, owing to the greater quantization mismatch between the DPCM quantizer and the Wyner-Ziv quantizer. . . . .	47
3.6	The end-to-end distortion $\mathcal{D}$ is evaluated for the case where source data $X$ are generated by a first-order Gauss-Markov process with $\rho = 0.75$ and $\sigma_W^2 = 5$ . For a fixed error resilience bit rate, SLEP provides graceful quality degradation over a wider range of erasure probabilities than FEC. . . . .	48
3.7	The end-to-end distortion $\mathcal{D}$ is evaluated for the case where source data $X$ are generated by a first-order Gauss-Markov process with $\rho = 0.75$ and $\sigma_W^2 = 5$ . If the maximum erasure probability is fixed, then SLEP allocates a larger fraction of the total bit rate $\mathcal{R}$ to source coding, incurring less distortion than FEC in the erasure-free case. . . . .	50
4.1	FMO Type 2 or <i>Foreground with Leftover</i> allows the encoder to perform segmentation of the frame to be encoded, and enables unequal protection strategies for the different segments. . . . .	54

4.2	Implementation of a SLEP system using H.264/AVC redundant slices and FMO-based region-of-interest determination. 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. . . . .	55
4.3	During Wyner-Ziv encoding, RS codes are applied across the redundant slices and only the parity slices are transmitted to the decoder. To each parity slice is appended helper information about the quantization parameter (QP) used in the redundant slices, and the shapes of the redundant slices. The parity slices, together with the helper information, constitute the Wyner-Ziv bit stream. . . . .	57
4.4	During Wyner-Ziv decoding, redundant slices corresponding to received primary slices are obtained by requantization, while those corresponding to the lost primary slices are treated as erasures. These are recovered by erasure decoding, using the parity slices and helper information received in the Wyner-Ziv bit stream. These recovered redundant slices are then decoded and displayed in lieu of the lost primary slices. . . . .	58
4.5	A redundant slice can be transmitted along with each primary slice, and can be decoded if the primary slice is lost. The residual distortion depends upon the difference in the quantization step sizes used in the redundant and primary slices. . . . .	60
4.6	Instead of transmitting the redundant slices, it is efficient to transmit parity symbols which enable the receiver to recover the redundant slices. The prediction structure is <i>I-B-P-B-P</i> and more parity slices are generated for the intra (I) frames than for the predictively encoded (P and B) frames. Since the parity bit rate is approximately constant at 40 kb/s, coarsely quantized redundant descriptions have stronger error protection. . . . .	61

4.7	Using a redundant description with QP=40 provides the best overall tradeoff between the quantization mismatch and error resilience. Lower QPs result in worse error resilience. Higher QPs result in high quantization mismatch. . . . .	61
4.8	Cut-outs of a video frame from the <i>Foreman</i> CIF sequence. The primary description is encoded at 408 kb/s, while the error resilience bit rate is fixed at 40 kb/s. Robustness increases with the quantization step in the redundant slice (See fig. 4.2). However, with QP=48, the increased quantization mismatch reduces the decoded picture quality.	63
4.9	A Region Of Interest (ROI) is determined at the encoder by finding the mean absolute error between the current image and its locally error concealed version. The error image is then thresholded to obtain the ROI. . . . .	64
4.10	A SLEP scheme is designated by two numbers. The first number, <b>R</b> , expresses the encoding bit rate of the redundant slices as a percentage of the bit rate of the primary video signal. The second number, <b>W</b> , expresses the Wyner-Ziv bit rate as a percentage of the bit rate of the primary video signal. The redundant description is not transmitted, but reconstructed at the decoder. . . . .	72
4.11	Comparison of FEC with SLEP schemes in which the redundant slices are encoded at 50 % and 25 % of the bit rate of the primary slices. The error resilience bit rate for all schemes, except decoder based error concealment, marked “EC” above, is 10 % of the source coding bit rate of the primary slices. When a coarsely quantized redundant description is used, the error robustness increases at the expense of an increased quantization mismatch between the primary and redundant descriptions at the decoder. . . . .	75

4.12	When coarse quantization is used in the redundant description, there is a small reduction in the decoded frame PSNR compared to the error-free case. In return, drastic reduction in picture quality is avoided. At a high packet loss rate of 10%, SLEP-25-10 provides the smallest instantaneous fluctuation in frame PSNR, followed by SLEP-50-10 followed by FEC. . . . .	76
4.13	Decoded frames of the <i>Bus</i> CIF sequence encoded at 1024 kb/s, when the parity bit rate is 10% of the primary source coding bit rate. With FEC (left), decoding fails for some portions of the frame, reducing the PSNR to 20.2 dB. With SLEP scheme for a redundant description encoded at 25% of the primary bit rate (right), successful Wyner-Ziv decoding results in a PSNR of 30.3 dB, much closer to the error-free PSNR of 32.2 dB. . . . .	77
4.14	Decoded frames of the <i>Mobile</i> SIF sequence encoded at 768 kb/s, when the parity bit rate is 10% of the primary source coding bit rate. With FEC (left), decoding fails for some portions of the frame, reducing the PSNR to 17.3 dB. With SLEP scheme for a redundant description encoded at 25% of the primary bit rate (right), successful Wyner-Ziv decoding results in a PSNR of 25.3 dB, close to the error-free PSNR of 25.9 dB. . . . .	77
4.15	Decoded frames of the <i>Coastguard</i> CIF sequence encoded at 512 kb/s, when the parity bit rate is 10% of the primary source coding bit rate. With FEC (left), decoding fails for some portions of the frame, reducing the PSNR to 22.7 dB. With SLEP scheme for a redundant description encoded at 25% of the primary bit rate (right), successful Wyner-Ziv decoding results in a PSNR of 31.2 dB, much closer to the error-free PSNR of 32.9 dB. . . . .	78
4.16	With a fixed redundant description, increasing the Wyner-Ziv bit rate results in an increase in error resilience. FEC with the same parity bit rate is displayed for comparison. . . . .	79



4.17	Decoded frames of the <i>Akiyo</i> CIF sequence encoded at 200 kb/s, when the parity bit rate is 10% of the primary source coding bit rate, and the redundant slices in the Wyner-Ziv codec are encoded at 40 kb/s. When SLEP is applied to the entire picture, there are smearing artifacts when intra-coded macroblocks must be replaced by their redundant versions. If the same bit rate is concentrated inside the ROI, then the picture quality after Wyner-Ziv decoding does not suffer from smearing artifacts.	80
4.18	Applying SLEP to the ROI results in superior decoded picture quality because it allows finer quantization in the ROI. Further, this usually results in fewer redundant slices, and stronger Wyner-Ziv protection as for the case of frame no. 72 from the trace shown in Fig. 4.17.	81
5.1	A few trial encodings (data points) are used to find the parametric rate-distortion curves for the primary and redundant descriptions of the <i>Foreman</i> CIF sequence. The parameters for the redundant description depend upon the primary description used as reference.	94
5.2	The end-to-end average PSNR calculated by the model (solid lines) of Section 5.1.1 closely approximates that obtained by experimental simulation (data points). The Wyner-Ziv bit rate, i.e., the bit rate of the parity slices generated by the Reed-Solomon Slepian-Wolf encoder is fixed at 10% percent of bit rate of the primary slices. For both modeling and simulation, the average PSNR on the vertical axis is calculated from the average MSE of the sequence.	95
5.3	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.	99

5.4	The model derived in Section 5.1.1 is used at the encoder to choose the primary video coding bit rate, the bit rate of the redundant description, and the Wyner-Ziv bit rate (equivalently, the strength of the Reed-Solomon Slepian-Wolf code). When compared with a fixed <i>a priori</i> assignment of bit rates, the optimized scheme provides superior average picture quality over all erasure probabilities. . . . .	100
5.5	When the erasure probability is known <i>a priori</i> , an optimized SLEP scheme and an optimized FEC scheme provide approximately the same video quality. Recall however, that when the erasure probability changes for a given bit allocation, SLEP provides graceful degradation compared to FEC, as plotted in Chapter 4, Fig. 4.11. . . . .	103
A.1	Embedded quantization (successive degradation) of $W$ with $m = \Delta_2/\Delta_1 = 7$ . Embedding increases the MSE by a factor of $(m^2 - 1)$ . . . . .	112
B.1	The Wyner-Ziv decoder uses a decoded error-concealed video waveform as side information in a systematic lossy source/channel coding setup. With an embedded Wyner-Ziv codec, graceful degradation of video quality is obtained without a layered video representation in the systematic transmission. . . . .	115
B.2	Implementation of systematic lossy error protection by combining MPEG coding and Reed-Solomon codes across slices. In the Wyner-Ziv encoder, multiple redundant descriptions are generated by embedded quantization. . . . .	116
B.3	Error resilience improves when unequal Wyner-Ziv protection is assigned to the I, P, B frames in a single redundant description. The transmitted Wyner-Ziv bit rate is 222 kb/s for each curve. . . . .	120
B.4	To achieve graceful degradation of video quality, a coarse redundant description encoded at 500 kb/s is embedded inside a finer redundant description encoded at 1 Mb/s. The available error resilience bit rate of 222 kb/s is then shared among the two descriptions. . . . .	121

B.5 At an error probability of  $2 \times 10^{-4}$ , SLEP with a finely quantized redundant description alone cannot provide adequate robustness to signal loss, while SLEP with the coarsely quantized redundant description alone incurs more distortion due to coarse quantization. With the bit rates apportioned as in Fig. B.4, the frame PSNR is at least as high as that achieved by the coarse redundant description. . . . . 121

# List of Abbreviations

Abbreviation	Full Form
ACK packet	Acknowledgment packet
AVC	Advanced Video Coding
CABAC	Context Adaptive Binary Arithmetic Coding
CAVLC	Context Adaptive Variable Length Coding
CBR	Constant Bit Rate
CIF	Common Interchange Format
CoDiO	Congestion-Distortion Optimized (Streaming)
DCT	Discrete Cosine Transform
DPCM	Differential Pulse Code Modulation
FEC	Forward Error Correction
FMO	Flexible Macroblock Ordering
GF	Galois Field
GOP	Group Of Pictures
IEC	International Electrotechnical Commission
i.i.d	Independent and Identically Distributed
IP	Internet Protocol
ISI	Inter-Symbol Interference
ISO	International Standards Organization
ITU-T	International Telecommunication Union - Telecommunication Standardization Sector
JPEG	Joint Photographic Experts Group
JVT	Joint Video Team
JM 11	Joint Test Model Version 11 (Video Compression Source Code)
LA-RDO	Loss-Aware Rate-Distortion Optimization
LCUEP	Layered Coding with Unequal Error Protection
LDPC code	Low Density Parity Check code

Abbreviation	Full Form
MPEG	Motion Picture Experts Group
MAE	Mean Absolute Error
MMSE	Minimum Mean Squared Error
MSE	Mean Squared Error
NACK packet	Negative Acknowledgment packet
PCP	Primary Coded Picture
PPS	Picture Parameter Set
PSNR	Peak Signal to Noise Ratio
QP	Quantization Parameter
RaDiO	Rate-Distortion Optimized (Streaming)
ROI	Region Of Interest
RPS	Reference Picture Selection
RS code	Reed-Solomon code
RTP	Real Time Protocol
SEI	Supplemental Enhancement Information
SIF	Standard Interchange Format
SLEP	Systematic Lossy Error Protection
SNR	Signal to Noise Ratio
SVC	Scalable Video Coding
UDP	Universal Datagram Protocol

# Chapter 1

## Introduction

The generation and consumption of video content have increased many fold over the past decade. Users all over the world access video content on diverse platforms primarily for entertainment, information and education. Broadcast television is now making the transition from analog to digital. The wired Internet has spawned a large number of video applications, from streaming of pre-stored or live content to interactive applications such as gaming and video conferencing. Recently, with the development of low-power video codec chipsets, it is possible to access video on mobile devices such as cellphones and PDAs.

In the vast majority of applications, the raw size of the video data is too large compared to the bandwidth available for the transmission. Therefore, video compression is essential and much effort has been invested in improving the rate-distortion performance of video compression algorithms. The central feature of video compression algorithms is motion-compensated predictive coding, which encodes the difference between the actual pixel in a video frame and a predicted value which is calculated using motion information. While the specific tools for transform coding, entropy coding and motion estimation have evolved over the years, motion compensated predictive coding has been retained in all the compression algorithms that have been standardized, viz., H.261, H.263, MPEG-1,2,4, and most recently H.264/AVC.

Though motion-compensated predictive coding provides impressive coding gains, it sacrifices robustness to transmission errors, since errors occurring in one frame

can now propagate to the subsequent frames because of the differential nature of the encoding. Indeed, as encoders become more and more efficient at removing the redundancy from the video signal, the potential impact of channel losses and error propagation becomes more and more severe. The newer video codecs have incorporated some tools to mitigate error propagation, most notably: Intra coded (I) frames, video slices, data partitioning and loss-aware rate-distortion optimization. However, in order to preserve compression efficiency, these tools must be used sparingly. In most video applications, a selection of these coding tools is combined with an error resilience scheme in order to ensure robust video delivery across error-prone channels.

Error-resilient video coding schemes can be broadly classified into methods which employ retransmissions, and methods which perform Forward Error Correction (FEC). By means of feedback from the receiver to the sender, error-prone or lost packets can be retransmitted so long as the delay associated with feedback is permissible. If the delay from the feedback-based schemes is prohibitive, forward error correction may be performed, in which a portion of the total transmitted bit rate is dedicated to parity information which enables the decoder to correct transmission errors in the received bit stream. FEC ensures acceptable picture quality as long as the number of channel errors is below the error correction capability of the code. However, when the error probability increases suddenly, for example, when the wireless channel is in a deep fade, the number of symbol errors in a block code is too high and error correction decoding fails. In this case, the decoder is compelled to use some local error concealment techniques to conceal the lost portions of the video signal. Decoder-based error concealment is seldom perfect and leaves concealment artifacts which considerably degrade the picture quality. This rapid reduction in picture quality due to the failure of FEC is termed as the *cliff effect*.

This thesis presents a new scheme for error-resilient video transmission, known as Systematic Lossy Error Protection (abbreviated as “SLEP” throughout the text). To protect the waveform of a compressed video signal, SLEP transmits an additional bit stream that is generated by a process known as *Wyner-Ziv coding*. In the distributed source coding literature, Wyner-Ziv coding refers to lossy compression of a signal with the help of side-information that is present at the decoder only. SLEP uses Wyner-Ziv

coding within a joint source-channel coding framework, in which an error-prone video signal furnishes side-information for Wyner-Ziv decoding. The robustness properties of SLEP are evaluated by theoretical analysis and by performing experimental simulation with the H.264/AVC and MPEG-2 video codecs, with synthetic and actual realizations of error-prone channels.

## 1.1 Research Contributions

The major contributions of this work are summarized below:

- The concept of Systematic Lossy Error Protection (SLEP) is proposed for error-resilient transmission of video signals. The systematic portion of the transmission consists of a compressed video signal which is sent to the decoder without channel coding. For error resilience, a supplementary bit stream generated by Wyner-Ziv encoding of the video signal is also transmitted to the receiver. The Wyner-Ziv bit stream allows the decoding of a coarsely quantized redundant video description which can be used in lieu of the lost or error-prone portions of the systematic signal. The resulting error protection scheme is based on a flexible tradeoff between the coarseness (quality) of the redundant video description and the error robustness provided by that redundant description.
- The rate-distortion tradeoffs involved in the design of a SLEP system are studied using high rate quantization theory. A simple scheme is considered in which SLEP is applied to a predictively encoded first-order Markov source. The distortion in the received signal is expressed in closed form, as a function of the error probability of the channel, the quality of the systematic and Wyner-Ziv descriptions, and the temporal correlation in the source. The derived rate-distortion tradeoffs are used to study the properties of lossy error protection, viz., graceful degradation of signal quality with increasing error probability, and superior robustness compared to traditional “lossless” correction methods such as FEC.



- A SLEP scheme based on H.264/AVC redundant slices is implemented, and experimental simulations of packetized video transmission are conducted to demonstrate the superiority of this error resilience scheme to traditional methods such as FEC. Using Flexible Macroblock Ordering (FMO), a feature supported by the H.264/AVC video coding standard, SLEP is applied to a region-of-interest within a video signal and performance enhancements are reported for low-motion sequences.
- An embedded SLEP scheme is proposed to provide error robustness even when there are large variations in the probability of transmission errors. In this scheme, the available error resilience bit budget is unequally allocated among several embedded redundant representations of the video signal. For MPEG-2 broadcast applications, the superiority of this scheme over FEC-based systems is demonstrated.
- We derive a model which expresses the average received video quality as a function of the rate-distortion functions of the systematic and redundant descriptions and the Wyner-Ziv bit rate. The model closely predicts the results obtained by experimental simulation of a SLEP system, and is used to explain and quantify the design trade-offs in the SLEP scheme. Finally, the model is used to find the optimum rate allocation between the primary video signal, redundant description and the Wyner-Ziv protection in order to maximize the received picture quality.

## 1.2 Organization

This thesis is organized as follows:

- In Chapter 2, we review the three intersecting areas connected to the present work. Section 2.1 reviews the state-of-the-art in hybrid video compression schemes. The intent is to provide a self-contained explanation of the basic structure of a video coding scheme, and to highlight the error resilience tools available in modern video codecs. Section 2.2 contains a discussion of schemes that

have been proposed for robust transmission of compressed video bit streams. These are classified into Forward Error Correction (FEC) schemes, layered video coding with unequal error protection, and feedback-based error control. In Section 2.3, the area of distributed source coding is reviewed. In particular, lossy and lossless compression, (respectively termed as Slepian-Wolf coding and Wyner-Ziv coding after the seminal researchers in this area) in the presence of side information at the decoder only. An overview is provided of the advances in the emerging field of distributed video coding. More pertinent to this work, the information-theoretic framework of systematic lossy source/channel coding is described.

- In Chapter 3, we present a closed-form mathematical analysis of a very simple SLEP system. Section 3.1 introduces the concept of SLEP, essentially explaining how systematic source/channel coding can be used to protect the waveform of a compressed video signal. Section 3.2 considers the transmission of a first-order Markov source which is compressed using a DPCM-style encoder and protected by additionally transmitting a Wyner-Ziv coded representation of the source. The end-to-end rate-distortion tradeoffs for this simple SLEP system are derived in Section 3.3. In Section 3.4, the derived tradeoffs are used to study the beneficial properties of SLEP, namely, the capability to provide increased robustness and achieve graceful quality degradation compared to conventional FEC-based systems.
- In Chapter 4, we present an implementation of a SLEP system for robust transmission of compressed video. Section 4.1 describes in detail the main tools from the H.264/AVC video coding standard which have been used in the implementation. In Section 4.2, a Wyner-Ziv codec is constructed by applying Reed-Solomon codes across redundant video descriptions. SLEP encoding and decoding operations are described in a step-by-step manner. Section 4.3 presents an enhancement of the SLEP scheme in which error protection is applied to a Region-Of-Interest (ROI) within a video frame. This is accomplished by means of a standard-compliant tool known as Flexible Macroblock Ordering (FMO).

Section 4.4 describes the results of experimental simulations which investigate the behavior of the video quality delivered by a SLEP system when the channel erases packets during transmission. The average video quality and instantaneous fluctuations in the quality delivered by SLEP are examined against those delivered by FEC and by decoder-based error concealment.

- In Chapter 5, we present model-based optimization of a SLEP system implemented using a standard video codec, such as the one described in Chapter 4. Section 5.1 describes a model for the end-to-end MSE distortion of the video signal, which depends upon the distortion-rate tradeoff of the original video signal, the distortion-rate tradeoff of the redundant video description, the Wyner-Ziv coding scheme and the probability with which video packets are lost during transmission. Section 5.2.2 discusses the tradeoff between the error resilience provided by a redundant description and the degradation in picture quality that results from using the redundant signal in lieu of the lost primary signal. In Section 5.3, the model is used for optimal bit rate selection in a video simulation, i.e., model parameters are estimated over a specified time window and the bit rates of the main video description, the redundant description and the strength of the Reed-Solomon code used in the Wyner-Ziv codec are determined such that the average distortion over that window is minimized.

Note: For ready reference, a list of abbreviations has been compiled before the beginning of this chapter (pages xx and xxi after the Table of Contents).

# Chapter 2

## Background

The objective of a practical video coder is to compress the video signal to a bit rate which is as close as possible to its theoretical rate-distortion bound. On the other hand, a channel coding scheme inserts redundant information in the form of parity symbols to enable the receiver to recover portions of the video signal that are lost or corrupted. According to Shannon's channel coding theorem [156], so long as the total transmitted bit rate is below the information theoretic capacity of the channel, there exists a channel encoder and decoder that can reproduce the transmitted signal at the distortion prescribed by the source coder.

The source and channel coders have conflicting aims: the former tries to remove redundancy while the latter tries to introduce it. Further, the data compression algorithm is independent of the channel characteristics while the channel coding algorithm is independent of the source distribution [41]. The fact that we can still design an optimal communication system by cascading a source coder with an independently designed channel coder is a consequence of Shannon's source/channel separation theorem [156]. This has enabled source coding researchers to design efficient compression algorithms for different types of signals, viz., data, speech, audio, images, and video, while at the same time channel coding experts have designed efficient error correcting codes for robust transmission across different types of channels. However, Shannon's separation theorem holds only in the limit of very long codeword lengths, which translates to a very high decoding delay. Moreover, it is applicable only if the statistics of

the channel are known. This means that it is relevant only for point-to-point communication scenarios [41] but not for multi-user and broadcast applications, the domain in which many video coding schemes reside. For this reason, joint source/channel coding is relevant in the context of video transmission, and extensive research has been performed in this area [183, 22, 105, 184]. After a suitable pair of source and channel coders have been chosen for a particular application, a practical way of applying joint source/channel coding is to perform joint parameter optimization for the source coder-channel coder pair [163].

In Section 2.1, we review the state of the art in video compression. The essential components of a video compression algorithm are described. Emphasis is placed on H.264/AVC, which provides the most powerful suite of video compression tools today. We discuss error resilience tools available within the standard, such as redundant slices and flexible macroblock ordering (FMO), which have been used in this research.

In Section 2.2, we briefly discuss robust video transmission. Schemes fall into two major categories: Forward Error Correction (FEC) and priority encoding transmission in conjunction with layered video coding. In addition, we review multiple description coding and feedback-based strategies for error control.

This thesis introduces and studies the performance of Systematic Lossy Error Protection (SLEP), a new scheme for robust video transmission. In a departure from previous error correction schemes, SLEP abandons the quest for perfect error correction of video bit streams. Instead, it performs imperfect, i.e., *lossy* error protection of the *waveform* of the coded video signal. In the subsequent chapters of this thesis, it will be shown that, in return for a small and usually imperceptible loss in the quality of the corrected picture, SLEP offers improved error resilience properties compared to traditional methods such as FEC. SLEP leverages information-theoretic results from distributed source coding, which refers to compression of a signal in the presence of side information that is available to the decoder, but not to the encoder. We prepare the foundation for understanding the concept of SLEP by reviewing the area of distributed source coding in Section 2.3. This covers information-theoretic results, practical implementation of codes for distributed source coding, advances in distributed

video coding and the role of distributed source coding within the information-theoretic framework of systematic lossy source/channel coding.

## 2.1 Hybrid Video Compression

State-of-the-art video codecs achieve high compression ratios by using appropriate combinations of a very large number of tools. For a comprehensive treatment of waveform-based video coding, please refer to [182]. In the following discussion, we focus only on those tools which are necessary to provide a self-contained explanation and which are relevant to this work:

- **DPCM and Motion Compensation:** The basic configuration of a video encoder and decoder is shown in Fig. 2.1. The encoder uses the principle of Differential Pulse Code Modulation (DPCM) [80], and first obtains the difference between the current frame and a predicted version of the current frame. This is advantageous and efficient owing to the high temporal correlation found in video data. Video compression is thus a predictive coding scheme in which the goal is to obtain a predictor that leads to a residual having the minimum entropy<sup>1</sup>. Rather than directly using the previous frame (or a weighted combination of previous frames) as a predictor for the current frame, a video encoder obtains a more accurate prediction using motion estimation. This reduces the energy in the prediction residual, resulting in dramatic improvements in the rate-distortion performance versus simple intraframe encoding [63]. Specifically, motion information is extracted from the current frame and one or more previous frames, and a prediction is formed by warping the previous frame(s) according to the motion information. The most common technique for motion estimation is to perform block matching. For integer-pel motion estimation, a search is performed in the reference frame for an  $M \times N$  pixel block which is closest in mean squared error (or mean absolute error) to the  $M \times N$  block being predicted in the current frame. A 2-dimensional motion vector specifies its horizontal and vertical displacement from the position of the block

---

<sup>1</sup>This sense of defining what constitutes the “best” predictor was formalized by Elias [46, 47]

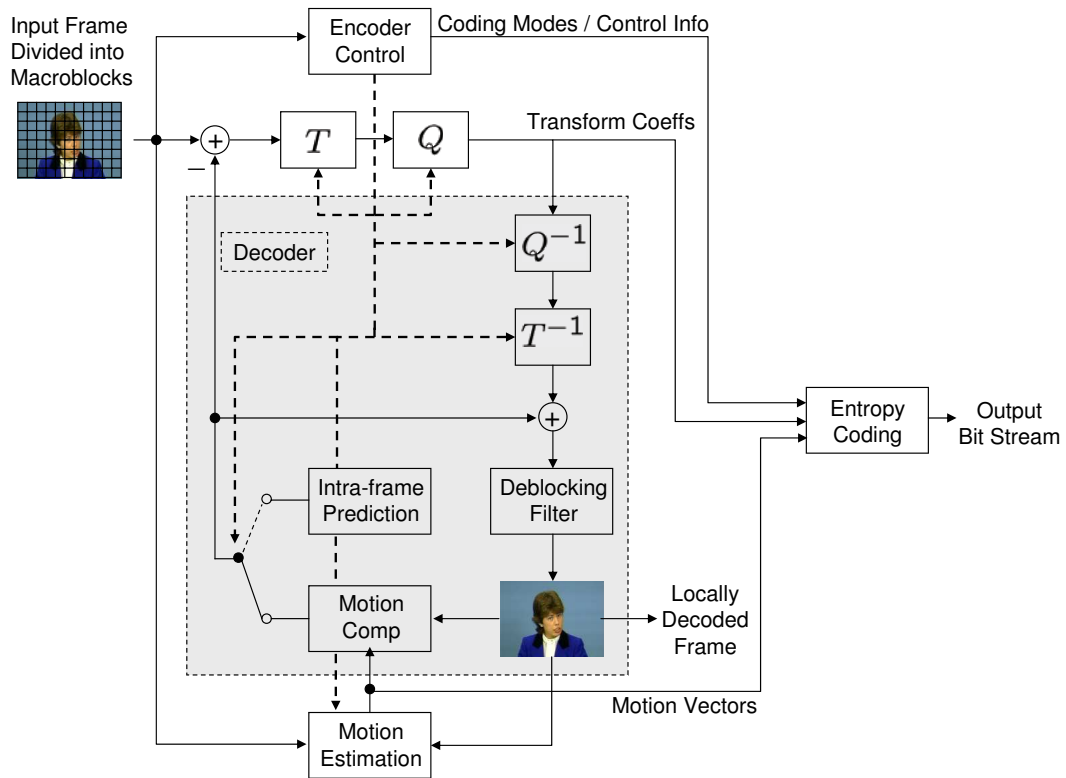


Figure 2.1: A video encoder consists of a motion-compensated predictive coding loop and performs transform coding of the quantized prediction residual. The encoder contains a local copy of the decoder (shown in grey), and uses the locally decoded frames for motion estimation.

being predicted. Prediction performed in this way, along the direction of the motion, is called motion-compensated prediction. State-of-the-art codecs like H.264/AVC provide for fractional-pel motion compensation [64], variable block sizes [36, 193, 167], multi-hypothesis motion-compensated prediction [50, 96], overlapped block motion compensation (OBMC) [109], and other tools to maximize the benefits from motion-compensated prediction.

According to the above procedure, motion estimation may result in the finding of a single  $16 \times 16$  predictor block, or a number of  $16 \times 16$  blocks which can be blended into a single predictor block. Alternatively, the macroblock in the current frame may be split (spatially) into smaller parts, with predictors and motion vectors being calculated for each part. The encoder adopts the motion

compensation strategy which incurs the lowest distortion for a given cost of encoding the motion vectors and the prediction residual.

- **Transform Coding of the Prediction Residual:** The residual error from motion-compensated prediction is transformed using a 2-dimensional spatial block transform, such as a discrete cosine transform [13]. As in conventional JPEG image compression, this compacts the energy of the prediction residual signal into a few spatial low-frequency coefficients [67]. The high-frequency coefficients contain a lower proportion of the signal's energy, and can be neglected if bit rate savings are desired. Unlike previous standards such as H.263 [108] and MPEG-2 [78], which used an  $8 \times 8$  2-D DCT, H.264/AVC uses a  $4 \times 4$  2-D integer transform whose coefficients approximate those of a  $4 \times 4$  DCT [97, 198]. The integer transform is marginally less efficient in its energy compaction properties compared to the DCT, but has the advantage of being perfectly reversible.
- **Quantization and Rate Control:** The transformed prediction residual is quantized by a scalar quantizer, i.e., the frequency components are scanned in a zigzag fashion, and each component is quantized separately by a scalar uniform quantizer and entropy coding is performed on the quantization indices. In most video codecs, the quantization step-size is specified by the quantization parameter(QP). In H.264/AVC, QP is an integer from 1 to 51, such that there is an increase of 12.5% in the quantizer step-size for every unit increase in QP [188]. The choice of the quantization parameter can be driven by a rate control algorithm, which estimates the number of bits required to losslessly encode the quantization indices, motion vectors and macroblock mode decisions and compares this estimate to a bit budget that has been prescribed by the user or the application. Please refer to [189] for a survey of rate-distortion optimization schemes used in video codecs.
- **Entropy Coding:** Early entropy coders used variable length coding (VLC) or arithmetic coding if additional complexity is allowed. H.264/AVC offers the option of using different VLC tables or different arithmetic codes for different contexts. For the syntax element being encoded, e.g., a motion vector, a context



is a subset of previously encoded syntax elements. Based on past encoded values, the probability distribution of the syntax element is updated and a suitable entropy code is chosen for the current syntax element. In H.264/AVC, these methods are called Context Adaptive VLC (CAVLC) [107, 190] and Context Adaptive Binary Arithmetic Coding (CABAC) [98] respectively. Since the VLC table or the probability distribution of the syntax elements is tailored to the context, these methods provide superior compression performance.

As seen in Fig. 2.1, the decoder reverses the quantization and transform operations, and adds the motion-compensated prediction to the recovered prediction residual, thus obtaining the decoded version of the current frame. Note that the encoder also performs this process, and houses a locally decoded version of the video frame. The locally decoded frames reside in a buffer at the encoder, and are used for motion compensation while encoding the future frames. In the absence of transmission errors, prediction mismatch between encoder and decoder is avoided, since both use exactly the same reference frames for motion-compensated prediction.

- **Error Resilience Tools:** In the presence of transmission errors, the loss of the prediction residual results in an error in the decoded frame, and this error propagates due to motion compensation. Video codecs provide a number of methods to prevent or mitigate the effect of channel errors, a small subset of which we shall discuss here.

1. *Intra Macroblock Refresh:* Most macroblocks in a video sequence are encoded from the past frames or from a set of past and future frames. However, to mitigate error propagation due to the loss of the prediction residual, some macroblocks are encoded independently. Periodic or pseudo-random insertion of intra macroblocks results in a reduction in compression efficiency, but improves the error resilience [72].

2. *Loss-Aware Rate-Distortion Optimization*: When a video encoder operates in a rate-distortion optimized mode, it selects the coding modes (I, P, or B<sup>2</sup>) and motion vectors which result in a minimization of a cost function  $J = D + \lambda R$ , where  $D$  is the distortion that must be incurred while encoding the prediction residual, coding modes and motion vectors at rate  $R$ . In Loss-Aware Rate-Distortion Optimization (LA-RDO), the mode decisions also take into account the severity of the error propagation if the macroblock under consideration were to be lost [162]. Thus, when the expected packet loss probability is high, LA-RDO forces a larger percentage of macroblocks to be coded in the intra (I) mode, resulting in reduced error propagation at the decoder.
3. *Redundant Slices*: A redundant slice is an alternative encoded description of a video slice<sup>3</sup>. In this work, we will refer to the originally encoded slice (picture) as the primary coded slice (picture). Redundant slices result in an expansion of the transmitted bit rate but are useful if the primary slices are lost. Much flexibility is allowed as regards the encoding mode and quality of the redundant slices, and we exploit this feature in our error resilience scheme. Systematic Lossy Error Protection (SLEP) uses redundant slices to perform Wyner-Ziv video coding, and the resulting Wyner-Ziv bit stream is used for error resilience. The details of the use of redundant slices in an implementation of a Wyner-Ziv codec will be presented in Chapter 4.
4. *Flexible Macroblock Ordering*: In earlier video coding standards a video slice had to consist of a collection of consecutive macroblocks, for e.g.,

---

<sup>2</sup>I: Intra coding involves encoding the samples from the current frame either without using prediction, or using predictions from the current frame. P: Predictive coding involves the use of prediction from temporally previous reference frame, B: Bi-predictive coding involves the use of predictions from two (or more) reference frames. In H.264/AVC, B frames may use multiple predictions from a single reference frame

<sup>3</sup>A video slice is defined as a collection of macroblocks. It is the smallest independently decodable unit of a video bit stream, in the sense that one slice can be decoded independently of others. In H.264/AVC, a video frame may consist of a number of slices, or may fit entirely inside a single slice. The standard does not allow a slice to contain parts of two different frames.

MPEG-2 requires a slice to consist of a row of macroblocks. Flexible Macroblock Ordering (FMO) relaxes this restriction in H.264/AVC, allowing various macroblock-to-slice mapping functions to suit the application. For example, decoder-based concealment of a video slice is significantly improved if the macroblocks are ordered in a checkerboard pattern, which ensures that horizontally and vertically adjacent macroblocks are never placed in the same slice. Alternatively, a (possibly discontinuous) region of interest within the video frame may be encoded as a separate slice, and stronger error protection can be applied to this slice in comparison to other less important slices [71]. We exploit the latter scheme by applying SLEP to a region-of-interest, as will be explained in Chapter 4.

5. *Decoder-Based Error Concealment*: Even after using retransmissions (where permissible) or error-correction information, it is possible that a video packet is lost or arrives after its decoding deadline. In such cases, the video decoder must have a post-processing strategy to conceal the lost portions of the video frame. The objective of a decoder-based error concealment scheme is to mitigate the degradation in visual quality caused by errors or losses in the current video frame and to mitigate the error propagation that results from motion-compensated temporal prediction. Such schemes can be as simple as previous-frame error concealment, but are usually much more sophisticated depending upon the computational complexity that can be tolerated at the decoder [184]. A wide array of error concealment schemes have been proposed which use, for example, estimation of coding modes and motion vectors [24, 123, 85], spatial interpolation [75, 14], projections onto convex sets [168], and Bayesian estimation of missing macroblocks [145], to name just a few. In this work, we utilize the non-normative error concealment scheme provided in H.264/AVC decoders [180]. This scheme uses weighted pixel value averaging for concealing lost intra-coded macroblocks. To conceal lost predictively encoded macroblocks, motion vectors of the neighboring available macroblocks are interpolated to obtain a candidate vector for motion compensation.

## 2.2 Robust Video Transmission

Having described the salient features of video compression algorithms, we now consider schemes for robust transmission of compressed video over lossy channels. Compressed video data is placed into packets prior to transmission, and hence the effect of channel errors is that a video packet is either corrupted or lost. It is assumed that detection mechanisms are in place to ascertain whether the received video packets are lost, corrupted or received correctly. Packet corruption can occur, for example, due to intersymbol interference (ISI) [119] when video is transmitted over a wireless link. Packet loss can occur when, for example, a network router chooses to drop one or more packets to mitigate congestion.

Typically, in video transmission over the wired Internet, if the data packet received at the transport layer is found to be corrupted, then it is not forwarded to the application layer. Thus, there are two possibilities at the application layer (1) the packet is received without error (2) the packet is not received either because it was dropped by a congested router, or because it was corrupted. In this work, a simplifying assumption is made that, whatever the cause of packet loss, the sequence number of the lost packet is known. In broadcast video, for example, the sequence number is available from the MPEG transport stream. The receiver sees an erasure channel at the output of which a packet is either received perfectly, or is erased completely. To recover from erasures, there are then two possible strategies: (1) Apply an erasure correcting channel code, (2) Use a feedback channel, if available. Under the first strategy, we review Forward Error Correction (FEC) and Layered Coding with Unequal Error Protection (LCUEP). Under the second strategy, we review channel-adaptive source coding and packet scheduling.

### 2.2.1 Forward Error Correction of Video Signals

Forward Error Correction (FEC) involves the addition of redundant information to the video bit stream in the form of parity symbols. These are generated using channel coding, and enable the receiver to detect and correct errors in the video bit stream. FEC is useful when the video application has low-delay requirements that cannot be

met by feedback-based retransmissions, or when feedback is unavailable, as is the case in broadcast television.

A widely used channel code used to provide FEC is the systematic Reed-Solomon (RS) code [141], in which parity symbols are appended to the video bit stream prior to transmission. Other channel codes which have been used to provide FEC include turbo codes [30], convolutional codes [51], Low Density Parity Check (LDPC) codes [54] and Fountain (Raptor) codes [146]. We focus only on RS codes, since these are employed in this work.

A Reed-Solomon (RS) code is a non-binary cyclic code on a Galois Field,  $\text{GF}(2^m)$ , with  $m > 2$ . It is possible to construct a RS code which encodes  $K$   $m$ -bit input symbols to  $N$   $m$ -bit output symbols, so long as  $0 < K < N \leq 2^m + 1$  [157]<sup>4</sup>. Among all linear codes with message length  $K$  and output codeword length  $N$ , RS codes achieve the largest possible minimum distance, given by  $d_{\min} = N - K + 1$ . The RS code can correct up to  $T$  errors such that

$$T = \lfloor \frac{d_{\min} - 1}{2} \rfloor = \lfloor \frac{N - K}{2} \rfloor$$

where  $\lfloor x \rfloor$  is the smallest integer less than or equal to  $x$ . When the error locations are known prior to RS decoding, they are called erasures. The RS code can correct up to  $E = d_{\min} - 1 = N - K$  symbol erasures. Combined error and erasure decoding is possible so long as

$$2T + E < N - K$$

In many cases, it is convenient to employ “systematic” RS codes in which  $N - K$  parity symbols are appended to  $K$  message symbols to construct a codeword containing  $N$  symbols<sup>5</sup>. In this work, we shall consider only systematic RS codes with

---

<sup>4</sup>Most RS codes in practice have  $N \leq 2^m - 1$ , though extended RS codes can have  $N = 2^m$  and  $N = 2^m + 1$

<sup>5</sup>Systematic Lossy Error Protection (SLEP) is conceptually independent of the choice of the systematic channel code used to perform Slepian-Wolf coding. Because of their conceptual simplicity, RS codes are used to perform Slepian-Wolf coding in this work. Implementation-wise, this amounts to applying the RS code in the usual manner, but transmitting only the parity symbols and discarding the systematic (message) symbols. This is described in detail in Chapter 4.

