Multiple description coding of visual data

Gabriella Olmo

Politecnico di TORINO
Department of Electronics
Corso Duca degli Abruzzi 24
10129 Torino
Gabriella.olmo@polito.it
www.telematica.polito.it/sas-ipl/

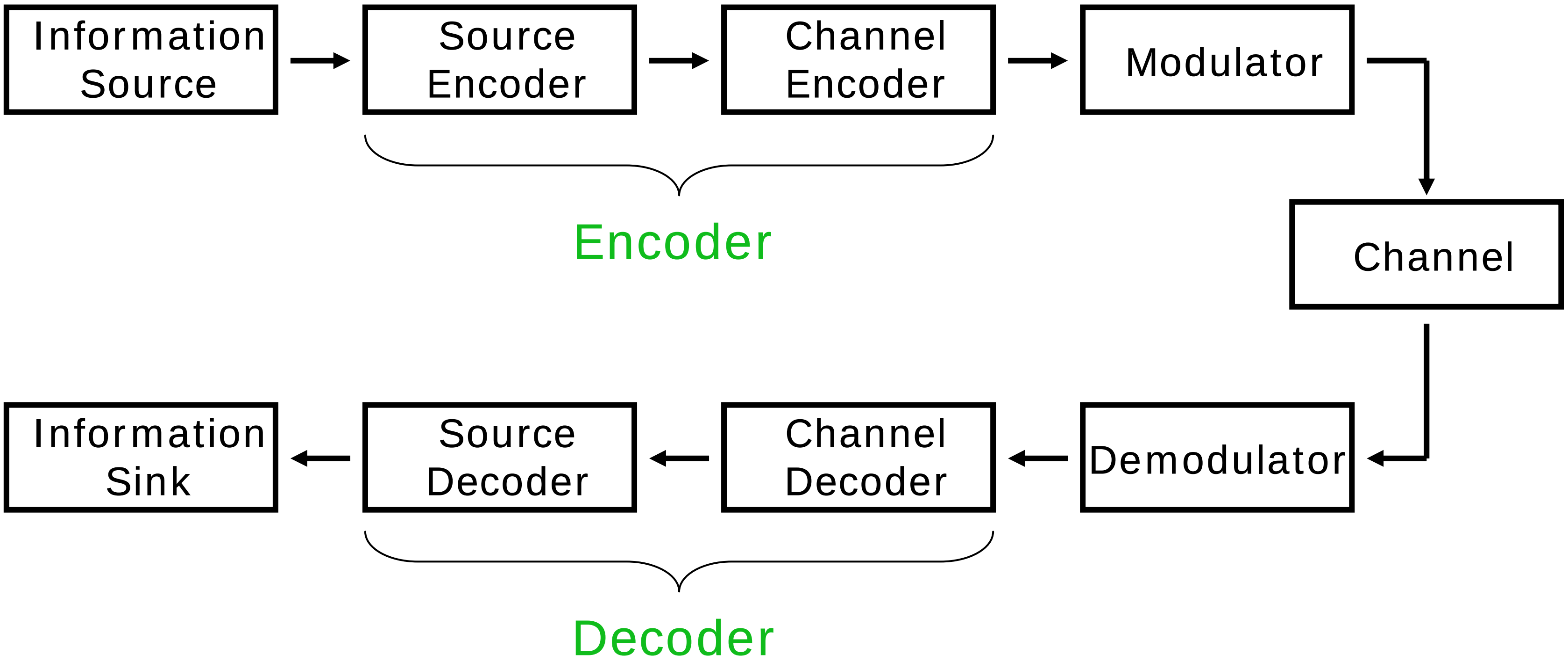
Outline

- **Introduction**
- **MDC rationale**
- **Classical methods:**
  - MD with UEP
  - MD with scalar quantization
  - MD correlating transforms
- **MD for still images**
  - Main topics
  - Oversampling based methods
  - RD optimized methods
- **MD for video**
  - The drift problem
  - Proposed solutions

Introduction

(When and why one should not to be happy with classical separated source and channel coding)

A single-sender communication system



Source Coding

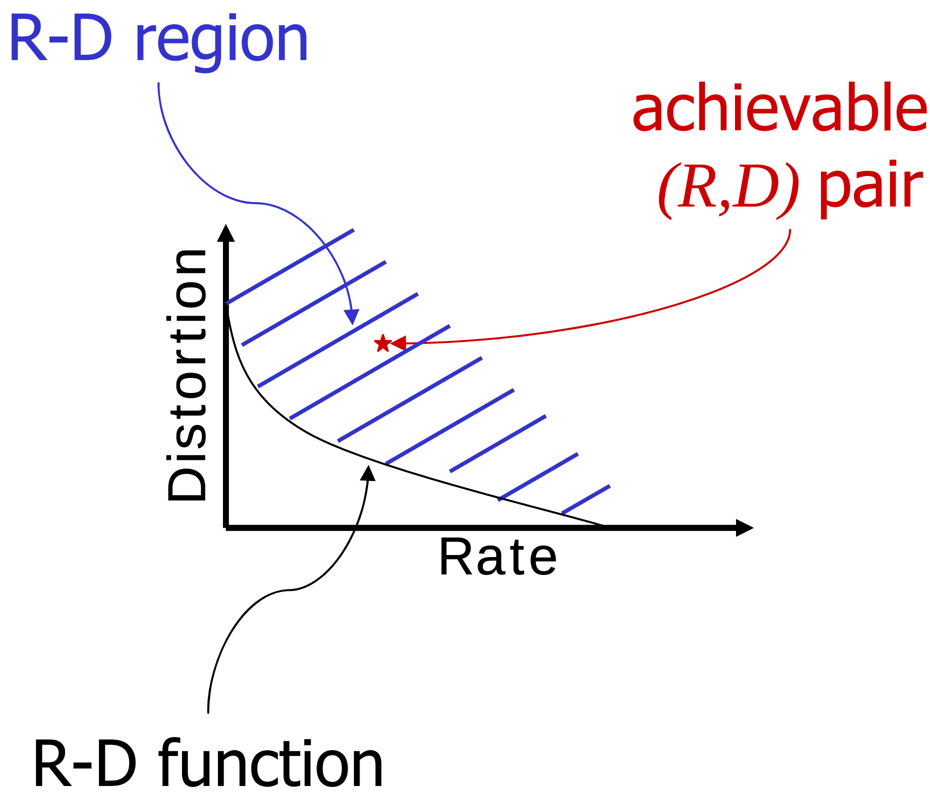
- **Source coding aims to:**
  - Represent the source information with minimum amount of bits → lossless compression, the limits are given by the source entropy

  OR

  - Minimize the distortion due to the source mapping → lossy compression, the limits are given by the R-D theory (→ see next slides)
- **Source coding does not deal with:**
  - How to combat errors (conceal, correct) due to transmission, storage, etc.
  - How to minimize distortion due to such errors

Rate Distortion theory

- A rate distortion pair *(R,D)* is ***achievable*** if there exists a source code with rate *R* *bits/sample* and distortion *D* for block length $n \rightarrow \infty$
- The closure of the achievable R-D pairs is the ***R-D region***, and is a feature of the source (difficult to determine!)
- The ***R-D function*** *D(R)* is the min. distortion such that *(R,D)* is in the R-D region



Rate Distortion bounds

- For a Gaussian source:

$$D(R) = \sigma^2 2^{-2R}$$

- For a source with probability density *f(x)* and variance σ^2:

$$\frac{1}{2\pi e} 2^{2h} 2^{-2R} \leq D(R) \leq \sigma^2 2^{-2R}$$

where $h = -\int_{-\infty}^{+\infty} f(x) \log_2 f(x) dx$ is the differential entropy.