$m = 8$ , i.e., the symbols of the RS code are one byte in length and reside in  $\text{GF}(2^8)$ . A common way of implementing FEC is to group the bits of a compressed video signal into  $K$  byte-long symbols and then to apply a systematic  $(N, K)$  RS code where  $K < N \leq 255$ . For example, the MPEG transport stream consists of 188-byte video packets which are protected by a  $(204, 188)$  RS code. Starting with a given  $(N, K)$  Reed-Solomon code, it is possible to change the code rate, and hence the error protection capability, by removing some of the parity symbols (puncturing), adding parity symbols (extending), removing a message symbol (shortening), or adding message symbols (lengthening).

For transmission across bursty channels, the message symbols are interleaved prior to applying the channel code. Upon deinterleaving the symbols at the receiver, the error bursts are spread out and appear as isolated errors to the channel decoder. Sometimes, two channel codes are concatenated with an interleaver placed in between. For instance, in MPEG-2 transport, an outer Reed-Solomon code is concatenated with an inner punctured convolutional code. At the receiver, the inner code first attempts to correct bit errors. This code cannot correct messages containing a large number of errors, and outputs an error burst in such cases. The error burst is then deinterleaved into isolated errors (or erasures) fed to the outer channel decoder. Thus, an interleaver/de-interleaver pair provides resilience to bursty errors at the expense of increased computational complexity and buffering delay prior to channel encoding and decoding.

### 2.2.2 Layered Video Coding and Priority Encoding Transmission

One of the disadvantages of using FEC to protect video bit streams is that catastrophic signal degradation occurs when error correction fails, a phenomenon known in the channel coding community as the “cliff effect.” So long as the number of errors is less than the error correction capability of the FEC code, a constant and acceptable video quality is maintained. However, when the number of errors becomes too large, the video decoder has no option other than to conceal the corrupted packets

using a local error concealment scheme. Decoder-based concealment is generally a last resort since the decoder does not possess enough information to perfectly conceal the transmission errors. Besides, in a predictive coding structure, the error concealment artifacts propagate to the subsequent frames, resulting in unacceptably large distortion.

One of the methods to mitigate the FEC cliff is to use layered video coding with unequal error protection [16, 55, 102, 77, 103, 164, 74, 82, 176, 175]. In layered coding schemes, the image or video signal is encoded into a base layer and one or more enhancement layers. Then, using priority encoding transmission [15, 32, 70], the more significant base layer is protected with a stronger channel code, while the less significant enhancement layers are protected with a weaker channel code. If the error probability increases, the decoding of the enhancement layers fails, but the base layer can still be decoded, limiting the maximum degradation that can occur. Despite this beneficial property of graceful degradation, layered video coding schemes with priority encoding transmission have hitherto not been applied in commercial broadcast scenarios. This is because of the rate-distortion inefficiency of standardized scalable video codecs<sup>6</sup> such as the SNR scalable versions of MPEG-2 [1], MPEG-4 [68], and H.263 [108]. The newly developed H.264/SVC [191] standard can provide temporal and SNR scalability without severely reducing the rate-distortion efficiency, but this requires complicated tuning of the coding parameters across the video layers.

### 2.2.3 Multiple Description Coding and Path Diversity

Multiple Description (MD) coding [66, 19] of multimedia signals is a robust video transmission technique that exploits the availability of multiple paths between the source and the receiver. This scenario arises in video transmission over the Internet or over wireless ad hoc networks. In this framework, a signal is encoded into multiple descriptions which are independently decodable but may be combined to generate a

---

<sup>6</sup>This does not necessarily mean that layered video coding schemes are not useful. Indeed, in transmission scenarios that require the video signal to be encoded once, and transmitted to several Internet users with different download bandwidths and device capabilities, layered coding is a very attractive option.

signal of higher quality. These descriptions are transmitted to the receiver on multiple paths, all of which may be error-prone. In the most general setup, depending upon the error rate and capacity of each available path, a joint source/channel bit allocation scheme may be used over all the available paths. In general, the receiver has access to a subset of the total number of descriptions and obtains a decoded signal by suitably combining the available descriptions [17, 20, 18]. The set of achievable rates for multiple description coding was derived by El Gamal and Cover [2]. Various MD coding schemes have been proposed for robust video transmission, including the use of B frames to construct multiple descriptions, rate-distortion optimal selection of temporal splitting, spatial splitting and repetition coding modes [76], matching pursuits for generating multiple descriptions [169, 106], using pairwise correlating transforms [181], and using different prediction paths while performing motion compensation in individual descriptions [142].

#### **2.2.4 Feedback-Based Error Control**

A feedback channel, if available, can be used to improve the error resilience of a video transmission. There is a very large suite of such techniques in the literature [21, 184], and only a small representative sampling is presented here. The simplest use of a feedback message is to notify the encoder about whether a certain portion of a video signal has been received or not. At its most basic, this is done via ACK (acknowledgment) or NAK (negative acknowledgment) packets which are transmitted at a very low bit rate. The syntax for the feedback messages is not a part of the video coding specification, but belongs to a separate layer in the protocol stack which is used for exchanging control information. For example, the location of macroblocks in a H.263-coded video bit stream may be fed back to the encoder using ITU-T Recommendation H.245 [69].

#### **Channel-Adaptive Source Coding**

In channel-adaptive source coding, feedback messages are taken into account in the decisions taken by the source coder. The error tracking approach proposed in [105]



uses a NACK message to reconstruct the error distribution in a video frame. Since motion information is available at the encoder, the macroblocks affected by error propagation can also be determined, and coded in the intra mode. This method effectively stops error propagation, but unlike simple retransmission, it does not introduce extra delay.

Another channel-adaptive source coding technique is to use Reference Picture Selection (RPS) [53, 31]. In RPS, as in error tracking, a feedback message is used to inform the encoder of the spatio-temporal location of the error. Then, while encoding a subsequent frame, the encoder chooses as reference an older frame which has been correctly received at the decoder. This reduces the efficiency of motion-compensated prediction, but stops error propagation. This approach can be combined with long-term memory motion compensated prediction [192]. In network-adaptive video coding [90], the reference picture is selected such that the expected distortion at the receiver is minimized given a rate constraint. This scheme avoids the need for retransmission of lost pictures, while incurring a very low delay of the order of a few hundred milliseconds, making it attractive for video conferencing applications.

Related approaches include the usage of multiple prediction threads [185] and the use of dual frame video encoding, which uses two motion vectors per macroblock [87]. In H.264/AVC, there is an option to respond to the feedback message by transmitting SP and SI frames which switch to a lower bit rate so that the packets can be transmitted more reliably [154, 81].

### **Packet Scheduling**

As an alternative to using the ACK/NAK messages to manipulate the decisions of the video encoder, these messages can also be used to schedule the transmission of video packets in such a way that the effect of losses is mitigated. For example, in Rate-Distortion Optimized (RaDiO) streaming, feedback messages are used to determine which packets have been received and lost at a particular stage of the video transmission. Each successfully transmitted packet contributes to a known reduction in distortion, and a corresponding increment in cost (measured in terms of the bit rate). This observation is used to compute a policy, which determines which video packets

ought to be transmitted, and at which transmission opportunity, such that the overall Lagrangian cost  $J = D + \lambda R$  is minimized [38]. For video, the functionality provided by simple ACK/NACK messages has been augmented using rate-distortion preambles [124], rich acknowledgments [35] and rate-distortion hint tracks [34]. Recently, a related approach named CoDiO (Congestion-Distortion Optimization) has been proposed in [44]. This approach recognizes that rate-distortion optimization alone may be insufficient when packets are dropped due to congestion at the network routers, and therefore incorporates the effect of congestion in the Lagrangian optimization. CoDiO has been applied for peer-to-peer video transmission.

## 2.3 Foundations of Distributed Source Coding

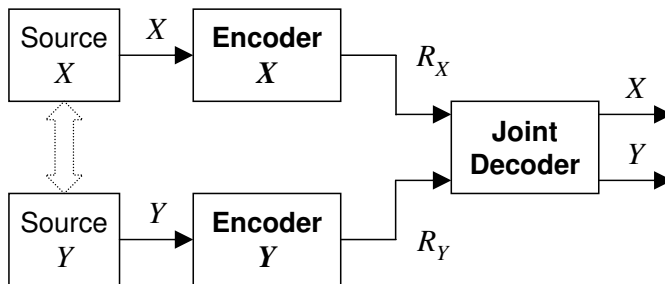


Figure 2.2: Distributed compression of two statistically dependent random processes,  $X$  and  $Y$ . The decoder jointly decodes  $X$  and  $Y$  and thus may exploit their mutual dependence.

Distributed compression refers to the coding of two or more dependent random sequences, but with the special twist that a separate encoder is used for each, as shown in Fig. 2.2. Each encoder sends a separate bit stream to a single decoder which may operate jointly on all incoming bit streams and thus exploit the statistical dependencies. This section contains a review of the main information-theoretic results in distributed source coding, methods used to implement distributed codes, and the state-of-the-art in distributed video coding. Excluding minor updates, this section is excerpted from our review paper on distributed video coding [65].

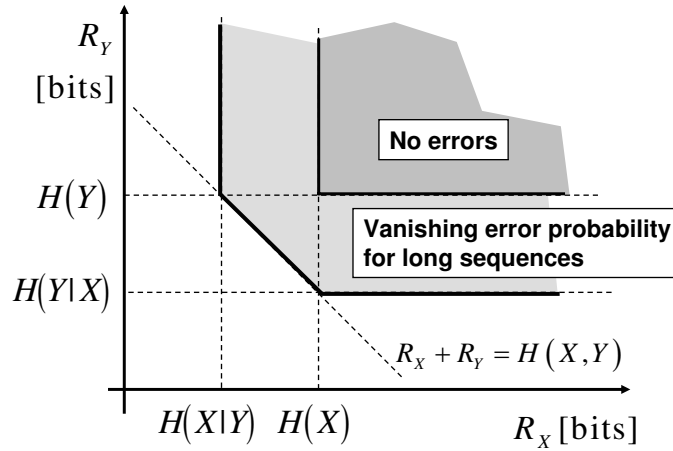


Figure 2.3: Slepian-Wolf Theorem: Achievable rate region for distributed compression of two statistically dependent i.i.d. sources  $X$  and  $Y$  [158].

### 2.3.1 Slepian-Wolf Theorem for Lossless Distributed Coding

Consider two statistically dependent, finite-alphabet random sequences  $X$  and  $Y$ . Samples from each sequence are independent and identically distributed (i.i.d.). With separate conventional entropy encoders and decoders one can achieve  $R_X \geq H(X)$  and  $R_Y \geq H(Y)$ , where  $H(X)$  and  $H(Y)$  are the entropies of  $X$  and  $Y$ , respectively. Interestingly, with separate encoding but joint decoding, better compression efficiency can be achieved if we are content with a residual error probability for recovering  $X$  and  $Y$  that can be made arbitrarily small (but, in general, not zero) for encoding long sequences. In this case, the Slepian-Wolf Theorem [158] establishes the rate region shown in Fig. 2.3:

$$R_X + R_Y \geq H(X, Y)$$

$$R_X \geq H(X|Y), R_Y \geq H(Y|X)$$

Surprisingly, the sum of rates,  $R_X + R_Y$ , can achieve the joint entropy  $H(X, Y)$ , just as for joint encoding of  $X$  and  $Y$ , despite separate encoders for  $X$  and  $Y$ .

Compression with decoder side information (Fig. 2.4) is a special case of the distributed coding problem (Fig. 2.2). The source produces a sequence  $X$  with statistics

that depend on side information  $Y$ . We are interested in the case where this side information  $Y$  is available at the decoder, but not at the encoder. Since  $R_Y = H(Y)$  is achievable for conventionally encoding  $Y$ , compression with receiver side information corresponds to one of the corners of the rate region in Fig. 2.3, and hence  $R_X \geq H(X|Y)$ , regardless of the encoder's access to side information  $Y$ . The type of side information available to the decoder depends upon the application. For example, in distributed compression of sensor data, the side information may consist of readings obtained from other sensors in a dense sensor network. In low-complexity video encoding, the side information consists of a collection of previously decoded video frames. In this work, the side information consists of a video signal afflicted by packet loss.

### 2.3.2 Practical Slepian-Wolf Coding

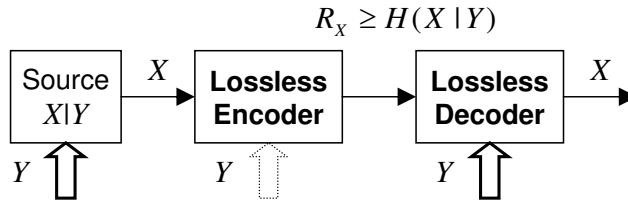


Figure 2.4: Compression of a sequence of random symbols  $X$  using statistically related side information  $Y$ . We are interested in the distributed case, where  $Y$  is only available at the decoder, but not at the encoder.

Although Slepian and Wolf's theorem dates back to the 1970s, it was only in the last few years that emerging applications have motivated serious attempts at practical techniques. However, it was understood already 30 years ago that Slepian-Wolf coding is a close kin to channel coding [194]. To appreciate this relationship, consider i.i.d. binary sequences  $X$  and  $Y$  in Fig. 2.4. If  $X$  and  $Y$  are similar, a hypothetical "error sequence"  $\Delta = X \oplus Y$  consists of 0's, except for some 1's that mark the positions where  $X$  and  $Y$  differ. To "protect"  $X$  against errors  $\Delta$ , we could apply a

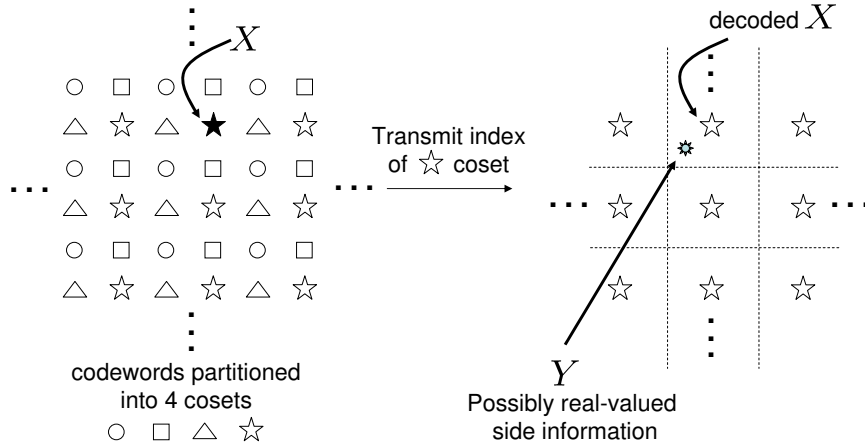


Figure 2.5: In the coset interpretation, the Slepian-Wolf decoder chooses that codeword in the coset of interest which is most likely given the value of the side information.

systematic channel code and only transmit the resulting parity bits<sup>7</sup>. At the decoder, one would concatenate the parity bits and the side information  $Y$  and perform error correction decoding. If  $X$  and  $Y$  are very similar indeed, only a few parity bits would have to be sent, and significant compression results. We emphasize that this approach does not perform forward error correction to protect against errors introduced by the transmission channel, but instead by a virtual “correlation channel” that captures the statistical dependence of  $X$  and the side information  $Y$ . In an alternative interpretation, the alphabet of  $X$  is divided into cosets and the encoder sends the index of the coset to which  $X$  belongs [194]. As shown in Fig. 2.5, the receiver decodes by choosing the codeword in that coset which is most probable<sup>8</sup> in light of the side information  $Y$ . It is easy to see that the coset and parity interpretations are equivalent. With the parity interpretation, we send a binary row vector  $X_p = XP$ , where  $G = [I|P]$  is the generator matrix of a systematic linear block code  $\mathbb{C}_p$ . With the coset interpretation,

<sup>7</sup>Parity is defined in the conventional way with the understanding that  $X$  is the transmitted bit sequence and  $Y$  is the bit sequence received at the end of a hypothetical channel. Then, a parity bit is simply a bit set to zero or one depending upon whether even or odd parity is used to verify the integrity of some or all bits of  $Y$ .

<sup>8</sup>Note that if the side information is corrupted, then the value of  $Y$  might cause the decoder to choose a different codeword in the coset of  $X$ . This case would constitute a decoding error.

we send the “syndrome”  $S = XH$ , where  $H$  is the parity check matrix of a linear block code  $\mathbb{C}_s$ . If  $P = H$ , the transmitted bit streams are identical.

Most distributed source coding techniques today are derived from proven channel coding ideas. The wave of recent work was initiated in 1999 by Pradhan and Ramchandran [114]. Initially, they addressed the asymmetric case of source coding with side information at the decoder for statistically dependent binary and Gaussian sources using scalar and trellis coset constructions. Their later work [115, 117, 116, 113] considers the symmetric case where  $X$  and  $Y$  are encoded with the same rate. Wang and Orchard [179] used an embedded trellis code structure for asymmetric coding of Gaussian sources and showed improvements over the results in [114].

Since then, more sophisticated channel coding techniques have been adapted to the distributed source coding problem. These often require iterative decoders, such as Bayesian networks or Viterbi decoders. While the encoders tend to be very simple, the computational load for the decoder, which exploits the source statistics, is much higher. García-Frías and Zhao [56, 58], Bajcsy and Mitran [27, 100], and Aaron and Girod [3] independently proposed compression schemes where statistically dependent binary sources are compressed using turbo codes. It has been shown that the turbo code-based scheme can be applied to compression of statistically dependent non-binary symbols [208, 207] and Gaussian sources [3, 99] as well as compression of single sources [58, 101, 209, 59]. Iterative channel codes can also be used for joint source-channel decoding by including both the statistics of the source and the channel in the decoding process [57, 209, 3, 101, 59, 94]. Liveris *et al.* [92, 93, 94], Schonberg *et al.* [147, 148, 149], and other authors [39, 159, 86, 61] have suggested that Low-Density Parity-Check (LDPC) codes might be a powerful alternative to turbo codes for distributed coding. With sophisticated turbo codes or LDPC codes, when the code performance approaches the capacity of the correlation channel, the compression performance approaches the Slepian-Wolf bound.

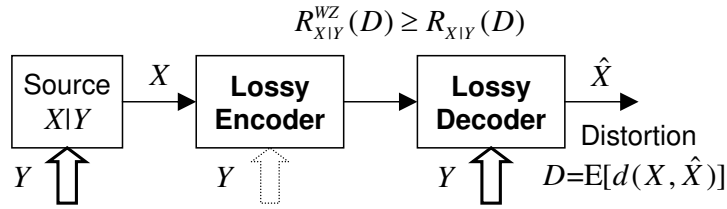


Figure 2.6: Lossy compression of a sequence  $X$  using statistically related side information  $Y$ .

### 2.3.3 Rate-Distortion Theory for Lossy Compression with Receiver Side Information

Shortly after Slepian and Wolf’s seminal paper, Wyner and Ziv [197, 195, 196] extended this work to establish information-theoretic bounds for lossy compression with side information at the decoder. More precisely, let  $X$  and  $Y$  represent samples of two i.i.d. random sequences, of possibly infinite alphabets  $\mathcal{X}$  and  $\mathcal{Y}$ , modeling source data and side information, respectively. The source values  $X$  are encoded without access to the side information  $Y$ , as shown in Fig. 2.6. The decoder, however, has access to  $Y$ , and obtains a reconstruction  $\hat{X}$  of the source values in alphabet  $\mathcal{X}$ .

Define the acceptable distortion as  $D = E[d(X, \hat{X})]$ . The Wyner-Ziv rate-distortion function  $R_{X|Y}^{WZ}(D)$  then is the achievable lower bound for the bit rate for a distortion  $D$ . We denote by  $R_{X|Y}(D)$  the rate required if the side information were available at the encoder as well.

Wyner and Ziv proved that, unsurprisingly, a rate loss  $R_{X|Y}^{WZ}(D) - R_{X|Y}(D) \geq 0$  is incurred when the encoder does not have access to the side information. However, they also showed that  $R_{X|Y}^{WZ}(D) - R_{X|Y}(D) = 0$  in the case of Gaussian memoryless sources and mean squared error distortion [197, 196]. This result is the dual of Costa’s “dirty paper” theorem for channel coding with sender-only side information [40, 165, 28, 111]. As Gaussian-quadratic cases, both lend themselves to intuitive sphere-packing interpretations.  $R_{X|Y}^{WZ}(D) - R_{X|Y}(D) = 0$  also holds for source sequences  $X$  that are the sum of arbitrarily distributed side information  $Y$  and independent Gaussian noise [111]. For general statistics and a mean-squared error distortion measure, Zamir [204] proved that the rate loss is less than 0.5 bits/sample.

### 2.3.4 Practical Wyner-Ziv Coding

As with Slepian-Wolf coding, efforts towards practical Wyner-Ziv coding schemes have been undertaken only recently. The first attempts to design quantizers for reconstruction with side information were inspired by the information-theoretic proofs. Zamir and Shamai [205, 206] proved that, under certain circumstances, linear codes and nested lattices may approach the Wyner-Ziv rate-distortion function, in particular if the source data and side information are jointly Gaussian. This idea was further developed and applied by Pradhan *et al.* [114, 84, 113], and Servetto [153], who published heuristic designs and performance analysis focusing on the Gaussian case, based on nested lattices. Xiong *et al.* [200, 91] implemented a Wyner-Ziv encoder as a nested lattice quantizer followed by a Slepian-Wolf coder, and in [203], a trellis-coded quantizer was used instead (see also [199]).

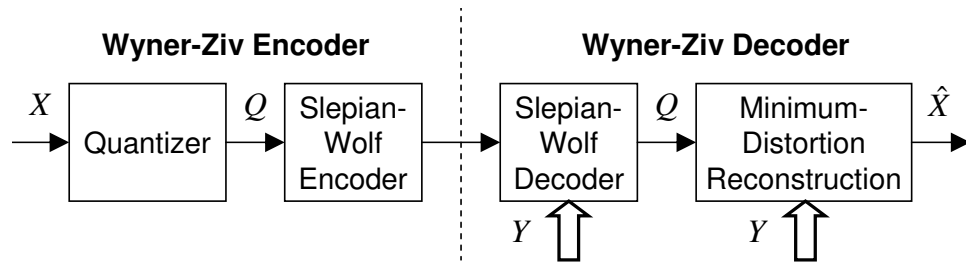


Figure 2.7: A practical Wyner-Ziv coder is obtained by cascading a quantizer and a Slepian-Wolf encoder.

In general, a Wyner-Ziv coder can be thought to consist of a quantizer followed by a Slepian-Wolf encoder, as illustrated in Fig. 2.7. The quantizer divides the signal space into cells, which, however, may consist of non-contiguous subcells mapped into the same quantizer index  $Q$ . This setting was considered, e.g., by Fleming, Zhao and Effros [49], who generalized the Lloyd algorithm [95] for locally optimal fixed-rate Wyner-Ziv vector quantization design. Later, Fleming and Effros [48] included rate-distortion optimized vector quantizers in which the rate measure is a function of the quantization index, for example, a codeword length. An efficient algorithm for finding globally optimal quantizers among those with contiguous code cells was provided in [104]. Unfortunately, it has been shown that code cell contiguity precludes



optimality in general [45]. Cardinal and Van Asche [33] considered Lloyd quantization for ideal Slepian-Wolf coding, without side information.

An independent, more general extension of the Lloyd algorithm appears in [140]. A quantizer is designed assuming that an ideal Slepian-Wolf coder is used to encode the quantization index. The introduction of a rate measure that depends on both the quantization index and the side information divorces the dimensionality of the quantizer from the block length of the Slepian-Wolf coder, a fundamental requirement for practical system design. In [137] it is shown that at high rates, under certain conditions, optimal quantizers are lattice quantizers, disconnected quantization cells need not be mapped into the same index, and there is asymptotically no performance loss by not having access to the side information at the encoder. The problem of Wyner-Ziv coding of noisy observations is considered in [139, 138]. It is shown that, at high rates, the optimal Wyner-Ziv coding scheme is one which obtains the minimum mean squared error (MMSE) estimate of the source given the noisy observation, and then proceeds to perform lattice quantization of the MMSE estimate. As in the noiseless case, there is no need for index repetition and there is asymptotically no performance loss if the side information is not available at the encoder.

### 2.3.5 Low Complexity Distributed Video Encoding

Implementations of the video compression standards discussed in Section 2.1 require much more computation for the encoder than for the decoder; typically, the encoder is 5 to 10 times more complex than the decoder. This asymmetry is well-suited for broadcasting or for streaming video-on-demand systems where video is compressed once and decoded many times. However, some applications may require the dual system, i.e., low-complexity encoders, possibly at the expense of high-complexity decoders. Examples of such systems include wireless video sensors for surveillance, wireless PC cameras, mobile cameraphones, disposable video cameras, and networked camcorders. In all of these cases, compression must be implemented at the camera where memory and computation are scarce.

The Wyner-Ziv theory [197, 204, 111] discussed in Sec. 2.3.3, suggests that an unconventional video coding system, which encodes individual frames independently, but decodes them conditionally, is viable. In fact, such a system might achieve a performance that is closer to conventional interframe coding (e.g., MPEG) than to conventional intraframe coding (e.g., Motion-JPEG). In contrast to conventional hybrid predictive video coding where motion-compensated previous frames are used as side information, in the proposed system, previous frames are used as “side information” at the decoder only.

Such a Wyner-Ziv video coder would have a great cost advantage, since it compresses each video frame by itself, requiring only intraframe processing. The corresponding decoder in the fixed part of the network would exploit the statistical dependence between frames, by much more complex interframe processing. Beyond shifting the expensive motion estimation and compensation from the encoder to the decoder, the desired asymmetry is also consistent with the Slepian-Wolf and Wyner-Ziv coding algorithms, discussed in Sec. 2.3.2 and 2.3.4, which tend to have simple encoders, but much more demanding decoders.

First implementations involving pixel-domain implementation of Wyner-Ziv video codecs appeared in [12, 9, 10]. In these schemes, the video frames are divided into Wyner-Ziv coded frames and conventionally coded “key frames”. The key frames are assumed to be available at the Wyner-Ziv decoder. The Wyner-Ziv encoder quantizes the non-key frames, and applies a turbo or LDPC code (which performs the function of a Slepian-Wolf code) to the block of quantization indices. Either the parity or syndrome bits generated by the code are sent to the receiver. The Wyner-Ziv decoder recovers the quantization indices using the key frames as side information. The rate-distortion performance of such a codec can be improved by first applying a 2-D block transform to the video frame before Wyner-Ziv coding [137, 8, 120, 122, 121]. Aaron and Girod [11] reported further improvements by performing transform-domain Wyner-Ziv encoding of the difference between the current and previous frame, as opposed to the current frame itself.

One of the main challenges in the design of systems with low-complexity video encoding is to find an efficient decoder-based motion estimation algorithm. Unlike the

situation in conventional video codecs, motion estimation must be performed without the availability of the frame being encoded. It has been proposed that additional information about the current frame (termed as a “hash”) transmitted at a low bit rate can significantly improve the decoder’s estimate of the current frame [6, 4]. For example, the hash may be constructed by performing coarse quantization and entropy coding of a few low-frequency DCT coefficients. A hash constructed in this way differs from a conventional cryptographic hash in that small changes in the video frame will not significantly alter the hash.

Wyner-Ziv coding also lends itself to low-complexity distributed encoding of light-fields [5], large camera arrays [210] and spherical images of 3-D scenes obtained from catadioptric cameras [171]. Yet another related application is the distributed compression of the plenoptic function in camera sensor networks [60]. Most recently, Varodayan *et al.* have reported encouraging results on lossless distributed compression of stereo images, in which the estimation of the disparity between the stereo images at the decoder appears as an analogue of motion compensation in video coding [173, 174].

### 2.3.6 Error-Resilient Video Compression using Distributed Source Coding

Wyner-Ziv coding can be thought of as a technique which generates parity information to correct the “errors” of the correlation channel between source sequence and side information, up to a distortion introduced by quantization. Wyner-Ziv coding thus lends itself naturally to robust video transmission as a “lossy” channel coding technique. It is straightforward to use a stronger Slepian-Wolf code which not only corrects the discrepancies of the correlation channel, but additionally corrects errors introduced during transmission of the source sequence. Experiments have been reported for the PRISM codec [120, 122, 121] that compare the effect of frame loss with that observed in a conventional predictive video codec (H.263+). With H.263+, displeasing visual artifacts are observed due to interframe error propagation. With PRISM the decoded video quality is minimally affected and there is no drift between

encoder and decoder. Sehgal *et al.* [151] have also proposed a Wyner-Ziv coding scheme based on turbo codes to combat interframe error propagation. In their scheme, Wyner-Ziv coding is applied to certain “peg frames,” while the remaining frames are encoded by a conventional predictive video encoder. This ensures that any decoding errors in the predictive video decoder can only propagate until the next peg frame. Jagmohan *et al.* [152] applied Wyner-Ziv coding to each frame to design a “state-free” video codec in which the encoder and decoder need not maintain precisely identical states while decoding the next frame. The state-free codec performs only 1 to 2.5 dB worse than a state-of-the-art standard video codec. Xu and Xiong used LDPC codes and nested Slepian-Wolf quantization to construct a layered Wyner-Ziv video codec [201]. This codec approaches the rate-distortion performance of a conventional codec with Fine-Granular Scalability (FGS) coding with the added advantage that the LDPC code provides resilience to transmission errors.

### 2.3.7 Systematic Lossy Source/Channel Coding

We now discuss the use of distributed source coding within a mathematical framework known as systematic lossy source/channel coding. Consider a source signal which is transmitted over an analog channel without channel coding. Owing to channel errors, the viewer at the output of the analog channel receives a degraded version of the original signal waveform. To provide error resilience, an additional encoded version of the source signal is sent over a digital channel as enhancement information. A receiver at the output of the digital channel can decode this encoded description using the degraded output of the analog channel as side information. As shown in Fig. 2.8, this “side information decoding” enhances the quality of the received signal. In this way, the degradation introduced by the analog channel is mitigated.

The term “systematic coding” has been introduced as an extension of systematic error-correcting channel codes to refer to a partially uncoded transmission. Further, the second description is decoded in the presence of side information at the receiver, which is, in fact, the operation described earlier as Wyner-Ziv coding. Shamai, Verdú, and Zamir established information-theoretic bounds and conditions for optimality

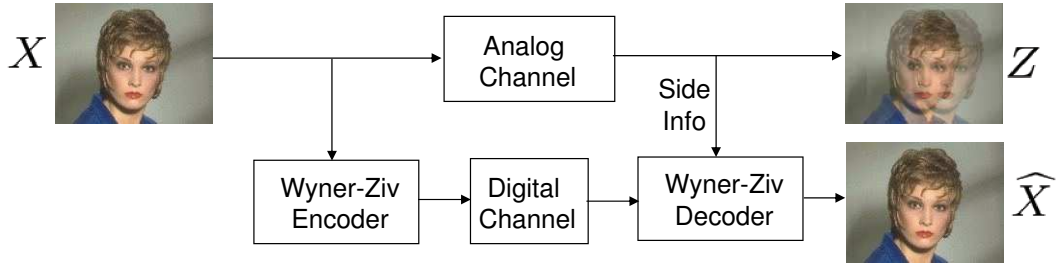


Figure 2.8: Digitally enhanced analog transmission.

of such a configuration in [155]. Optimality conditions refer to the requirements that must be satisfied by a systematic lossy source/channel coding system so that it achieves the same rate-distortion performance as a system with a single channel whose capacity equals the sum of the capacities of the analog and digital channels. Here, we reproduce the optimality conditions derived in [155] for the systematic lossy source coding scenario in Fig. 2.8:

1. The source maximizes the mutual information of the analog (uncoded) channel, i.e.,  $I(X; Z) = C_A$ , where  $C_A$  is the capacity of the analog channel.
2. The degraded output of the analog channel is not needed at the encoder. In other words, Wyner-Ziv coding has the same rate-distortion function as conditional coding, i.e.,  $R_{X|Z}^{WZ}(D) = R_{X|Z}(D)$ .
3. The output of the analog channel is maximally useful at the source decoder, i.e.,  $R_X(D) = R_{X|Z}(D) + I(X; Z)$ .

This framework was applied by Pradhan and Ramchandran [112] to enhance analog image transmission using digital side information. A scheme using distributed source coding on an auxiliary channel, similar to the error protection scheme proposed and investigated in this thesis, has been concurrently studied by Wang *et al.* for error resilient transmission of H.263 video over lossy networks [177] and for MPEG-2 broadcast [178]. In an extension of the work by Shamai, Verdú and Zamir on systematic lossy source/channel coding, Steinberg and Merhav derived information-theoretic optimality conditions for the case of layered Wyner-Ziv coding [160] and hierarchical joint source/channel coding [161].

## 2.4 Summary

This chapter reviews the state-of-the-art in robust video transmission. Motion-compensated predictive video coding is described with emphasis on error resilience strategies used at the encoder and the decoder. The compressed bit streams are protected against channel errors by a number of methods, chief among them being either forward error correction or layered coding with unequal error protection. Other methods to achieve robustness include exploiting path diversity by means of multiple description coding, and utilizing a feedback channel (wherever possible) to inform the video encoder about lost packets, thus changing the coding modes used for the temporally successive frames.

Further, a review [65] is provided of the area of distributed source coding, i.e., source coding with decoder-based side information. Slepian-Wolf and Wyner-Ziv coding are introduced from an information-theoretic perspective, and recently proposed strategies to implement them in practice have been discussed.

Finally, the information-theoretic framework of systematic lossy source/channel coding is described, which uses Wyner-Ziv coding to provide error resilience. This framework will be used in the subsequent chapters as a basis for the design of a Systematic Lossy Error Protection (SLEP) scheme for robust transmission of video signals.

## Chapter 3

# Systematic Lossy Error Protection

In the previous chapter, it was shown that Wyner-Ziv coding can be used to provide error resilience in a systematic lossy source/channel coding framework. We now apply this framework to Systematic Lossy Error Protection (SLEP) of video signals. The concept of SLEP is explained in Section 3.1. In Section 3.2, we apply SLEP to the transmission of a Markov source which is compressed lossily using a DPCM (Differential Pulse Code Modulation) encoder and transmitted over an error prone channel. We use high-rate approximations to derive the end-to-end distortion experienced by this system. Since a video codec, in its most basic form, resembles a DPCM-type system, this theoretical treatment provides insights into the design of a practical SLEP system and the degrees of freedom involved.

### 3.1 Concept of SLEP

The application of systematic lossy source-channel coding to error-resilient digital video transmission is illustrated in Fig. 3.1. At the transmitter, the input video  $S$  is compressed independently by a hybrid video encoder and a Wyner-Ziv encoder. As shown, the compressed video signal transmitted over the error-prone channel constitutes the systematic portion of the transmission. For robustness, the systematic portion is augmented by the Wyner-Ziv bit stream. The Wyner-Ziv bit stream can be thought of as a second description of  $S$ , but with coarser quantization. Thus the

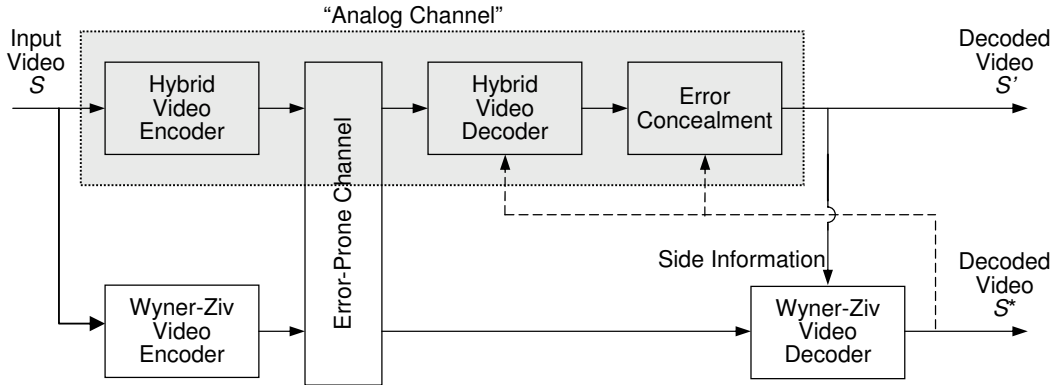


Figure 3.1: A video transmission system in which the video waveform is protected by a Wyner-Ziv bit stream, in a systematic source/channel coding configuration. The receiver decodes the Wyner-Ziv bit stream using the error-prone received video signal as side information.

Wyner-Ziv bit stream contains a low quality description of the original video signal. We refer to the scheme in Fig. 3.1 as Systematic Lossy Error Protection (SLEP)<sup>1</sup>.

Without transmission errors, the Wyner-Ziv description is fully redundant, i.e., it can be regenerated bit-by-bit at the decoder, using the decoded video  $S'$ . When transmission errors occur, the receiver performs error concealment, but some portions of  $S'$  might still have unacceptably large errors. In this case, Wyner-Ziv bits allow reconstruction of the second description, using the decoded waveform  $S'$  as side information. This coarser second description and side information  $S'$  are combined to yield an improved decoded video  $S^*$ . In portions of  $S'$  that are unaffected by transmission errors,  $S^*$  is essentially identical to  $S'$ . However, in portions of  $S'$  that are degraded by transmission errors, the coarser second representation limits the maximum degradation that can occur in the current decoded frame. This “repaired” frame is then fed back to the video decoder to serve as a more accurate reference for the motion-compensated decoding of the subsequent frames. Since digital video transmission is being considered, there is no analog/digital channel separation, as was the case in Fig. 2.8. However, the role played by the hybrid video codec and the

<sup>1</sup>In the early work on SLEP [7, 126, 125, 133], we have used the term Forward Error Protection (FEP). This term was replaced by SLEP, because FEP can be easily confused with classic Forward Error Correction (FEC).



error-prone channel in Fig. 3.1 is analogous to the role played by the analog channel in Fig. 2.8.

Thus, SLEP is essentially a scheme which efficiently transmits an alternative representation of the video signal which may be used when portions of the main signal (also referred to as the “primary” description) are lost or corrupted due to channel errors. Since this alternative representation is coarsely quantized, each successful instance of Wyner-Ziv decoding results in a quantization mismatch between the high-quality primary description and the low-quality second description. Owing to motion-compensated decoding of predictively coded frames, this quantization mismatch propagates to the subsequent frames. The more frequently channel errors occur, the more frequently Wyner-Ziv decoding is invoked and the larger is the quantization mismatch. To ensure that the ensuing video quality degradation is graceful, the quantization mismatch must be controlled, i.e., the quantization levels in the second description must be selected appropriately. This adds a degree of freedom in the design of a SLEP system, compared to the design of traditional FEC-based systems. For optimizing FEC, the designer determines the percentage of the available bit rate that must be allocated for channel coding. In SLEP, it is not sufficient to determine the percentage of the available bit rate that must be allocated to Wyner-Ziv coding. Indeed, it is also essential to determine the quality (equivalently, the source coding bit rate) of the second description which travels in the Wyner-Ziv bit stream. These issues are considered from a more theoretical standpoint in the next section, wherein we use SLEP for robust transmission of a compressed first-order Markov source, and derive the resulting end-to-end distortion.

## 3.2 SLEP of a First-Order Markov source

In this section, we study a SLEP scheme which is simple enough for a closed-form mathematical analysis. We consider robust transmission of a first-order Markov source over an erasure channel. The Markov source is compressed by a first-order DPCM coder. The prediction residual is quantized, entropy-coded and transmitted over an erasure channel. For error resilience, we requantize the prediction residual

and use Wyner-Ziv coding to mitigate the effect of erasures on the distortion in the transmitted signal. We derive expressions for the total rate and the end-to-end distortion in the decoded sequence. The theoretical analysis in the remainder of this chapter has been partially presented in [136].

### 3.2.1 DPCM Source Coding Scheme

We now describe the encoding and decoding scheme for the systematic transmission. In addition, we detail the assumptions on the source data and the coding operations, which will be used to obtain the expressions for rate and distortion:

1. **Source data:** The encoding scheme is shown in Fig. 3.2. The source data is represented by  $(X_n)_{n \in \mathbb{Z}}$ , a zero-mean, stationary, first-order Markov process.

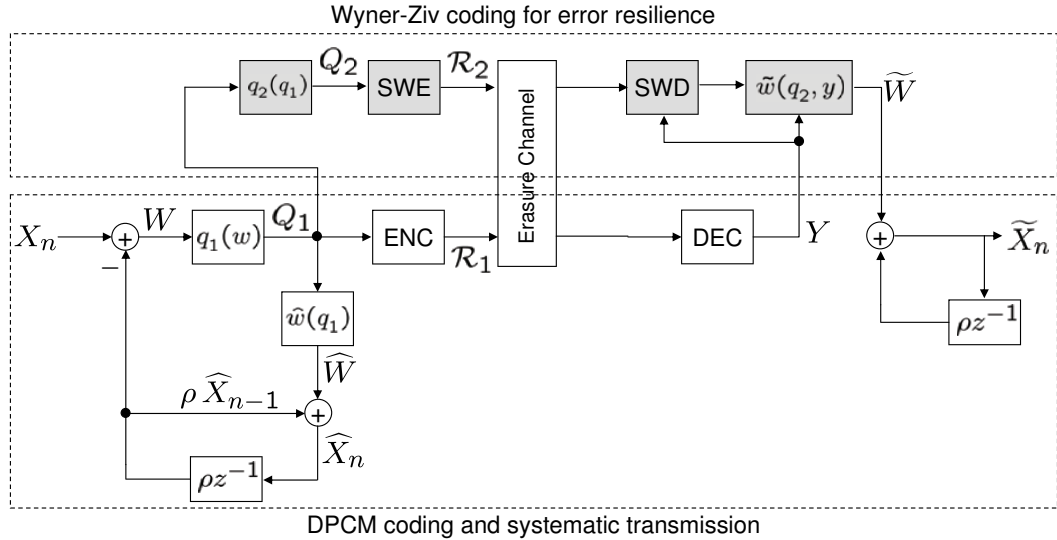


Figure 3.2: Systematic lossy error protection applied to the prediction residual signal of a DPCM coding scheme.

2. **Prediction residual:** We consider a simple linear predictor  $X_n = \rho \hat{X}_{n-1} + W_n$ , where  $|\rho| < 1$ , and  $W_n$  represents the unpredictable component, i.e., the prediction residual. In this example,  $\rho$  is the correlation coefficient between  $X_n$  and  $X_{n-1}$ , and  $\rho X_{n-1}$  is the best linear unbiased estimate of  $X_n$  given  $X_{n-1}$ . Note that in the DPCM encoder, we predict  $X_n$  from the reconstructed sample

$\widehat{X}_{n-1}$  and not from  $X_{n-1}$ . At high rates, the quantization of  $W_n$  is fine enough so that  $X_n \simeq \widehat{X}_n$ . Therefore, we immediately become less formal and say that the  $W_n$  are i.i.d. and independent of the past values of the source data  $X_{n-1}, X_{n-2}, \dots$ . This situation occurs, for example, when the source data are produced by a first-order Gauss-Markov process. Note that whenever a variable, or a difference of variables, is identically distributed, we will drop the time index  $n$ .

3. **Quantization of prediction residual:** The quantizer  $q_1(w)$  maps the prediction error  $W$  into the quantization index  $Q_1$ , which is compressed by an ideal entropy coder. Thus, the source coding bit rate is  $\mathcal{R}_1 \triangleq H(Q_1)$ . The codewords generated by the entropy coder are transmitted across an error-prone channel. The reconstruction of  $W$  corresponding to the index  $Q_1$  is  $\widehat{W} = E[W|Q_1]$ . Mean squared error (MSE) is used as the distortion measure, thus the expected source coding distortion in  $W$  is  $\mathcal{D}_1 \triangleq E(W - \widehat{W})^2$ .
4. **Using local reconstructions as reference samples:** The encoder's local reconstruction of  $X_n$ , to be used for predictive encoding of the future samples, is given by  $\widehat{X}_n = \rho \widehat{X}_{n-1} + \widehat{W}_n$ . Note that, in the absence of channel errors, the receiver would recover the quantization indices and obtain  $\widehat{X}_n$  exactly, and there would be no mismatch between encoder and decoder. i.e.,  $E(X - \widehat{X})^2 = E(W - \widehat{W})^2 = \mathcal{D}_1$ .

### 3.2.2 Wyner-Ziv Coding of the Prediction Residual

We assume that the codewords generated by the entropy coder are erased with probability  $p$ . The process causing the erasures is assumed to be independent of the source statistics. At the receiver, reversing the entropy coding operation yields either the quantization index  $Q_1$ , or an erasure (denoted by the symbol  $e$ ). Thus, the side information for the Wyner-Ziv decoder is:

$$Y = \begin{cases} Q_1 & \text{w.p. } 1 - p \\ e & \text{w.p. } p \end{cases} \quad (3.1)$$

With no error protection in the case of an erasure, the best possible reconstruction of  $W$  is  $E[W|e] = EW = 0$ , which would result in a MSE in  $W$  of  $\sigma_W^2$ , the variance of  $W$ . Because of the predictive coding structure, this error energy will propagate to the subsequently decoded samples. To mitigate this error propagation, SLEP transmits additional symbols generated by distributed coding of the prediction residual. The Wyner-Ziv coding procedure is as follows:

1. **Quantization:** First, the prediction residual is requantized. Specifically, let the quantizer  $q_2(q_1)$  map the quantization index  $Q_1$  from Fig. 3.2 into the quantization index  $Q_2$ . Thus,  $q_1(w)$  is embedded inside  $q_2(q_1(w))$ . The corresponding reconstruction levels for  $W$  are given by  $\widehat{W} = E[W|Q_2]$ .
2. **Slepian-Wolf coding:** Now, ideal lossless encoding of the quantization indices  $Q_2$  is performed assuming the presence of side information  $Y$  at the decoder. Note that the statistics of  $Y$  are known to the Slepian-Wolf encoder, but the actual value of  $Y$  is unknown. With ideal Slepian-Wolf encoding [158], the bit rate required would be  $H(Q_2|Y)$ . However, the Slepian-Wolf code is transmitted over an erasure channel. In order to ensure that the Slepian-Wolf codewords can be recovered in spite of these erasures, the bit rate must be increased to  $\mathcal{R}_2 > H(Q_2|Y)$ .  $\mathcal{R}_2$  will also be referred to as the error resilience bit rate or the Wyner-Ziv bit rate. At the receiver, Slepian-Wolf decoding returns the quantization index  $Q_2$ .
3. **SLEP decoding:** Let  $\widetilde{W}$  denote the output of the SLEP decoder. We now define the operation of the SLEP decoder, i.e., its response to erasures that may occur on both the systematic and the Wyner-Ziv transmissions. If there is no erasure on the systematic transmission, it means that the side information  $Y = Q_1$  and no error has occurred. In this case, the output is defined to be  $\widetilde{W} = \widehat{W} = E[W|Q_1]$ . If there is an erasure on the systematic transmission, Wyner-Ziv decoding must be performed and the output is given by  $\widetilde{W} = \widehat{W} = E[W|Q_2, e] = E[W|Q_2]$ , because the erasure provides no information about  $W$ .

To summarize, the output of the SLEP decoder is given by:

$$\widetilde{W} = \mathbb{E}[W|Q_2, Y] = \begin{cases} \widehat{W} & \text{if } Y = Q_1 \\ \widehat{\widehat{W}} & \text{if } Y = e \end{cases} \quad (3.2)$$

We emphasize that, owing to requantization, the Wyner-Ziv representation has lower quality compared to the main transmitted signal, and will only be called upon when the main prediction error signal is lost. Owing to the embedding of the quantizers,  $\mathbb{E}[W|Q_1, Q_2] = \mathbb{E}[W|Q_1]$ . This justifies the above decoding strategy, since  $\widetilde{W}$  is the optimal reconstruction of  $W$  in the MSE sense. In this simple setup, SLEP is the same as unequal error protection of the prediction error, in which the higher significant bit-planes in the binary representation of  $W$  are protected, while the lower significant bit-planes are not. Since the error process of the channel is independent of the prediction and quantization operations,  $\widetilde{W}_n$  is i.i.d. and the subscript  $n$  has been omitted. The MSE distortion in  $W$ , after SLEP decoding, is  $\mathcal{D}_2 \triangleq \mathbb{E}(W - \widetilde{W})^2$ .

### 3.3 Rate-Distortion Tradeoffs in SLEP

As shown in Fig. 3.2, the final goal is to reproduce  $X_n$ . This final reproduction, denoted by  $\widetilde{X}_n$ , is obtained by reversing the prediction process at the encoder. Thus,  $\widetilde{X}_n = \rho\widetilde{X}_{n-1} + \widetilde{W}_n$ . Our goal is to obtain an expression for the total rate, defined as  $\mathcal{R} \triangleq \mathcal{R}_1 + \mathcal{R}_2$  and the end-to-end distortion, defined as  $\mathcal{D} \triangleq \mathbb{E}(X - \widetilde{X})^2$ . Please refer to Lemmas 4 and 5 in Appendix A for an explanation of why  $X_n$ ,  $\widetilde{X}_n$  and the difference  $X_n - \widetilde{X}_n$  are all identically distributed.

We assume that  $W$  is encoded at high rates. The results in this section hold if the Bennett assumptions<sup>2</sup> [29] apply to the probability density function  $f_W(w)$ .

---

<sup>2</sup>The Bennett assumptions require that (1) the number of quantization bins is very large, (2)  $f_W(w)$  is smooth so that Reimann sums may be approximated by Reimann integrals, (3) the widths of the quantization bins are very small, (4) the reconstruction codewords are the Lloyd centroids of their respective quantization bins and (5) the total overload distortion is negligible even when  $f_W(w)$  has infinite support.

We consider, in turn, the rate-distortion relation for the source coder of  $W$ , the rate-distortion relation for the Wyner-Ziv coder of  $W$ , and the final expression for end-to-end distortion in  $X$ .

Suppose that the statistics of  $W$  are such that the differential entropy  $h(W)$  is defined and finite. By a direct application of high rate quantization theory [62], an asymptotically optimal scalar quantization strategy for the prediction residual  $W$  is to perform uniform quantization with step-size  $\Delta_1$ , which satisfies, for large  $\mathcal{R}_1$ :

$$\begin{aligned}\mathcal{R}_1 &\simeq h(W) - \log_2 \Delta_1 \\ \mathcal{D}_1 &\simeq \frac{\Delta_1^2}{12} \\ \mathcal{D}_1 &\simeq \frac{1}{12} 2^{2h(W)} 2^{-2\mathcal{R}_1}\end{aligned}\tag{3.3}$$

Note that, since  $W$  is encoded at high rates,  $\Delta_1 \ll \sigma_W$ . We now obtain a rate-distortion relation for the Wyner-Ziv coder.

**Proposition 1.** *Suppose that the statistics of  $W$  are such that the differential entropy  $h(W)$  is defined and finite. Suppose also that asymptotically optimal scalar quantization has been used in the systematic transmission. Then, an asymptotically optimal scalar quantization strategy for the SLEP decoding procedure described in Section 3.2.2 is to perform uniform quantization of  $Q_1$  with step-size  $m$ , which satisfies, for large  $\mathcal{R}_2$ , and  $\Delta_1 \rightarrow 0$ :*

$$\begin{aligned}\mathcal{R}_2 &\simeq \frac{p}{1-p} (\mathcal{R}_1 - \log_2 m) \\ \mathcal{D}_2 &\simeq (1-p+pm^2) \mathcal{D}_1 \\ \mathcal{D}_2 &\simeq \frac{(1-p+pm^2)}{m^2} \frac{1}{12} 2^{2h(W)} 2^{-2\mathcal{R}_2 \frac{1-p}{p}}\end{aligned}\tag{3.4}$$

*Proof.* Since  $Q_2$  is obtained via requantization of the indices  $Q_1$ , knowledge of  $Q_1$  unambiguously determines  $Q_2$ . If there were no erasures in the Slepian-Wolf transmission, then the error resilience bit rate would be given by the Slepian-Wolf theorem:

$$H(Q_2|Y) = (1 - p) H(Q_2|Q_1) + p H(Q_2|e) = 0 + p H(Q_2) \quad (3.5)$$

The last term simplifies to  $H(Q_2)$  because an erasure in the side information  $Y$  is independent of the Wyner-Ziv quantization process and thus provides no information about  $Q_2$ . Now, if there are erasures in the Slepian-Wolf transmission, then the Slepian-Wolf theorem cannot be used directly because it assumes error-free transmission of the Slepian-Wolf code<sup>3</sup>. To find  $\mathcal{R}_2$ , we use the analogy between the Slepian-Wolf code and the parity portion of a systematic channel code. Consider a systematic channel code in which both the source and the parity symbols are erased with probability  $p$ . Let the parity portion of a capacity-achieving channel code be used as a Slepian-Wolf code. Then, the parity bit rate, which equals  $\mathcal{R}_2$  in the present problem, is given by:

$$\mathcal{R}_2 = \frac{p}{1-p} H(Q_2) \simeq \frac{p}{1-p} (\log_2 W - \log_2(m\Delta_1)) = \frac{p}{1-p} (\mathcal{R}_1 - \log_2 m) \quad (3.6)$$

Here, requantization to obtain  $Q_2$  is asymptotically equivalent to transcoding  $W$  using a uniform quantizer with step-size  $\Delta_2 = m\Delta_1$ ,  $m \in \mathbb{Z}^+$ . We further claim that there is no loss of optimality if  $m \in \mathbb{Z}^+$  (instead of the more general claim that  $m \in \mathbb{R}^+$ ). For a given distortion, since  $\Delta_1 \rightarrow 0$ , the increase in rate due to this introduced gradation is arbitrarily small. Such a gradation gives *points* on the  $\mathcal{R}_2(\mathcal{D}_2)$  curve, but these points are arbitrarily close at high rates, so we can take the rate-distortion function to be asymptotically continuous<sup>4</sup>. Further, (3.6) uses the

---

<sup>3</sup>Intuitively, the error resilience bit rate  $\mathcal{R}_2$  should be *higher* than the Slepian-Wolf bit rate because we want the Slepian-Wolf code to provide protection not only against erasures in the DPCM-coded transmission, but also against erasures in the transmission of the Slepian-Wolf codewords themselves.

<sup>4</sup>This result holds only if  $\mathcal{R}_2(\mathcal{D}_2)$  is itself asymptotically continuous as  $\Delta_1 \rightarrow 0$  and  $m \in \mathbb{R}^+$ . To verify that this is the case, assume that  $m \in \mathbb{R}^+$  and prove Proposition 1 exactly as above. Then, the system of equations (3.4) indicates that  $\mathcal{R}_2(\mathcal{D}_2)$  is asymptotically continuous.

result that a uniform quantizer of width  $\Delta_2 = m\Delta_1$ , without index repetition, is asymptotically optimal as shown in [138].

The MSE distortion at the output of the Wyner-Ziv decoder is then given by:

$$\mathcal{D}_2 = \mathbb{E}(W - \widetilde{W})^2 = (1 - p) \mathbb{E}(W - \widehat{W})^2 + p \mathbb{E}(W - \widehat{\widehat{W}})^2 \quad (3.7)$$

$$\simeq (1 - p) \frac{\Delta_1^2}{12} + p \frac{m^2 \Delta_1^2}{12} \simeq (1 - p + pm^2) \mathcal{D}_1 \quad (3.8)$$

where (3.7) is obtained by iterated expectation on the side information  $Y$ , and (3.8) uses the distortions observed at high rates for quantizers with step sizes  $\Delta_1$  and  $m\Delta_1$ .  $\square$

For  $p = 0$ , we have  $\mathcal{D}_2 = \mathcal{D}_1, \mathcal{R}_2 = 0$ , confirming that no bits need to be spent on error resilience for the error-free case. We now derive an expression for  $\mathcal{D}$ , the effective distortion in  $X$  as a result of the distortion in  $W$ , accounting for the effect of error propagation from previously decoded samples.

**Theorem 2.** *Consider a SLEP system in which the systematic transmission has a rate-distortion relation given by (3.3) and the Wyner-Ziv transmission has a rate-distortion relation given by Proposition 1. Then, the end-to-end mean squared error distortion in  $X$  is given by:*

$$\mathcal{D} \simeq \left(1 + p \frac{m^2 - 1}{1 - \rho^2}\right) m^{-2p} \frac{1}{12} 2^{\mathsf{h}(W)} 2^{-2\mathcal{R}(1-p)} \quad (3.9)$$

*Proof.* Consider the error in the reconstruction of  $X$  at the decoder:

$$X_n - \widetilde{X}_n = (\rho \widehat{X}_{n-1} + W_n) - (\rho \widetilde{X}_{n-1} + \widetilde{W}_n) = \rho (\widehat{X}_{n-1} - \widetilde{X}_{n-1}) + (W_n - \widetilde{W}_n) \quad (3.10)$$

From Lemmas 4 and 5 in Appendix A, the differences  $W_n - \widetilde{W}_n, X_n - \widetilde{X}_n, \widehat{X}_n - \widetilde{X}_n$  are stationary, and we can drop the time indices while writing the distortions. Moreover, since  $W$  is i.i.d., the difference  $W_n - \widetilde{W}_n$  is independent of  $\widehat{X}_{n-1} - \widetilde{X}_{n-1}$



Then, from (3.10),

$$\begin{aligned} \mathcal{D} &= \mathbb{E}(X - \tilde{X})^2 = \rho^2 \mathbb{E}(\hat{X} - \tilde{X})^2 + \mathbb{E}(W - \tilde{W})^2 + 2\rho \mathbb{E}(\hat{X} - \tilde{X}) \mathbb{E}(W - \tilde{W}) \\ &= \rho^2 \mathbb{E}(\hat{X} - \tilde{X})^2 + \mathcal{D}_2 + 0 \end{aligned} \quad (3.11)$$

where the last term vanishes because, by iterated expectation,  $\mathbb{E} \tilde{W} = \mathbb{E} \mathbb{E}[W|Q_2, Y] = \mathbb{E} W$ . Now consider the difference,

$$V_n = \hat{X}_n - \tilde{X}_n = \rho(\hat{X}_{n-1} - \tilde{X}_{n-1}) + (\widehat{W}_n - \tilde{W}_n) = \rho V_{n-1} + U_n \quad (3.12)$$

Thus, the new random process  $V_n$  is obtained by passing a strict sense stationary zero-mean random process  $U_n$  through a LTI filter<sup>5</sup>. Then, from Lemma 6, we have,

$$\mathbb{E}(\hat{X} - \tilde{X})^2 = \sigma_V^2 = \frac{1}{1 - \rho^2} \sigma_U^2 = \frac{1}{1 - \rho^2} \mathbb{E}(\widehat{W} - \tilde{W})^2 \quad (3.13)$$

where the MSE in the right hand side can be evaluated as follows:

$$\mathbb{E}(\widehat{W} - \tilde{W})^2 = (1 - p) \mathbb{E}(\widehat{W} - \widehat{W})^2 + p \mathbb{E}(\widehat{W} - \widehat{W})^2 \simeq 0 + p(m^2 - 1)\mathcal{D}_1 \quad (3.14)$$

The last term in (3.14) is the MSE between the reconstruction levels of the source quantizer and Wyner-Ziv quantizer. For any  $m \in \mathbb{Z}^+$ , this MSE evaluates to  $(m^2 - 1)\mathcal{D}_1$ . This calculation is worked out in Proposition 7 in Appendix A.

Substituting the expressions of (3.13) and (3.14) into (3.11), the end-to-end MSE distortion in  $X$  is given by

$$\mathcal{D} \simeq \frac{\rho^2}{1 - \rho^2} p(m^2 - 1)\mathcal{D}_1 + \mathcal{D}_2 = \left(1 + p \frac{m^2 - 1}{1 - \rho^2}\right) \mathcal{D}_1 \quad (3.15)$$

This equation may be reduced to the form in the theorem statement by expressing  $\mathcal{D}_1$  in terms of  $\mathcal{R}_1$ , and finally expressing  $\mathcal{R}_1$  in terms of the total rate  $\mathcal{R}$ . For  $p = 0$ , the

---

<sup>5</sup>The stationarity of  $U_n = \widehat{W}_n - \tilde{W}_n$  arises from the initial assumptions on  $W$  and  $X$  and is a consequence of Lemma 4 in Appendix A.

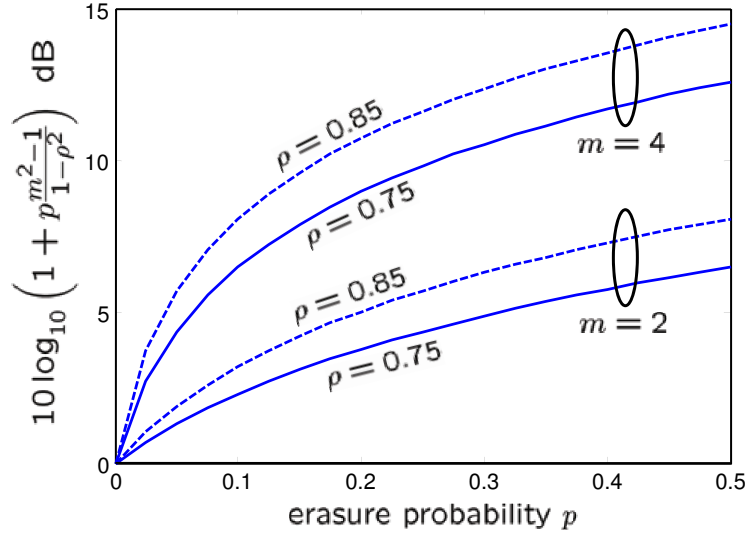


Figure 3.3: Residual distortion after Wyner-Ziv decoding increases when the erasure probability increases.

familiar high-rate result is obtained, with  $\mathcal{D}$  reducing by 6.02 dB/bit. For non-zero  $p$ ,  $\mathcal{D}$  falls at the rate of  $6.02(1 - p)$  dB/bit.  $\square$

It is useful to express the MSE distortion in  $X$  according to (3.15) rather than the expression in the statement of Theorem 2. By expressing  $\mathcal{D}$  in terms of the distortion  $\mathcal{D}_1$  introduced by the source coder, it is easy to find the excess distortion in the signal after Wyner-Ziv decoding. We make the following observations about the residual distortion after Wyner-Ziv decoding, from (3.15):

1. When the erasure probability increases, the residual distortion in the decoded signal increases. This is because an increase in the erasure probability results in an increase in the frequency with which a quantization mismatch is introduced in the prediction residual  $\tilde{W}$ , and by error propagation, into the decoded signal  $\tilde{X}$ . This is depicted in Fig. 3.3.
2. When the correlation coefficient  $\rho$  between the current and previous sample of the signal increases, the residual distortion increases. In the simple DPCM coding scheme,  $\rho$  is used as the prediction coefficient. From (3.15), it is clear

that an increase in  $\rho$  results in an increase in the fraction of the quantization mismatch error energy that propagates to the current frame from the previous frame. The influence of  $\rho$  on the residual distortion is shown in Fig. 3.4.

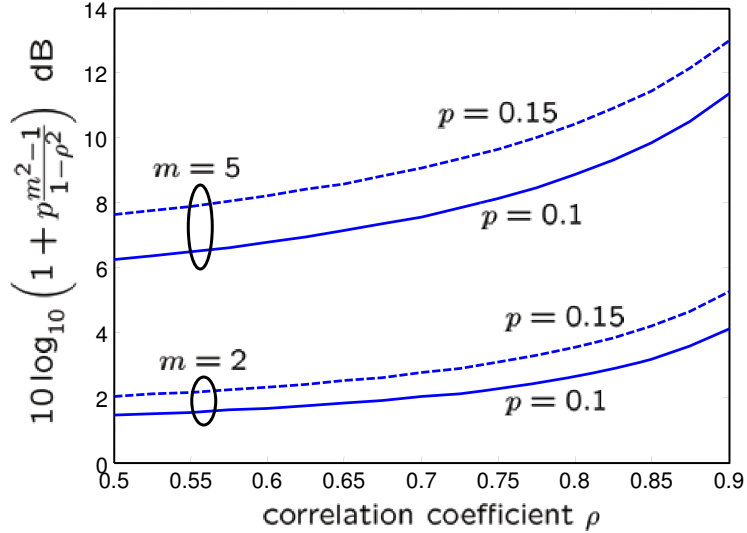


Figure 3.4: Residual distortion after Wyner-Ziv decoding increases as the prediction coefficient  $\rho$  approaches unity.

3. When  $m$  is increased, i.e., when the step-size  $\Delta_2 = m\Delta_1$  of the Wyner-Ziv quantizer is increased, the residual distortion increases. This is because an increase in  $\Delta_2$  results in an increase in the quantization mismatch energy, derived in (A.2). The behavior of the residual distortion with increasing  $m$  is shown in Fig. 3.5. It will be shown in the subsequent section that, in return for the increased residual distortion, a coarser Wyner-Ziv quantizer allows erasure protection at higher erasure probabilities.

Further, substituting  $m = 1$  in (3.15) gives  $\mathcal{D} = \mathcal{D}_1$ , which is the distortion-rate tradeoff for lossless erasure protection. This indicates that, so long as error protection succeeds, the signal quality with lossless error protection remains constant with respect to the erasure probability  $p$ , whereas that with lossy protection degrades according to (3.15). Thus SLEP is a generalized error protection scheme which includes

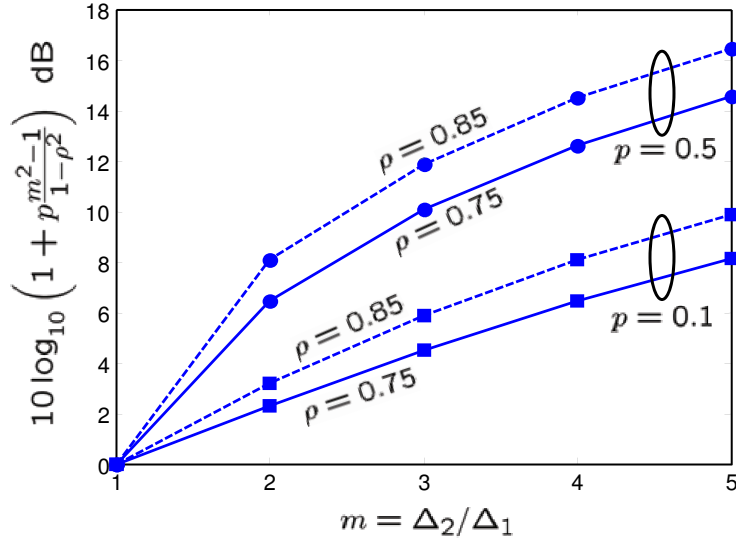


Figure 3.5: Residual distortion after Wyner-Ziv decoding increases when the step-size of the Wyner-Ziv quantizer is increased, owing to the greater quantization mismatch between the DPCM quantizer and the Wyner-Ziv quantizer.

a lossless correction scheme, such as FEC, as a special case. This is examined in further detail in the next section.

### 3.4 Observations on Lossy Versus Lossless Protection

The treatment in the earlier sections assumed that the erasure probability is known. Now consider the case in which  $\mathcal{R}_2$  is set to allow error protection for any erasure probability  $p \leq p_{\text{cliff}}$ . In that case, we can write the overall distortion in  $X$  as:

$$\mathcal{D} \simeq \begin{cases} \mathcal{D}_1 \left( 1 + p \frac{m^2 - 1}{1 - \rho^2} \right) & \text{if } p \leq p_{\text{cliff}} \\ \mathcal{D}_1 \left( 1 + p \frac{(\sigma_W^2/\mathcal{D}_1) - 1}{1 - \rho^2} \right) & \text{if } p > p_{\text{cliff}} \end{cases} \quad (3.16)$$

where the distortion for  $p \leq p_{\text{cliff}}$  is obtained from (3.15). The distortion for  $p > p_{\text{cliff}}$  can be obtained by repeating the steps in the proof of Theorem 2 noting that erasure

protection fails for  $p > p_{\text{cliff}}$ , and so the minimum MSE reconstruction of  $W$  is not  $\widehat{W}$  but  $EW$ . From (3.16), notice that  $\mathcal{D}$  has a discontinuity at  $p = p_{\text{cliff}}$  because  $m^2 \mathcal{D}_1 \ll \sigma_W^2$  at high rates. Now we compare SLEP ( $m > 1$ ) against lossless forward error correction ( $m = 1$ ) in two scenarios:

- The bit rates  $\mathcal{R}_1$  and  $\mathcal{R}_2$  are fixed. Let  $p_{\text{cliff},m}$ , indexed by  $m = \Delta_2/\Delta_1 \in \mathbb{Z}^+$ , be the maximum erasure probability at which the system can provide error protection. Using (3.4), we have  $p_{\text{cliff},m} \geq p_{\text{cliff},1}$ . Thus, SLEP provides erasure protection over a wider range of erasure probabilities compared to FEC. As shown in Fig. 3.6, the distortion for FEC is constant for  $p \leq p_{\text{cliff},1}$  and increases rapidly for  $p > p_{\text{cliff},1}$  owing to the failure of the channel code. This is the familiar ‘‘cliff effect’’. In SLEP, the distortion increases gracefully owing to coarse quantization, as long as  $p \leq p_{\text{cliff},m}$ . Moreover, the cliff in SLEP is pushed further to the right, as compared to FEC. The larger the value of  $m$ , the greater the robustness of the error protection scheme.

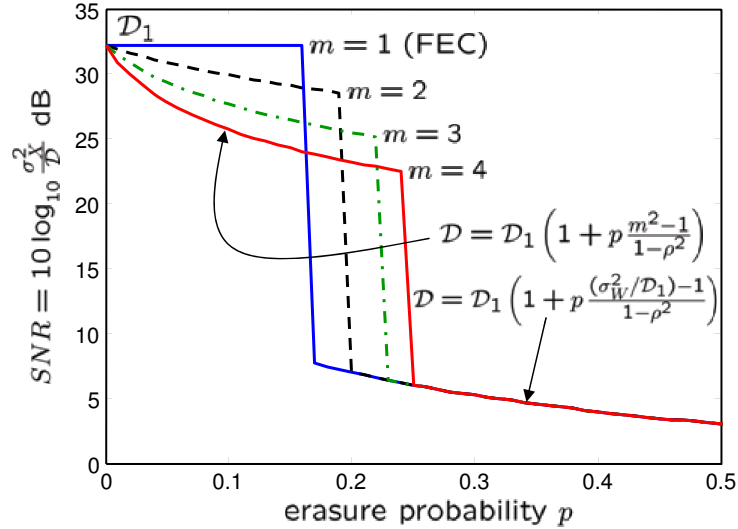


Figure 3.6: The end-to-end distortion  $\mathcal{D}$  is evaluated for the case where source data  $X$  are generated by a first-order Gauss-Markov process with  $\rho = 0.75$  and  $\sigma_W^2 = 5$ . For a fixed error resilience bit rate, SLEP provides graceful quality degradation over a wider range of erasure probabilities than FEC.

2. The total bit rate  $\mathcal{R}$  is fixed and the system is designed to tolerate a fixed maximum erasure probability  $p_{\text{cliff}}$ . Let  $\mathcal{R}_{1,m}$  and  $\mathcal{R}_{2,m}$  be the optimally chosen source coding bit rate and error resilience bit rate, depending upon the value of  $m$ . From (3.4) and the total bit rate constraint,  $\mathcal{R}_{1,m} \geq \mathcal{R}_{1,1}$  and  $\mathcal{R}_{2,m} \leq \mathcal{R}_{2,1}$ . Thus, SLEP allocates more bits to the source code than FEC. For  $p = 0$ , the erasure-free case, the SNR with SLEP is higher than that with FEC by  $20p_{\text{cliff}} \log_{10} m$  dB. This is proved in Appendix A. As shown in Fig. 3.7 for  $0 \leq p \leq p_{\text{cliff}}$ , FEC incurs constant distortion, while the distortion of SLEP increases with  $p$ . The system design ensures that the cliff occurs at probability  $p = p_{\text{cliff}}$  for both FEC and SLEP. It can be shown that the distortion plots for FEC and SLEP must cross at:

$$p_{\text{cross}} = \frac{(1 - \rho^2)(m^{2p_{\text{cliff}}} - 1)}{m^2 - 1} < p_{\text{cliff}} \text{ for } m > 1 \quad (3.17)$$

A detailed proof of this result is provided in Appendix A. It is also evident from Fig. 3.7 that as the decoded signal quality in the erasure-free case increases, the crossover probability given by (3.17) reduces. In other words, the better the quality of the signal compressed by the DPCM coder, the faster the degradation that occurs due to erasures.

### 3.5 Summary

The concept of Systematic Lossy Error Protection (SLEP) has been explained, using compressed video transmission as an example. This scheme involves the transmission of a Wyner-Ziv bit stream which provides lossy protection, by allowing lost or error-prone portions of the video signal to be concealed by a coarsely quantized video description.

A simple error-resilient codec has been analyzed in which a first order Markov source is predictively encoded and transmitted over an erasure channel. In addition, a bit stream generated by Wyner-Ziv coding is used to provide lossy error protection. Using high-rate quantization theory, closed form expressions for rate and distortion

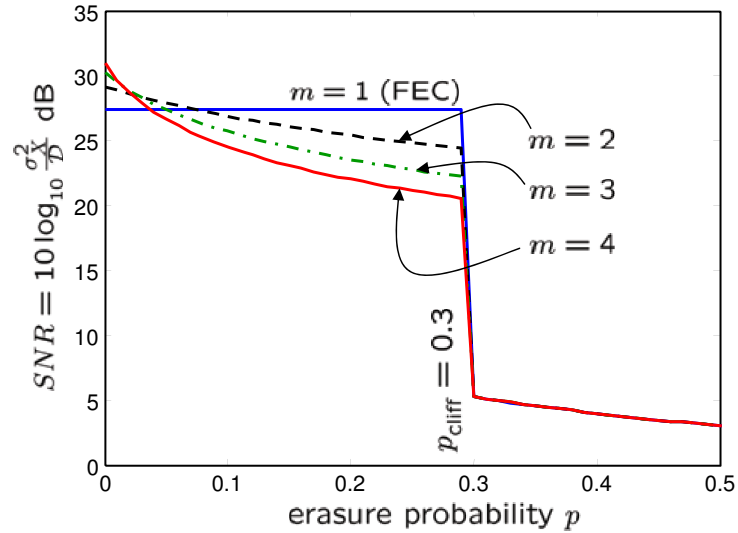


Figure 3.7: The end-to-end distortion  $\mathcal{D}$  is evaluated for the case where source data  $X$  are generated by a first-order Gauss-Markov process with  $\rho = 0.75$  and  $\sigma_W^2 = 5$ . If the maximum erasure probability is fixed, then SLEP allocates a larger fraction of the total bit rate  $\mathcal{R}$  to source coding, incurring less distortion than FEC in the erasure-free case.

have been derived for the encoding of the prediction residual and the overall encoding of the Markov source. Using these relations, it is shown that the lossy error protection property can be used to provide graceful degradation over a wider range of erasure probabilities compared to a lossless error correction approach like FEC.

# Chapter 4

## A SLEP Scheme based on H.264/AVC Redundant Slices

In the previous chapter, the principle of SLEP was explained in the context of robust video transmission. Then, SLEP was used for error resilient transmission of a compressed first-order Markov source and, using high-rate quantization theory, we derived the end-to-end distortion-rate tradeoff achieved by this system. Recall that, in this simple SLEP implementation, Wyner-Ziv coding is applied to the prediction error signal rather than to the original signal itself. This is similar to the manner in which traditional FEC systems are constructed: The residual signal is obtained by motion compensated prediction, it is transformed and compressed and then a channel code is applied to the compressed prediction residual, not the original video signal. Different from FEC, however, SLEP involves the application of Wyner-Ziv coding to the prediction residual, and provides greater error resilience in exchange for a small loss in the decoded picture quality.

In this chapter, we present the design and implementation of a SLEP scheme using the state-of-the-art H.264/AVC standard. Section 4.1 discusses H.264/AVC standard support for redundant slices and Flexible Macroblock Ordering (FMO), the tools that are leveraged in our SLEP implementation. In Section 4.2, these tools are used in conjunction with a Reed-Solomon coder to construct a Wyner-Ziv video bit stream that provides error protection when the primary video signal is lost. As explained in



Section 2.3.4, Wyner-Ziv coding can be implemented in practice by combining coarse quantization with Slepian-Wolf coding. In the ensuing description, redundant slices will be used to generate coarsely quantized descriptions, while the Reed-Solomon code will perform the function of Slepian-Wolf coding. In Section 4.3, FMO will be used in a variation of SLEP in which error protection is applied only to a Region-of-Interest (ROI) within the video frame. The use of FMO enhances the performance of SLEP for video sequences that contain slow motion and/or a static background. In Section 4.4, the performance of SLEP is evaluated experimentally by performing channel simulations over a range of packet erasure probabilities. Specifically, the quantization in the redundant slices and the Wyner-Ziv protection are varied, and the average and instantaneous video quality provided by SLEP is compared with that provided by traditional FEC and decoder-based error concealment.

## 4.1 H.264/AVC Tools

### 4.1.1 Standard Support for Redundant Slices

A *video slice* is a container which houses a compressed representation of a subset of macroblocks from a video frame. A slice - be it I, P or B - can be decoded independently from other slices belonging to the same picture. A *redundant slice* is an alternate or redundant representation of an already encoded video slice. The main or primary slices, taken together, constitute a *Primary Coded Picture* (PCP). Similarly, the redundant slices, taken together, constitute a *redundant picture*. Redundant slices are included in the *Baseline* and *Extended* profiles of the standard.

In H.264/AVC redundant slices are considered as optional components of an *Access Unit*. An access unit consists of one *Primary Coded Picture* (PCP) with some optional additional information, that may include one or more redundant pictures. The specification poses the following constraints on the content of the redundant slices:

1. The redundant representations must follow the corresponding PCP in the *same access unit*.

2. The redundant picture cannot use a different sampling structure (frame/field) than the primary picture.
3. The redundant picture cannot use as reference for motion compensation, a frame that does not belong to the list of reference frames that is assigned to the PCP.

On the other hand, much freedom is left to the encoder about the actual content of the redundant slices. For example, a redundant slice may have different coding modes and quantization parameters than those used for the corresponding PCP. Redundant slices can have different shapes from those in the PCP. They may contain a different number of macroblocks than the slices of the PCP. Further, decoded slices within a redundant picture need not cover the entire picture area. As described in Section 4.3, flexible macroblock ordering (FMO) can be used to signal a redundant picture that contains macroblocks that belong only to a region-of-interest in the video frame, as opposed to the entire frame.

### 4.1.2 Standard Support for FMO

*Flexible Macroblock Ordering* (FMO) is another error resilience tool available in the *Baseline* and *Extended* profiles of H.264/AVC. This tool involves dividing a frame into a number of macroblock partitions, each of which is called a *slice group*. The shapes of these slice groups are specified in a special data structure called a *Picture Parameter Set* (PPS) which travels as part of the video bit stream [79]. The standard contains 7 predefined macroblock-to-slice group mapping functions [42], labeled FMO Type 0 through Type 6.