The Gaussian sources are the most difficult to compress…

Visual impact of rate on distortion

Rate = 0.1 bits per pixel

Rate = 1.0 bits per pixel

Source and channel coding

- **Situation:**
  - Source coder has suppressed the redundancy
- **Problem:**
  - If an error occurs or a packet is lost, in principle there is no redundancy which can help detecting errors and correcting them
- **Solution:**
  - (*classical*) After source coding: introducing a certain amount of redundancy → Channel coding.
  - (*joint*) In the source coder: keeping or inserting a controllable amount of redundancy → MDC, UEP,…

A simple example: layered source

- Information is partitioned in two or more layers
  - Base layer: carries the basic information
  - Enhancement layer(s): carry the refinement information
  - Base layer by itself provides acceptable quality of the data
  - Base layer plus enhancement layer(s), yield fine quality reconstruction(s)
- Enhancement layer(s) are useless if base layer is lost
- **Assigning the same FEC or ARQ strategy to all symbols is nonsense!!**

The crisis of the separation principle

- If the data characteristics are taken into account
  - Progressivity, subjective importance
- If real application constraints are taken into account
  - Finite block length
  - Limited delay and computational complexity
  - Actual impact of errors, possibility of error concealment
- **The separation principle is not true any more. Redundancy should be traded between source and channel coders**

"Dear Source Coders,

Don't compress too much and consequently require error-free transmission. It makes the task for us channel coders very difficult. We might have to add more redundancy for error protection than you removed!"

Professor Joachim Hagenauer, 1996

Multiple description coding rationale

(How to by-pass many problems generating many others)

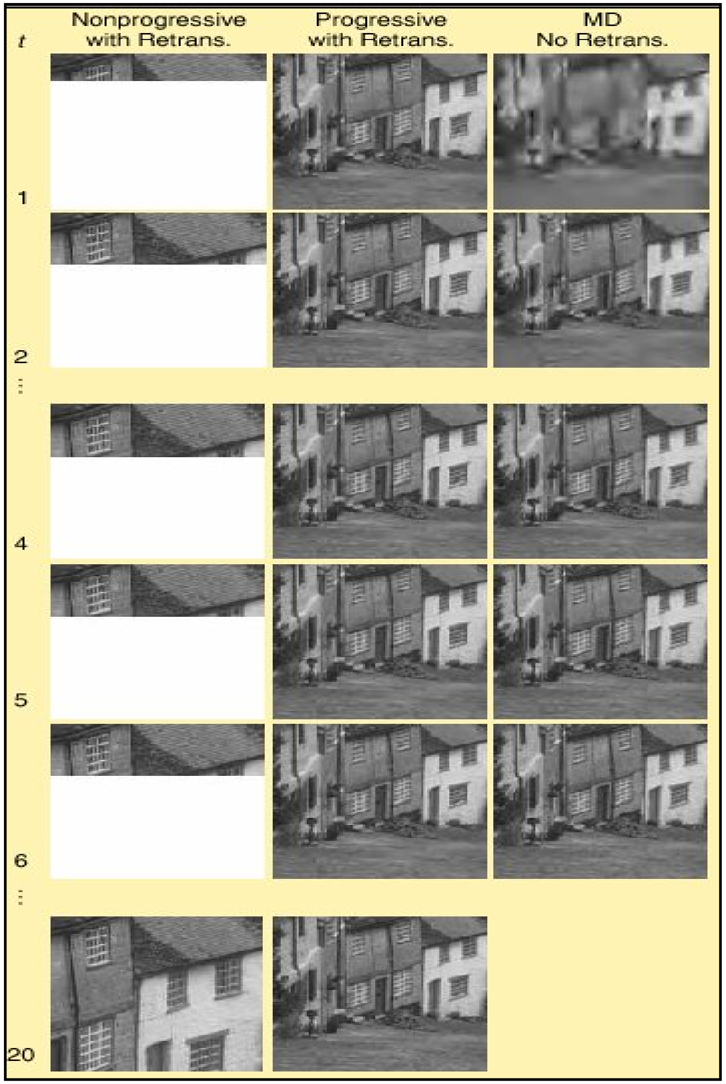
MDC basics

- Two or more representations *(descriptions)* of the same data are generated
- Descriptions should be independently decodable (opposite to layered coding)
- Several quality levels are obtained depending on the number of received descriptions.
- Full quality is obtained when decoding all descriptions

Downloading an image from a server using a web browser

- The image is represented using 6 packets. Third packet is lost
- The progressive (SNR) and the non-progressive representations use TCP as the transport layer → retransmissions
- The Multiple description representation uses UDP → no retransmissions