A slice is then defined as a sequence of macroblocks from a slice group, taken in raster scan order. Of course, a slice group can consist of one or more slices depending upon the restrictions placed on the maximum allowable size for a slice. In this work, we exploit the FMO Type 2 mapping, also called *Foreground with Leftover*, that allows to define a maximum of seven rectangular slice groups for the foreground, plus one for the background. As noted above, the slice groups are specified in the PPS by 7 sets of macroblock coordinates, signaling the `upper-left` and `bottom-right` corners of each rectangle. Macroblocks belonging to overlapping slice groups are assigned to the

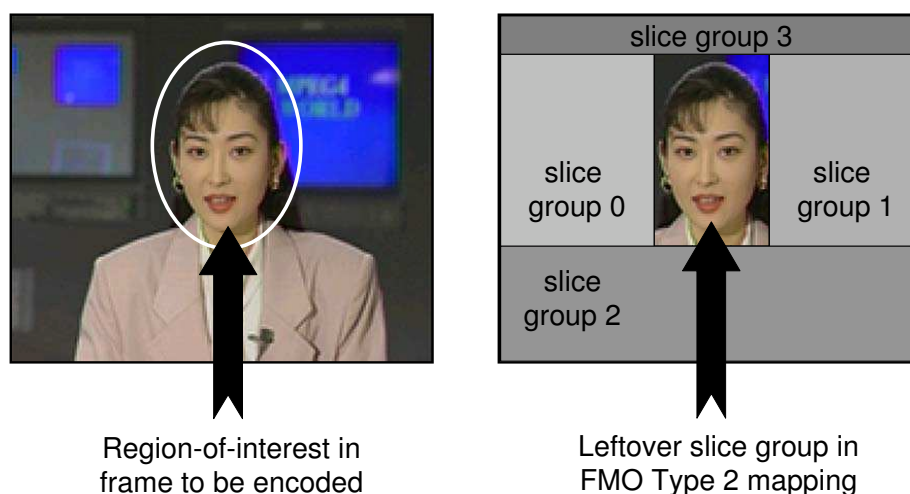


Figure 4.1: FMO Type 2 or *Foreground with Leftover* allows the encoder to perform segmentation of the frame to be encoded, and enables unequal protection strategies for the different segments.

slice group with the lowest slice group number. As shown in Fig. 4.1, this mapping allows the encoder to distinguish the 7 slice groups in the foreground region of the image from the *leftover* slice group in the background. This segmentation enables the underlying transmission layers to adopt different protection policies for the two regions.

## 4.2 SLEP Implementation in H.264/AVC

### 4.2.1 Wyner-Ziv Video Encoding

Our first implementation of a SLEP system consisted of a simple scheme in which pixel-domain Wyner-Ziv coding was applied to a video frame [9, 7]. This was followed by a SLEP implementation for MPEG-2 broadcasting, which exploited the coding efficiency of the DPCM-style predictive structure of standard video codecs [125, 126, 133, 127, 134, 211]. The present implementation, designed using tools supported under the H.264/AVC standard, has been presented in part in [135, 128, 25]. The SLEP scheme is shown in Fig. 4.2. From here onward, the video slices in the systematic

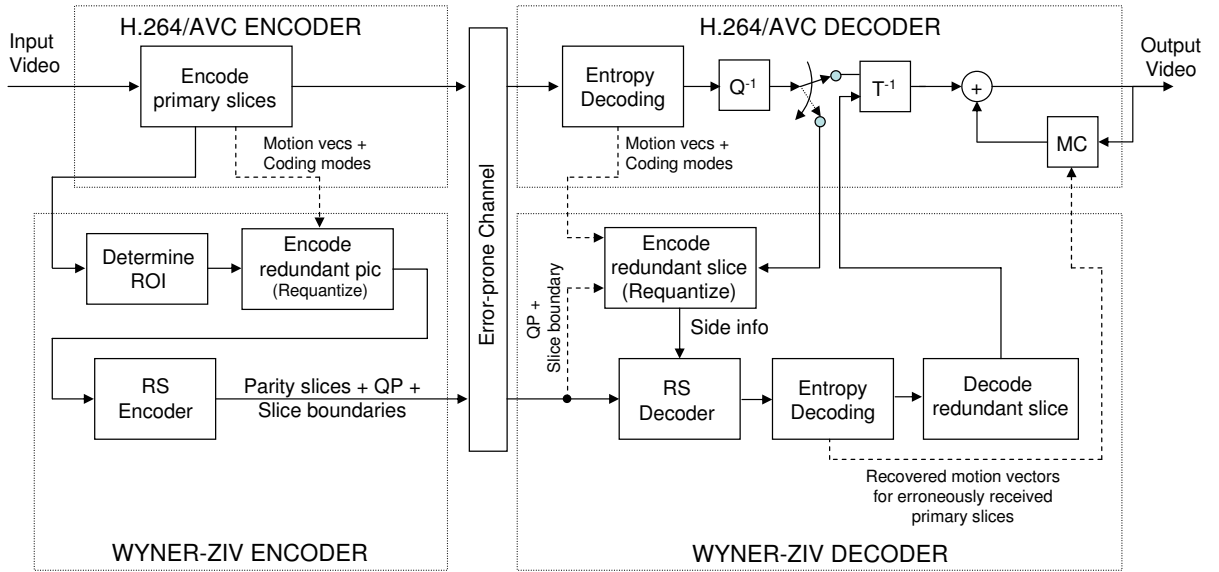


Figure 4.2: Implementation of a SLEP system using H.264/AVC redundant slices and FMO-based region-of-interest determination. Reed Solomon codes applied across the redundant slices play the role of Slepian-Wolf codes in distributed source coding. At the receiver, the Wyner-Ziv decoder obtains the correct 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.

transmission will be referred to as *primary slices* to distinguish them from the redundant slices that are used to generate the Wyner-Ziv bit stream. The following operations are performed on the encoder side:

1. *ROI determination*: The image is analyzed to check for the existence of a Region-of-Interest (ROI). We determine the portions that do not need protection because decoder-based error concealment would reconstruct them with an acceptable distortion. This process is described in detail in Section. 4.3 and may result in the generation of a PPS that specifies the FMO mapping for encoding the redundant slices. Note that the FMO mapping is used to specify slice groups only for the redundant picture and not for the primary picture.
2. *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. This restriction is not imposed by H.264/AVC, but we use it in order to simplify the decoder implementation. As explained later, in Section 4.2.2, this restriction also increases the robustness when Wyner-Ziv decoding fails.

3. *Reed-Solomon encoding*: Reed-Solomon (RS) codes perform the role of Slepian-Wolf coding in this system<sup>1</sup>. A Reed-Solomon code over  $\text{GF}(2^8)$  is applied across the redundant slices to generate parity slices as shown in Fig. 4.3. 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. This is reminiscent of the analogy between Slepian-Wolf coding and traditional channel coding described in Chapter 2, in which parity symbols corresponding to the source were transmitted in order to correct the errors in the side information.
4. *Wyner-Ziv bit stream generation*: In addition to the parity slices resulting from the previous step, we encode for each redundant slice:
  - (a) The number of the first macroblock
  - (b) The number of macroblocks encoded
  - (c) The difference between the quantization parameters (QPs) used for encoding the primary and redundant slices. This difference could be also specified per macroblock instead of per slice, if the rate control algorithm is instructed to change the QP on a macroblock-by-macroblock basis.

This extra “helper information” is appended to the parity slices generated in the above step. In the context of the H.264/AVC video coding standard, this helper information could travel in a special container known as a Supplemental

---

<sup>1</sup>As noted in Chapter 2, any other systematic channel coder could be used depending upon the application and the characteristics of the channel. An example of a SLEP implementation using turbo codes as Slepian-Wolf codes has recently appeared in [89].



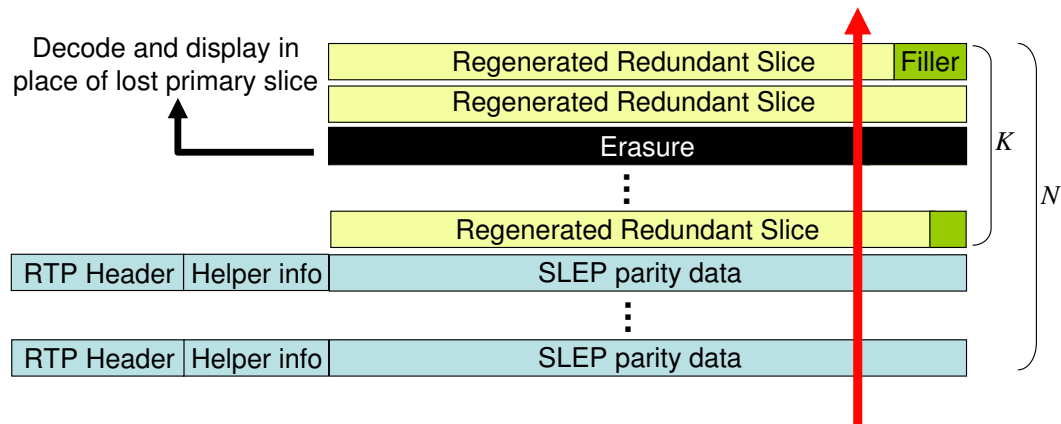


Figure 4.4: During Wyner-Ziv decoding, redundant slices corresponding to received primary slices are obtained by requantization, while those corresponding to the lost primary slices are treated as erasures. These are recovered by erasure decoding, using the parity slices and helper information received in the Wyner-Ziv bit stream. These recovered redundant slices are then decoded and displayed in lieu of the lost primary slices.

those portions of the frame where the primary bit stream has not experienced channel errors. The redundant slices corresponding to the error-prone portions are treated as erasures. Since the coding modes for the redundant macroblocks are identical to those in the primary bit stream, the requantization procedure is straightforward.

Note that this simplification is a slight departure from the conceptual SLEP system of Fig. 3.1, which would require a full redundant encoding of the error-concealed primary video signal. This would have a large complexity cost because motion estimation would have to be re-performed at the decoder, in order to generate the redundant bit stream. Since the redundant description inherits the motion vectors from the primary description, and only performs requantization of the prediction residual, some coding efficiency is sacrificed. However, the redundant description is now generated at very low complexity compared to a full re-encoding. Further, this method is more robust to Wyner-Ziv decoder failure, because the requantization process, followed by entropy coding, would generate the same bit stream irrespective of whether an error occurred during

Wyner-Ziv decoding of the reference frame(s) in the past. In contrast, if full re-encoding of the decoded video signal had been performed, past errors would result in a redundant bit stream which is different from the one used at the encoder. This could potentially cause catastrophic failure of the Wyner-Ziv decoder. Hence, we resort to the much simpler procedure of requantizing the prediction residual at the encoder and mimicking the process at the decoder.

2. *Reed-Solomon Slepian-Wolf decoding:* The parity slices received in the Wyner-Ziv bit stream are now combined with the redundant slices, and erasure decoding is performed to recover the slices which were erased from the redundant bit stream, as shown in Fig. 4.4. 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 redundant prediction error signal and the motion vectors recovered from Wyner-Ziv decoding. The coarse fall-back operation results in a quantization mismatch that propagates to the future frames, but a drastic reduction in picture quality is avoided.

From a joint source/channel coding perspective, the advantage of SLEP over traditional FEC can be understood using the following example: Coarsely quantized redundant descriptions occupy fewer bits than the corresponding finely quantized primary description. This is depicted in Fig. 4.5(a) for the *Foreman* CIF sequence encoded at 408 kb/s. This is done simply by setting the value of the quantization parameter (QP) to 28. If some of the primary slices are lost, and replaced by redundant slices, then the average PSNR decreases, with the largest reduction occurring for the redundant slice with the coarsest quantization, as shown in Fig. 4.5(b). Thus, a naive error protection scheme would transmit all the redundant slices, which can be decoded if the corresponding primary slices are lost. This is inefficient because it



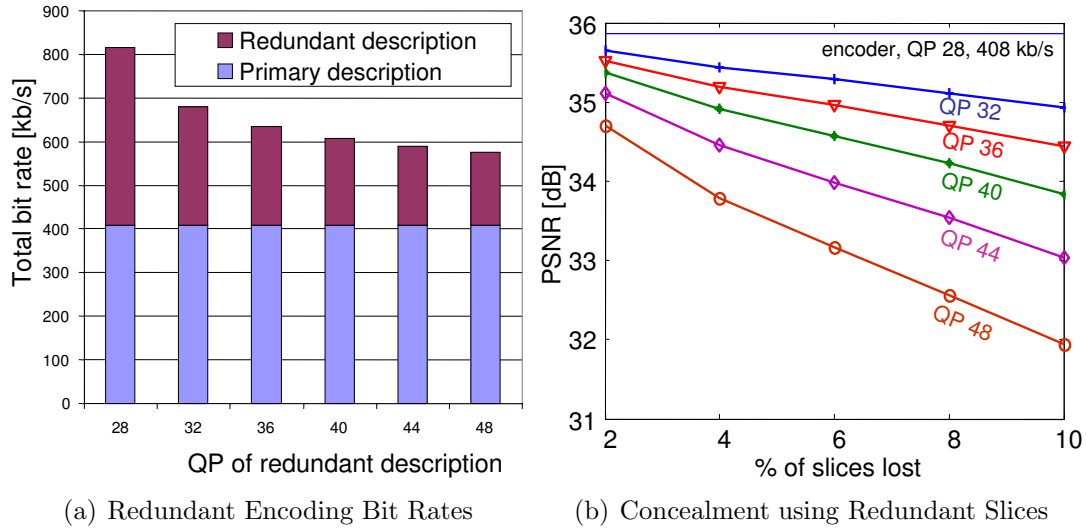


Figure 4.5: A redundant slice can be transmitted along with each primary slice, and can be decoded if the primary slice is lost. The residual distortion depends upon the difference in the quantization step sizes used in the redundant and primary slices.

dramatically increases the transmitted bit rate. SLEP solves this problem by transmitting, not the redundant slices, but parity symbols corresponding to the redundant slices. The bit rate overhead for transmitting the parity slices versus transmitting the entire redundant description is compared in Fig. 4.6. It is evident from Fig 4.6 that, if bit rate of the parity symbols at the output of the Reed-Solomon encoder is fixed, then coarser redundant descriptions get stronger error protection. Of course, decoding a coarser redundant description also results in a greater quantization mismatch at the decoder. This tradeoff is depicted in the PSNR vs. frame number trace in Fig. 4.7. For visual quality inspection, a cut-out of frame no. 81 is shown in Fig. 4.8 for a test video sequence. Without any excess bit rate for error protection, the only option available to the decoder is to perform error concealment. In this case, the picture quality is unacceptable due to error concealment artifacts. As the quantization parameter (QP) of the redundant description is increased from 28 to 40, the robustness of the SLEP scheme increases and the quality of the decoded video frame improves to within 1.5 dB of the error-free case. For very coarse quantization in the redundant video signal (QP = 48), the increased robustness is compensated by the

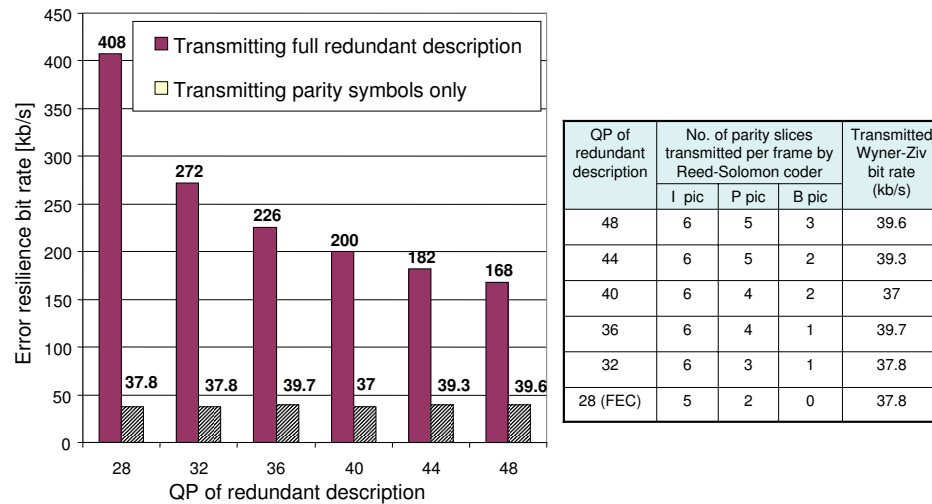


Figure 4.6: Instead of transmitting the redundant slices, it is efficient to transmit parity symbols which enable the receiver to recover the redundant slices. The prediction structure is  $I-B-P-B-P$  and more parity slices are generated for the intra (I) frames than for the predictively encoded (P and B) frames. Since the parity bit rate is approximately constant at 40 kb/s, coarsely quantized redundant descriptions have stronger error protection.

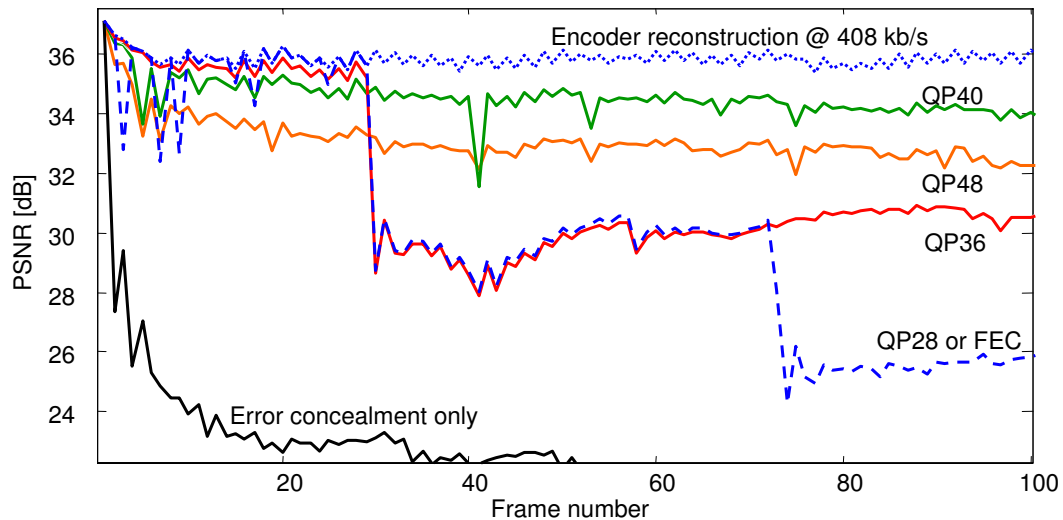


Figure 4.7: Using a redundant description with QP=40 provides the best overall tradeoff between the quantization mismatch and error resilience. Lower QPs result in worse error resilience. Higher QPs result in high quantization mismatch.

reduction in picture quality owing to the quantization mismatch<sup>2</sup>. This tradeoff will be discussed in greater detail in the experimental results in Section 4.4 and in the analysis and modeling of SLEP in Chapter 5.

### 4.3 Applying SLEP to a Region-of-Interest

In this section we discuss the rationale for applying error protection to a Region-of-Interest (ROI), as opposed to the entire picture. Then, we describe the procedure to determine the ROI and to specify it in the Wyner-Ziv bit stream using FMO Type 2.

#### 4.3.1 Rationale for ROI-based SLEP

As seen in the previous section, the error resilience of a SLEP scheme with a high bit rate redundant description is less than that for a low bit rate redundant description. Thus, in order to provide error resilience at high error rates, one must reduce the encoding bit rate of the redundant slices, i.e., one must generate redundant descriptions with a very coarse quantization step.

The distortion introduced by a coarsely quantized redundant description can be large, especially for intra-coded macroblocks. Visually, this large distortion would appear as a “smearing” of the portions of the image that have been protected by SLEP. To mitigate this effect, we avoid redundant encoding of “unimportant” regions of the image. This includes regions which experience zero motion or constant motion, which can be satisfactorily concealed using a simple decoder-based error concealment scheme. For these regions, the residual error after decoder-based error concealment is smaller than the residual error after SLEP. These regions can therefore be excluded from the ROI, which is protected by SLEP.

More importantly, the above classification of ROI and non-ROI ensures that the bit rate of the redundant slices is concentrated only in the ROI. Thus, instead of using coarse quantization for all macroblocks of the frame, the same bit rate can be obtained

---

<sup>2</sup>In this example, for simplicity, the QP value has been held constant for the entire video sequence. This need not be the case. In the simulations at the end of the chapter, a bit rate target is specified and the rate control algorithm changes the QP when the scene complexity changes.



Figure 4.8: Cut-outs of a video frame from the *Foreman* CIF sequence. The primary description is encoded at 408 kb/s, while the error resilience bit rate is fixed at 40 kb/s. Robustness increases with the quantization step in the redundant slice (See fig. 4.2). However, with QP=48, the increased quantization mismatch reduces the decoded picture quality.

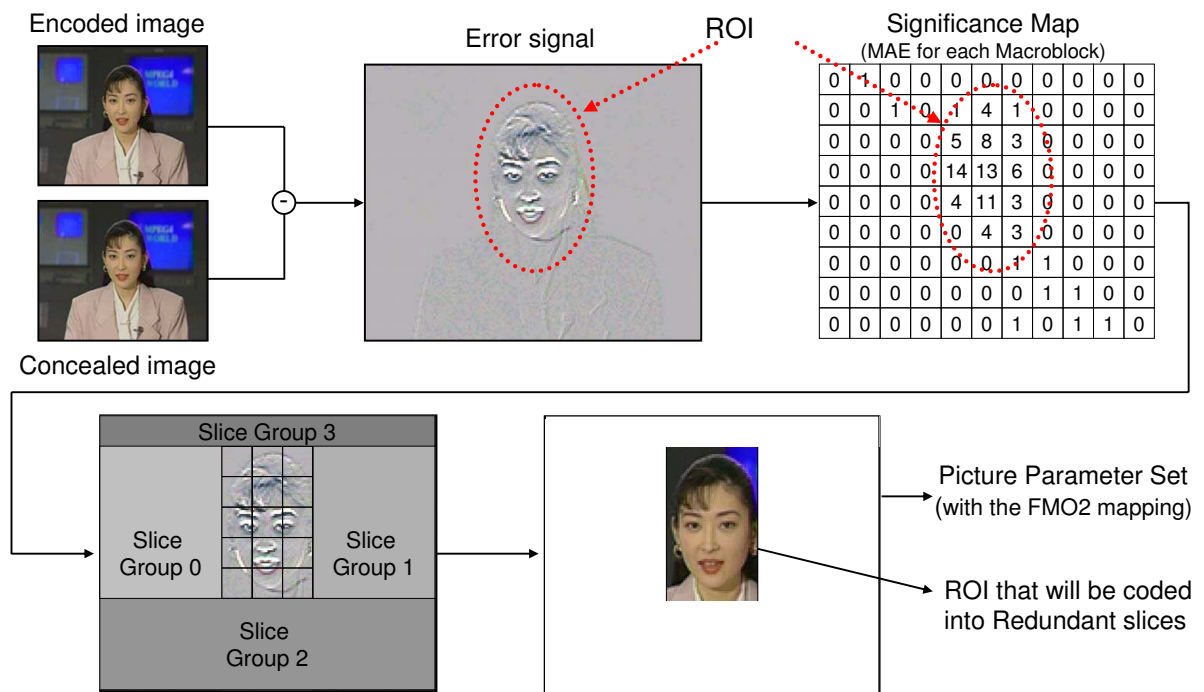


Figure 4.9: A Region Of Interest (ROI) is determined at the encoder by finding the mean absolute error between the current image and its locally error concealed version. The error image is then thresholded to obtain the ROI.

with finer quantization for the macroblocks belonging to the ROI only. Alternatively, by keeping the same quantization step-size but neglecting the macroblocks outside the ROI, the bit rate of the redundant slices can be reduced.

### 4.3.2 Determination of ROI

The procedure to determine the ROI appears in Fig. 4.9. First, we evaluate the impact of the loss of each slice, simulating the previous frame error concealment process at the encoder. A more complicated concealment scheme may also be used, if available. The error signal, obtained as the difference between the current encoded frame and its concealed version, provides a measure of the expected distortion in case of losses. For each macroblock, we compute the “Mean Absolute Error” (MAE), thus producing a *Significance Map* for the whole image, in which larger values indicate macroblocks of higher significance. From this matrix of MAE values, the ROI is

obtained by thresholding. Note that the resulting ROI may have arbitrary shapes, and may consist of a number of disconnected regions.

To further clarify that the distortion resulting from the loss of portions of the (unprotected) non-ROI region is negligible, consider that the threshold used for differentiating between the ROI and the non-ROI is  $T$ , a low value of the MAE between the current frame and its concealed version. Thus, a macroblock with  $\text{MAE} \geq T$  is included in the ROI, while one with  $\text{MAE} < T$  is not included [23]. Typically, the MAE for macroblocks belonging to the ROI is much higher than  $T$ . By construction, this means that, even if the entire non-ROI portion is lost during transmission, the MAE of the non-ROI region cannot exceed  $T$ . By choosing  $T$  conservatively at the encoder, the sender can guarantee that the non-ROI region will have a distortion that is negligible in comparison to the distortion that would result if portions of the ROI were lost.

### 4.3.3 Specification of ROI in Wyner-Ziv Bit Stream

As explained in Section 4.1.2, FMO Type 2 allows 7 rectangular slice groups to cover a region of the frame which is labeled as the “Foreground”, while the area corresponding to the remaining uncovered macroblocks is termed as “Leftover”. We use FMO Type 2, but reverse the roles of the foreground and the leftover. Thus, the non-ROI region is treated as the foreground and we attempt to cover it with 7 non-overlapping rectangles. The remaining uncovered area is the ROI. The top-left and bottom-right co-ordinates of the 7 rectangles are specified in the Picture Parameter Set (PPS) for each frame.

It would be prohibitively complex to determine a scheme in which 7 rectangles can optimally cover the non-ROI region. Instead, a dichotomic (binary) search is performed in the space of possible rectangles, to cover as much of the background as possible. This search proceeds as follows: An initial list of rectangles is constructed, and the largest rectangle is sub-divided into two. Then, the Mean Absolute Error (MAE) of the macroblocks contained in each rectangle is compared to a threshold, to determine whether that rectangle contains a portion of the ROI or not. If not, the

rectangle is not divided further. If the rectangle does contain a portion of the ROI, it is further subdivided into two rectangles. This procedure is iterated until 7 slice groups covering the greatest possible number of unimportant macroblocks are found. For further details of the dichotomic search procedure, please refer to [23].

In this way, FMO Type 2 is used to specify the ROI indirectly, by specifying the coordinates of 7 rectangles which cover the non-ROI region of a video frame. As required by the standard, these coordinates are included in a small data structure called a Picture Parameter Set (PPS). A PPS is transmitted for every picture with a non-trivial ROI, i.e., a ROI which is not equal to the whole frame. Using the PPS, the decoder can determine which macroblocks belong to the ROI.

## 4.4 SLEP Experimental Results

We now describe, in detail, the experimental settings used for the video codec, the video sequences used for testing, and the channel simulation tools. These experimental settings are based on the recommendations obtained from the Joint Video Team, during the course of the standardization effort for SLEP. For a detailed account of the SLEP standardization proposal and to see the results of the ensuing core experiments, please refer to [131, 132, 129, 26, 130].

### 4.4.1 Video Codec Settings

- *Video sequences and coding structure:* We use JVT version JM 11 [166] of the H.264/AVC video codec for our simulations. The experiments are carried out on the SIF ( $352 \times 240$ ) and CIF ( $352 \times 288$ ) resolution video sequences listed in Table 4.1. The encoding bit rates for the primary (systematic) video signal are chosen based on the amount of scene complexity and motion present in the sequence. Thus, the sequence *Football*, which has very high motion, is encoded at 1024 kb/s, while the sequence *Irene*, which features a predominantly static subject in front of a static background, is encoded at 384 kb/s. In keeping with H.264/AVC Baseline, the GOP structure used is *I-P-P-P...* To minimize the

effect of instantaneous fluctuations in the channel characteristics on the average picture quality, the sequences are mirrored and concatenated to a length of 4000 frames. In addition to characterizing the variation of the average picture quality with respect to the packet loss probability, it is also important to observe the instantaneous fluctuation in the picture quality resulting from Wyner-Ziv decoding or error concealment. This is done by observing a small simulation window of a few hundred frames.

- *Intra Macroblock Line Refresh*: To mitigate error propagation resulting from lost slices, one row of macroblocks shall be encoded using the intra mode in each frame. For a CIF frame, this is equivalent to a full intra refresh every 18 frames. We avoid encoding entire Intra frames because this results in a large spike in the bit rate and a correspondingly large buffering delay.
- *Rate control*: Constant Bit-Rate (CBR) coding is used in these simulations. To accomplish this, the rate control method provided in the H.264/AVC standard codec [107] is used at the encoder to determine the modulation of the quantization parameters while encoding the macroblocks in the video sequence. Once the encoding rates for the primary and redundant slices has been decided, the rate control algorithm for the primary picture proceeds independently from that for the redundant picture.
- *Redundant slice encoding*: The redundant slices are encoded at 25%, 50%, and 100% of the bit rate of the primary slices, respectively. This corresponds to SLEP schemes with different error protection capabilities. Note that these percentages refer to the bit rate of the redundant description and *not* to the transmitted Wyner-Ziv bit rate. The latter will be specified separately. To distinguish one SLEP scheme from another, based on the bit rate of the redundant slices and the bit rate of the Wyner-Ziv slices, we establish a naming convention in Section 4.4.3 and use it throughout the thesis.
- *Packetization and slices*: The *primary slices* are constrained to a fixed length (see below) at the beginning of the simulation. To this, we add the RTP, UDP



and IP headers which occupy a total of  $12+8+20=40$  bytes per slice. Smaller slice lengths result in a larger total number of slices, which provides greater error resilience but increases the header overhead. We do not optimize the slice lengths, but choose them so as to obtain a reasonable tradeoff between header overhead and error resilience. Since each video slice travels inside an IP packet, the words “slice” and “packet” will be used interchangeably. Thus, unlike our experiments with the MPEG-2 codec [126, 127], the slices are no longer constrained to contain the same number of macroblocks. In other words, the primary slices can have different shapes. The packetization for the redundant slices is dictated by the shapes of the primary slices. In other words, a redundant slice is constrained to contain the same number of macroblocks as its corresponding primary slice.

- *Decoder-based error concealment:* Wyner-Ziv decoding is carried out as explained in 4.2.2. If  $N$  and  $K$  are the parameters of the Reed-Solomon Slepian-Wolf code, then Wyner-Ziv decoding is successful if the number of lost primary and parity slices is less than or equal to  $N - K$ . However, at very high error probabilities, the number of lost primary slices is too large and Wyner-Ziv decoding fails for some or all video frames. In this case, the non-normative error concealment scheme [180] included in the reference JVT codec is used to conceal the lost primary slices.

Sequence	Resolution	Primary Bit Rate (kb/s)	Frame Rate (frames/s)	Number of Frames
<i>Football</i>	SIF	1024	30	4000
<i>Bus</i>	CIF	1024	30	4000
<i>Mobile</i>	SIF	768	30	4000
<i>Foreman</i>	CIF	512	30	4000
<i>Coastguard</i>	CIF	512	30	4000
<i>Irene</i>	CIF	384	30	4000
<i>Akiyo</i>	CIF	200	30	4000

Table 4.1: Video sequences used in the SLEP simulations. CIF sequences have a size of  $352 \times 288$  pixels, while SIF sequences have a size of  $352 \times 240$  pixels.

### 4.4.2 Channel Simulations

The video data travel to the decoder in the form of RTP packets [187, 150]. Independent of the type of the network - wireline or wireless - the application layer sees a packet erasure channel. To elaborate: It is assumed that the lower layers possess their own error control mechanisms for responding to lost or corrupted packets. Examples of such mechanisms include channel coding at the transport layer, or retransmissions at the link layer. The application layer is assumed to be unaware of the particulars of these error control mechanisms. If they succeed, then a transmitted packet is correctly received and forwarded to the application layer. If they fail, then the application layer receives a notification that the packet is lost. Thus, irrespective of whether a network packet is (1) corrupted, (2) arrives too late after its specified deadline, or (3) is dropped altogether due to congestion, the application layer simply treats the event as an erasure.

In order to simulate packet losses in a video transmission experiment, it is assumed that packets are erased (lost at the application layer) with probability  $p$ . Packet losses are independent of each other and occur according to a Bernoulli distribution. In the next chapter, which examines the performance of an optimized SLEP system, more realistic channel traces obtained from actual Internet experiments [186] will be used.

We now consider the effect of the block length of the channel code. Recall that we use a  $(N, K)$  Reed-Solomon code as the Slepian-Wolf code, and that the code is applied after buffering  $K$  redundant slices. From the explanation above, the Wyner-Ziv decoding operation is analogous to erasure decoding. If we allow for very large encoding and decoding delays, then the most efficient erasure code satisfies

$$\frac{N}{K} \simeq 1 + p \quad \text{for very large } N \text{ and } K \quad (4.1)$$

In other words, the fractional redundancy introduced by the Reed-Solomon (RS) code equals the erasure probability, if the system can tolerate very high channel coding complexity and very high buffering delay. In practice, there are limits on both the allowable channel coding complexity and on the number of redundant slices  $K$ , which can be buffered before RS encoding/decoding. For RS codes with  $m$ -bit

symbols, the RS codewords belong to the Galois Field  $\text{GF}(2^m)$ , and  $N \leq 2^m - 1$ . In our simulations, we use  $m = 8$ . Therefore, the Slepian-Wolf encoding/decoding complexity is dependent on the buffering delay, i.e., on the values chosen for  $N$  and  $K$ . Further, the efficiency of the Slepian-Wolf code, measured as the percentage of the total transmitted bit rate allocated to parity symbols, is also dependent on the values of  $N$  and  $K$ . The selection of the appropriate buffering delay is explained below.

From the perspective of buffering delay, it is best to buffer the redundant slices belonging to only one frame. However, this results in an inefficient Slepian-Wolf code. From the perspective of Slepian-Wolf coding efficiency, it is best to divide a frame into very small slices, so that  $K$  and  $N$  can be made large. However, having a large number of slices introduces a very large header overhead and compromises the source coding efficiency of the video coder. In practice, it is difficult to characterize the effect of the slice size and buffering delay on the decoded picture quality and hence to find their optimum values. Besides, our goal is not to optimize slice size and buffering delay but to compare the error resilience of SLEP with that of traditional FEC, in *which both schemes use exactly the same primary (systematic) video description*. Therefore, to resolve the buffering-delay/slice-size trade-off, while ensuring fairness between SLEP and FEC, the following settings<sup>3</sup> are used in all the experiments reported:

1. The buffering delay is 1/3 second. Thus, for a sequence encoded at 30 frames per second, the encoder (decoder) buffers the redundant slices belonging to 10 frames before performing RS encoding (decoding).
2. The primary slices are constrained to be 500 bytes long if the source coding bit rate is less than 512 kb/s. Otherwise, they are chosen to be 800 bytes long. As explained above, the redundant slices are forced to have the same shapes as that of the corresponding primary slices. Thus, while the lengths of the primary

---

<sup>3</sup>These settings were based on discussions that occurred during a core experiment conducted within the ITU-T/MPEG Joint Video Team (JVT) from April to October 2006. The experiment investigated standardization of SLEP under H.264/AVC, and is described in further detail in Chapter 6.

slices are constant (500 or 800), the lengths of the redundant slices depend upon the quantization parameters used while encoding the redundant description.

### 4.4.3 Designation of SLEP Schemes

The SLEP scheme used is designated as SLEP- $\mathbf{R}$ - $\mathbf{W}$ , where  $\mathbf{R}$  is the bit rate of the redundant slices expressed as a percentage of the bit rate of the primary slices, and  $\mathbf{W}$  is the Wyner-Ziv bit rate expressed as a percentage of the bit rate of the primary slices. This is explained pictorially in Fig. 4.10. Thus, for the *Football* sequence encoded at 1024 kb/s, SLEP-25-10 denotes a scheme in which the redundant slices are encoded at 256 kb/s, while the Reed-Solomon parity slices in the Wyner-Ziv bit stream constitute a bit rate of 102.4 kb/s. The total transmitted bit rate in this case is 1024 kb/s for the primary picture + 102.4 kb/s for the Wyner-Ziv bit stream. As explained in the system implementation in Section 4.2.1, the redundant slices themselves are not transmitted, but the bit rate at which the redundant slices are encoded is an important variable in the system design because it determines the quantization mismatch between the primary and redundant slices.

Again, if SLEP is implemented as in Fig. 4.2, SLEP-100- $\mathbf{W}$  is simply FEC since the redundant description is the same as the primary description.  $\mathbf{W}$  would then indicate the channel coding redundancy introduced by FEC. Thus, for our implementation, FEC is a special case of SLEP<sup>4</sup>.

For convenience, whenever an experiment is carried out in which the percentage of Wyner-Ziv bits  $\mathbf{W}$  is the same for all the schemes being compared,  $\mathbf{W}$  is specified in the results but omitted from the scheme designation. To avoid any confusion,  $\mathbf{R}$  is always specified.

### 4.4.4 Comparison of SLEP and FEC

This section presents a comparison of the average and instantaneous picture quality delivered by the following schemes:

---

<sup>4</sup>There are other ways of implementing a SLEP system [9, 125] in which FEC would not be a special case of SLEP and hence, would not be designated as SLEP-100- $\mathbf{W}$ .

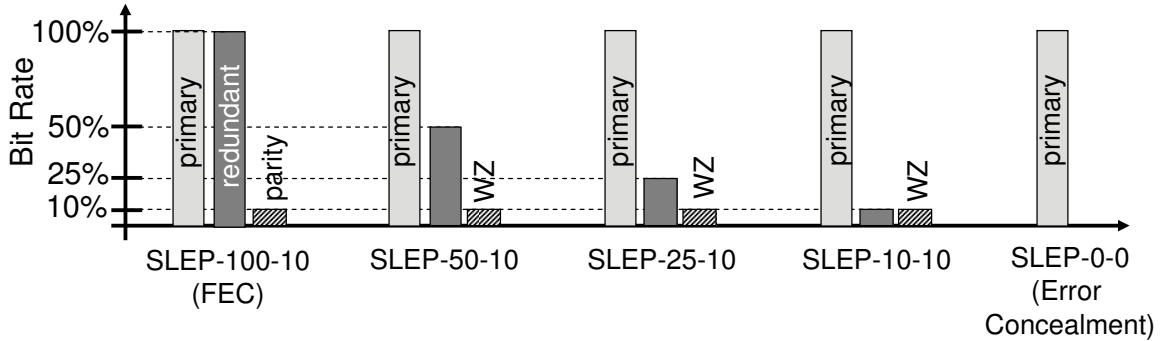


Figure 4.10: A SLEP scheme is designated by two numbers. The first number,  $\mathbf{R}$ , expresses the encoding bit rate of the redundant slices as a percentage of the bit rate of the primary video signal. The second number,  $\mathbf{W}$ , expresses the Wyner-Ziv bit rate as a percentage of the bit rate of the primary video signal. The redundant description is not transmitted, but reconstructed at the decoder.

1. A variety of SLEP schemes which have different redundant descriptions but the same Wyner-Ziv bit rate, i.e., the same error resilience bit rate. Specifically, in the designation SLEP- $\mathbf{X}$ - $\mathbf{Y}$ , the values of  $\mathbf{X}$  used are 25%, and 50%, while the value of  $\mathbf{Y}$  is kept constant at 10%. Therefore, as noted in Section 4.4.3, the schemes are referred to as SLEP- $\mathbf{X}$  (with  $\mathbf{X} = 25\%$  or  $50\%$ ) with the implicit understanding that the Wyner-Ziv bit rate used for each SLEP scheme is the same, i.e., 10% of the primary bit rate.
2. An FEC scheme, in which Reed-Solomon coding is directly applied to the primary video signal to generate parity slices. The bit rate of the parity slices is 10% of that of the primary slices. For the current implementation, this FEC scheme is a special case of SLEP in which the primary and redundant descriptions are identical.
3. The non-normative decoder-based error concealment algorithm [180] provided as part of the H.264/AVC video codec. This scheme does not transmit any extra bits for error resilience and relies solely on information available at the decoder. Whenever possible, it estimates motion vectors for the lost video slices and conceals the lost portions of the video frame using one or more temporally

previous frames<sup>5</sup>. The mode decisions are the same as those used in the competing SLEP and FEC schemes, and error resilience tools such as forced intra coded macroblocks are not used.