Upon receiving the lost packet



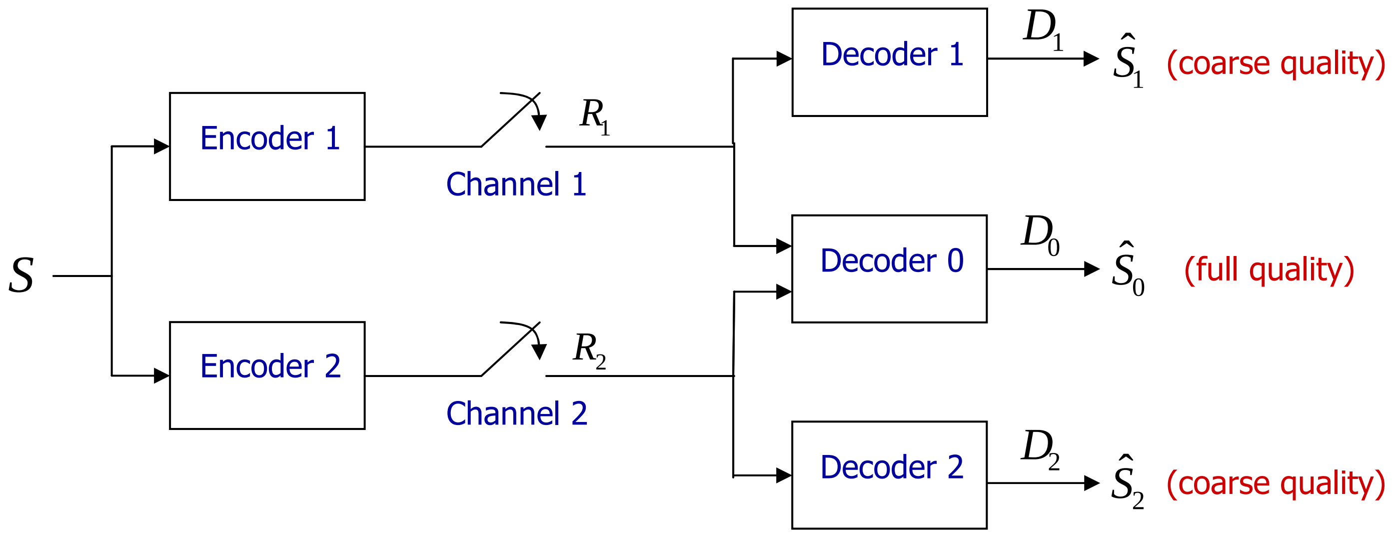
Pros and cons at a glance

- **PROS:**
  - All pieces of information are useful (in contrast to layered coding)
  - No ARQ is needed (suitable for data highly sensitive to delays - voice and video)
  - Able to exploit path diversity (actually *requires* physical or logical path diversity)
  - Matched to multimedia data, when several levels of quality may be acceptable
- **CONS:**
  - Redundancy (correlation among descriptions)
  - Not suitable for data transfer

Applications

- Internet
  - Streaming applications from multiple servers
  - Broadcasting (e.g. live video)
- Sensitive wired application
  - Service outages cannot be tolerated (security …)
- Exploitation of diversity
  - Multi antenna systems
- Distributed data storage
- Management of users with different bandwidth requirements

Two description scheme



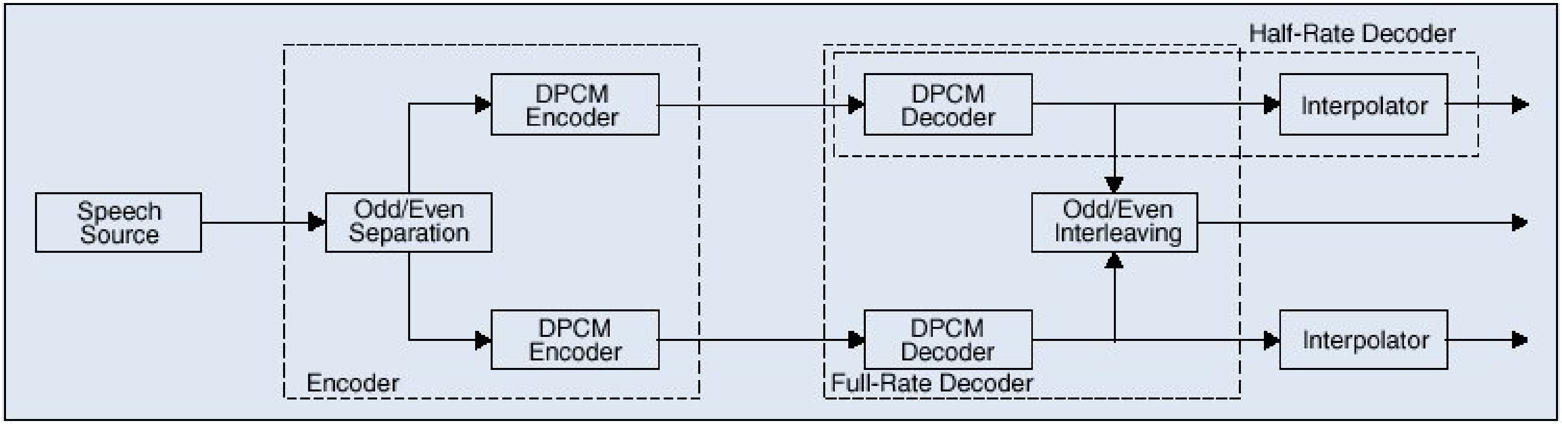
$$(R_1, R_2, D_0, D_1, D_2)$$

Balanced descriptions (mostly addressed):

$$\mathbf{D_1 = D_2,\ R_1 = R_2}$$

History of MDC (1)

- Investigated since the '70 to improve the reliability of speech over the telephone network, without using standby links
- Problem: find how to split the information and send it over two channels such that receiving only the information from one channel still yields acceptable quality
- **Jayant** proposed to sample the original voice signal @ 12 KHz, to separately (DPCM) compress odd and even samples and send them over two separate channels



History of MDC (2)

- **Witsenhausen:** If the source is described with two descriptions, and given a total rate, what are the limitations on the quality of such descriptions taken separately and jointly?
- **Theoretical handling**
  - Preliminary results presented on the Information Theory Workshop in **1979.** Since then, this problem has been known as the Multiple Description Coding (MDC) problem.
  - At first, MDC was considered only as an information theoretic problem.

The problem was clearly understood:

- A *deeply compressed* representation at rate R_1+R_2 cannot be split into two useful descriptions (we cannot estimate a missing description from the received ones)
- On the other hand, it is useless to have *very good descriptions over both channels* (too much rate is lost when both descriptions are received. If we have good descriptions at rates R_1 and $R_{2,}$ combining them at Decoder 0 does not yield much better results)
- So we should make descriptions that are *similar (to help mutual estimation)*, but *not too much* (we cannot waste too much rate)!

MD region for two descriptions

What is the set of achievable 5-tuples (R_1, R_2, D_1, D_2, D_0)?

Ozarow bound

$$D_i \geq 2^{-2R_i}, \qquad i = 1, 2$$
$$D_0 \geq 2^{-2(R_1+R_2)} \cdot \gamma(D_1, D_2, R_1, R_2)$$

$$\gamma = \frac{1}{1 - \left(\sqrt{(1-D_1)(1-D_2)} - \sqrt{D_1 D_2 - 2^{-2(R_1+R_2)}}\right)^2}$$

for $D_1 + D_2 < 1 + 2^{-2(R_1+R_2)}$ and $\gamma = 1$ otherwise.

Two examples:

- The best possible performance for side distortions:

  $D_1 = 2^{-2R1}$ and $D_2 = 2^{-2R2}$ → $D_0 \geq \min\{D_1, D_2\}/2$

- The best possible performance for central distortion:

  $D_0 = 2^{-2(R1+R2)}$ → one of the side distortions needs to be 1

Formalization of the problem

- We want to distribute the rate $R = R_1 + R_2$ between the descriptions in order to:
  - Min. D_1 given a bound on D_0 and total rate $R \leq R^*$
  - Min. average distortion taking losses into account:
  - $D_{ave} = (1\text{-}p)^2 D_0 + 2p(1\text{-}p) D_1 + p^2 \sigma^2$

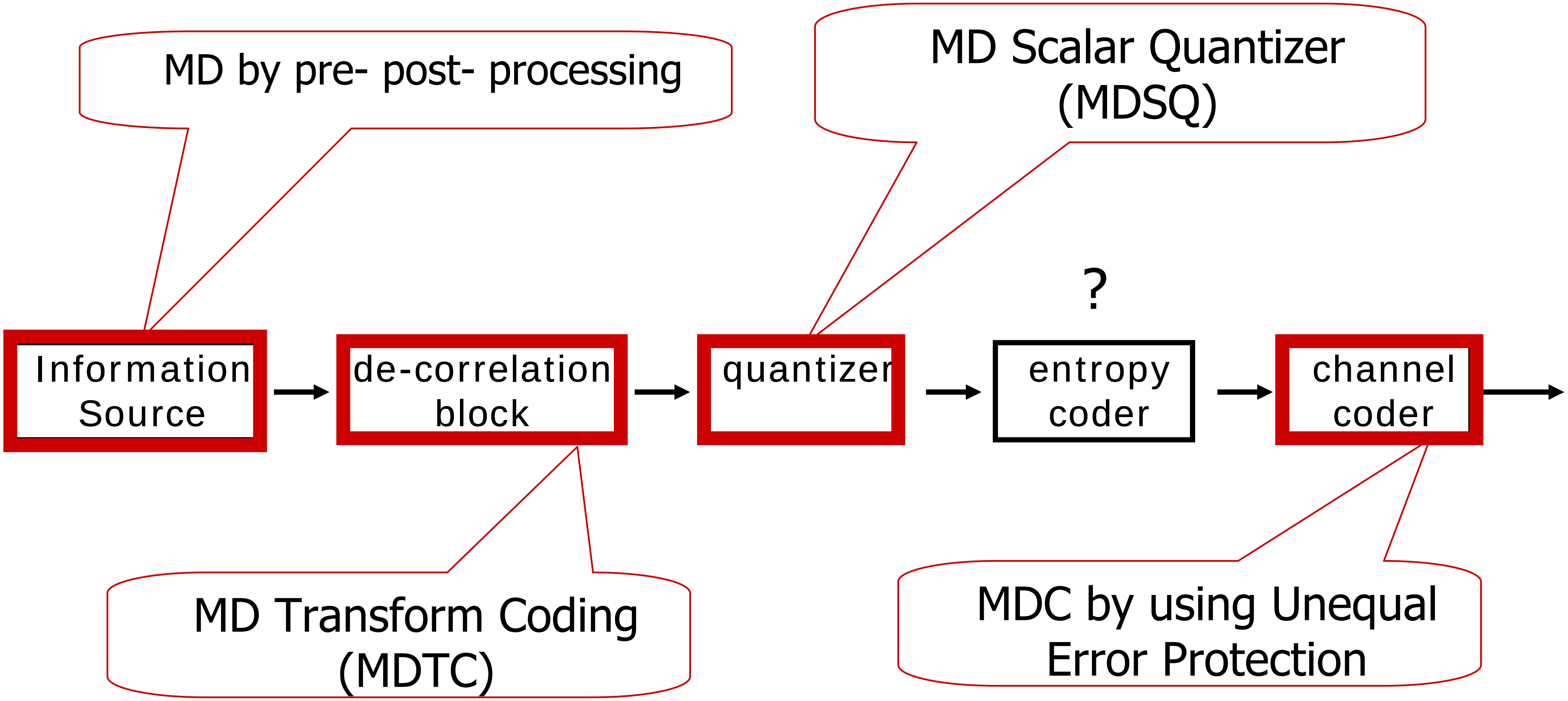
    subject to $R \leq R^*$

    p = probability of description loss
- $R = r + \rho$
  - r = basic rate necessary to achieve D_0 in single description coding (SDC)
- ρ = extra rate or redundancy

Please note that….

- The correlation between the descriptions is beneficial in the case of single description reception *(help estimating lost description)*
- When both description are received, the redundancy impairs the R-D performance of the system, as an extra rate must be accounted for in the overall system performance.
- **ρ is a critical parameter that should be carefully tuned**

Typical approaches to realize MDC



MDC with Unequal Error Protection

(How to re-use channel codes to create MDC)

The Mohr proposal (2000)

- Reference coder: SPIHT - still images, fully progressive. Also applicable to JPEG 2000
- The bit stream is divided into fragments
- The length (expressed in bytes) of each fragment is a result of the optimization procedure
- FEC is allocated to each fragment
- Fragment + FEC = stream
- Streams are accommodated on the rows of a matrix whose columns are called packets

The Mohr structure

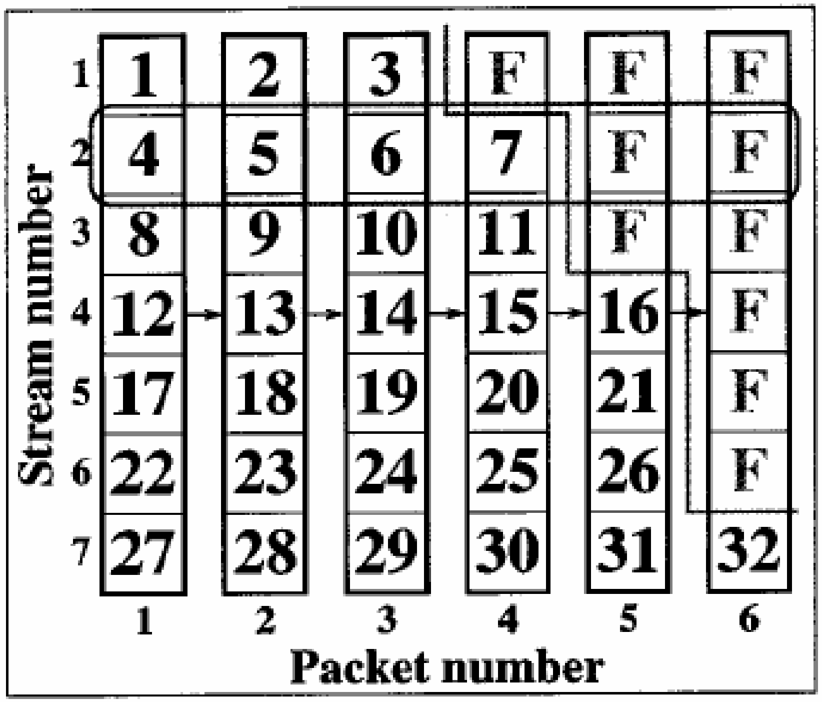


Fig. 2. Each of the rows is a stream and each of the columns is a packet. A stream contains 1 byte from each packet. The numbers 1–32 are data, and the symbol F is FEC.