The bit rate used to encode the primary slices is specified in Table 4.1.

First, consider the variation of the average Peak Signal-to-Noise Ratio (PSNR) with increasing percentage of packet loss, as shown in Fig. 4.11. In all simulations, the following relations are used to determine the PSNR for frame  $n$ , and the average PSNR for a video sequence of 4000 frames:

$$\begin{aligned} \text{PSNR}[n] &= 10 \log_{10} \frac{255^2}{\text{MSE}[n]} \\ \text{Average MSE} &= \frac{1}{4000} \sum_{n=1}^{4000} \text{MSE}[n] \\ \text{Average PSNR} &= 10 \log_{10} \frac{255^2}{\text{Average MSE}} \end{aligned}$$

where  $\text{MSE}[n]$  is the mean squared error between the decoded version and the original uncoded version of frame  $n$ . In general, when a significant portion of a video frame is lost, decoder-based error concealment is unable to conceal the packet losses completely. Besides, due to the predictive coding structure, the concealment artifacts propagate to the succeeding frames. Therefore, as seen in Fig. 4.11, error concealment provides the worst average PSNR among all the schemes considered. More importantly, the following trends are observed in the comparison of FEC and SLEP schemes:

1. When the packet loss percentage is low, the end-to-end distortion is dominated by the quantization mismatch between the primary and redundant descriptions. Thus FEC, which has zero quantization mismatch, provides the highest average PSNR while SLEP-25, which has the highest quantization mismatch provides the lowest PSNR among the SLEP schemes.

---

<sup>5</sup>When the number of parity symbols is insufficient for erasure decoding, the receivers in schemes (1) and (2) employ the decoder-based error concealment scheme as a last resort.

2. As the packet loss percentage increases, the end-to-end distortion is dominated by error-concealment artifacts resulting from the failure of FEC and Wyner-Ziv decoding. Recall that while both FEC and SLEP have the same error resilience bit rate, FEC applies error protection to the primary slices, while SLEP applies error protection to the redundant slices which are coarsely quantized. Therefore, for the given bit budget for error resilience, Wyner-Ziv protection in SLEP is stronger than conventional FEC protection. Thus, at high packet loss percentage, the average PSNR of SLEP-25 is higher than that of SLEP-50, which, in turn, is higher than that of FEC.

Next, consider the instantaneous variation of the PSNR of the decoded video frame. In Fig. 4.12, the PSNR is plotted against the frame number for an experiment carried out with 10% packet loss. The simulations in Fig. 4.12 were carried over video sequences of 4000 frames. An arbitrarily chosen 200-frame window, consisting of frame numbers 651-850 from each simulation, is displayed in the plots. All transmission schemes are afflicted by the same error trace, but the loss patterns in these instantaneous PSNR plots may appear uncorrelated because the three schemes have different Wyner-Ziv protection (or parity protection in case of FEC). Note that, in the case of Reed-Solomon codes with infinitely long block lengths, allocating 10% of the bit rate for parity information would be sufficient to provide FEC protection at 10% packet loss. However, owing to the use of finite block lengths in practical transmission systems, this allocated parity bit rate cannot always provide erasure protection. Thus, FEC fails in some instances, and this results in a rapid reduction in the frame PSNR. In contrast, as explained earlier, SLEP has stronger error protection at the same error resilience bit rate, because it uses smaller redundant descriptions. Due to the quantization mismatch between the primary and redundant slices, the frame PSNR after successful SLEP decoding is slightly lower than that in the error-free case, but a drastic reduction in picture quality is avoided.

For visual comparison, a video frame from each of the three sequences is shown in Figs. 4.13, 4.14 and 4.15. It is observed that the subjective degradation associated with the quantization mismatch from Wyner-Ziv decoding is not as severe as the degradation associated with error concealment artifacts.

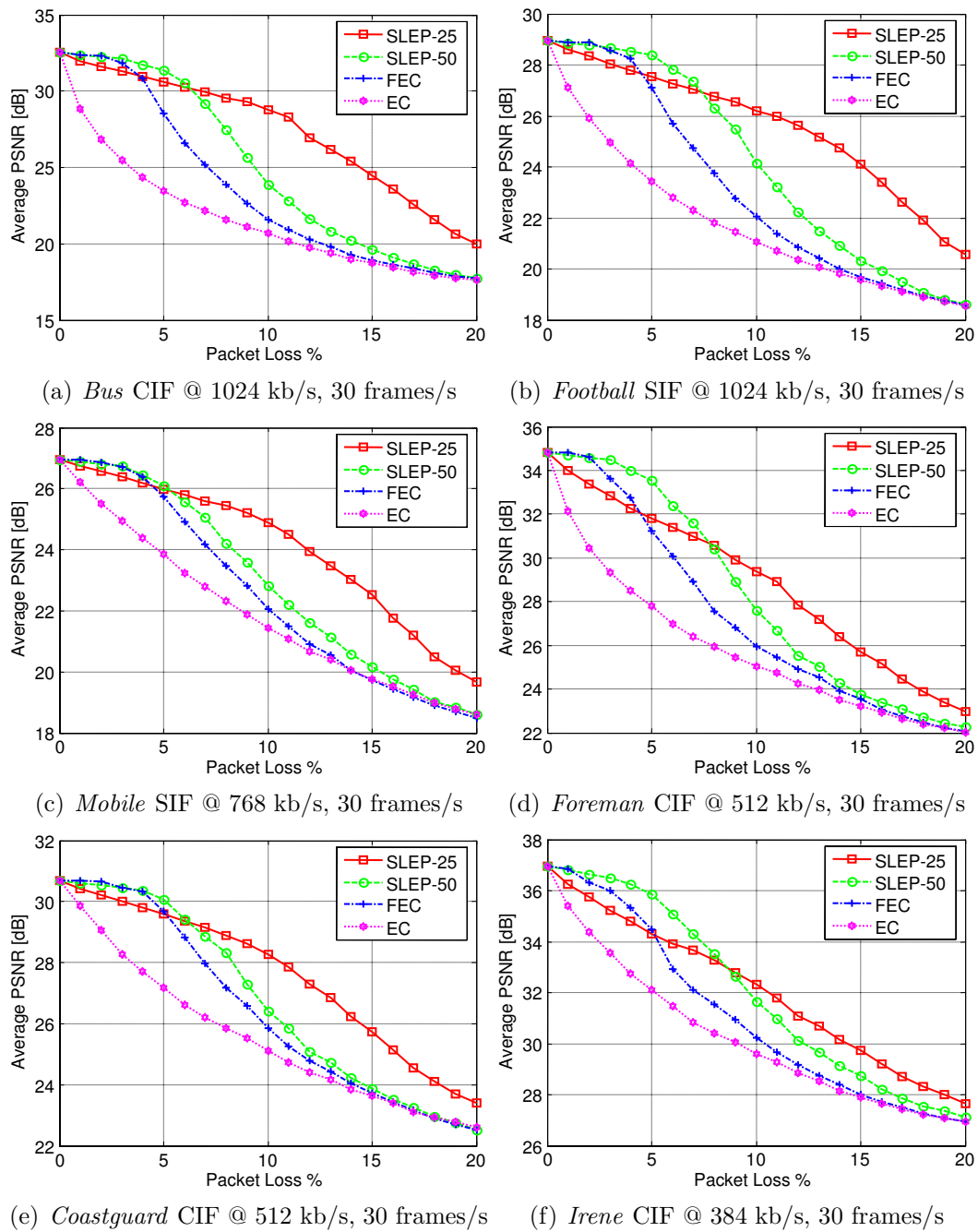


Figure 4.11: Comparison of FEC with SLEP schemes in which the redundant slices are encoded at 50 % and 25 % of the bit rate of the primary slices. The error resilience bit rate for all schemes, except decoder based error concealment, marked “EC” above, is 10 % of the source coding bit rate of the primary slices. When a coarsely quantized redundant description is used, the error robustness increases at the expense of an increased quantization mismatch between the primary and redundant descriptions at the decoder.



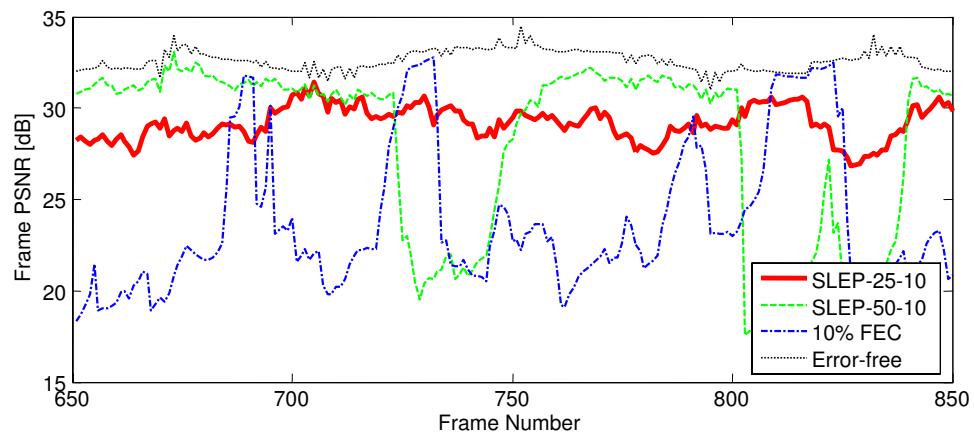
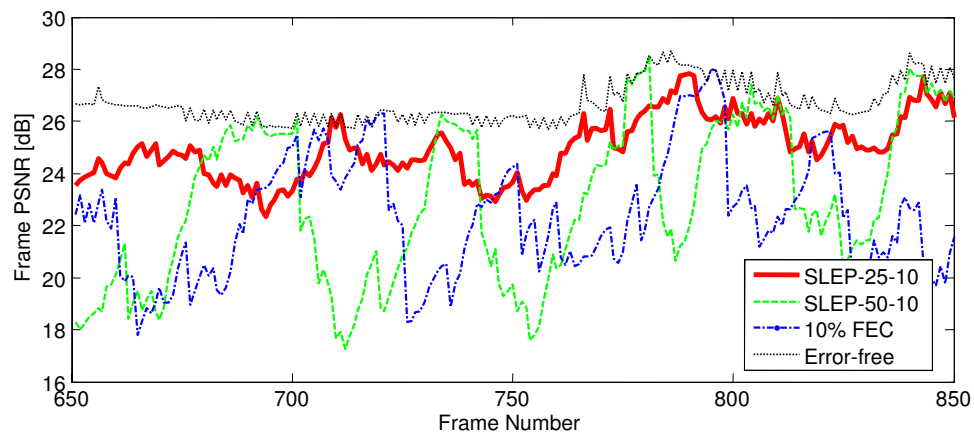
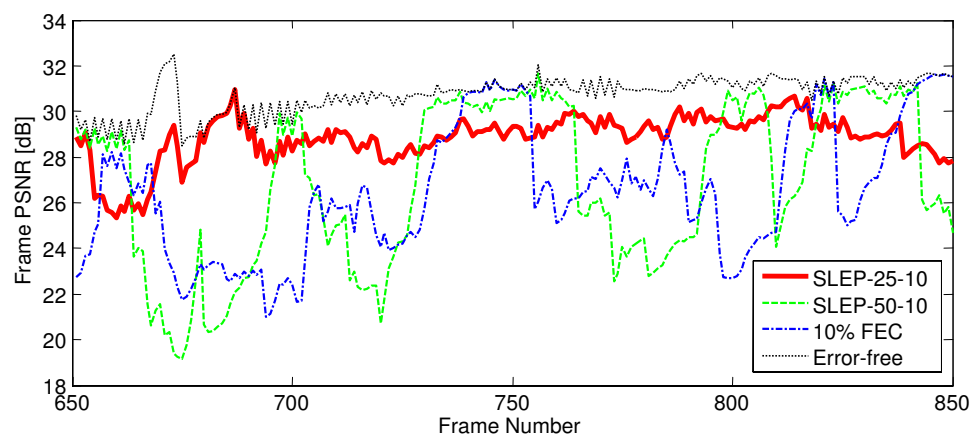
(a) *Bus* CIF @ 1024 kb/s, 30 frames/s(b) *Mobile* SIF @ 768 kb/s, 30 frames/s(c) *Coastguard* CIF @ 512 kb/s, 30 frames/s

Figure 4.12: When coarse quantization is used in the redundant description, there is a small reduction in the decoded frame PSNR compared to the error-free case. In return, drastic reduction in picture quality is avoided. At a high packet loss rate of 10%, SLEP-25-10 provides the smallest instantaneous fluctuation in frame PSNR, followed by SLEP-50-10 followed by FEC.



Figure 4.13: Decoded frames of the *Bus* CIF sequence encoded at 1024 kb/s, when the parity bit rate is 10% of the primary source coding bit rate. With FEC (left), decoding fails for some portions of the frame, reducing the PSNR to 20.2 dB. With SLEP scheme for a redundant description encoded at 25% of the primary bit rate (right), successful Wyner-Ziv decoding results in a PSNR of 30.3 dB, much closer to the error-free PSNR of 32.2 dB.



Figure 4.14: Decoded frames of the *Mobile* SIF sequence encoded at 768 kb/s, when the parity bit rate is 10% of the primary source coding bit rate. With FEC (left), decoding fails for some portions of the frame, reducing the PSNR to 17.3 dB. With SLEP scheme for a redundant description encoded at 25% of the primary bit rate (right), successful Wyner-Ziv decoding results in a PSNR of 25.3 dB, close to the error-free PSNR of 25.9 dB.

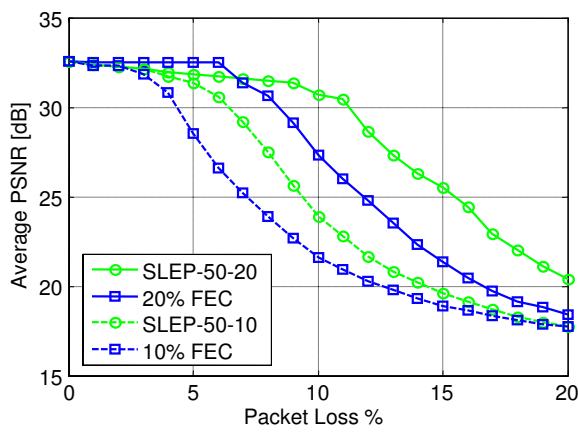


Figure 4.15: Decoded frames of the *Coastguard* CIF sequence encoded at 512 kb/s, when the parity bit rate is 10% of the primary source coding bit rate. With FEC (left), decoding fails for some portions of the frame, reducing the PSNR to 22.7 dB. With SLEP scheme for a redundant description encoded at 25% of the primary bit rate (right), successful Wyner-Ziv decoding results in a PSNR of 31.2 dB, much closer to the error-free PSNR of 32.9 dB.

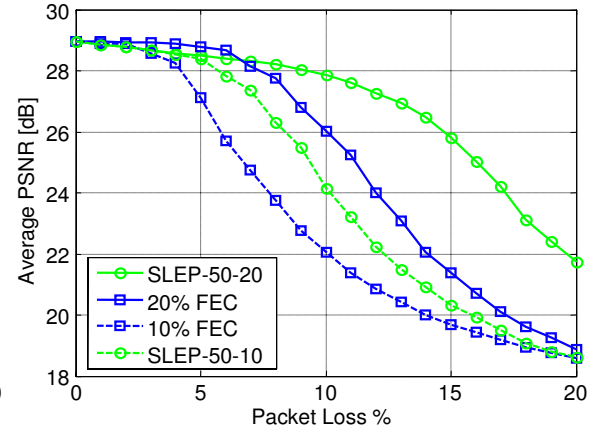
#### 4.4.5 Effect of increasing Wyner-Ziv Bit Rate

In the previous subsection, the Wyner-Ziv bit rate was kept constant at 10%, and the error resilience of SLEP was investigated for various redundant descriptions. In this subsection, the redundant description is kept constant, while the Wyner-Ziv bit rate is increased from 10% to 20%. As expected, the error resilience improves and superior average PSNR is obtained at higher packet erasure percentages (Fig. 4.16). The amount of improvement in error resilience is dependent on the video content. For example, the increase in robustness is more pronounced for video sequences which have higher scene activity and irregular motion fields (such as the *Football* sequence) than for sequences with relatively less motion (such as the *Coastguard* sequence).

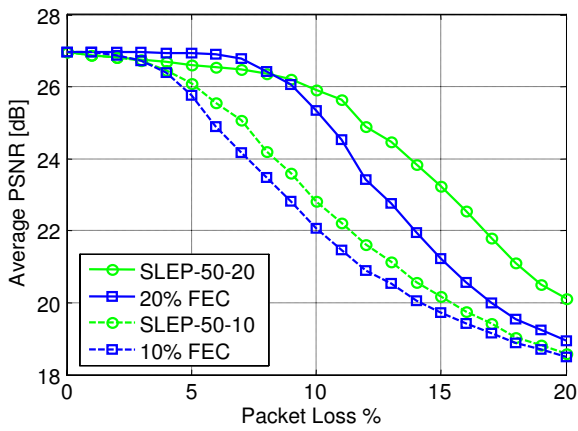
From the discussion of these results, it is clear that the Wyner-Ziv bit rate (i.e., the error resilience bit rate) and the source coding bit rate of the redundant description must be selected judiciously in order to achieve high error resilience while ensuring that the quantization mismatch between the primary and redundant descriptions is



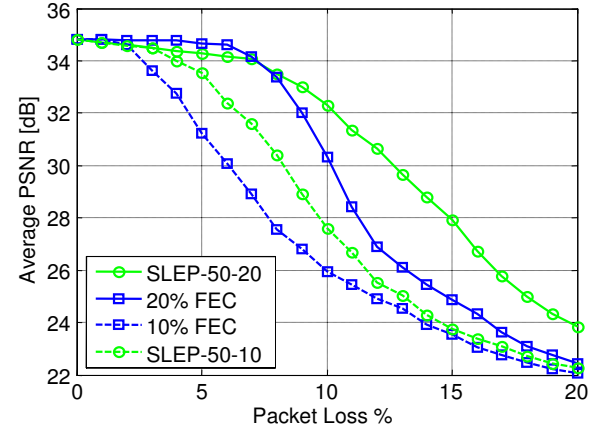
(a) *Bus* CIF @ 1024 kb/s, 30 frames/s



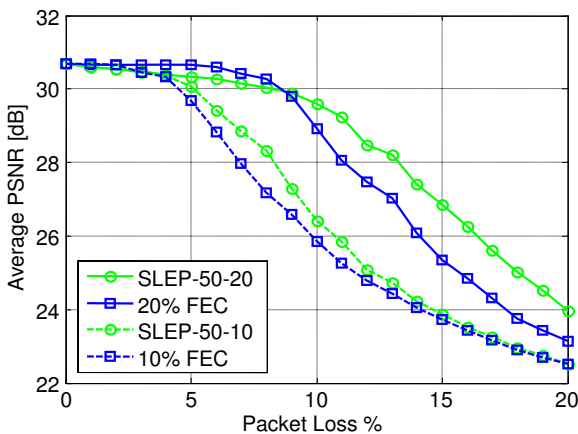
(b) *Football* SIF @ 1024 kb/s, 30 frames/s



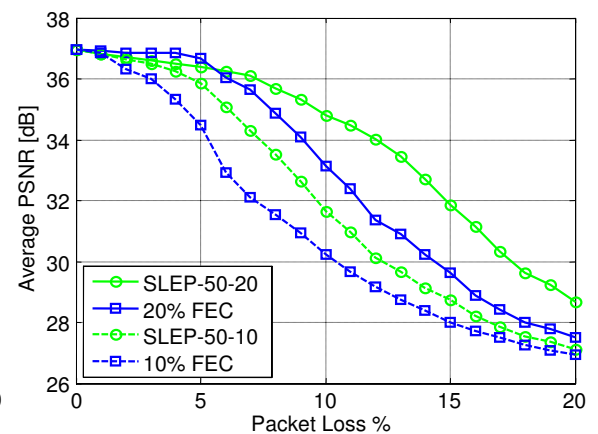
(c) *Mobile* SIF @ 768 kb/s, 30 frames/s



(d) *Foreman* CIF @ 512 kb/s, 30 frames/s



(e) *Coastguard* CIF @ 512 kb/s, 30 frames/s



(f) *Irene* CIF @ 384 kb/s, 30 frames/s

Figure 4.16: With a fixed redundant description, increasing the Wyner-Ziv bit rate results in an increase in error resilience. FEC with the same parity bit rate is displayed for comparison.

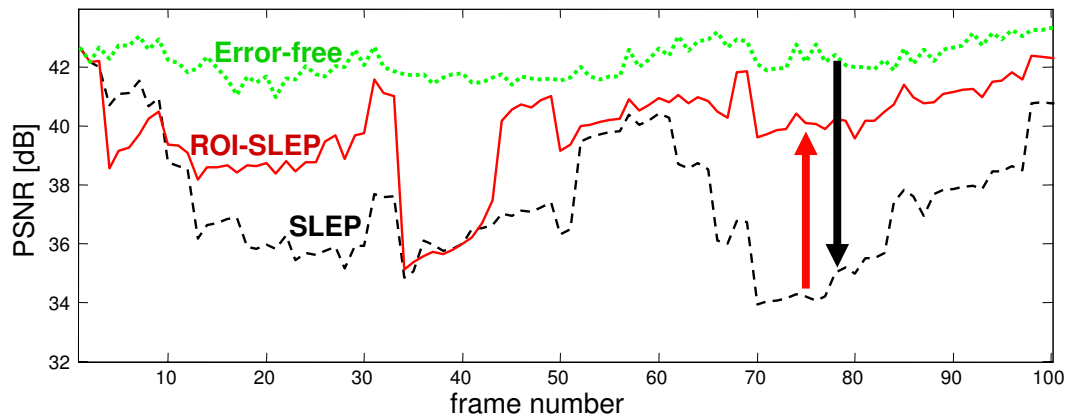


Figure 4.17: Decoded frames of the *Akiyo* CIF sequence encoded at 200 kb/s, when the parity bit rate is 10% of the primary source coding bit rate, and the redundant slices in the Wyner-Ziv codec are encoded at 40 kb/s. When SLEP is applied to the entire picture, there are smearing artifacts when intra-coded macroblocks must be replaced by their redundant versions. If the same bit rate is concentrated inside the ROI, then the picture quality after Wyner-Ziv decoding does not suffer from smearing artifacts.

visually acceptable. The optimal selection of these bit rates is the subject matter of Chapter 5.

#### 4.4.6 Benefit of ROI-Based SLEP

As explained in Section 4.3, the quantization mismatch associated with Wyner-Ziv decoding can be mitigated by performing a redundant encoding and error protection of the ROI only. Fig. 4.17 shows an instantaneous PSNR trace for the decoded frames of the *Akiyo* sequence. When SLEP is applied to the entire video frame, the quantization mismatch between the primary signal (encoded at 200 kb/s) and the redundant signal (encoded at only 40 kb/s) results in a large drop in PSNR. When SLEP is applied only to the ROI, this large reduction in frame PSNR is avoided.

Two frames from the decoded trace of Fig. 4.17 are displayed in Fig. 4.18. The smearing artifacts associated with coarsely quantized, redundant, intra coded macroblocks are significantly mitigated by using the ROI-based scheme.

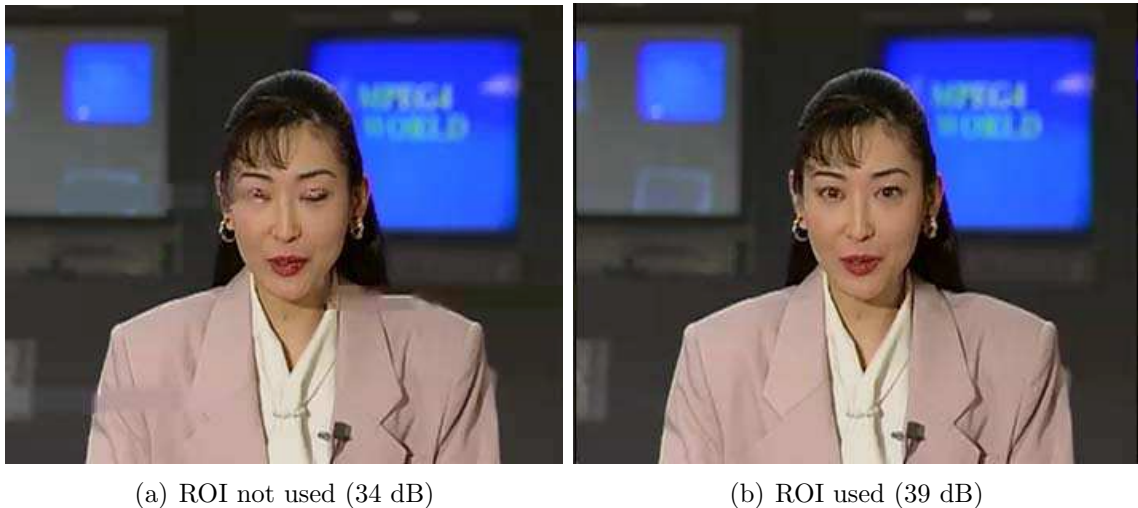


Figure 4.18: Applying SLEP to the ROI results in superior decoded picture quality because it allows finer quantization in the ROI. Further, this usually results in fewer redundant slices, and stronger Wyner-Ziv protection as for the case of frame no. 72 from the trace shown in Fig. 4.17.

#### 4.4.7 SLEP with Multiple Redundant Descriptions

The SLEP scheme described in this chapter can be extended to use multiple redundant descriptions. In such a system, several redundant descriptions are encoded and the available error resilience bit rate is divided among multiple Wyner-Ziv bit streams. A coarsely quantized redundant description would be assigned stronger Reed-Solomon protection to guarantee a minimum decoded picture quality for some high erasure probability. Redundant descriptions with finer quantization can be protected with a weaker Reed-Solomon code so that they can be recovered only at lower erasure probabilities. This can enable the receiver to exploit the tradeoff between error resilience and picture quality better than in the case with a single redundant description. Appendix B contains experimental simulations of a SLEP system with multiple embedded redundant descriptions applied to robust MPEG-2 video transmission.

## 4.5 Summary

A SLEP scheme has been implemented using H.264/AVC redundant slices and Reed-Solomon codes. The Wyner-Ziv bit stream is constructed by generating a redundant video description, applying a Reed-Solomon code, and transmitting only the parity symbols. For low-motion sequences, it is beneficial to apply SLEP to a region of interest in a video frame, which can be specified efficiently by means of Flexible Macroblock Ordering (FMO). Experimental simulations show that the robustness of a SLEP scheme increases when coarsely quantized redundant description are used. This increase in robustness is in exchange for some residual distortion after Wyner-Ziv decoding, which is caused by the quantization mismatch between the redundant and the primary descriptions.

The designation of a SLEP scheme gives an idea about its error resilience and about the quantization mismatch introduced due to Wyner-Ziv decoding. For example, the error resilience of SLEP-50-10 is greater than that of FEC (equivalently SLEP-100-10) but less than that of SLEP-25-10. This is because, for a constant Wyner-Ziv bit rate, the scheme with the smaller redundant slices has stronger Wyner-Ziv protection. At the same time, the quantization mismatch introduced by SLEP-50-10 is greater than that introduced by FEC (which introduces no quantization mismatch) but less than that of SLEP-25-10.

## Chapter 5

# SLEP Modeling and Optimization

In the previous chapter, a SLEP scheme was implemented for the robust transmission of H.264/AVC compressed video. The Wyner-Ziv bit stream is generated by applying a channel code to a redundant description of the original video signal. To use SLEP for robust video transmission, it is necessary to optimally split the available bit rate into two portions: (1) the source coding bit rate of the video signal, and (2) the Wyner-Ziv bit rate, which, in the implementation of Chapter 4, is the bit rate of the parity symbols obtained after applying the channel code. This is reminiscent of the bit allocation problem in joint source/channel coding. However, as noted in Section 3.1, there is a third degree of freedom in the optimization of a SLEP system, which is the bit rate used for encoding the redundant slices. These three bit rates can be optimized for the given packet loss probability only if their combined impact on the overall rate-distortion performance of the SLEP system is known. In other words, we would like to investigate the dependence of the end-to-end video quality on the source coding bit rate of the primary slices, the source coding bit rate of the redundant slices and the strength of the Wyner-Ziv code. In this chapter, a model is developed for this purpose.

In Section 5.1, we model motion compensated predictive coding of a pixel in the current video frame, and motion compensated decoding based on the reconstructed past pixel values and the received prediction residual. Assumptions are specified on the prediction residual, quantization errors, and on the erasures introduced by



the channel. A recursive relation is obtained for the evolution of the end-to-end distortion in a decoded video packet. The rate-distortion functions for the primary and redundant descriptions are described using a parametric model.

In Section 5.2, the MSE distortion predicted by the model is compared to that obtained from the simulations using the setup of Chapter 4. The end-to-end rate distortion tradeoff of the entire SLEP system is studied in further detail. Assuming that Wyner-Ziv decoding is always successful, a closed form expression is derived for the final end-to-end distortion. The overall distortion is expressed as a function of the packet loss probability and the source coding distortions of the primary and redundant descriptions. Finally, in Section 5.3, we incorporate the model into the H.264/AVC video encoder and use it to optimize the bit rates of the primary and redundant descriptions at the given packet loss probability.

## 5.1 Distortion-Rate Modeling of SLEP

### 5.1.1 Motion Compensated Encoding and Decoding

Since we are concerned with modeling error propagation, the ensuing treatment assumes a video sequence consisting only of predictively encoded frames (i.e., P frames). Modifications for intra-coded (I) frames and bidirectionally predicted (B) frames are straightforward.

Let the original value of a pixel at location  $i$  in the  $n^{\text{th}}$  frame be denoted by a random variable  $X_n^i$ . This pixel is predicted from another pixel at location  $j$  in the encoder's local reconstruction of the previous frame, i.e., the previously reconstructed value,  $\hat{X}_{n-1}^j$  serves as the predictor for  $X_n^i$ . To prevent a source coding mismatch between the encoder and the decoder, the predictor used is the *reconstructed* pixel value  $\hat{X}_{n-1}^j$  and not the original pixel value  $X_{n-1}^j$ . Thus,

$$X_n^i = \hat{X}_{n-1}^j + V_n^i \quad (5.1)$$

where,  $V_n^i$  denotes the error in the prediction of the current pixel, which is transformed, quantized, entropy-coded and transmitted to the decoder. The encoder also

retains  $\widehat{V}_n^i$ , which is obtained after inverse quantization and inverse transformation of  $V_n^i$ . Finally,  $\widehat{X}_n^i$ , the locally reconstructed value of  $X_n^i$ , is obtained as

$$\widehat{X}_n^i = \widehat{X}_{n-1}^j + \widehat{V}_n^i \quad (5.2)$$

With an error-free channel, the decoder would receive  $\widehat{V}_n^i$  and precisely reconstruct  $\widehat{X}_n^i$ . The Mean Squared Error (MSE) distortion in the primary slices, resulting from the quantization of the prediction residual, is then given by  $\mathcal{D}_p = \text{E}(X_n^i - \widehat{X}_n^i)^2 = \text{E}(V_n^i - \widehat{V}_n^i)^2$ .

With an error-prone channel, the decoder would receive  $\widetilde{V}_n^i \neq \widehat{V}_n^i$  in general. Thus, the decoder's reconstruction of  $X_n^i$  is given by:

$$\widetilde{X}_n^i = \widetilde{X}_{n-1}^j + \widetilde{V}_n^i \quad (5.3)$$

where  $\widetilde{X}_{n-1}^j$  is a possibly error-prone reconstruction of  $X_{n-1}^j$ , the pixel used for motion compensation. In practice, motion estimation and motion compensation are performed using blocks of  $4 \times 4$ ,  $8 \times 8$ ,  $16 \times 16$ ,  $8 \times 16$  or  $16 \times 8$  pixels. We consider pixel-level motion compensation in the present analysis because this makes it convenient to write the expressions for MSE distortion.

Consider the encoding of the redundant slices in SLEP, in which the reference pixel is the same as that used for encoding the primary slices. A general encoding of a redundant slice need not be restricted in this way. However, as explained in our SLEP implementation in Section 4.2.1, we constrain the redundant slices to use the same reference pixels as the corresponding primary slices in order to mitigate the error propagation that would result when the redundant slice is decoded. Thus, the unquantized prediction error  $V_n^i$  for the redundant slices is the same as that in (5.1) above. However, due to coarser quantization, the redundant reconstructed prediction error is  $\widehat{\widetilde{V}}_n^i \neq \widehat{V}_n^i$  in general. Then,  $\widehat{\widetilde{X}}_n^i$ , the redundant locally reconstructed value of  $X_n^i$  at the encoder, is obtained as

$$\widehat{\widetilde{X}}_n^i = \widehat{\widetilde{X}}_{n-1}^j + \widehat{\widetilde{V}}_n^i \quad (5.4)$$

and the MSE distortion introduced due to the coarse quantization used in the redundant slices is given by  $\mathcal{D}_r = \mathbb{E}(X_n^i - \widehat{X}_n^i)^2 = \mathbb{E}(V_n^i - \widehat{V}_n^i)^2$ . According to the SLEP scheme described in Section 4.2,  $\widehat{V}_n^i$  can be recovered at the receiver if Wyner-Ziv decoding is successful for the given bit rate assignment.

While modeling the end-to-end MSE, the following simplifying assumptions are made about the prediction error, the quantization error processes, and the process which introduces erasures during transmission:

1. It is assumed that, at the pixel location  $i$ , the prediction residual,  $V_n^i$ , and its quantized versions,  $\widehat{V}_n^i$  and  $\widetilde{V}_n^i$ , have zero mean over the duration of the sequence. i.e.,

$$\mathbb{E} V_n^i = \mathbb{E} \widehat{V}_n^i = \mathbb{E} \widetilde{V}_n^i = 0 \quad (5.5)$$

2. The quantization errors in the current sample,  $V_n^i - \widehat{V}_n^i$  and  $V_n^i - \widetilde{V}_n^i$  are assumed to be respectively independent of  $X_{n-1}^i - \widehat{X}_{n-1}^i$  and  $X_{n-1}^i - \widetilde{X}_{n-1}^i$ , the errors in the past samples. These assumptions are similar to those used in the analysis of the DPCM codec for the first-order Markov source considered in Chapter 3.
3. The quantization errors  $V_n^i - \widehat{V}_n^i$  and  $V_n^i - \widetilde{V}_n^i$  are assumed to be independent of the errors  $\widehat{X}_n^i - \widetilde{X}_n^i$  and  $\widetilde{X}_n^i - \widehat{X}_n^i$ . Note that  $\widetilde{X}_n^i$  can contain error energy contributed by (a) the quantization mismatch in the current sample  $\widehat{V}_n^i - \widetilde{V}_n^i$  or (b) the erasure of both the current quantized prediction residuals  $\widehat{V}_n^i$  and  $\widetilde{V}_n^i$ . In addition they contain error energy propagating from the errors that have occurred previously. Therefore, when the primary video signal is received correctly,

$$\begin{aligned} \mathbb{E}(X_n^i - \widetilde{X}_n^i)^2 &= \mathbb{E}(X_n^i - \widehat{X}_n^i + \widehat{X}_n^i - \widetilde{X}_n^i)^2 = \mathbb{E}(V_n^i - \widehat{V}_n^i + \widehat{X}_n^i - \widetilde{X}_n^i)^2 \\ &= \mathbb{E}(V_n^i - \widehat{V}_n^i)^2 + \mathbb{E}(\widehat{X}_n^i - \widetilde{X}_n^i)^2 = \mathcal{D}_p + \mathbb{E}(\widehat{X}_n^i - \widetilde{X}_n^i)^2 \end{aligned} \quad (5.6)$$

Similarly, for the case in which the redundant slices are decoded instead of the primary slices,

$$E(X_n^i - \tilde{X}_n^i)^2 = \mathcal{D}_r + E(\hat{\tilde{X}}_n^i - \tilde{X}_n^i)^2 \quad (5.7)$$

The expressions used to describe the encoding and decoding process will be now used to find the end-to-end MSE distortion at the decoder. Apart from the fact that the predictor is obtained via motion compensation, the predictive coding scheme described here is similar to the compression scheme used for the Markov source in Chapter 3. The decoding scheme, however, is different because the Wyner-Ziv codec, as described in Section 4.2, buffers a number of redundant slices before encoding or decoding the Reed-Solomon (Slepian-Wolf) code. If more redundant slices are buffered, the channel code becomes more efficient, but the delay involved in Reed-Solomon encoding and decoding increases. Optimizing this tradeoff is outside the scope of this work. However, the inefficiency associated with short block lengths will be captured in the model in the next section.

### 5.1.2 Distortion in the Decoded Video Sequence

The decoder performs different actions depending on whether the primary video packets are received or erased, and whether the number of received parity packets are sufficient for recovering the redundant slices via Reed-Solomon decoding. Assume that packets are erased (lost), randomly and uniformly, with probability  $p$ . Assume also that the location of the macroblocks contained in the lost packet is known. Then, we have the following cases:

1. **Primary slices received correctly:** With probability  $1 - p$ , a packet is received and decoded correctly by the main (primary) decoder and Wyner-Ziv decoding is unnecessary. The only source of error energy in this case is the error propagation from previous frames owing to decoding errors in the past

frames. Thus, the mean-squared error in the decoded pixel value is given by:

$$\begin{aligned} \mathcal{D}_n^{EP} &= \mathbb{E}(X_n^i - \tilde{X}_n^i)^2 \\ &= \mathbb{E}(X_n^i - \hat{X}_n^i)^2 + \mathbb{E}(\hat{X}_n^i - \tilde{X}_n^i)^2 + 2 \mathbb{E}[(X_n^i - \hat{X}_n^i)(\hat{X}_n^i - \tilde{X}_n^i)] \\ &= \mathcal{D}_p + \mathbb{E}(\hat{X}_{n-1}^j - \tilde{X}_{n-1}^j)^2 + 2 \mathbb{E}[(V_n^i - \hat{V}_n^i)(\hat{X}_{n-1}^j - \tilde{X}_{n-1}^j)] \end{aligned} \quad (5.8)$$

$$\simeq \mathcal{D}_p + \mathbb{E}(X_{n-1}^j - \tilde{X}_{n-1}^j)^2 - \mathbb{E}(X_{n-1}^j - \hat{X}_{n-1}^j)^2 + 0 \quad (5.9)$$

$$= \mathcal{D}_p + \mathcal{D}_{n-1}^{EE} - \mathcal{D}_p = \mathcal{D}_{n-1}^{EE} \quad (5.10)$$

where  $\mathcal{D}_{n-1}^{EE}$  denotes the overall MSE distortion in the previous frame. The second term in (5.8) is split into two terms in (5.9), after noting that the error  $X_n^j - \hat{X}_n^j$  introduced solely by quantization is independent of the error  $\hat{X}_n^j - \tilde{X}_n^j$  introduced solely by channel erasures. The third term in (5.8) vanishes because of the zero-mean assumption on the prediction residuals and the independence of  $V_n^i - \hat{V}_n^i$  from the past sample difference  $\hat{X}_{n-1}^j - \tilde{X}_{n-1}^j$ .

2. **Successful Wyner-Ziv decoding:** Wyner-Ziv decoding is invoked only when the primary video slice is lost. The probability that Wyner-Ziv decoding succeeds, denoted by  $p_{wz}$ , depends on the parameters,  $N$  and  $K$ , of the Reed-Solomon code. It is assumed that the location of the lost packet is known to the Wyner-Ziv decoder. Thus, the Reed-Solomon decoder has to perform erasure decoding. Similar to traditional erasure codes, Wyner-Ziv decoding in the current SLEP system succeeds if at least  $K$  out of  $N$  packets are received, but not otherwise. Since the Reed-Solomon code is applied across  $K$  redundant slices (Fig. 4.4), we have:

$$p_{wz} = \sum_{m=K}^{m=N-1} \binom{N-1}{m} (1-p)^m p^{N-1-m} \quad (5.11)$$

In the case of successful Wyner-Ziv decoding, error energy is contributed by the coarser quantization in the Wyner-Ziv decoded packet as well as by error

propagation from the previous frames. The distortion contribution is given by:

$$\begin{aligned}\mathcal{D}_n^{WZ} &= \mathbb{E}[(X_n^i - \tilde{X}_n^i)^2] \\ &= \mathbb{E}(X_n^i - \hat{X}_n^i)^2 + \mathbb{E}(\hat{X}_n^i - \tilde{X}_n^i)^2 + 2 \mathbb{E}[(X_n^i - \hat{X}_n^i)(\hat{X}_n^i - \tilde{X}_n^i)] \\ &= \mathcal{D}_r + \mathbb{E}(\hat{X}_{n-1}^j - \tilde{X}_{n-1}^j)^2 + 2 \mathbb{E}[(V_n^i - \hat{V}_n^i)(\hat{X}_{n-1}^j - \tilde{X}_{n-1}^j)]\end{aligned}\quad (5.12)$$

$$\simeq \mathcal{D}_r + \mathbb{E}(X_{n-1}^j - \tilde{X}_{n-1}^j)^2 - \mathbb{E}(X_{n-1}^j - \hat{X}_{n-1}^j)^2 + 0 \quad (5.13)$$

$$= \mathcal{D}_r + \mathcal{D}_{n-1}^{EE} - \mathcal{D}_p \quad (5.14)$$

As above, the second term in (5.12) is split into two terms in (5.13), after noting that the error  $X_n^j - \hat{X}_n^j$  introduced solely by quantization is independent of the error  $\hat{X}_n^j - \tilde{X}_n^j$  introduced solely by channel erasures. The third term in (5.12) vanishes because of the zero-mean assumption on the prediction residuals and the independence of  $V_n^i - \hat{V}_n^i$  from the past sample difference  $\hat{X}_{n-1}^j - \tilde{X}_{n-1}^j$ .

3. **Decoder-Based Error Concealment:** A decoder-based error concealment scheme must be used if a packet is lost in the systematic transmission and Wyner-Ziv protection is insufficient to reconstruct a redundant (i.e., coarsely quantized) version of the lost video slice. In this case, we assume that the lost slice is concealed using its co-located slice in the previous frame. The error energy is now contributed by the process of error concealment of the current packet as well as by the error propagation from the previous frames. The distortion contribution is then given by:

$$\begin{aligned}\mathcal{D}_n^{EC} &= \mathbb{E}(X_n^i - \tilde{X}_n^i)^2 = \mathbb{E}(X_n^i - \tilde{X}_{n-1}^i)^2 \\ &= \mathbb{E}(X_n^i - \hat{X}_{n-1}^i)^2 + \mathbb{E}(\hat{X}_{n-1}^i - \tilde{X}_{n-1}^i)^2\end{aligned}\quad (5.15)$$

$$= \mathbb{E}(X_n^i - \hat{X}_n^i)^2 + \mathbb{E}(\hat{X}_n^i - \hat{X}_{n-1}^i)^2 + \mathcal{D}_{n-1}^{EE} - \mathcal{D}_p \quad (5.16)$$

$$\begin{aligned}&= \mathcal{D}_p + \text{MSE}(n, n-1) + \mathcal{D}_{n-1}^{EE} - \mathcal{D}_p \\ &= \text{MSE}(n, n-1) + \mathcal{D}_{n-1}^{EE}\end{aligned}\quad (5.17)$$

where  $\text{MSE}(n, n - 1)$  is the mean squared error between the reconstructed current and previous frames. The third equality assumes that the temporal pixel variations,  $X_n^i - \widehat{X}_{n-1}^i$ , are independent of the errors,  $\widehat{X}_n^i - \widetilde{X}_n^i$ , introduced by the channel. The fourth equality assumes that quantization errors,  $X_n^i - \widehat{X}_n^i$ , are independent of the pixel variations and have zero mean as before. At the decoder, it is possible to use error concealment schemes that are more advanced than simple previous frame concealment. However a more advanced, and necessarily more complex, error concealment scheme may not always be available locally at the encoder, where the modeling and optimization is carried out. If such an advanced scheme is indeed available, then the term  $\text{MSE}(n, n - 1)$  in (5.17) may be replaced by the true average error energy calculated when the encoder locally implements the advanced error concealment scheme.

In summary, the decoded video packet in the  $n^{\text{th}}$  frame has a decoded prediction residual given by:

$$\widetilde{V}_n^i = \begin{cases} \widehat{V}_n^i & \text{w.p. } 1 - p \\ \widehat{\widehat{V}}_n^i & \text{w.p. } p p_{wz} \\ \text{erasure} & \text{w.p. } p(1 - p_{wz}) \end{cases}$$

Each case results in a different MSE distortion which was evaluated above. Using (5.10), (5.14) and (5.17) with the corresponding probabilities, the end-to-end distortion in the  $n^{\text{th}}$  frame due to all of the above effects is then given by:

$$\mathcal{D}_n^{EE} = (1 - p) \mathcal{D}_n^{EP} + p p_{wz} \mathcal{D}_n^{WZ} + p(1 - p_{wz}) \mathcal{D}_n^{EC} \quad (5.18)$$

At this point, it is worthwhile to elaborate on the implications of using a decoder-based error concealment scheme that is more sophisticated than previous frame error concealment. An example of such a scheme is the non-normative error concealment method [180] provided in H.264/AVC, which was used in the simulations in Chapter 4. A sophisticated concealment algorithm may, for example, obtain an estimate of the motion vectors of the lost macroblock by interpolating the motion vectors of its available neighbors. This reduces the magnitude of the term  $\mathcal{D}_n^{EC}$  in (5.18) and hence reduces the end-to-end distortion  $\mathcal{D}_n^{EE}$ . For the implementation under consideration,

this reduction is independent of the distortion due to the quantization mismatch in Wyner-Ziv decoding which is captured in the term  $\mathcal{D}_n^{WZ}$ . Therefore, any error resilience scheme, be it SLEP or FEC, benefits from superior error concealment. However, the trends that were observed in Chapter 4 are still preserved, i.e., at high packet loss rates, the error resilience of SLEP is superior to that of FEC which, in turn, is superior to that of decoder-based error concealment acting alone.

Note that an intra coded video slice stops the propagation of the error energy associated with the quantization mismatch and the concealment artifacts. Thus, if a macroblock is intra-refreshed every  $M$  frames, the average distortion over the  $M$  frames is given by:

$$\mathcal{D} = \frac{1}{M} \sum_{n=1}^M \mathcal{D}_n^{EE} \quad (5.19)$$

### 5.1.3 Encoder Distortion-Rate Model

In this section, we model the distortion-rate tradeoff for the primary and redundant slices. A number of such models have been developed for the purpose of rate-control in standardized video codecs [43, 73, 37, 143, 52, 170]. The encoding bit rate is controlled by manipulating the quantizer step-size, with the model being used to predict the distortion that would result if the rate was changed.

Following the analysis of [163, 83], the rate distortion performance of a video encoder is modeled by the following parametric equation:

$$D = D_m + \frac{\theta}{R - R_m} \quad (5.20)$$

where  $D_m$ ,  $R_m$ , and  $\theta$  are parameters to be determined from trial encodings.  $R$  can be measured in bits, or bits per second, or bits per frame, and the appropriate scaling factor can be lumped into the value of  $\theta$ .  $D$  is the MSE distortion for the number of frames for which the bit rate  $R$  is calculated.

To obtain the parameters  $D_m$ ,  $R_m$ , and  $\theta$ , a minimum of three trial encodings are necessary. With three or more rate-distortion pairs, the parameters are obtained



by least squares curve fitting. Note that the parameters depend not only on the sequence being coded, but also on the encoding parameters and mode decisions, such as the frequency with which intra frames are inserted, the number of B frames used, the number of reference frames used for predicting the current frame, etc. If these quantities change, the parameter values also change. For a given video sequence and motion compensation strategy, if the encoding parameters change by a small amount, the corresponding change in the model parameters may be obtained by linear interpolation [83].

One set of parameters ( $D_m$ ,  $R_m$ , and  $\theta$ ) may be obtained for the entire video sequence. However, from a practical standpoint, it is beneficial to calculate these parameters at short intervals, such as a Group of Pictures (GOP), or one or several seconds of video. Repeatedly updating the encoder model ensures that rate and distortion values in a temporal window reflect the scene content within that window.

Using (5.20), we can find the rate-distortion pairs for the primary slices, i.e.,  $(\mathcal{R}_p, \mathcal{D}_p)$ , as well as the redundant slices, i.e.,  $(\mathcal{R}_r, \mathcal{D}_r)$ . It is important to remember that the transmitted Wyner-Ziv bit rate, denoted by  $\mathcal{R}_{WZ}$ , is different from the bit rate of the redundant slices,  $\mathcal{R}_r$  (See Fig. 4.2). Wyner-Ziv encoding involves coarse quantization followed by Slepian-Wolf encoding. Coarse quantization reduces the encoding bit rate to  $\mathcal{R}_r \leq \mathcal{R}_p$ . Slepian-Wolf encoding, which is implemented using a channel encoder transmitting only parity symbols (or syndromes in other implementations) further reduces the bit rate to  $\mathcal{R}_{WZ} \leq \mathcal{R}_r$ . In fact, for the implementation of Chapter 4, we have

$$R_{WZ} = \frac{N - K}{K} R_r \quad (5.21)$$

Thus, for a chosen primary slice bit rate  $\mathcal{R}_p$  or redundant slice bit rate  $\mathcal{R}_r$ , the model can be used to calculate the MSE distortion,  $\mathcal{D}_p$  or  $\mathcal{D}_r$  respectively for the video sequence or a portion of the sequence. We assume that, for the portion of the video sequence over which the rate distortion modeling is performed at the encoder,

the rate and distortion are constant. With this assumption, we have:

$$\mathcal{D}_p = \text{E}[(X_n^i - \widehat{X}_n^i)^2] = D_{0p} + \frac{\theta_p}{\mathcal{R}_p - R_{0p}} \quad (5.22)$$

$$\mathcal{D}_r = \text{E}[(X_n^i - \widehat{X}_n^i)^2] = D_{0r} + \frac{\theta_r}{\mathcal{R}_r - R_{0r}} \quad (5.23)$$

where the encoding parameters,  $(D_{0p}, R_{0p}, \theta_p)$  and  $(D_{0r}, R_{0r}, \theta_r)$  are determined from trial encodings at the encoder. Since the redundant slices are coarsely quantized versions of the primary slices, it might be tempting to reuse the same parameters of the primary for the redundant slices. However, this would be incorrect, because the redundant description uses the higher quality reference frames from the primary description. For instance, as shown in Fig. 5.1, the redundant slices which are encoded at 500 kb/s use, for prediction, the locally decoded versions of a primary description encoded at 1 Mb/s. Therefore, by virtue of using better quality reference frames for prediction, the redundant description encoded at 500 kb/s can have a higher quality than a primary description encoded at 500 kb/s. Hence, the parameter values are different for the redundant and primary descriptions. The above discussion also indicates that the redundant and primary descriptions are coupled. Thus, for each point on the rate-distortion curve of the primary description, there is a new rate-distortion curve (hence a new set of parameters) for the redundant description. This is depicted for the *Foreman* sequence in Fig. 5.1.

Even though the redundant slices based on the 1 Mb/s primary slices uses superior quality reference pictures compared to the redundant slices based on the 384 kb/s primary slices, the latter encoding has higher quality at low bit rates. This is manifested as a crossing of the R-D curves in Fig. 5.1. We now explain why this is the case. Let  $\mathcal{R}_{p,\text{ref}}$  be the bit rate of the primary slices, and let the redundant slices be encoded at bit rate  $\mathcal{R}_r$ . According to the implementation of Fig. 4.2, the redundant slices encoded at rate  $\mathcal{R}_r$  must use the same reference pictures, motion vectors and coding modes as those used in the primary slices encoded at bit rate  $\mathcal{R}_{p,\text{ref}}$ . Let  $\delta = \mathcal{R}_{p,\text{ref}} - \mathcal{R}_r$ . Since the redundant description is coarsely quantized,

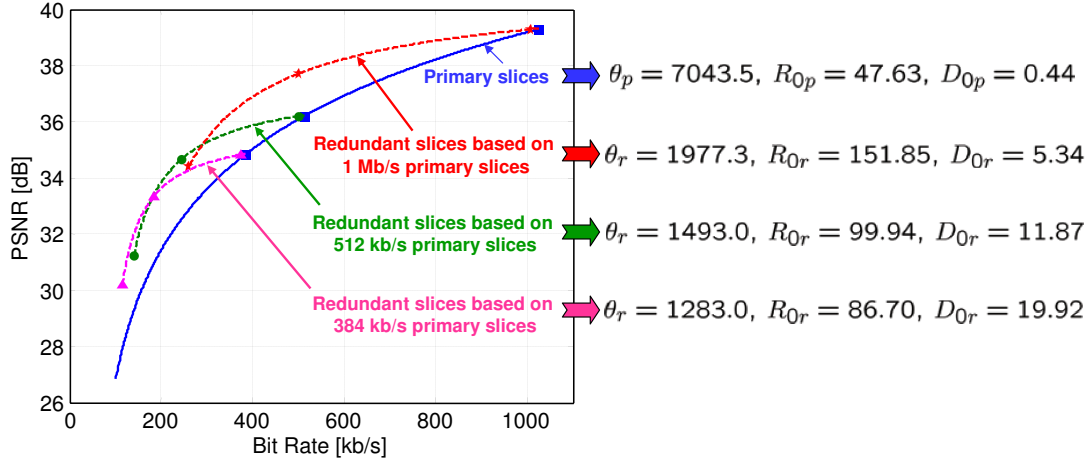


Figure 5.1: A few trial encodings (data points) are used to find the parametric rate-distortion curves for the primary and redundant descriptions of the *Foreman* CIF sequence. The parameters for the redundant description depend upon the primary description used as reference.

$\delta \geq 0$ . For large values of  $\delta$ , the motion vectors and coding modes selected for encoding the primary slices are not the best choices for encoding the redundant slices. In other words, at very low bit rates, the motion vectors and coding modes based on the 384 kb/s primary slices are more rate-distortion efficient compared with those based on the 500 kb/s primary slices, which in turn are more rate-distortion efficient than those based on the 1 Mb/s primary slices.

## 5.2 Resilience-Quality Tradeoff in SLEP

### 5.2.1 Overall Distortion for a Fixed Bit Rate Allocation

The model derived in the previous section is now used to study the resilience-quality tradeoffs associated with a SLEP system. To evaluate the accuracy of the model, the bit rates  $\mathcal{R}_p$ ,  $\mathcal{R}_r$  and  $\mathcal{R}_{WZ}$  are fixed, and the average MSE,  $\mathcal{D}$ , predicted by the model is compared with that achieved from simulations. The experimental setup is identical to that described in Section 4.4, with *I-P-P-P* coding structure, intra

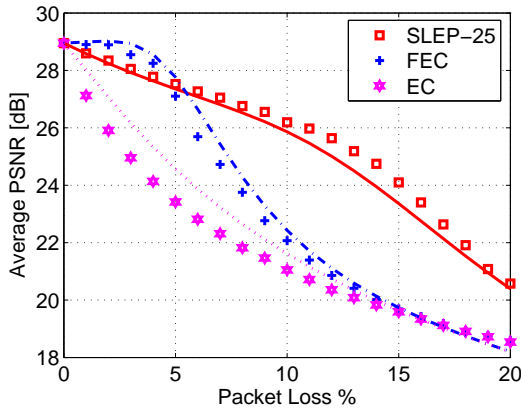
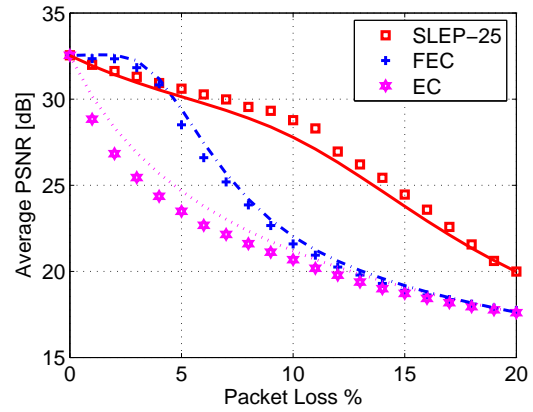
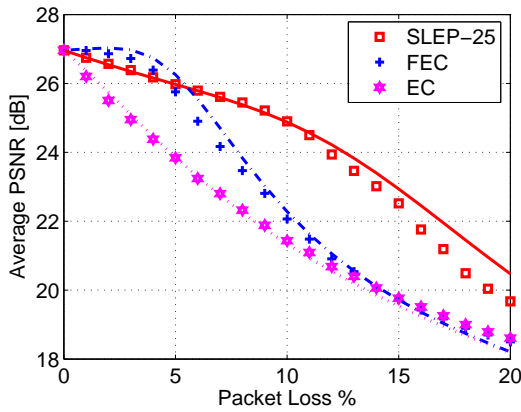
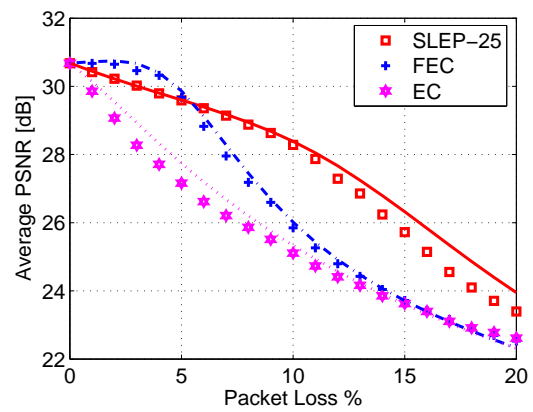
(a) *Football SIF* @ 1024 kb/s, 30 frames/s(b) *Bus CIF* @ 1024 kb/s, 30 frames/s(c) *Mobile SIF* @ 768 kb/s, 30 frames/s(d) *Coastguard CIF* @ 512 kb/s, 30 frames/s

Figure 5.2: The end-to-end average PSNR calculated by the model (solid lines) of Section 5.1.1 closely approximates that obtained by experimental simulation (data points). The Wyner-Ziv bit rate, i.e., the bit rate of the parity slices generated by the Reed-Solomon Slepian-Wolf encoder is fixed at 10% percent of bit rate of the primary slices. For both modeling and simulation, the average PSNR on the vertical axis is calculated from the average MSE of the sequence.

macroblock line refresh, encoder-based rate control, and non-normative, decoder-based error concealment from H.264/AVC. To ensure that we have a sufficient number of data points for comparison, we simulate a simple channel which randomly and uniformly drops packets with a constant probability  $p$ . The average PSNR over 4000 frames obtained from experimental simulations with  $0 \leq p \leq 0.2$  is compared with that calculated by the model in Section 5.1.1. Fig. 5.2 plots these results when the Wyner-Ziv bit rate is constrained to be 10% of the bit rate used for encoding the primary slices. It is evident that the model closely follows the experimentally obtained distortion results for the range of erasure probabilities considered, and accounts for redundant descriptions encoded at different bit rates.

### 5.2.2 Residual Distortion after Wyner-Ziv Decoding

As observed in Section 5.1, changing the redundant description not only affects the error resilience of the SLEP scheme but also changes the residual distortion in the received signal after Wyner-Ziv decoding. We now evaluate the minimum increase in video distortion that must be tolerated after Wyner-Ziv decoding. We are interested in the distortion due to the quantization mismatch only, and not the distortion from error concealment. Therefore, in the following, it is assumed that the *average*<sup>1</sup> Wyner-Ziv bit rate is just large enough to ensure that Wyner-Ziv decoding is successful, at the erasure probability  $p$  encountered by the system.

With this assumption,  $p_{WZ} \rightarrow 1$  and  $p_{EC} \rightarrow 0$  in (5.18). Then, the average distortion for a GOP of length  $M$  frames is given by

$$\begin{aligned} \mathcal{D} &= \frac{1}{M} \sum_{n=1}^M \mathcal{D}_n^{EE} = \left(1 - \frac{M+1}{2} p\right) \mathcal{D}_p + \frac{M+1}{2} p \mathcal{D}_r \\ &= \mathcal{D}_p + \frac{M+1}{2} p (\mathcal{D}_r - \mathcal{D}_p) = \mathcal{D}_p + \Delta \end{aligned} \quad (5.24)$$

---

<sup>1</sup>The Wyner-Ziv bit rate, i.e., the bit rate of the parity slices, changes slightly over the duration of a video sequence because the number of redundant slices per frame is not constant. In this section, we consider the average Wyner-Ziv bit rate to simplify the analysis.

where  $\Delta$  is the residual error energy due to the quantization mismatch. Clearly, to minimize the quantization mismatch between the primary and redundant descriptions, the encoding bit rate  $\mathcal{R}_r$  of the redundant description must be as close as possible to the primary description bit rate  $\mathcal{R}_p$ . The remaining bit rate,  $R_T - \mathcal{R}_p$  is then allocated to the Wyner-Ziv bit stream. Thus,

$$\mathcal{R}_{wz} = R_T - \mathcal{R}_p = \frac{N - K}{K} \mathcal{R}_r \simeq \frac{p}{1 - p} \mathcal{R}_r \quad (5.25)$$

where the third expression indicates that the Wyner-Ziv bit rate depends upon the parameters of the Reed-Solomon code and the encoding bit rate of the redundant description. The last expression above assumes that  $N$  and  $K$  are large enough to ensure that the Reed-Solomon code operates at its maximum efficiency<sup>2</sup>. Now, the maximum allowable bit rate for encoding the redundant description is given by:

$$\mathcal{R}_r = \min \left( (R_T - \mathcal{R}_p) \frac{1 - p}{p}, \mathcal{R}_p \right)$$

where the  $\min(\cdot, \cdot)$  operation prevents the encoder from choosing a redundant description that has finer quantization than the primary description. Thus, at packet erasure probability  $p$ , a redundant description encoded at bit rate  $\mathcal{R}_r$  increases the MSE distortion by:

$$\Delta = p \frac{M + 1}{2} (\mathcal{D}_r - \mathcal{D}_p) \quad (5.26)$$

where  $\mathcal{D}_r$  and  $\mathcal{D}_p$  depend on  $\mathcal{R}_r$  and  $\mathcal{R}_p$  through (5.22) and (5.23). The drop in video quality in dB, resulting from the usage of the redundant description rather than the

---

<sup>2</sup>This assumption is made only for this subsection, the goal being to investigate the tradeoff between the quantization mismatch and the robustness while being oblivious to other design considerations. In the remainder of this chapter, the inefficiency associated with the use of short block lengths in the Reed-Solomon coder is captured in the term  $p_{wz}$ , which is evaluated in (5.11).

primary description, is given by:

$$\begin{aligned}\Delta_{\text{dB}} &= 10 \log_{10} \frac{255^2}{\mathcal{D}_p} - 10 \log_{10} \frac{255^2}{\mathcal{D}_p + \Delta} = 10 \log_{10} \frac{\mathcal{D}_p + \Delta}{\mathcal{D}_p} \\ &= 10 \log_{10} \left( 1 + \frac{M+1}{2} p \left( \frac{\mathcal{D}_r}{\mathcal{D}_p} - 1 \right) \right)\end{aligned}$$

where  $\Delta$  is obtained from (5.26). Fig. 5.3 plots this loss in dB, at various packet erasure rates, for  $R_T = 1.1$  Mb/s,  $\mathcal{R}_p = 1$  Mb/s, and  $M = 15$  for the *Foreman CIF* sequence. The plots indicate that error resilience at high erasure probability is achieved at the price of increased distortion from the quantization mismatch between the redundant and primary descriptions. Observe from (5.26) that the residual distortion is directly proportional to the erasure probability, the quantization mismatch between the primary and redundant descriptions, and the number of frames over which the quantization mismatch propagates due to motion-compensated decoding. After  $M$  frames, the quantization error propagation is stopped by an intra coded video slice. This is reminiscent of the theoretical analysis of Chapter 3, in which the residual distortion after Wyner-Ziv decoding was expressed as a function of the erasure probability, the quantization mismatch and the error propagation resulting from the temporal correlation in the first-order Markov source.

### 5.3 Optimization of a Practical SLEP System

The model is now used to determine the bit allocation that results in the best average rate-distortion performance. Specifically, this bit allocation involves selecting the bit rates  $\mathcal{R}_p^*$ ,  $\mathcal{R}_r^*$ , and  $\mathcal{R}_{wz}^*$ , which result in the smallest MSE between the decoded video signal and the original video signal, given the erasure probability  $p$ , and the total bit rate constraint  $R_T$ . For the present implementation, determining  $\mathcal{R}_r^*$  and  $\mathcal{R}_{wz}^*$  amounts to fixing the parameters  $N$  and  $K$  of the Reed-Solomon code. Subject to a total bit rate constraint,  $\mathcal{R}_p + \mathcal{R}_{wz} \leq R_T$ , the optimum bit rates are selected by carrying out an exhaustive search in the space of available bit rates at the encoder. For each  $(\mathcal{R}_p, \mathcal{R}_t, \mathcal{R}_{wz})$  triplet, the encoder performs the following calculations:

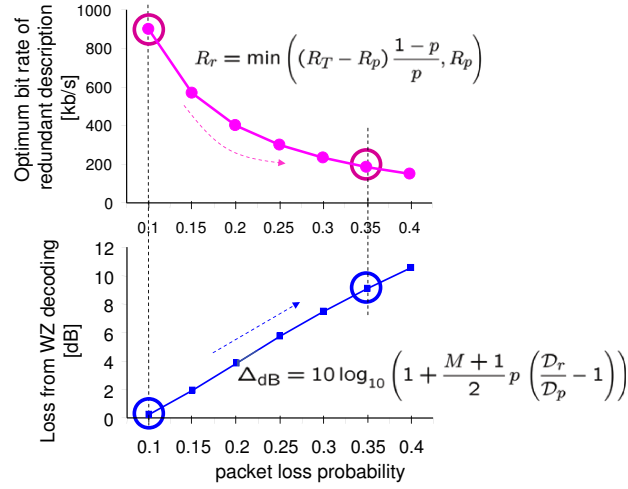


Figure 5.3: As the erasure probability increases, redundant descriptions encoded at a lower bit rate must be used to provide error robustness. The increased resilience is achieved at the cost of increased quantization mismatch after Wyner-Ziv decoding.

- The source coding distortion in the primary and redundant slices is calculated from (5.22) and (5.23).
- The parameters  $N$  and  $K$  of the Reed-Solomon code are chosen such that the Wyner-Ziv bit rate satisfies both (5.21) and the total bit rate constraint. For any  $(N, K)$ , the probability that Wyner-Ziv decoding succeeds is given by (5.11).
- The end-to-end distortion resulting from a combination of error concealment, quantization mismatch from Wyner-Ziv decoding, and error propagation is obtained via (5.18).

This optimization is carried out at intervals of 1 second, i.e., 30 frames, for the video sequences considered in this thesis. Therefore, the parameters of the rate-distortion models of (5.22) and (5.23) are updated based on the rate and distortion values collected over a window of 30 frames. This ensures that the rate-distortion functions used during the optimization reflect the changes in the scene content. Further, since the experimental simulations are very long (4000 frames each), it is practical to evaluate the model parameters at short intervals.



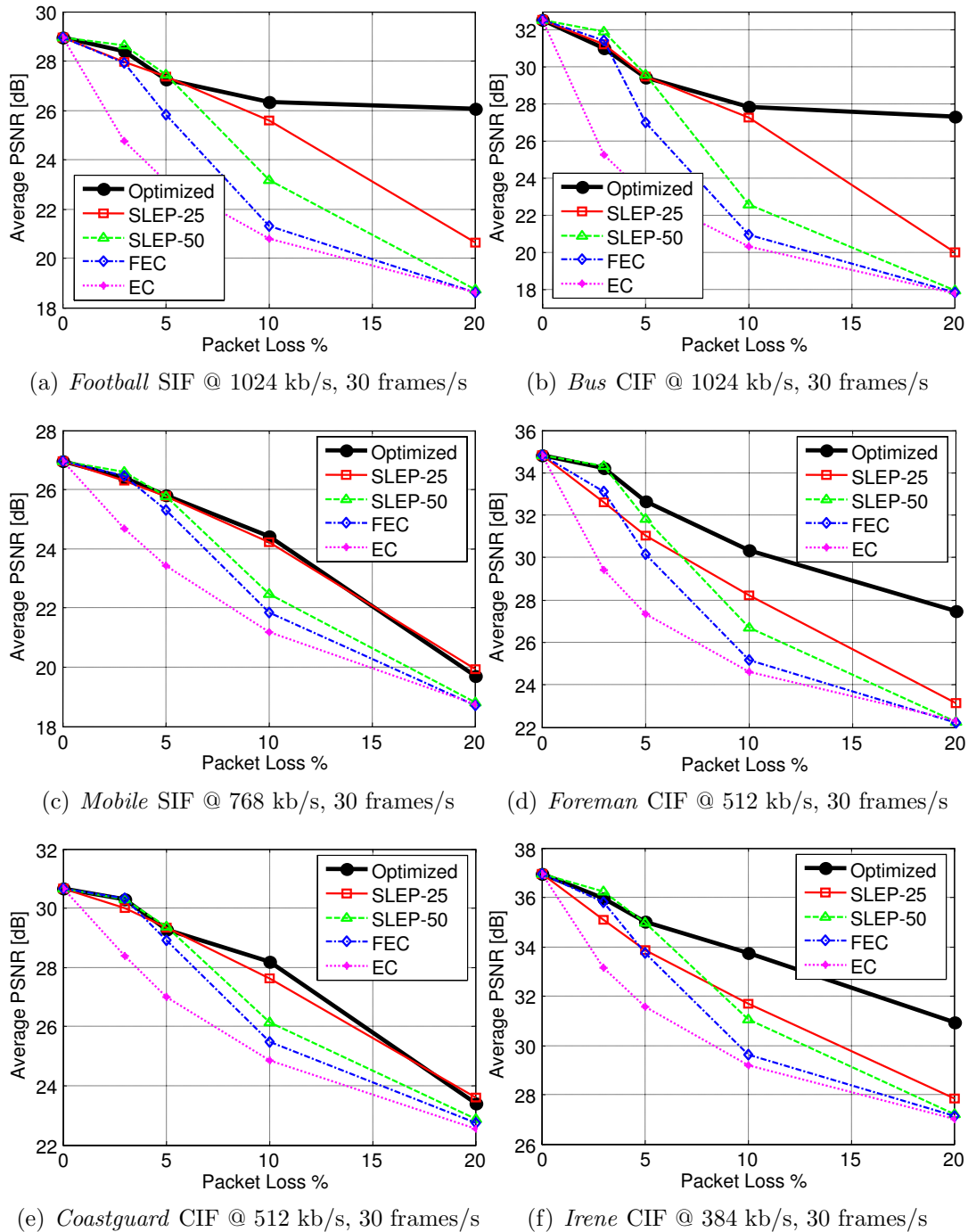


Figure 5.4: The model derived in Section 5.1.1 is used at the encoder to choose the primary video coding bit rate, the bit rate of the redundant description, and the Wyner-Ziv bit rate (equivalently, the strength of the Reed-Solomon Slepian-Wolf code). When compared with a fixed *a priori* assignment of bit rates, the optimized scheme provides superior average picture quality over all erasure probabilities.

As explained in Section 4.4.2, there is a delay during which redundant slices corresponding to one or more frames may have to be buffered before applying Reed-Solomon coding. This is especially true when the total bit rate  $R_T$  is low, and an entire video frame fits into a single video packet. Assuming that  $L$  frames need to be buffered ( $L \leq M$ ), the following optimization problem is solved for each group of  $M$  frames:

$$\begin{aligned}
 & \text{Minimize } \mathcal{D} \\
 & \text{such that } \mathcal{R}_p + \mathcal{R}_{wz} \leq R_T \\
 & \quad \mathcal{R}_{wz} = \frac{N - K}{K} \mathcal{R}_r \\
 & \quad N \geq K_L \geq 1, \mathcal{R}_p \geq \mathcal{R}_r
 \end{aligned} \tag{5.27}$$

where  $\mathcal{D}$  is the average distortion from (5.19),  $R_T$  is the total bit rate constraint, and  $K_L$  is the average number of redundant slices contained in  $L$  frames.

The video codec settings are identical to those used in Section 4.4.1. However, in the interest of conducting a realistic simulation, the simple channel used in Chapter 4 is not used. Instead of inserting erasures according to a uniform distribution, actual traces obtained from Internet measurements are used. These traces [186], available for erasures occurring with probability 3%, 5%, 10%, and 20%, are the common test conditions prescribed by the Joint Video Team for low-delay error resilience experiments.

The performance of a SLEP scheme optimized in this way is compared to that of multiple schemes in which the redundant descriptions and Wyner-Ziv bit rates are fixed *a priori*. The competing schemes all use an error resilience bit rate of 10% of the primary description bit rate, and have fixed redundant descriptions encoded at 100% (FEC), 50% and 25% of the primary bit rate. As shown in Fig. 5.4, the optimized scheme, which chooses the best combination of  $\mathcal{R}_p$ ,  $\mathcal{R}_r$  and  $\mathcal{R}_{wz}$ , outperforms the fixed schemes for all the erasure traces used in the experiment.

In the experiments conducted in this thesis, the erasure probability  $p$  and the total bandwidth constraint  $R_T$  remain constant. If feedback is unavailable, as is the case in

our experiments, the encoder must use conservative values for  $p$  and  $R_T$ . If feedback is available, then the sender can perform bandwidth estimation [118] and receive regular updates of the changes in  $p$  and  $R_T$ . Using the model-based optimization scheme described in this chapter, the encoder can update its bit allocation ( $\mathcal{R}_p$ ,  $\mathcal{R}_r$ , and  $\mathcal{R}_{WZ}$ ) in response to the changes in the channel, thus allowing the sender to have tighter control over the received video quality.

Recall that, for the SLEP implementation of Fig. 4.2, the distortion-rate tradeoff for an optimized FEC system can be obtained by imposing the constraint  $\mathcal{R}_r = \mathcal{R}_p$  in (5.27). Given a total rate constraint  $R_T$ , such a scheme determines the bit-allocation between source coding and FEC parity such that the average output picture quality is maximized. The average PSNR delivered by such an optimized FEC system is plotted against the erasure probability in Fig. 5.5. It is observed that, if the erasure probability is known *a priori*, then, an optimized SLEP scheme delivers approximately the same video quality as an FEC scheme optimized in the above sense. However, if the erasure probability is not known, the performance of SLEP degrades gracefully as the channel worsens, unlike FEC, as shown in the experiments of Chapter 4.

## 5.4 Summary

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. As explained in Chapter 4, SLEP involves transmission of a Wyner-Ziv bit stream to add error robustness to a compressed video signal. The model expresses the MSE distortion in the received signal as a function of the small distortion introduced due to the quantization mismatch from Wyner-Ziv decoding, the large distortion due to error concealment, and the distortion due to error propagation. The model closely approximates the observed performance of the SLEP system, which provides graceful degradation of video quality. It is shown that the residual distortion in the received signal, after Wyner-Ziv decoding, is directly proportional to the erasure probability, the difference in the MSE distortions of the primary and redundant descriptions, and the number of video frames over which the energy from

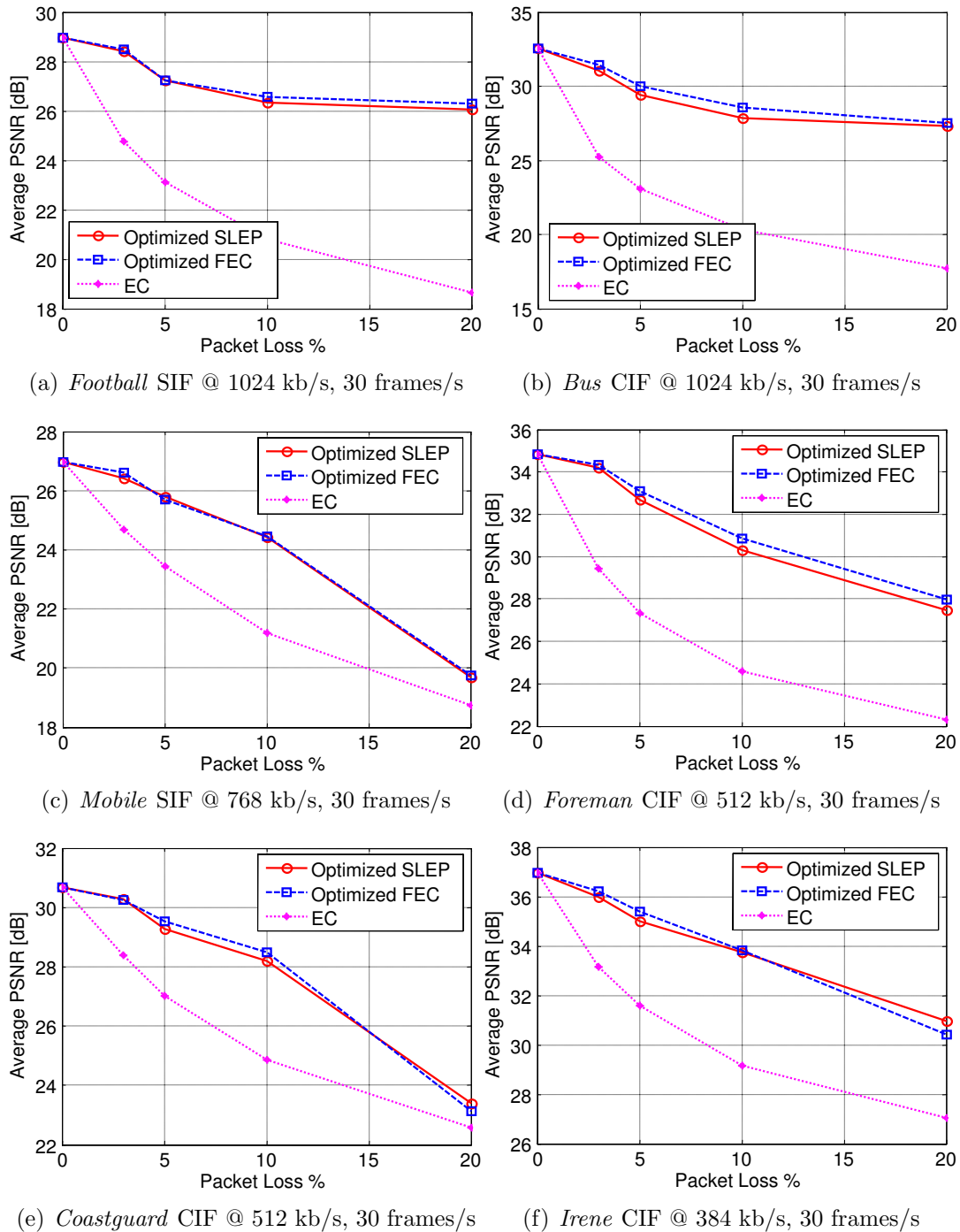


Figure 5.5: When the erasure probability is known *a priori*, an optimized SLEP scheme and an optimized FEC scheme provide approximately the same video quality. Recall however, that when the erasure probability changes for a given bit allocation, SLEP provides graceful degradation compared to FEC, as plotted in Chapter 4, Fig. 4.11.

the quantization mismatch propagates before being arrested by an intra coded video slice. Given the erasure probability and the total bit rate constraint, the model has been used at the encoder to find the combination of the primary description bit rate, the redundant description bit rate and the Wyner-Ziv protection which maximizes the average received video quality.