L=7 streams

N=6 packets

Budget of 10 FEC bytes to be allocated

32 data bytes

The Mohr principle

- Packets are the basic transmission unit
- If a given number of packets gets lost, the i-th stream will be decoded, provided that the number of lost packets does not exceed the correction capability of the i-th allocated FEC
- Given that the data are progressive, and given an overall budget of FEC bytes to be allocated, ***it makes sense to allocate a decreasing number of FEC bytes to successive streams***

Reed Solomon codes

- Reed Solomon (RS) codes optimized for erasures (i.e. the position of errors is known)
- (N,K) systematic RS code, N = block length, K= # of information bytes (heading bytes)
- Property: such a code is able to cope with (max) N-K erasures
- K info. bytes can be recovered from any subset of (at least) K bytes in the N-length block

Example: 1 packet gets lost

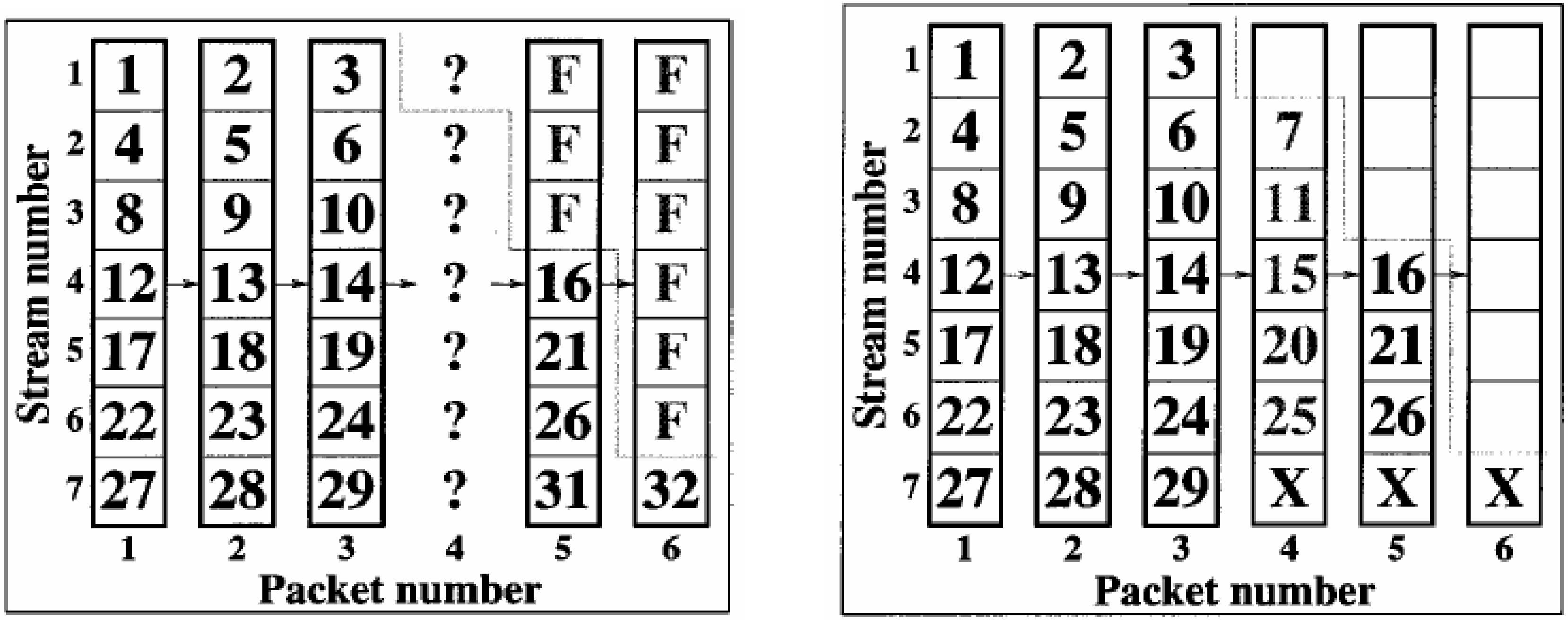


Fig. 3. Demonstration of how much data can be recovered when one of six packets is lost. Here, stream 1 is unaffected by the loss, streams 2–6 use FEC to recover from the loss, and in stream 7, only the bytes up to the lost packet are useful to the decoder.

Progressive quality: example 1

Image quality at 0.2 bits per pixel total rate for Unequal Loss Protection of Lenna over a channel that has an exponential loss profile with a mean of 20%.

- (a) 30% of packets lost.
- (b) 40% of packets lost.
- (c) 50% of packets lost.
- (d) 60% of packets lost.

Progressive quality: example 2

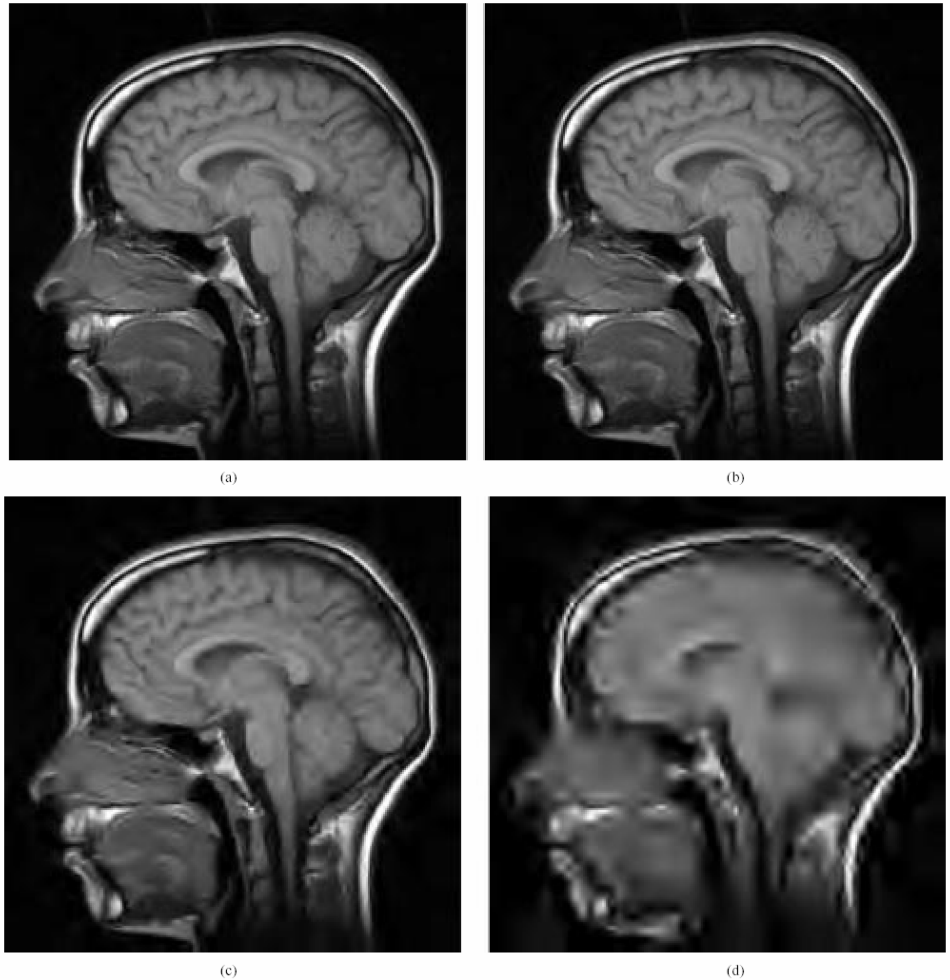


Image quality at 1.0 bit per pixel total rate for Unequal Loss Protection of a magnetic resonance image over a channel that has an exponential loss profile with a mean of 10%.

a) 10% of packets lost.

b) 20% of packets lost.

c) 30% of packets lost.

d) 40% of packets lost.

Problem statement

- Problem: how to divide data into fragments so as to maximize the average PSNR of the decoded image
- Average PSNR is a function of
  - Number of information data included into the stream (the more info, the fewer FEC)
  - Packet loss rate
- Many solutions proposed. Other metrics different from PSNR may be addressed as well

Bibliography

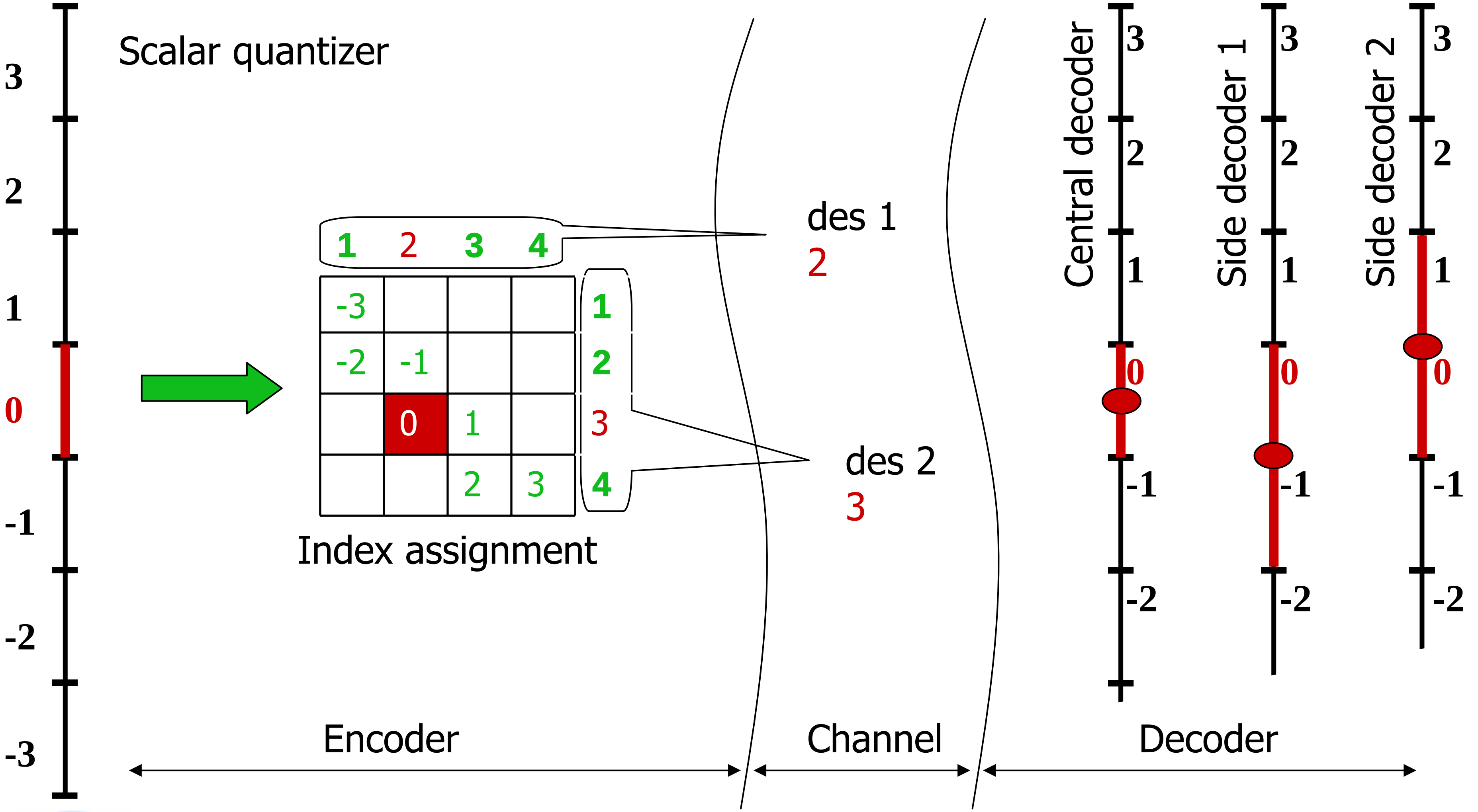
1. A. E. Mohr R. E. Ladner, E. A. Riskin, "Unequal Loss Protection: Graceful Degradation of Image Quality over Packet Erasure Channels Through Forward Error Correction," *IEEE Journal on Selected Areas in Comms*, vol. 18, no. 6, June 2000
2. A. E. Mohr R. E. Ladner, E. A. Riskin, "Approximately optimal assignment for unequal loss protection," *IEEE Intl. Conf. on Image processing* (ICIP), 2000
3. M. Grangetto, E. Magli, G. Olmo, "Ensuring Quality of Service for Image Transmission: Hybrid Loss Protection," *IEEE Trans. on image processing*, vol. 13, no. 6, June 2004

MDC with scalar quantizer

(How to make simple things tricky)

- **The idea:**
  - Quantize the coefficient with a scalar quantizer
  - Identify each quantization interval by two indexes
- **Side decoder:**
  - Received indexes may represent disjoint intervals → use the centroid of these intervals for reconstruction
- **Central decoder:**
  - Use both received indexes in order to identify the original quantization cell at the encoder

MDC with scalar quantizer



MDC with scalar quantizer

- Index assignment is complicated; several heuristics given by Vaishampayan
- N-dimensional extension is not trivial
- In the example, 4 bits are necessary to address 7 cells: redundancy depends on how much sparse the matrix is
- *Redundancy is not easy to tune*

1	2	3	4	
-3				1
-2	-1			2
	0	1		3
		2	3	4

Bibliography

1. V. A. Vaishampayan, “Design of multiple description scalar quantizer,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 821-834, May 1993
2. V. A. Vaishampayan, J. Domaszewicz, “Design of Entropy-Constrained Multiple-Description Scalar Quantizers,” *IEEE Trans. Inform. Theory*, vol. 40, No. 1, Jan. 1994
3. V. A. Vaishampayan, N. J. A. Sloane, S. D, Servetto, “Multiple-description vector quantization with lattice codebooks: design and analysis,” *IEEE Trans. Inform. Theory*, Volume 47, Issue 5, ,pp. 1718 – 1734, July 2001