# Chapter 6

## Conclusions

This thesis presents a robust video transmission scheme which we denote as Systematic Lossy Error Protection (SLEP). The scheme is based on the information theoretic framework of systematic lossy source/channel coding. The treatment of SLEP in this thesis comprises:

1. Analysis of SLEP applied to a first-order Gauss Markov source using high-rate quantization theory.
2. Implementation of a realistic SLEP scheme using standardized state-of-the-art video coding tools within the H.264/AVC specification.
3. Modeling the end-to-end rate-distortion performance of a SLEP scheme, and optimum bit allocation using the model.

The concept of SLEP was presented in Chapter 3 using transmission of a compressed video signal as an example. Following this, a theoretical analysis was carried out in order to study the properties of SLEP when it is applied to a predictively encoded source, while keeping the configuration simple enough to allow the derivation of closed-form mathematical expressions. Using the derived rate-distortion tradeoff it was shown that, at high rates, the robustness of SLEP increases when step-size of the Wyner-Ziv quantizer is increased, i.e., the loss probability at which the received signal quality starts degrading rapidly, is higher for the SLEP scheme than for the

competing FEC scheme. Moreover the signal quality degrades gracefully as the symbol loss probability increases. This property is of practical significance (in broadcast applications, for example) when the symbol (or packet) loss (or error) probability is unknown or fluctuating. SLEP offers the possibility of graceful degradation beyond the FEC cliff.

When the packet loss probability is known, the derived rate distortion functions indicate that the cliff occurs at the same loss probability for both FEC and SLEP. However, in the case of FEC, only one bit allocation between the source and parity symbols is possible, leading to a constant quality for all loss probabilities to the left of the cliff. When the system is designed for a high loss probability, this results in higher source coding distortion in those packets which are received error-free. In SLEP, a variety of bit allocations is possible depending upon the coarseness of the Wyner-Ziv quantizer relative to the source quantizer. In particular, SLEP allows a large percentage of the total bit rate to be allocated to the source coding, while retaining the same cliff probability as FEC. Therefore, error-free video is decoded at a higher quality, and the signal quality degrades gracefully when the loss probability increases. For the simplified SLEP scheme described in Chapter 3, and also for the video implementation in Chapter 4, SLEP is a generalization of FEC. In other words, if the Wyner-Ziv quantizer is removed, then SLEP reduces to FEC.

The second part of this thesis was concerned with a practical implementation of a SLEP system for error-resilient video transmission. For this purpose, a Wyner-Ziv codec was constructed using a redundant video representation in conjunction with Reed-Solomon coding. The Reed-Solomon code, applied across the redundant slices, plays the role of a Slepian-Wolf code. The observed behavior of the average decoded video quality versus the packet loss probability echoes the theoretical tradeoffs derived earlier for the first-order Markov source. When the quality of the redundant slices is lowered, the robustness to packet erasures is increased in exchange for a quantization mismatch between the primary and redundant slices. This quantization mismatch propagates during the decoding of the subsequent frames and results in a reduction in picture quality. If the quality of the redundant description is chosen appropriately, this loss is almost imperceptible. In contrast, failure of the competing FEC scheme

leaves the decoder no option other than local error concealment of the lost packets. Decoder-based error concealment is often unable to provide acceptable picture quality. Thus, a SLEP scheme with a suitably chosen redundant description provides higher instantaneous and average picture quality compared to the equivalent FEC scheme.

The final part of the thesis covers end-to-end modeling, analysis and optimization of a SLEP scheme used for robust video transmission. The average MSE distortion in the decoded video sequence is expressed as a function of bit rate of the compressed video sequence, the (untransmitted) bit rate of the redundant video description and the Wyner-Ziv bit rate. Thus, the model shows how the average decoded picture quality is affected by the rate-distortion function of the primary and redundant video encoders, as well as the strength of the Slepian-Wolf code. It is shown that the quality loss incurred due to Wyner-Ziv decoding is directly proportional to the packet loss probability, the quantization mismatch between the primary and redundant descriptions, and the number of frames over which the mismatch propagates before being arrested by an intra macroblock refresh. This loss in video quality, measured as a MSE distortion, increases linearly with packet loss probability, as opposed to the drastic MSE increase associated with artifacts from decoder-based error concealment. This is consistent with the relation between the distortion and the symbol erasure probability obtained from the theoretical analysis of the simple SLEP system in Chapter 3.

## 6.1 Standardization Effort for SLEP

The SLEP scheme was proposed for standardization within H.264/AVC at the Joint Video Team (JVT), composed of the ITU-T and ISO/IEC MPEG. The proposal was made at the JVT meeting in Geneva in April 2006 [131]. This consisted of the implementation presented in Chapter 4 along with a Supplementary Enhancement Information (SEI) message which specified the syntax of a data structure used to carry the parity symbols and quantization parameters, which enable Wyner-Ziv decoding at the receiver. A core experiment was instituted at this meeting [132] and the subsequent meeting in Klagenfurt [26]. The objective of the core experiment



was to compare the error resilience of SLEP with that of Loss-Aware Rate Distortion Optimization (LA-RDO), an encoder-based scheme that determines macroblock mode decisions to minimize error propagation. In addition, it was required that the redundant descriptions and the strength of the Wyner-Ziv code be chosen optimally, and the model proposed in Chapter 5 was used for this purpose. The results of the core experiments were presented in Klagenfurt in July 2006 [129] and Hangzhou in October 2006 [130]. It was demonstrated that, over all the video sequences tested, and all packet loss rates, SLEP provided an average video quality gain of 2.6 dB over LA-RDO. The (subjective) improvement in visual quality was also confirmed by the JVT delegates. The consensus among the delegates was that H.264/AVC is at a mature stage of deployment and that supporting SLEP would require some additions to the implementation of the video decoder. Therefore, SLEP has not been included in the H.264/AVC standard at the present time. It was suggested at the Hangzhou meeting, that transport-layer FEC [144] could be used to transmit the SLEP parity symbols. Our position has been that, since the scheme is conceptually independent of the video standard, it would be better from the point of view of realizing practical deployment of SLEP to revive the standardization effort when a Call for Proposals is issued for a new video coding standard which will succeed H.264/AVC.

## 6.2 Improvements and Extensions of SLEP

1. *Combining error concealment with Wyner-Ziv decoding:* In the SLEP scheme presented in Chapters 4 and 5, the decoder-based error concealment scheme is only used as a last resort when Wyner-Ziv decoding fails. A more sophisticated scheme could perform both Wyner-Ziv decoding and decoder-based error concealment, and blend the two reconstructions. This will require additional signal processing at the receiver, but can conceivably provide better decoded picture quality than using the redundant video description alone.
2. *Incorporating instantaneous quality fluctuations in rate-distortion optimization:* Instantaneous video quality fluctuations critically affect the viewing experience.

SLEP is able to reduce instantaneous fluctuations by strengthening the Slepian-Wolf code while using coarser quantization in the redundant slices. The end-to-end rate-distortion model presented in Chapter 5 predicts the average video quality at the receiver, and performs bit allocation between the primary and redundant descriptions. An optimization scheme that penalizes large instantaneous video quality fluctuations is expected to provide tighter control over the viewing experience. In particular, it will prevent the quantization step sizes in the redundant slices from becoming too large, and at the same time, disallow bit allocations which produce a very weak Slepian-Wolf code.

3. *Novel uses for the Wyner-Ziv stream:* Within the systematic source/channel coding framework, distributed source coding can be used for other purposes besides error resilience. Wyner-Ziv coding can, for example, be used to construct an enhancement layer bit stream [201] at the expense of increased Wyner-Ziv bit rate. It can also be used to perform authentication of the media signal in the systematic portion of the transmission [202].

A consequence of the systematic source/channel coding framework is that SLEP is backward-compatible with legacy broadcast systems. Legacy receivers may ignore the Wyner-Ziv bit stream while modern receivers can utilize it for error protection and provide improved picture quality. Traditional FEC schemes attempt to achieve the highest visual quality by performing source/channel bit allocation. By performing Wyner-Ziv coding instead of conventional channel coding, SLEP provides an additional degree of freedom: It inserts a small, bounded distortion in the protected signal, and gracefully trades off this distortion against the resilience to packet loss. Up to the present time, the only scheme that achieved this graceful degradation of received video quality was layered coding with unequal error protection. Our work has shown that, using Wyner-Ziv coding, it is possible to achieve graceful degradation without a layered video representation in the systematic portion of the transmission. Hence, the SLEP scheme does not incur the loss in rate-distortion performance associated with layered video codecs. Furthermore, a Wyner-Ziv codec can be constructed out of well-understood components: quantizers, entropy coders and channel coders, at

a small additional complexity cost compared with conventional FEC-based systems. Therefore, it is our hope that SLEP will provide a viable alternative to existing joint source/channel coding schemes, and facilitate robust low-delay video communication over broadcast channels and across the Internet.

# Appendix A

## Stationarity Relations and Successive Degradation Quantization

The results in Lemmas 4, 5 and 6 are well-known [110] and are provided for the sake of completeness in order to fill in the details in the proofs sketched in Chapter 3. All references to stationarity will mean stationarity in the strict sense.

**Definition 3.**  $(U_n)_n$  and  $(V_n)_n$  are defined to be jointly stationary processes if and only if the joint process  $(U_n, V_n)_n$  is stationary.

**Lemma 4.** If  $(U_n, V_n)_n$  is stationary, then  $(U_n - V_n)_n$  is stationary.

Recall, in the DPCM encoder,  $W_n$  is i.i.d. Then, by the above definition,  $W_n, \widehat{W}_n, \widetilde{W}_n$ , and  $\widetilde{\widetilde{W}}_n$  are jointly stationary. By Lemma 4, the differences  $W_n - \widehat{W}_n, W_n - \widetilde{W}_n, W_n - \widetilde{\widetilde{W}}_n$  are all stationary.

**Lemma 5.** If  $U_n$  is stationary, and  $V_n = h * U_n$ , where  $h$  is the impulse response of a stable Linear Time Invariant (LTI) system, then  $(U_n, V_n)_n$  is stationary.

By the above lemma  $(X_n, \widehat{X}_n, \widetilde{X}_n)_n = h * (W, \widehat{W}, \widetilde{W})_n$  is stationary, because  $h(n) = \rho^n u(n)$  with  $|\rho| < 1$  to ensure stability. By Lemma 4, this implies that  $X_n - \widehat{X}_n$  is also stationary. Similarly, it may be shown that the differences,  $\widehat{X}_n - \widetilde{X}_n$

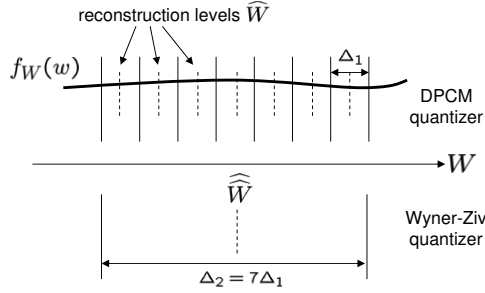


Figure A.1: Embedded quantization (successive degradation) of  $W$  with  $m = \Delta_2/\Delta_1 = 7$ . Embedding increases the MSE by a factor of  $(m^2 - 1)$

and  $X_n - \tilde{X}_n$  are stationary. Therefore, the functionals  $\mathcal{D}, \mathcal{D}_1, \mathcal{D}_2, \mathcal{R}, \mathcal{R}_1, \mathcal{R}_2$  may be defined by dropping the time index  $n$ .

**Lemma 6.** Let  $V_n = \rho V_{n-1} + U_n$ , where  $|\rho| < 1$  and  $(U_n)_n$  is a stationary zero mean process with  $U_n$  independent of the past values  $V_{n-k}$ ,  $k \in \mathbb{Z}^+$ . Then  $\mathbb{E}V = 0$  and  $\sigma_V^2 = \frac{\sigma_U^2}{1-\rho^2}$ .

**Proposition 7.** Consider the embedded quantization scheme for quantizing  $W$  in which  $m = \frac{\Delta_2}{\Delta_1} \in \mathbb{Z}^+$ . Then, the MSE between the reconstruction functions of the finer quantizer and the coarser quantizer is given by

$$\mathbb{E}(\widehat{W} - \widehat{\widehat{W}})^2 = (m^2 - 1) \frac{\Delta_1^2}{12} \simeq (m^2 - 1) \mathcal{D}_1 \quad (\text{A.1})$$

*Proof.* We prove the result for odd valued  $m$ . Note that the proof for even  $m$  follows the same method. By the high-rate assumption,  $W$  is approximately uniformly distributed over the width of the bins. Fig. A.1 shows the embedded quantization scenario for  $m = 7$ . In this case,

$$\mathbb{E}(\widehat{W} - \widehat{\widehat{W}})^2 = \frac{2}{m} \sum_{i=1}^{\frac{m-1}{2}} i^2 \Delta_1^2 = (m^2 - 1) \frac{\Delta_1^2}{12} \simeq (m^2 - 1) \mathcal{D}_1 \quad (\text{A.2})$$

□

**Proposition 8.** *For the SLEP scheme in Fig. 3.2, consider that the total bit rate  $\mathcal{R}$  is fixed and the system is designed to tolerate a fixed maximum erasure probability  $p_{\text{cliff}}$ . Let  $\mathcal{R}_{1,m}$  and  $\mathcal{R}_{2,m}$  be the optimally chosen source coding bit rate and error resilience bit rate, depending upon the value of  $m$ . Then:*

1. *For  $p = 0$ , the erasure-free case, the SNR with SLEP is higher than that with FEC by  $20p_{\text{cliff}} \log_{10} m$  dB.*
2. *In Fig. 3.7, the distortion plots for FEC and SLEP must cross at:*

$$p_{\text{cross}} = \frac{(1 - \rho^2)(m^{2p_{\text{cliff}}} - 1)}{m^2 - 1} < p_{\text{cliff}} \text{ for } m > 1 \quad (\text{A.3})$$

*Proof.* Let  $\mathcal{D}_{1,m}$  and  $\mathcal{D}_m$  be the source coding distortion of the DPCM coder and the end-to-end distortion for the chosen value of  $m$ , respectively. We have fixed  $\mathcal{R} = \mathcal{R}_{1,m} + \mathcal{R}_{2,m}$ . It was proved that, according to (3.4), the optimally chosen bit rates satisfy,

$$\begin{aligned} \mathcal{R}_{2,m} &\simeq \frac{p_{\text{cliff}}}{1 - p_{\text{cliff}}} (\mathcal{R}_{1,m} - \log_2 m) \\ \mathcal{R}_{1,m} &\simeq (1 - p_{\text{cliff}}) \mathcal{R} + p_{\text{cliff}} \log_2 m \end{aligned}$$

This means that the DPCM coder must incur a source coding distortion of

$$\begin{aligned} \mathcal{D}_{1,m} &\simeq \frac{1}{12} 2^{2h(W)} 2^{-2\mathcal{R}_1} \simeq \frac{1}{12} 2^{2h(W)} 2^{-2(1-p_{\text{cliff}})\mathcal{R}} 2^{2p_{\text{cliff}} \log_2 m} \\ &\simeq m^{2p_{\text{cliff}}} \frac{1}{12} 2^{2h(W)} 2^{-2(1-p_{\text{cliff}})\mathcal{R}} \end{aligned} \quad (\text{A.4})$$

As discussed in Section 3.4, (3.15) holds for all  $p \leq p_{\text{cliff}}$ . Thus,

$$\mathcal{D}_m \simeq \left(1 + p \frac{m^2 - 1}{1 - \rho^2}\right) \mathcal{D}_{1,m} \quad (\text{A.5})$$

From the above relation, the end-to-end distortion for SLEP at  $p = 0$  is given by  $\mathcal{D}_m = \mathcal{D}_{1,m}$  with  $m > 1$ . With  $m = 1$ , we obtain the end-to-end distortion for FEC at  $p = 0$  as  $\mathcal{D}_1 = \mathcal{D}_{1,1}$ . Both  $\mathcal{D}_{1,m}$  and  $\mathcal{D}_{1,1}$  can be obtained via direct substitution from

(A.4). Then, using (A.4), the difference in SNR between SLEP and FEC at  $p = 0$  is given by

$$\Delta_{\text{dB}} = 10 \log_{10} \frac{\mathcal{D}_{1,1}}{\mathcal{D}_{1,m}} = 10 \log_{10} m^{2p_{\text{cliff}}} = 20 p_{\text{cliff}} \log_{10} m$$

which proves the first result. To find the erasure probability,  $p_{\text{cross}}$ , at which the distortion due to SLEP crosses the distortion due to FEC in Fig. 3.7, we evaluate (A.5) at this crossover probability. Thus

$$\mathcal{D}_{m,\text{cross}} \simeq \left( 1 + p_{\text{cross}} \frac{m^2 - 1}{1 - \rho^2} \right) \mathcal{D}_{1,m} \quad (\text{A.6})$$

where  $\mathcal{D}_{1,m}$  is given by (A.4). Recall that the end-to-end distortion for FEC at the crossover probability is obtained simply by putting  $m = 1$  in (A.6). Then,  $\mathcal{D}_{m,\text{cross}} = \mathcal{D}_{1,\text{cross}}$  gives the crossover probability as

$$p_{\text{cross}} = \frac{(1 - \rho^2)(m^{2p_{\text{cliff}}} - 1)}{m^2 - 1} < p_{\text{cliff}} \text{ for } m > 1$$

□

# Appendix B

## SLEP with Multiple Embedded Redundant Descriptions

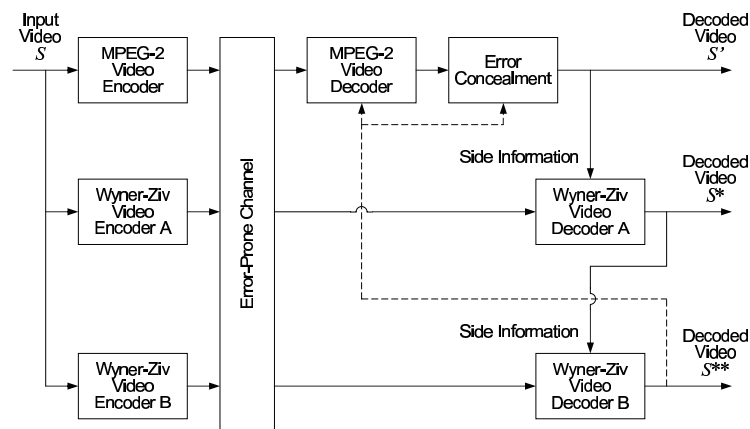


Figure B.1: The Wyner-Ziv decoder uses a decoded error-concealed video waveform as side information in a systematic lossy source/channel coding setup. With an embedded Wyner-Ziv codec, graceful degradation of video quality is obtained without a layered video representation in the systematic transmission.

The tradeoff between distortion due to transmission errors and Wyner-Ziv bit rate can be exploited to construct an embedded Wyner-Ziv code that achieves graceful degradation of the decoded video when the error rate of the channel increases [127, 134]. Such a system is shown in Fig. B.1 for the case of 2 quality levels. Wyner-Ziv



encoder A employs a coarser redundant representation that is embedded in the finer redundant representation of Wyner-Ziv encoder B. Since Wyner-Ziv encoder A has a coarser quantizer, its bit stream is easier to decode and, therefore, has stronger error protection capability. It is decoded first, using decoded video  $S'$  as side information to yield improved decoded video  $S^*$ . If the transmission errors are not too severe, then the Wyner-Ziv stream B can also be decoded using the decoded output symbols from Wyner-Ziv decoder A, and the side information  $S'$ . This yields a further improved decoded video signal  $S^{**}$ .

### B.1 Embedded Wyner-Ziv Codec

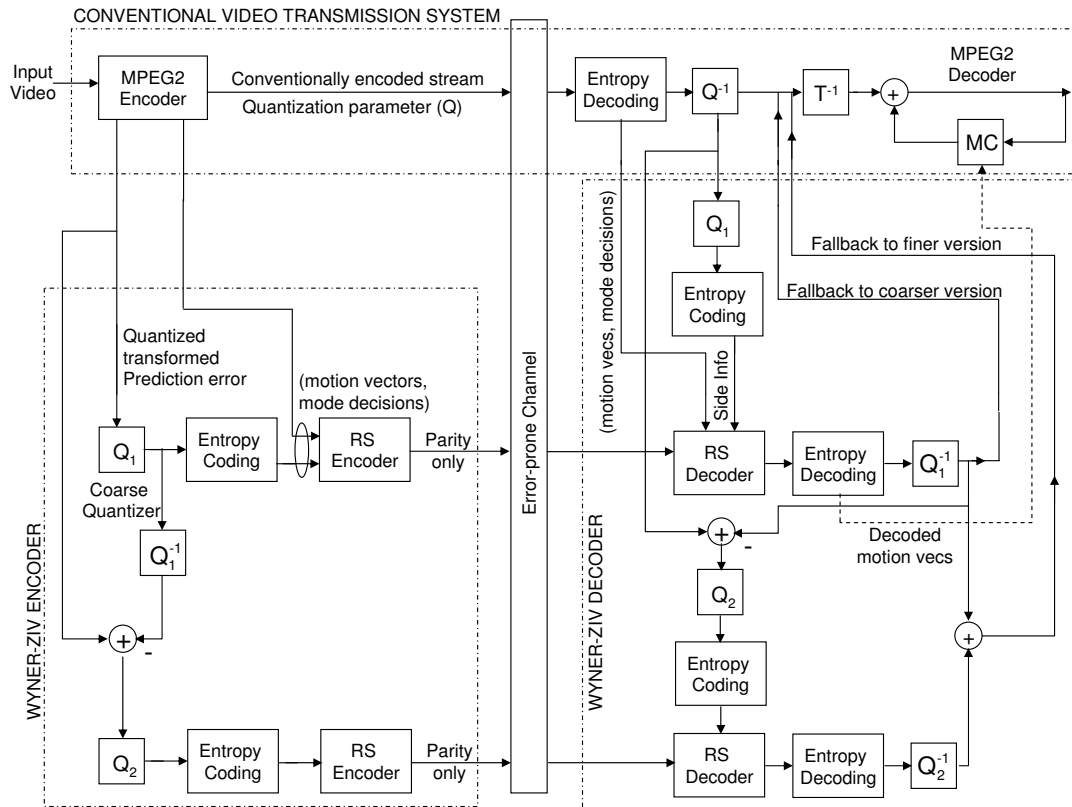


Figure B.2: Implementation of systematic lossy error protection by combining MPEG coding and Reed-Solomon codes across slices. In the Wyner-Ziv encoder, multiple redundant descriptions are generated by embedded quantization.

### B.1.1 Unequal Error Protection for I, P and B Slices

To exploit the tradeoff between coding efficiency and error resilience, the MPEG-2 bit stream consists of I slices (most significant, highest bit rate), P slices and B slices (least significant, smallest bit rate). For improved error resilience, this coding structure must be taken into account when the Wyner-Ziv bit stream is constructed, i.e., The Reed-Solomon encoder must output varying amounts of parity symbols for I, P, and B slices. For a video slice of length  $l$  symbols, at a symbol error probability  $p$  of a memoryless channel, the probability that the slice is corrupted is given by  $1 - (1 - p)^l \approx lp$  for small  $p$ , i.e., the probability of losing a slice is proportional to its length. Thus, an I slice is  $s_I = L_I/L_B$  times more likely to be lost, than a B slice, where  $L_I, L_B$  are the average lengths of I and B slices in the main video description. Therefore, for every one parity slice appended to a B frame, our system appends  $s_I$  parity slices to the I frame at the beginning of the GOP. Similarly, for every one parity slice appended to a B frame, we append  $s_P = L_P/L_B$  parity slices to each P frame in the GOP. Let  $l_I, l_P$ , and  $l_B$  be the average lengths of I, P, and B slices in the redundant description and let  $m_I, m_P$ , and  $m_B$  be the number of I, P, and B frames in one GOP. Based on the priorities assigned above, let the number of parity slices for the I, P, and B frames be  $s_I x, s_P x, s_B x$ , where  $s_B = L_B/L_B = 1$ . Thus,  $x$  may be defined as the minimum allowable length for a B slice, and all other slice lengths are multiples of  $x$ . Then, since only the parity slices are transmitted in the Wyner-Ziv bit stream, the Wyner-Ziv bit rate is given by  $\mathcal{R}_{WZ} = (m_I l_I s_I + m_P l_P s_P + m_B l_B s_B) x$ , which can be solved for  $x$ , because all other quantities are known. In this way, unequal Wyner-Ziv protection is assigned within a single redundant description.

### B.1.2 Embedded Redundant Descriptions

Now consider the generation of a second Wyner-Ziv bit stream, which contains a video description with finer quantization than that described in the preceding section. For this, the difference between the original transformed prediction error and the coarsely quantized redundant description is obtained and then finely quantized and entropy-coded, as shown in Fig. B.2. The resulting bit stream is input to a Slepian-Wolf

encoder which applies Reed-Solomon coding across the video slices, and transmits only the parity symbols. This method of generating the embedded redundant descriptions is reminiscent of SNR-scalable video coding with fine granular scalability [88, 172]. The difference is that only the parity bit streams corresponding to the redundant descriptions are transmitted, while the systematic portions are regenerated (possibly with errors) at the decoder. Just like SNR-scalable video coding, the decoder can only recover the second finer redundant description if the first coarser redundant description has been successfully decoded, and not otherwise. Now consider allocation of the available bit rate among the finer redundant description and the embedded coarser redundant description. Clearly, a larger share of the Wyner-Ziv bit rate must be allocated to the more significant coarser redundant description. This is done in such a way that the number of R-S parity slices for a given frame-type, is midway between the number of parity slices possible if only one of the two descriptions were available, and Wyner-Ziv protection was carried out as in Section B.1. E.g., in Section B.1, if  $n_p^c$  parity slices were used for P frame of the coarse description alone, and  $n_p^f$  parity slices were used for P frames of the fine description alone, then in the embedded scheme,  $(n_p^c + n_p^f)/2$  parity slices are used for P frames of the embedded coarse redundant description. The remaining bit rate is allocated to the finer redundant description. There are other ways to divide the Wyner-Ziv bit rate among the coarser and finer redundant descriptions. Experimentally, the *ad hoc* approach described above provided the best tradeoff between the error resilience and the average received video quality over the chosen range of symbol error probabilities.

## B.2 Experimental Results

We now describe the results of applying the SLEP system to error-resilient MPEG-2 video broadcasting. In our experiment the systematic transmission consists of the *Foreman.CIF* sequence encoded at 2 Mb/s. For error resilience, an additional bit rate of 222 kb/s is available, i.e., for the system of Fig. B.2, the sum of the parity bit rates transmitted by the Reed-Solomon encoders is 222 kb/s. Note that, different from the simulations of Chapter 4, MPEG-2 transport is used across a memoryless channel

which causes symbol errors (1 symbol = 1 byte). There are no resynchronization markers in the video bit stream except at the slice boundaries. Thus, even if there is a single symbol error inside a video slice, the entire slice is discarded<sup>1</sup>.

### B.2.1 Unequal Protection for I, P, and B Slices

First, consider the advantages of exploiting the coding structure of the video bit stream, as described in Section B.1. In Fig. B.3, the dashed curves indicate the variation in PSNR for the system in [126], in which the I, P, and B slices are not treated differently. The solid curves describe the performance of the proposed system. The bit rate for error resilience is the same in each case. Thus, for all schemes in Fig. B.3,  $\mathcal{R}_{WZ} = 222$  kb/s and  $\mathcal{R}_p = 2$  Mb/s. The untransmitted bit rate of the redundant description,  $\mathcal{R}_r$ , takes three different values depending upon the experiment: 2 Mb/s, 1 Mb/s and 500 kb/s. When only a 500 kb/s redundant description is available, the Reed-Solomon codes for I, P and B frames are (36,18), (29,18), and (22,18) respectively. When only a 1 Mb/s redundant description is available, they are (27,18), (23,18), and (20,18) respectively. As expected, protecting the I, P and B slices after taking into account their lengths and the number of times they occur in the sequence, yields superior error resilience.

### B.2.2 Embedded Wyner-Ziv Coding

In the second part of the experiment, we allocate the available 222 kb/s bit rate among two Wyner-Ziv streams, which are generated according to the procedure described in Section B.1.2. In the experiment, the first coarser redundant description has a source coding bit rate of 500 kb/s, and a transmitted Wyner-Ziv bit rate of 166 kb/s. The second finer redundant description has a source coding bit rate of 1 Mb/s, and a transmitted Wyner-Ziv bit rate of 56 kb/s. Within each Wyner-Ziv bit stream, the rates of the Reed-Solomon codes for the I, P and B frames are decided according to the procedure described in Section B.1. The RS codes used are (32,18), (26,18),

---

<sup>1</sup>This is a conservative assumption, and it may be possible to correctly decode those macroblocks in the error-prone slice, which are earlier in the decoding order with respect to the error-prone macroblock.

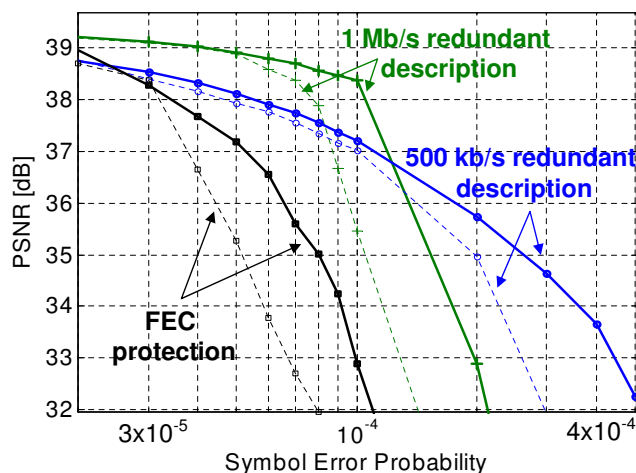


Figure B.3: Error resilience improves when unequal Wyner-Ziv protection is assigned to the I, P, B frames in a single redundant description. The transmitted Wyner-Ziv bit rate is 222 kb/s for each curve.

(21,18) for the coarser redundant description, and (23,18), (20,18), (19,18) for the finer redundant description. As shown in Fig. B.4, at low symbol error rates, the decoded video quality provided by the embedded Wyner-Ziv coding scheme is close to that obtained by using the (finer) 1 Mb/s redundant description. At high symbol error rates, the decoded video quality is closer to that obtained using the (coarser) embedded 500 kb/s redundant description. Thus the tradeoff between resilience to transmission errors, and the residual distortion resulting from Wyner-Ziv quantization is exploited to obtain better overall performance than using any of the two redundant descriptions alone. Since Fig. B.4 contains average PSNR values, it demonstrates the graceful degradation property of embedded Wyner-Ziv coding, but does not show the instantaneous effects of decoder failure. To appreciate the advantage of using embedded Wyner-Ziv coding from the point of view of mitigating error propagation within a video sequence, refer to Fig. B.5, which shows the variation of PSNR with time for a simulation trace at a symbol error probability of  $2 \times 10^{-4}$ .

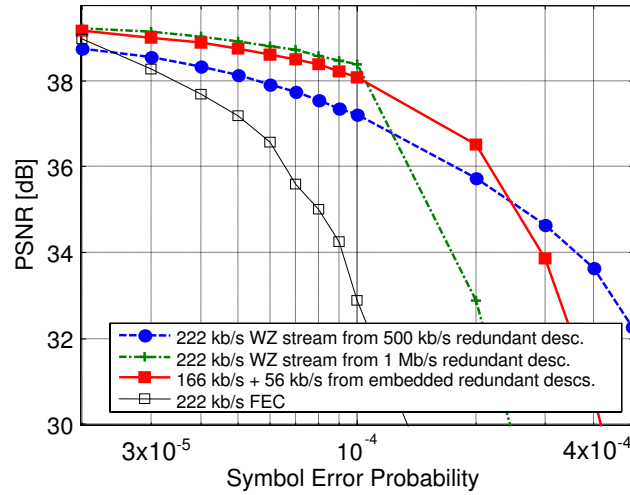


Figure B.4: To achieve graceful degradation of video quality, a coarse redundant description encoded at 500 kb/s is embedded inside a finer redundant description encoded at 1 Mb/s. The available error resilience bit rate of 222 kb/s is then shared among the two descriptions.

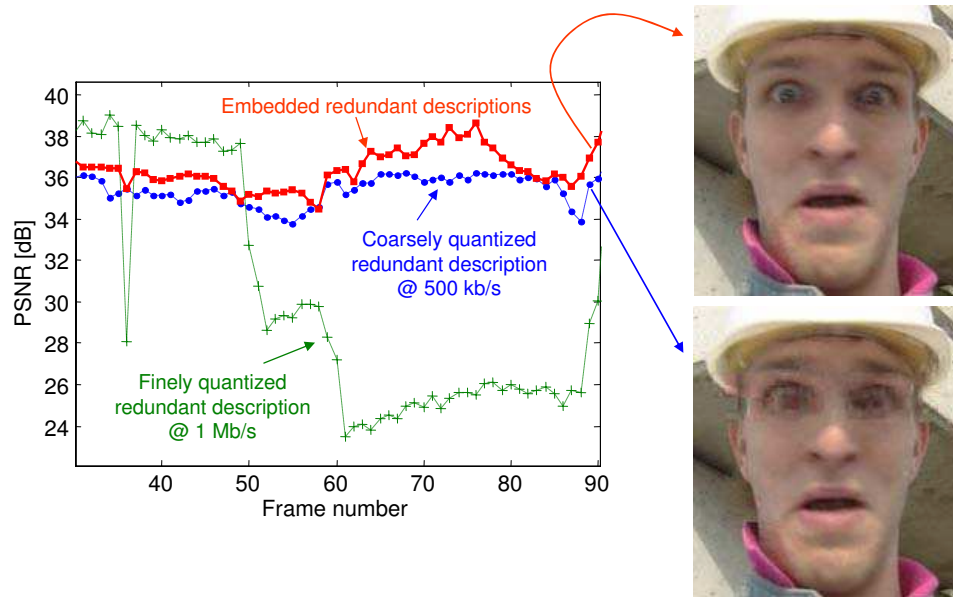


Figure B.5: At an error probability of  $2 \times 10^{-4}$ , SLEP with a finely quantized redundant description alone cannot provide adequate robustness to signal loss, while SLEP with the coarsely quantized redundant description alone incurs more distortion due to coarse quantization. With the bit rates apportioned as in Fig. B.4, the frame PSNR is at least as high as that achieved by the coarse redundant description.

# Bibliography

- [1] ISO/IEC 13818. *Generic Coding of Moving Pictures and Associated Audio*. April 1993.
- [2] A. El Gamal and T. Cover. Achievable Rates for Multiple Descriptions. *IEEE Transactions on Information Theory*, 28(6):851–857, November 1982.
- [3] A. Aaron and B. Girod. Compression with Side Information using Turbo Codes. In *Proc. IEEE Data Compression Conference*, pages 252–261, Snowbird, UT, April 2002.
- [4] A. Aaron and B. Girod. Wyner-Ziv Video Coding with Low Encoder Complexity. In *Proc. Picture Coding Symposium*, San Francisco, CA, December 2004.
- [5] A. Aaron, P. Ramanathan, and B. Girod. Wyner-Ziv Coding of Light Fields for Random Access. In *Proc. IEEE International Workshop on Multimedia Signal Processing*, pages 323–326, Siena, Italy, September 2004.
- [6] A. Aaron, S. Rane, and B. Girod. Wyner-Ziv Video Coding with Hash-based Motion Compensation at the Receiver. In *Proc. IEEE International Conference on Image Processing*, volume 5, pages 3097–3100, Singapore, October 2004.
- [7] A. Aaron, S. Rane, D. Rebollo-Monedero, and B. Girod. Systematic Lossy Forward Error Protection of Video Waveforms. In *Proc. IEEE International Conference on Image Processing*, volume 1, pages 609–612, Barcelona, Spain, September 2003. Invited Paper.

- [8] A. Aaron, S. Rane, E. Setton, and B. Girod. Transform-domain Wyner-Ziv Codec for Video. In *Proc. SPIE Visual Communications and Image Processing*, volume 5308, pages 520–528, San Jose, CA, January 2004.
- [9] A. Aaron, S. Rane, R. Zhang, and B. Girod. Wyner-Ziv Coding for Video: Applications to Compression and Error Resilience. In *Proc. IEEE Data Compression Conference*, pages 93–102, Snowbird, UT, March 2003.
- [10] A. Aaron, E. Setton, and B. Girod. Towards Practical Wyner-Ziv Coding of Video. In *Proc. IEEE International Conference on Image Processing*, volume 3, pages 869–872, Barcelona, Spain, September 2003.
- [11] A. Aaron, D. Varodayan, and B. Girod. Wyner-Ziv Residual Coding of Video. In *Proc. Picture Coding Symposium*, Beijing, China, April 2006.
- [12] A. Aaron, R. Zhang, and B. Girod. Wyner-Ziv Coding of Motion Video. In *Proc. Asilomar Conference on Signals and Systems*, volume 1, pages 240–244, Pacific Grove, CA, November 2002.
- [13] N. Ahmed, T. Natarajan, and K. R. Rao. Discrete Cosine Transform. *IEEE Transactions on Computers*, 23(1):90–93, January 1974.
- [14] S. Aign and K. Fazel. Temporal and Spatial Error Concealment Techniques for Hierarchical MPEG-2 Video Codec. In *Proc. IEEE International Conference on Communications*, volume 3, pages 1778–1783, Seattle, WA, June 1995.
- [15] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan. Priority Encoding Transmission. *IEEE Transactions on Information Theory*, 42(6):1737–1744, November 1996.
- [16] M. Andronico, A. Lombardo, S. Palazzo, and G. Schembra. Performance Analysis of Priority Encoding Transmission of MPEG Video Streams. In *Proc. Global Telecommunications Conference (GLOBECOM '96)*, pages 267–271, London, UK., November 1996.