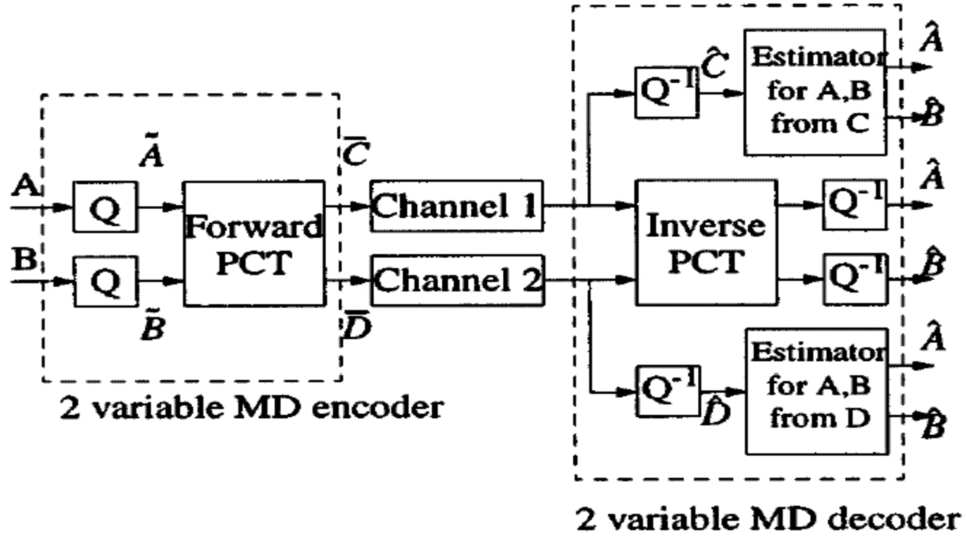
MDC with correlating transforms

(How to revolution the transform coding paradigm)

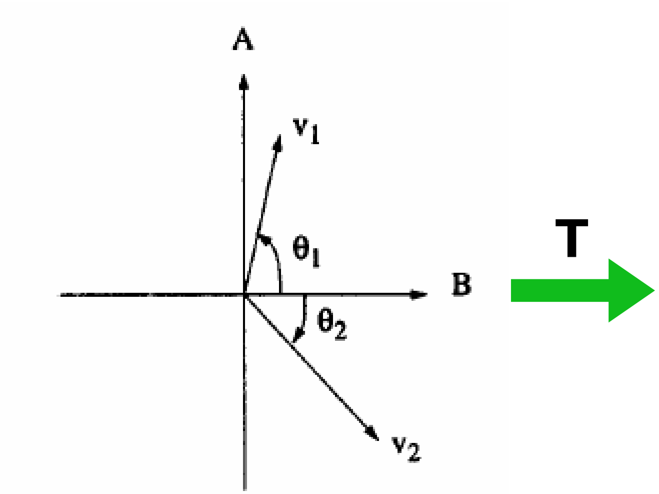
How does it work: encoder

- The uncorrelated input pair [A,B] is uniformly quantized
- The coefficients are subject to a 2X2 transform allowing reversible integer-to-integer mapping. The resulting variables are no more uncorrelated
- The transformed variables are independently entropy coded and sent over the two separate channels

- The **central decoder** simply applies the inverse transform to recover the data, which are then dequantized
- At the **side decoders**, the lost component is **estimated** from the received one. Then the inverse transform is applied to get the original output



- For a general transform T and [A, B] independent Gaussian variables:



$$\mathbf{T}^{-1} = \begin{bmatrix} r_1 \sin\theta_1 & r_2 \sin\theta_2 \\ r_1 \cos\theta_1 & r_2 \cos\theta_2 \end{bmatrix} = [\mathbf{v}_1 \quad \mathbf{v}_2]$$

$$\mathbf{T} = \begin{bmatrix} r_2 \cos\theta_2 & -r_2 \sin\theta_2 \\ -r_1 \cos\theta_1 & r_1 \sin\theta_1 \end{bmatrix}$$

To support lossless integer implementation of **T** the determinant should be 1→

$$r_1 r_2 = \frac{1}{\sin\Delta\theta}$$

Mathematical analysis

Given that [A,B] have variance σ_A^2 and σ_B^2 , the variance after the transform is:

$$\xrightarrow{\mathbf{T}}\quad \begin{aligned} \sigma_C^2 &= r_2^2(\cos^2\theta_2\sigma_A^2 + \sin^2\theta_2\sigma_B^2) \\ \sigma_D^2 &= r_1^2(\cos^2\theta_1\sigma_A^2 + \sin^2\theta_1\sigma_B^2) \end{aligned}$$

Taking into account that:

$$\mathbf{R}_{AB} = \begin{bmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{bmatrix} \qquad \mathbf{R}_{CD} = \mathbf{T}\mathbf{R}_{AB}\mathbf{T}^T$$

$$\underbrace{\mathbf{R}_{CD} = \begin{bmatrix} \sigma_C^2 & \sigma_C\sigma_D\cos\phi \\ \sigma_C\sigma_D\cos\phi & \sigma_D^2 \end{bmatrix}}$$

$$\sigma_C^2\sigma_D^2\sin^2\phi = \sigma_A^2\sigma_B^2$$

$\sigma_C\,\sigma_D \geq \sigma_A\,\sigma_B \rightarrow$ It is clear that the rate increases!

Mathematical analysis

Using optimum rate allocation → $R = \frac{1}{2}\log_2 \frac{\sigma_C \sigma_D}{D_0} + K, \quad R^* = \frac{1}{2}\log_2 \frac{\sigma_A \sigma_B}{D_0} + K$

Thus the redundancy per variable is

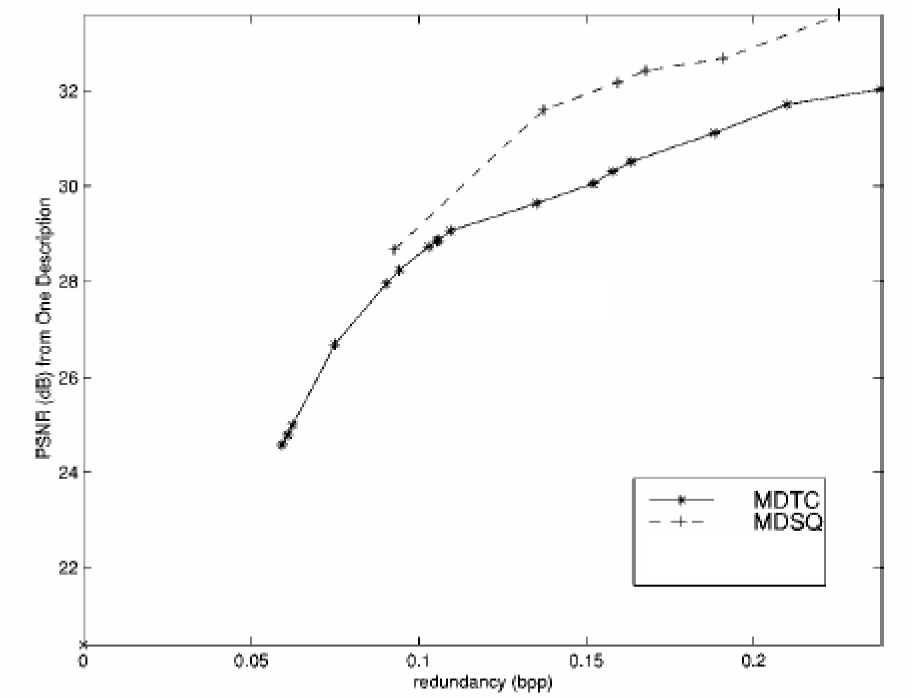
$$\boxed{\rho = R - R^* = \frac{1}{2}\log_2 \frac{\sigma_C \sigma_D}{\sigma_A \sigma_B} = -\frac{1}{2}\log_2 \sin\phi.}$$

- **Optimality conditions:**
  - $\theta_1 = \theta_2$
  - Given *N* coefficient the pairing strategy should pair the *i* th largest coefficient with the *(N-i)*th largest one → pairing problems.

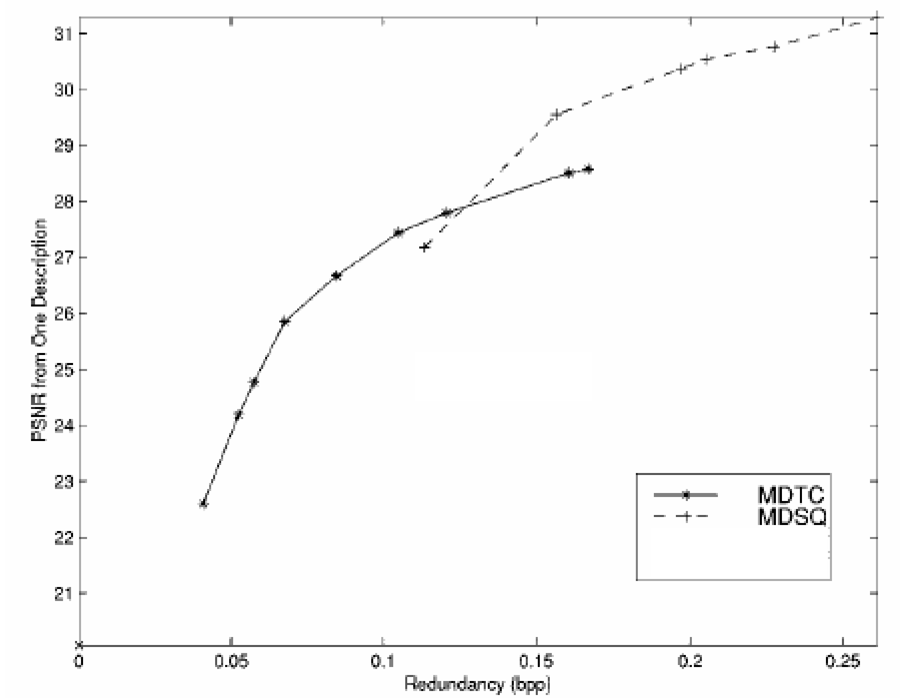
Application to image coding (2 desc.)

- 8X8 Block based DCT (JPEG-like) of the image is performed
- Uniform quantization is applied to DCT coefficients
- Coefficients are ordered so as to have decreasing variance.
- The first L coefficients having significant variance are paired such that i^{th} largest coefficient is paired with the $(L-i)^{th}$
- The remaining (N-L) coefficients are split between two descriptions (even/odd ones).
- JPEG like entropy coding is applied

Performance (image coding, 2 desc.)



RRD performance for image *Lena*. R=0.63 bpp, D_0 = 35.78 dB

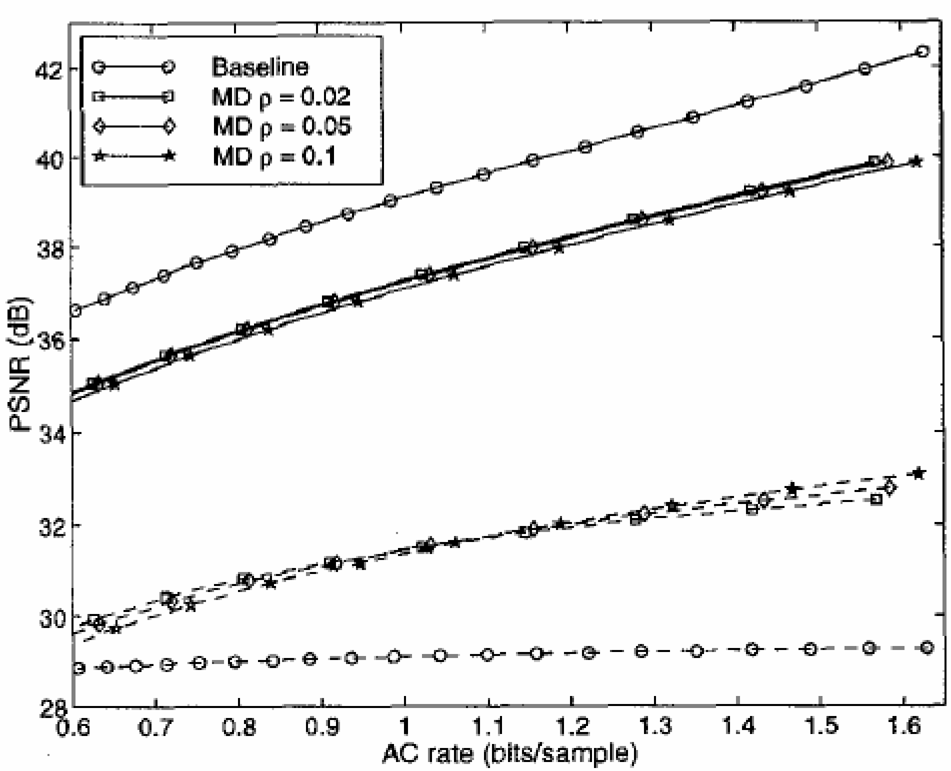


RRD performance for image *Horse*. R=0.70bpp, D_0 = 33.51 dB

Application to image coding (4 desc.)

- 8×8 block DCT
- Uniform quantization of DCT coefficients
- Vectors of length 4 created by taking pixels separated in frequency and space. The spatial separation is maximized → uncorrelated coefficients.
- Correlating transform applied to these vectors (2X2 transform cascaded)
- Entropy coding

Performance (image coding, 4 desc.)



PSNR for correlation transform applied on image *Lenna*. Either all or 3 out of 4 descriptions received.

Bibliography

1. Y. Wang, M. T. Orchard, V. Vaishampayan, A. R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Trans. on Image Processing*, vol. 10, no. 3, pp. 351-366, March 2001
2. V. K. Goyal, J. Kovacevic, R. Arean, M. Vetterli, "Multiple description transform coding of images," *ICIP 98*