- [17] J. Apostolopoulos. Reliable Video Communication over Lossy Packet Networks using Multiple State Encoding and Path Diversity. In *Proc. SPIE Conference on Visual Communications and Image Processing*, volume 1, pages 392–409, San Jose, CA, January 2001.
- [18] J. Apostolopoulos and M. Trott. Path Diversity for Enhanced Media Streaming. *IEEE Communications Magazine*, 42(8):80–87, August 2004. Special Issue on Proxy Support for Streaming on the Internet.
- [19] J. Apostolopoulos, M. Trott, and W. Tan. Path Diversity for Media Streaming. In M. van der Schaar and P. Chou, editors, *Multimedia over IP and Wireless Networks: Compression, Networking, and Systems*, pages 559–590. Elsevier Academic Press, March 2007.
- [20] J. Apostolopoulos and S. Wee. Unbalanced Multiple Description Video Communication using Path Diversity. In *IEEE International Conference on Image Processing*, volume 1, pages 966–969, Rochester, NY, October 2001.
- [21] B. Girod and N. Färber. Feedback-based Error Control for Mobile Video Transmission. *Proceedings of the IEEE, Special Issue on Video for Mobile Multimedia*, 87(10):1707–1723, October 1999.
- [22] B. Girod and N. Färber. Wireless Video. In A. Reibman and M.-T. Sun, editors, *Compressed Video over Networks*, pages 465–512. Marcel Dekker, 2000.
- [23] P. Baccichet. *Solutions for the Protection and Reconstruction of H.264/AVC Video Signals for Real-time Transmission over Lossy Networks*. Ph.D. dissertation, University of Milan, 2006.
- [24] P. Baccichet, D. Bagni, A. Chimienti, L. Pezzoni, and F. Rovati. Frame Concealment for H.264/AVC Decoders. *IEEE Transactions on Consumer Electronics*, 51(1):227–233, February 2005.
- [25] P. Baccichet, S. Rane, and B. Girod. Systematic Lossy Error Protection using H.264/AVC Redundant Slices and Flexible Macroblock Ordering. *Journal of Zhejiang University, Science*, 7(5):727–736, May 2006.

- [26] P. Baccichet, S. Rane, D.Y. Suh, and B. Girod. Core Experiments on Error Resilience Schemes for H.264/AVC and SVC. In *Output documents of 20th ITU-T/MPEG JVT Meeting*, number JVT-T309, Klagenfurt, Austria, July 2006.
- [27] J. Bajcsy and P. Mitran. Coding for the Slepian-Wolf Problem with Turbo Codes. In *Proc. IEEE Global Communications Conference*, volume 2, pages 1400–1404, San Antonio, TX, November 2001.
- [28] R. J. Barron, B. Chen, and G. W. Wornell. The Duality between Information Embedding and Source Coding with Side Information and some Applications. *IEEE Transactions on Information Theory*, 49(5):1159–1180, May 2003.
- [29] W. R. Bennett. Spectra of Quantized Signals. *Technical Journal* 27, Bell Systems, July 1948.
- [30] C. Berrou, A. Glavieux, and P. Thitimajshima. Near-Shannon Limit Error-Correcting Coding: Turbo Codes. In *Proc. IEEE International Conference on Communications*, pages 1064–1070, Geneva, Switzerland, May 1993.
- [31] G. Bjontegaard. An Error-Resilient Method based on Back-Channel Signalling and FEC. *ITU-T/SG15/WP15/1, Input Document No. LBC-96-033*, January 1996.
- [32] S. Boucheron and M. R. Salamatian. About Priority Encoding Transmission. *IEEE Transactions on Information Theory*, 46(2):699–705, March 2000.
- [33] J. Cardinal and G. Van Asche. Joint Entropy-Constrained Multiterminal Quantization. In *Proc. IEEE International Symposium on Information Theory*, page 63, Lausanne, Switzerland, June 2002.
- [34] J. Chakareski, J. Apostolopoulos, S. Wee, W. Tan, and B. Girod. Rate-Distortion Hint Tracks for Adaptive Video Streaming. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(10):1257–1269, October 2005.

- [35] J. Chakareski and B. Girod. Computing Rate-Distortion Optimized Policies for Streaming Media with Rich Acknowledgments. In *Proc. Data Compression Conference (DCC)*, volume 1, pages 202–211, Snowbird, UT, March 2004.
- [36] M. Chan, Y. Yu, and A. Constantinides. Variable-Size Block Matching Motion Compensation with Applications to Video Coding. *IEE Proceedings on Communications, Speech and Vision*, 137(4):205–212, August 1990.
- [37] T. Chiang and Y. Q. Zhang. A New Rate Control Scheme using Quadratic Rate-Distortion Model. *IEEE Transactions on Circuits and Systems for Video Technology*, 7(1):246–250, February 1997.
- [38] P. Chou and Z. Miao. Rate-Distortion Optimized Streaming of Packetized Media. *IEEE Transactions on Multimedia*, 8(2):390–404, April 2006.
- [39] T. P. Coleman, A. H. Lee, M. Medard, and M. Effros. On some New Approaches to Practical Slepian-Wolf Compression Inspired by Channel Coding. In *Proc. IEEE Data Compression Conference*, pages 282–291, Snowbird, UT, March 2004.
- [40] M. Costa. Writing on Dirty Paper. *IEEE Transactions on Information Theory*, 29(3):439–441, May 1983.
- [41] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.
- [42] Y. Dhondt and P. Lambert. Flexible Macroblock Ordering: An Error Resilience Tool in H.264/AVC. *Abstracts of the Fifth FTW Ph.D. Symposium*, (106), December 2004.
- [43] W. Ding and B. Liu. Rate Control of MPEG Video Coding and Recording by Rate-Quantization Modeling. *IEEE Transactions on Circuits and Systems for Video Technology*, 6(1):12–20, February 1996.
- [44] E. Setton. *Congestion-Aware Video Streaming over Peer-to-Peer Networks*. Ph.D. dissertation, Stanford University, December 2006.

- [45] M. Effros and D. Muresan. Codecell Contiguity in Optimal Fixed-Rate and Entropy-Constrained Network Scalar Quantizers. In *Proc. IEEE Data Compression Conference*, pages 312–321, Snowbird, UT, April 2002.
- [46] P. Elias. Predictive Coding - I. *IEEE Transactions on Information Theory*, 1(1):16–24, March 1955.
- [47] P. Elias. Predictive Coding - II. *IEEE Transactions on Information Theory*, 1(1):24–33, March 1955.
- [48] M. Fleming and M. Effros. Network Vector Quantization. In *Proc. IEEE Data Compression Conference*, pages 13–22, Snowbird, UT, March 2001.
- [49] M. Fleming, Q. Zhao, and M. Effros. Network Vector Quantization. *IEEE Transactions on Information Theory*, 50(8):1584–1604, 2004.
- [50] M. Flierl, T. Wiegand, and B. Girod. Rate-Constrained Multihypothesis Prediction for Motion-Compensated Video Compression. *IEEE Journal on Circuits and Systems for Video Technology*, 12(11):957–969, November 2002.
- [51] G. David Forney. Convolutional Codes - I, Algebraic Structure. *IEEE Transactions on Information Theory*, 16(6):720–738, November 1970.
- [52] E. Fromout, J. Biemond, and R. Lagendijk. Forward Rate Control for MPEG Recording. *Proceedings of the SPIE Visual Communications and Image Processing*, 2094(1):184–194, October 1993.
- [53] S. Fukunaga, T. Nakai, and H. Inoue. Error-Resilient Video Coding by Dynamic Replacing of Reference Pictures. In *Proc. IEEE Global Telecommunications Conference*, volume 3, pages 1503–1508, November 1996.
- [54] R. Gallager. Low-Density Parity-Check Codes. *IEEE Transactions on Information Theory*, 8(1):21–28, January 1962.
- [55] M. Gallant and F. Kossentini. Rate-Distortion Optimized Layered Coding with Unequal Error Protection for Robust Internet Video. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(3):357–372, March 2001.

- [56] J. García-Frías. Compression of Correlated Binary Sources using Turbo Codes. *IEEE Communications Letters*, 5(10):417–419, October 2001.
- [57] J. García-Frías. Joint Source-Channel Decoding of Correlated Sources over Noisy Channels. In *Proc. IEEE Data Compression Conference*, pages 283–292, Snowbird, UT, March 2001.
- [58] J. García-Frías and Y. Zhao. Data Compression of Unknown Single and Correlated Binary Sources using Punctured Turbo Codes. In *Proc. Allerton Conference on Communication, Control, and Computing*, Monticello, IL, October 2001.
- [59] J. García-Frías and Y. Zhao. Compression of Binary Memoryless Sources using Punctured Turbo Codes. *IEEE Communications Letters*, 6(9):394–396, September 2002.
- [60] N. Gehrig and P. Dragotti. Distributed Compression of the Plenoptic Function. In *Proceedings of the IEEE International Conference on Image Processing*, volume 1, pages 529–532, Singapore, October 2004.
- [61] N. Gehrig and P. L. Dragotti. Symmetric and Asymmetric Slepian-Wolf Codes with Systematic and Non-Systematic Linear Codes. *IEEE Communications Letters*, 9(1):61–63, January 2005.
- [62] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, Boston, MA, 1992.
- [63] B. Girod. The Efficiency of Motion-Compensating Prediction for Hybrid Coding of Video Sequences. *IEEE Journal on Selected Areas in Communications*, 5(7):1140–1154, August 1987.
- [64] B. Girod. Motion-Compensating Prediction with Fractional-Pel Accuracy. *IEEE Transactions on Communications*, 41(4):604–612, April 1993.

- [65] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero. Distributed Video Coding. *Proceedings of the IEEE, Special Issue on Advances in Video Coding and Delivery*, 93(1):71–83, January 2005.
- [66] V. Goyal. Multiple Description Coding: Compression meets the Network. *IEEE Signal Processing Magazine*, 18(5):74–93, September 2001.
- [67] V. Goyal. Theoretical Foundations of Transform Coding. *IEEE Signal Processing Magazine*, 18(5):9–21, September 2001.
- [68] H. Radha, M. van der Schaar and Y. Chen. The MPEG-4 Fine-Grained Scalable Video Coding Method for Multimedia Streaming over IP. *IEEE Transactions on Multimedia*, 3(1):53–68, March 2001.
- [69] ITU-T Recommendation H.245. *Control Protocol for Multimedia Communication*. 1996.
- [70] J. Hagenauer. Rate-Compatible Punctured Convolutional Codes (RCPC codes) and their Applications. *IEEE Transactions on Communications*, 36(4):389–400, April 1988.
- [71] M. Hannuksela, Y. K. Wang, and M. Gabbouj. Sub-picture: ROI Coding and Unequal Error Protection. In *Proc. IEEE International Conference on Image Processing (ICIP)*, volume 3, pages 537–540, Rochester, NY, September 2002.
- [72] M. Hannuksela, Y. K. Wang, and M. Gabbouj. Isolated Regions in Video Coding. *IEEE Transactions on Multimedia*, 6(2):259–267, April 2004.
- [73] Z. He and S. Mitra. Optimum Bit Allocation and Accurate Rate Control for Video Coding via Rho-domain Source Modeling. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(10):840–849, October 2002.
- [74] W. Heinzelman, M. Budagavi, and R. Talluri. Unequal Error Protection of MPEG-4 Compressed Video. In *Proc. IEEE International Conference on Image Processing*, volume 2, pages 503–504, Kobe, Japan, October 1999.

- [75] S. Hemami and T. Meng. Transform-coded Image Reconstruction Exploiting Interblock Correlation. *IEEE Transactions on Image Processing*, 4(7):1023–1027, July 1995.
- [76] B. Heng, J. Apostolopoulos, and J. Lim. End-to-End Rate-Distortion Optimized MD Mode Selection for Adaptive Multiple Description Video Coding. *EURASIP Journal on Applied Signal Processing*, 2006:Article ID 32592, 12 pages, October 2006.
- [77] U. Horn, K. Stuhlmüller, M. Link, and B. Girod. Robust Internet Video Transmission based on Scalable Coding and Unequal Error Protection. *Image Communication, Special Issue on Real-time Video over the Internet*, 15(1-2):77–94, September 1999.
- [78] ISO/IEC. *MPEG-2 Test Model 5 (TM5) ISO/IEC JTC1/SC29/WG11/93-225b*. April 1993.
- [79] ITU-T Video Coding Experts Group and ISO/IEC Moving Picture Experts Group. Study of Final Committee Draft of Joint Video Specification, (ITU-T Rec. H.264, ISO/IEC 14496-10 AVC). March 2003.
- [80] N. Jayant and P. Noll. *Digital Coding of Waveforms*. Prentice-Hall, Englewood Cliffs, NJ, 1984.
- [81] M. Karczewicz and R. Kurceren. The SP and SI Frames Design for H.264/AVC. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):637–644, July 2003. Special Issue on the H.264/AVC Video Coding Standard.
- [82] J. H. Kim, R. M. Mersereau, and Y. Altunbasak. Bitplane-wise Unequal Error Protection for Internet Video Applications. In *Proc. IEEE International Conference on Communications*, volume 4, pages 2508–2512, New York, NY, April 2002.
- [83] Klaus Stuhlmüller. *Modeling and Optimization of Video Transmission Systems*. Ph.D. dissertation, Technischen Fakultät der Universität Erlangen-Nürnberg, 2000.

- [84] J. Kusuma, L. Doherty, and K. Ramchandran. Distributed Compression for Sensor Networks. In *Proc. IEEE International Conference on Image Processing*, volume 1, pages 82–85, Thessaloniki, Greece, October 2001.
- [85] W. Lam, A. Reibman, and B. Liu. Recovery of Lost or Erroneously Received Motion Vectors. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 417–420, Minneapolis, MN, April 1993.
- [86] C. F. Lan, A. D. Liveris, K. Narayanan, Z. Xiong, and C. Georghiades. Slepian-Wolf Coding of Multiple M-ary Sources using LDPC Codes. In *Proc. IEEE Data Compression Conference*, page 549, Snowbird, UT, March 2004.
- [87] A. Leontaris and P. C. Cosman. Dual Frame Video Encoding with Feedback. In *Proc. 37th Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1514–1518, November 2003.
- [88] W. Li. Overview of Fine Granularity Scalability in MPEG-4 Video Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(3):301–317, March 2001.
- [89] L. Liang, P. Salama, and E. Delp. Unequal Error Protection using Wyner-Ziv Coding for Error Resilience. In *Proc. SPIE Conference on Visual Communications and Image Processing*, volume 6508, pages 65080S1–65080S9, San Jose, CA, January 2007.
- [90] Y. Liang. *Robust Low-Latency Voice and Video Communication over Best-Effort Networks*. Ph.D. dissertation, Stanford University, 2003.
- [91] Z. Liu, S. Cheng, A. Liveris, and Z. Xiong. Slepian-Wolf Coded Nested Quantization (SWC-NQ) for Wyner-Ziv Coding: Performance Analysis and Code Design. In *Proc. IEEE Data Compression Conference*, pages 322–331, Snowbird, UT, March 2004.
- [92] A. Liveris, Z. Xiong, and C. Georghiades. Compression of Binary Sources with Side Information at the Decoder using LDPC Codes. *IEEE Communications Letters*, 6(10):440–442, October 2002.



- [93] A. Liveris, Z. Xiong, and C. Georghiades. Compression of Binary Sources with Side Information at the Decoder using LDPC Codes. In *Proc. IEEE Global Communications Symposium*, volume 2, pages 1300–1304, Taipei, Taiwan, November 2002.
- [94] A. Liveris, Z. Xiong, and C. Georghiades. Joint Source-Channel Coding of Binary Sources with Side Information at the Decoder using IRA Codes. In *Proc. IEEE Workshop on Multimedia Signal Processing*, pages 53–56, St. Thomas, US Virgin Islands, December 2002.
- [95] S. P. Lloyd. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, IT-28:129–1373, March 1982.
- [96] M. Flierl. *Video Coding with Superimposed Motion-Compensated Signals*. Ph.D. dissertation, Technischen Fakultät der Universität Erlangen-Nürnberg, 2003.
- [97] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky. Low-Complexity Transform and Quantization in H.264/AVC. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):598–603, July 2003.
- [98] D. Marpe, H. Schwarz, and T. Wiegand. Context-based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):620–636, July 2003.
- [99] P. Mitran and J. Bajcsy. Coding for the Wyner-Ziv Problem with Turbo-like Codes. In *Proc. IEEE International Symposium on Information Theory*, page 91, Lausanne, Switzerland, June 2002.
- [100] P. Mitran and J. Bajcsy. Near Shannon-Limit Coding for the Slepian-Wolf Problem. In *Proc. Biennial Symposium on Communications*, pages 95–99, Kingston, Ontario, June 2002.
- [101] P. Mitran and J. Bajcsy. Turbo Source Coding: A Noise-Robust Approach to Data Compression. In *Proc. IEEE Data Compression Conference*, page 465, Snowbird, UT, April 2002.

- [102] A. Mohr, E. Riskin, and R. Ladner. Unequal Loss Protection: Graceful Degradation of Image Quality over Packet Erasure Channels through Forward Error Correction. *IEEE Journal on Selected Areas in Communications*, 18(6):819–828, June 2000.
- [103] A. E. Mohr, R. E. Ladner, and E. A. Riskin. Approximately Optimal Assignment for Unequal Loss Protection. In *Proc. IEEE International Conference on Image Processing (ICIP 2000)*, volume 1, pages 367–370, Vancouver, BC, Canada, September 2000.
- [104] D. Muresan and M. Effros. Quantization as Histogram Segmentation: Globally Optimal Scalar Quantizer Design in Network Systems. In *Proc. IEEE Data Compression Conference*, pages 302–311, Snowbird, UT, March 2002.
- [105] N. Färber. *Feedback-based Error Control for Robust Video Transmission*. Ph.D. dissertation, Technischen Fakultät der Universität Erlangen-Nürnberg, 2000.
- [106] T. Nguyen and A. Zakhor. Matching Pursuits-based Multiple Description Video Coding for Lossy Environments. In *Proc. IEEE International Conference on Image Processing*, volume 1, pages 57–60, Barcelona, Spain, September 2003.
- [107] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG. *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, May 2003.
- [108] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG. *ITU-T Recommendation H.263: Video Coding for Low Bit Rate Communication (Prepublished Version)*. ITU-T VQEG, April 2005.
- [109] M. Orchard and G. Sullivan. Overlapped Block Motion Compensation - An Estimation-Theoretic Approach. *IEEE Transactions on Image Processing*, 3(5):693–699, September 1994.

- [110] A. Papoulis and S. U. Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill Publishing, fourth edition, 2002.
- [111] S. Pradhan, J. Chou, and K. Ramchandran. Duality between Source Coding and Channel Coding and its Extension to the Side Information Case. *IEEE Transactions on Information Theory*, 49(5):1181–1203, May 2003.
- [112] S. Pradhan and K. Ramchandran. Enhancing Analog Image Transmission Systems using Digital Side Information: A New Wavelet-based Image Coding Paradigm. In *Proc. IEEE Data Compression Conference*, pages 63–72, Snowbird, UT, March 2001.
- [113] S. S. Pradhan, J. Kusuma, and K. Ramchandran. Distributed Compression in a Dense Microsensor Network. *IEEE Signal Processing Magazine*, 19(2):51–60, March 2002.
- [114] S. S. Pradhan and K. Ramchandran. Distributed Source Coding using Syndromes (DISCUS): Design and Construction. In *Proc. IEEE Data Compression Conference*, pages 158–167, Snowbird, UT, March 1999.
- [115] S. S. Pradhan and K. Ramchandran. Distributed Source Coding: Symmetric Rates and Applications to Sensor Networks. In *Proc. IEEE Data Compression Conference*, pages 363–372, Snowbird, UT, March 2000.
- [116] S. S. Pradhan and K. Ramchandran. Geometric Proof of Rate-Distortion Function of Gaussian Sources with Side Information at the Decoder. In *Proc. IEEE International Symposium on Information Theory*, page 351, Sorrento, Italy, June 2000.
- [117] S. S. Pradhan and K. Ramchandran. Group-Theoretic Construction and Analysis of Generalized Coset Codes for Symmetric/Asymmetric Distributed Source Coding. In *Proc. Conference on Information Sciences and Systems*, Princeton, NJ, March 2000.

- [118] R. Prasad, C. Dovrolis, M. Murray, and K. Claffy. Bandwidth Estimation: Metrics, Measurement Techniques, and Tools. *IEEE Network Magazine*, 17(6):27–35, November 2003.
- [119] J. Proakis. *Digital Communications*. McGraw-Hill Science/Engineering/Math, fourth edition, 2000. ISBN 0072321113.
- [120] R. Puri and K. Ramchandran. PRISM: A New Robust Video Coding Architecture based on Distributed Compression Principles. In *Proc. Allerton Conference on Communication, Control, and Computing*, Allerton, IL, October 2002.
- [121] R. Puri and K. Ramchandran. PRISM: A ‘Reversed’ Multimedia Coding Paradigm. In *Proc. IEEE International Conference on Image Processing*, volume 1, pages 617–620, Barcelona, Spain, September 2003.
- [122] R. Puri and K. Ramchandran. PRISM: An Uplink-Friendly Multimedia Coding Paradigm. In *Proc. International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 856–859, Hong Kong, April 2003.
- [123] E. Quacchio, E. Magli, G. Olmo, P. Baccichet, and A. Chimienti. Enhancing Whole-Frame Error Concealment with an Intra Motion Vector Estimator in H.264/AVC. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2005)*, volume 2, pages 329–332, Philadelphia, PA, March 2005.
- [124] P. Ramanathan and B. Girod. Receiver-driven Rate-Distortion Optimized Streaming of Light Fields. In *Proc. IEEE International Conference on Image Processing*, volume 3, pages 25–28, Genoa, Italy, September 2005.
- [125] S. Rane, A. Aaron, and B. Girod. Systematic Lossy Forward Error Protection for Error-Resilient Digital Video Broadcasting. In *Proc. SPIE Visual Communications and Image Processing*, volume 5308, pages 520–528, San Jose, CA, January 2004.

- [126] S. Rane, A. Aaron, and B. Girod. Systematic Lossy Forward Error Protection for Error-Resilient Digital Video Broadcasting - A Wyner-Ziv Coding Approach. In *Proc. IEEE International Conference on Image Processing*, volume 5, pages 3101–3104, Singapore, October 2004.
- [127] S. Rane, A. Aaron, and B. Girod. Error-Resilient Video Transmission using Multiple Embedded Wyner-Ziv Descriptions. In *Proc. IEEE International Conference on Image Processing*, volume 2, pages 666–669, Genoa, Italy, September 2005.
- [128] S. Rane, P. Baccichet, and B. Girod. Modeling and Optimization of a Systematic Lossy Error Protection System. In *Picture Coding Symposium*, Beijing, China, April 2006.
- [129] S. Rane, P. Baccichet, and B. Girod. Systematic Lossy Error Protection based on H.264/AVC Redundant Slices. In *Input documents to 20th ITU-T/MPEG JVT Meeting*, number JVT-T093, Klagenfurt, Austria, July 2006.
- [130] S. Rane, P. Baccichet, and B. Girod. Systematic Lossy Error Protection based on H.264/AVC Redundant Slices. In *Input documents to 21st ITU-T/MPEG JVT Meeting*, number JVT-U057, Hangzhou, China, October 2006.
- [131] S. Rane, P. Baccichet, and B. Girod. Systematic Lossy Error Protection based on H.264/AVC Redundant Slices and Flexible Macroblock Ordering. In *Input documents to 19th ITU-T/MPEG JVT Meeting*, number JVT-S025, Geneva, Switzerland, April 2006.
- [132] S. Rane, P. Baccichet, D. Y. Suh, and B. Girod. Core Experiments for Parity-Based Schemes for Error Resilience in H.264/AVC and SVC. In *Output documents of 19th ITU-T/MPEG JVT Meeting*, number JVT-S306, Geneva, Switzerland, April 2006.
- [133] S. Rane and B. Girod. Analysis of Error-Resilient Video Transmission based on Systematic Source-Channel Coding. In *Picture Coding Symposium*, San Francisco, CA, December 2004.

- [134] S. Rane and B. Girod. Systematic Lossy Error Protection versus Layered Coding with Unequal Error Protection. In *Proc. SPIE Image and Video Communications and Processing*, volume 5685, pages 663–671, San Jose, CA, January 2005.
- [135] S. Rane and B. Girod. Systematic Lossy Error Protection based on H.264/AVC Redundant Slices. In *Proc. SPIE Visual Communications and Image Processing*, volume 6077, pages Q2–Q9, San Jose, CA, January 2006.
- [136] S. Rane, D. Rebollo-Monedero, and B. Girod. High-Rate Analysis of Systematic Lossy Error Protection of a Predictively Encoded Source. In *Proc. IEEE Data Compression Conference*, pages 263–272, Snowbird, UT, March 2007.
- [137] D. Rebollo-Monedero, A. Aaron, and B. Girod. Transforms for High-Rate Distributed Source Coding. In *Proc. Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 850–854, Pacific Grove, CA, November 2003.
- [138] D. Rebollo-Monedero, S. Rane, A. Aaron, and B. Girod. High-Rate Quantization and Transform Coding with Side Information at the Decoder. *EURASIP Signal Processing Journal, Special Issue on Distributed Source Coding*, 86:3160–3179, April 2006.
- [139] D. Rebollo-Monedero, S. Rane, and B. Girod. Wyner-Ziv Quantization and Transform Coding of Noisy Sources at High Rates. In *Proc. Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 2084–2088, Pacific Grove, CA, November 2004.
- [140] D. Rebollo-Monedero, R. Zhang, and B. Girod. Design of Optimal Quantizers for Distributed Source Coding. In *Proc. IEEE Data Compression Conference*, pages 13–22, Snowbird, UT, March 2003.
- [141] I. S. Reed and G. Solomon. Polynomial Codes over Certain Finite Fields. *Journal of the Society for Industrial and Applied Mathematics*, 8(2):300–304, June 1960.

- [142] A. Reibman, H. Jafarkhani, Y. Wang, M. Orchard, and R. Puri. Multiple Description Video Coding using Motion-Compensated Temporal Prediction. 12(3):193–204, March 2002.
- [143] J. Ribas-Corbera and S. Lei. Rate Control in DCT Video Coding for Low-Delay Communications. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(1):172–185, February 1999.
- [144] J. Rosenberg and H. Schulzrinne. An RTP Payload Format for Generic Forward Error Correction. IETF, RFC 2733, December 1999.
- [145] P. Salama, N. Shroff, and E. Delp. A Bayesian Approach to Error Concealment in Encoded Video Streams. In *Proc. IEEE International Conference on Image Processing*, volume 1, pages 49–52, Lausanne, Switzerland, September 1996.
- [146] P. Salama, N. Shroff, and E. Delp. Raptor Codes. *IEEE Transactions on Information Theory*, 52(6):2551 – 2567, June 2006.
- [147] D. Schonberg, S. S. Pradhan, and K. Ramchandran. LDPC Codes can approach the Slepian-Wolf Bound for General Binary Sources. In *Proc. Allerton Conference on Communication, Control, and Computing*, Champaign, IL, October 2002.
- [148] D. Schonberg, S. S. Pradhan, and K. Ramchandran. Distributed Code Constructions for the Entire Slepian-Wolf Rate Region for Arbitrarily Correlated Sources. In *Proc. Asilomar Conference on Signals and Systems*, volume 1, pages 835–839, Pacific Grove, CA, November 2003.
- [149] D. Schonberg, S. S. Pradhan, and K. Ramchandran. Distributed Code Constructions for the Entire Slepian-Wolf Rate Region for Arbitrarily Correlated Sources. In *Proc. IEEE Data Compression Conference*, pages 292–301, Snowbird, UT, March 2004.
- [150] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: A Transport Protocol for Real-time Applications. IETF, RFC 3550, July 2003.

- [151] A. Sehgal and N. Ahuja. Robust Predictive Coding and the Wyner-Ziv Problem. In *Proc. IEEE Data Compression Conference*, pages 103–112, Snowbird, UT, March 2003.
- [152] A. Sehgal, A. Jagmohan, and N. Ahuja. A Causal State-free Video Encoding Paradigm. In *Proc. IEEE International Conference on Image Processing*, volume 1, pages 605–608, Barcelona, Spain, September 2003.
- [153] S. D. Servetto. Lattice Quantization with Side Information. In *Proc. IEEE Data Compression Conference*, pages 510–519, Snowbird, UT, March 2000.
- [154] E. Setton and B. Girod. Rate-Distortion Analysis and Streaming of SP and SI Frames. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(6):733–743, June 2006.
- [155] S. Shamai, S. Verdú, and R. Zamir. Systematic Lossy Source/Channel Coding. *IEEE Transactions on Information Theory*, 44(2):564–579, March 1998.
- [156] C. E. Shannon. A Mathematical Theory of Communication. *Bell Systems Technical Journal*, 27:379–423, July 1948.
- [157] B. Sklar. *Digital Communications: Fundamentals and Applications*. Prentice-Hall, second edition, 2001. ISBN 0130847887.
- [158] D. Slepian and J. Wolf. Noiseless Coding of Correlated Information Sources. *IEEE Transactions on Information Theory*, IT-19:471–480, July 1973.
- [159] V. Stankovic, A. D. Liveris, Z. Xiong, and C. N. Georghiades. Design of Slepian-Wolf Codes by Channel Code Partitioning. In *Proc. IEEE Data Compression Conference*, pages 302–311, Snowbird, UT, March 2004.
- [160] Y. Steinberg and N. Merhav. On Successive Refinement for the Wyner-Ziv Problem. *IEEE Transactions on Information Theory*, 50(8):1636–1654, August 2004.



- [161] Y. Steinberg and N. Merhav. On Hierarchical Joint Source-Channel Coding with Degraded Side Information. *IEEE Transactions on Information Theory*, 52(3):886–903, March 2006.
- [162] T. Stockhammer, D. Kontopodis, and T. Wiegand. Rate-Distortion Optimization for JVT/H.26L Video Coding in Packet Loss Environment. In *Proceedings of the 2002 Packet Video Workshop*, Pittsburgh, PA, April 2002.
- [163] K. Stuhlmüller, N. Färber, M. Link, and B. Girod. Analysis of Video Transmission over Lossy Channels. *IEEE Journal on Selected Areas in Communications*, 18(6):1012–1032, June 2000.
- [164] K. Stuhlmüller, M. Link, B. Girod, and U. Horn. Scalable Internet Video Streaming with Unequal Error Protection. In *Proc. Packet Video Workshop*, New York, NY, April 1999.
- [165] J. K. Su, J. J. Eggers, and B. Girod. Illustration of the Duality between Channel Coding and Rate Distortion with Side Information. In *Proc. 2000 Asilomar Conference on Signals and Systems*, volume 2, pages 1841–1845, Pacific Grove, CA, October 2000.
- [166] K. Suehring. JVT JM reference software home page (online). <http://iphome.hhi.de/suehring/tml/>.
- [167] G. Sullivan and R. Baker. Rate-Distortion Optimized Motion Compensation for Video Compression with Fixed or Variable Block Sizes. In *Proc. IEEE Global Telecommunications Conference*, pages 85–90, Phoenix, AZ, December 1991.
- [168] H. Sun and W. Kwok. Concealment of Damaged Block Transform-coded Images using Projections onto Convex Sets. *IEEE Transactions on Image Processing*, 4(4):470–477, April 1995.
- [169] X. Tang and A. Zakhor. Matching Pursuits Multiple Description Coding for Wireless Video. In *Proc. IEEE International Conference on Image Processing*, volume 1, pages 926–929, Thessaloniki, Greece, October 2001.

- [170] B. Tao, H. Peterson, and B. Dickinson. A Rate-Quantization Model for MPEG Encoders. In *Proc. IEEE International Conference on Image Processing*, volume 1, pages 338–341, Santa Barbara, CA, October 1997.
- [171] T. Tillo, B. Penna, P. Frossard, and P. Vandergheynst. Distributed Coding of Spherical Images with Jointly Refined Decoding. In *Proc. Multimedia Signal Processing Workshop*, pages 1–4, Shanghai, China, October 2005.
- [172] M. van der Schaar and H. Radha. A Hybrid Temporal-SNR Fine-Granular Scalability for Internet Video. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(3):318–331, March 2001.
- [173] D. Varodayan, A. Mavlankar, M. Flierl, and B. Girod. Distributed Coding of Random Dot Stereograms with Unsupervised Learning of Disparity. In *Proc. IEEE International Workshop on Multimedia Signal Processing*, pages 5–8, Victoria, Canada, October 2006.
- [174] D. Varodayan, A. Mavlankar, M. Flierl, and B. Girod. Distributed Grayscale Stereo Image Coding with Unsupervised Learning of Disparity. In *Proc. IEEE Data Compression Conference*, pages 143–152, Snowbird, UT, March 2007.
- [175] G. Wang, Q. Zhang, and W. Zhu. Channel-Adaptive Error Protection for Scalable Video over Channels with Bit Errors and Packet Erasures. In *Proc. IEEE International Symposium on Circuits and Systems*, volume 2, pages 712–715, Scottsdale, AZ, May 2002.
- [176] G. Wang, Q. Zhang, W. Zhu, and Y.-Q. Zhang. Channel-Adaptive Unequal Error Protection for Scalable Video Transmission over Wireless Channel. In *Proc. SPIE Conference on Visual Communications and Image Processing*, volume 4310, pages 648–655, SanJose, CA, January 2001.
- [177] J. Wang, A. Majumdar, and K. Ramchandran. Robust Video Transmission over a Lossy Network using a Distributed Source Coded Auxiliary Channel. In *Proc. Picture Coding Symposium (PCS)*, San Francisco, CA, December 2004.

- [178] J. Wang, A. Majumdar, and K. Ramchandran. On Enhancing MPEG Video Broadcast over Wireless Networks with an Auxiliary Broadcast Channel. In *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pages 1101–1104, Philadelphia, PA, March 2005.
- [179] X. Wang and M.T. Orchard. Design of Trellis Codes for Source Coding with Side Information at the Decoder. In *Proc. IEEE Data Compression Conference*, pages 361–370, Snowbird, UT, March 2001.
- [180] Y. Wang, M. Hannuksela, V. Varsa, A. Hourunranta, and M. Gabbouj. The Error Concealment Feature in the H.26L Test Model. In *Proc. IEEE International Conference on Image Processing*, volume 2, pages 729–732, Rochester, NY, September 2002.
- [181] Y. Wang, M. Orchard, V. Vaishampayan, and R. Reibman. Multiple Description Coding using Pairwise Correlating Transforms. *IEEE Transactions on Image Processing*, 10(3):351–366, March 2001.
- [182] Y. Wang, J. Ostermann, and Y. Zhang. *Video Processing and Communications*. Prentice-Hall, NJ, first edition, 2002.
- [183] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos. Error-Resilient Video Coding Techniques. *IEEE Signal Processing Magazine, Special Issue on Multimedia Communications over Networks*, 17(4):61–82, July 2000.
- [184] Y. Wang and Q. Zhu. Error Control and Concealment for Video Communications: A Review. *Proceedings of the IEEE, Special Issue on Multimedia Signal Processing*, 86(5):974–977, May 1998.
- [185] S. Wenger. Video Redundancy Coding in H.263+. In *Proc. International Workshop on Audio-Visual Services over Packet Networks*, Aberdeen, UK, September 1997.
- [186] S. Wenger. Common Test Conditions for Wireline Low Delay IP/UDP/RTP Packet Loss-Resilient Testing. ITU-T SG16, VQEG-N79r1, 2001.

- [187] S. Wenger, T. Stockhammer, and M. Hannuksela. RTP Payload Format for H.264 Video. IETF, RFC 3984, Feb 2005.
- [188] T. Wiegand, H. Schwartz, A. Joch, F. Kossentini, and G. Sullivan. Rate-Constrained Coder Control and Comparison of Video Coding Standards. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):688–703, July 2003.
- [189] T. Wiegand and G. Sullivan. Rate-Distortion Optimization for Video Compression. *IEEE Signal Processing Magazine*, 15(6):74–90, November 1998.
- [190] T. Wiegand, G. Sullivan, G. Bjøntegaard, and A. Luthra. Overview of the H.264/AVC Video Coding Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):560–576, July 2003.
- [191] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, and M. Wien. *Joint Draft 5: Scalable Video Coding (in integrated form with ITU-T Rec. H.264, ISO/IEC 14995-10)*. ISO/IEC MPEG & ITU-T, January 2006.
- [192] T. Wiegand, X. Zhang, and B. Girod. Long-Term Memory Motion-Compensated Prediction. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(2):70–84, February 1999.
- [193] M. Wien. Variable Block-Size Transforms for H.264/AVC. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):604–613, July 2003.
- [194] A. Wyner. Recent Results in the Shannon Theory. *IEEE Transactions on Information Theory*, 20(1):2–10, January 1974.
- [195] A. Wyner. On Source Coding with Side Information at the Decoder. *IEEE Transactions on Information Theory*, IT-21(3):294–300, May 1975.
- [196] A. Wyner. The Rate-Distortion Function for Source Coding with Side Information at the Decoder—II: General Sources. *Information and Control*, 38(1):60–80, July 1978.

- [197] A. Wyner and J. Ziv. The Rate-Distortion Function for Source Coding with Side Information at the Decoder. *IEEE Transactions on Information Theory*, IT-22(1):1–10, January 1976.
- [198] J. Xin, A. Vetro, and H. Sun. Converting DCT Coefficients to H.264/AVC Transform Coefficients. In *Proc. 5th Pacific Rim Conference on Multimedia*, volume 3332, pages 939–946, Tokyo, Japan, November 2004. Archive: Springer Lecture Notes in Computer Science, ISBN 978-3-540-23977-2.
- [199] Z. Xiong, A. Liveris, and S. Cheng. Distributed Source Coding for Sensor Networks. In *IEEE Signal Processing Magazine*, volume 21, pages 80–94, September 2004.
- [200] Z. Xiong, A. Liveris, S. Cheng, and Z. Liu. Nested Quantization and Slepian-Wolf Coding: A Wyner-Ziv Coding Paradigm for I.I.D. Sources. In *Proc. IEEE Workshop on Statistical Signal Processing*, pages 399–402, St. Louis, MO, September 2003.
- [201] Q. Xu and Z. Xiong. Layered Wyner-Ziv Video Coding. *IEEE Transactions on Image Processing*, 15(12):3791–3803, December 2006.
- [202] Y. -C. Lin and B. Girod. Image Authentication Based on Distributed Source Coding. In *Proc. IEEE International Conference on Image Processing*, San Antonio, Texas, September 2007. Invited Paper, Special Session on Distributed Video and Image Coding.
- [203] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao. Wyner-Ziv Coding based on TCQ and LDPC Codes. In *Proc. Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 825–829, Pacific Grove, CA, November 2003.
- [204] R. Zamir. The Rate Loss in the Wyner-Ziv Problem. *IEEE Transactions on Information Theory*, 42(6):2073–2084, November 1996.
- [205] R. Zamir and S. Shamai. Nested Linear/Lattice Codes for Wyner-Ziv Encoding. In *Proc. Information Theory Workshop*, pages 92–93, Killarney, Ireland, June 1998.

- [206] R. Zamir, S. Shamai, and U. Erez. Nested Linear/Lattice Codes for Structured Multiterminal Binning. *IEEE Transactions on Information Theory*, 48(6):1250–1276, June 2002.
- [207] Y. Zhao and J. García-Frías. Data Compression of Correlated Non-Binary Sources using Punctured Turbo Codes. In *Proc. IEEE Data Compression Conference*, pages 242–251, Snowbird, UT, April 2002.
- [208] Y. Zhao and J. García-Frías. Joint Estimation and Compression of Correlated Non-Binary Sources using Punctured Turbo Codes. *IEEE Transactions on Communications*, 53(3):385–390, March 2005.
- [209] G. Zhu and F. Alajaji. Turbo Codes for Non-Uniform Memoryless Sources over Noisy Channels. *IEEE Communications Letters*, 6(2):64–66, February 2002.
- [210] X. Zhu, A. Aaron, and B. Girod. Distributed Compression for Large Camera Arrays. In *Proc. IEEE Workshop on Statistical Signal Processing*, pages 30–33, St. Louis, MO, September 2003.
- [211] X. Zhu, S. Rane, and B. Girod. Systematic Lossy Error Protection (SLEP) for Video Transmission over Wireless Ad Hoc Networks. In *Proc. SPIE Visual Communications and Image Processing*, volume 5960, pages 1849–1860, Beijing, China, July 2005.