# VIDEOTELEPHONY AND VIDEOCONFERENCE OVER ISDN

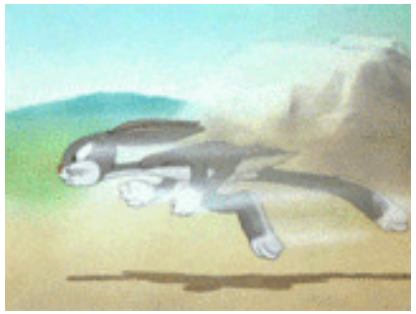
*Fernando Pereira*

*Instituto Superior Técnico*


# Digital Video

# Video versus Images

- **Still Image Services – No strong temporal requirements; no real-time notion.**
- **Video Services (moving images) – There is a need to strictly follow critical delay requirements to provide a good illusion of motion; essential to provide real-time performance.**



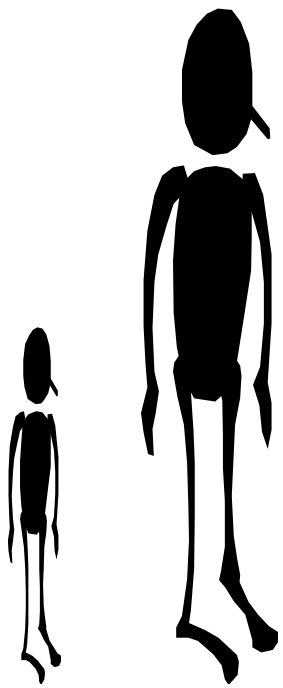
**For each image and video service, it is possible to associate a quality target (quality of service); the first impact of this target is the selection of the right spatial and temporal resolutions to use.**

# Why Does Video Information Have to be Compressed ?

**A video sequence is created and consumed as a set of images, happening at a certain temporal rate (F), each of them with a spatial resolution of M×N luminance and chrominance samples and a certain number of bits per sample (L)**

**This means the total number of (PCM) bits**

**- and thus the required bandwidth and memory – necessary to digitally represent a video sequence is HUGE !!!**



# Videotelephony: Just an Example

- **Resolution: 10 images/s with 288×360 luminance samples and 144 × 188 samples (4:2:0) for each chrominance, with 8 bit/sample**

  $[(360 \times 288) + 2 \times (180 \times 144)] \times 8 \times 10 = 12.44$ Mbit/s

- **Reasonable bitrate: e.g. 64 kbit/s for an ISDN B channel**

**=> Compression Factor: 12.44 Mbit/s/64 kbit/s ≈ 194**

**The usage or not of compression/source coding implies the possibility or not to deploy services and, thus, the existence or not of certain industries, e.g. DVD.**


# Digital Video: Why is it So Difficult ?

| Service | Luminance Spatial Resolution | Chrominance Spatial Resolution | Temporal resolution | Aspect ratio | PCM Bitrate |
|---|---|---|---|---|---|
| HDTV | 1152×1920 | 576×960 | 50 frames/s | 16/9 | 1.3 Gbit/s |
| SDTV | 576×720 | 576×360 | 25 frames/s | 4/3 | 166 Mbit/s |
| Video CD | 288×360 | 144×180 | 25 frames/s | 4/3 | 31 Mbit/s |
| Videotelephony | 288×360 | 144×180 | 10 frames/s | 4/3 | 12.4 Mbit/s |
| Mobile videotelephony | 144×180 | 72×90 | 5 frames/s | 4/3 | 1.6 Mbit/s |


# Video Coding/Compression: a Definition

**Efficient representation (this means with a smaller than the PCM number of bits) of a periodic sequence of (correlated) images, satisfying the relevant requirements, e.g. minimum acceptable quality, error robustness, random access.**

**And the service requirements change with the services/applications and the corresponding funcionalities ...**

# How Big Has to be the Compression 'Hammer' ?

| Service | Luminance Spatial Resolution | Chrominance Spatial Resolution | Temporal resolution | Aspect ratio | PCM Bitrate | Compressed Bitrate | Compression Factor |
|---|---|---|---|---|---|---|---|
| HDTV | 1152×1920 | 576×960 | 50 frames/s | 16/9 | 1.3 Gbit/s | 10 Mbit/s | 130 |
| SDTV | 576×720 | 576×360 | 25 frames/s | 4/3 | 166 Mbit/s | 2 Mbit/s | 83 |
| Video CD | 288×360 | 144×180 | 25 frames/s | 4/3 | 31 Mbit/s | 500 kbit/s | 62 |
| Videotelephony | 288×360 | 144×180 | 10 frames/s | 4/3 | 12.4 Mbit/s | 64 kbit/s | 194 |
| Mobile videotelephony | 144×180 | 72×90 | 5 frames/s | 4/3 | 1.6 Mbit/s | 20 kbit/s | 80 |

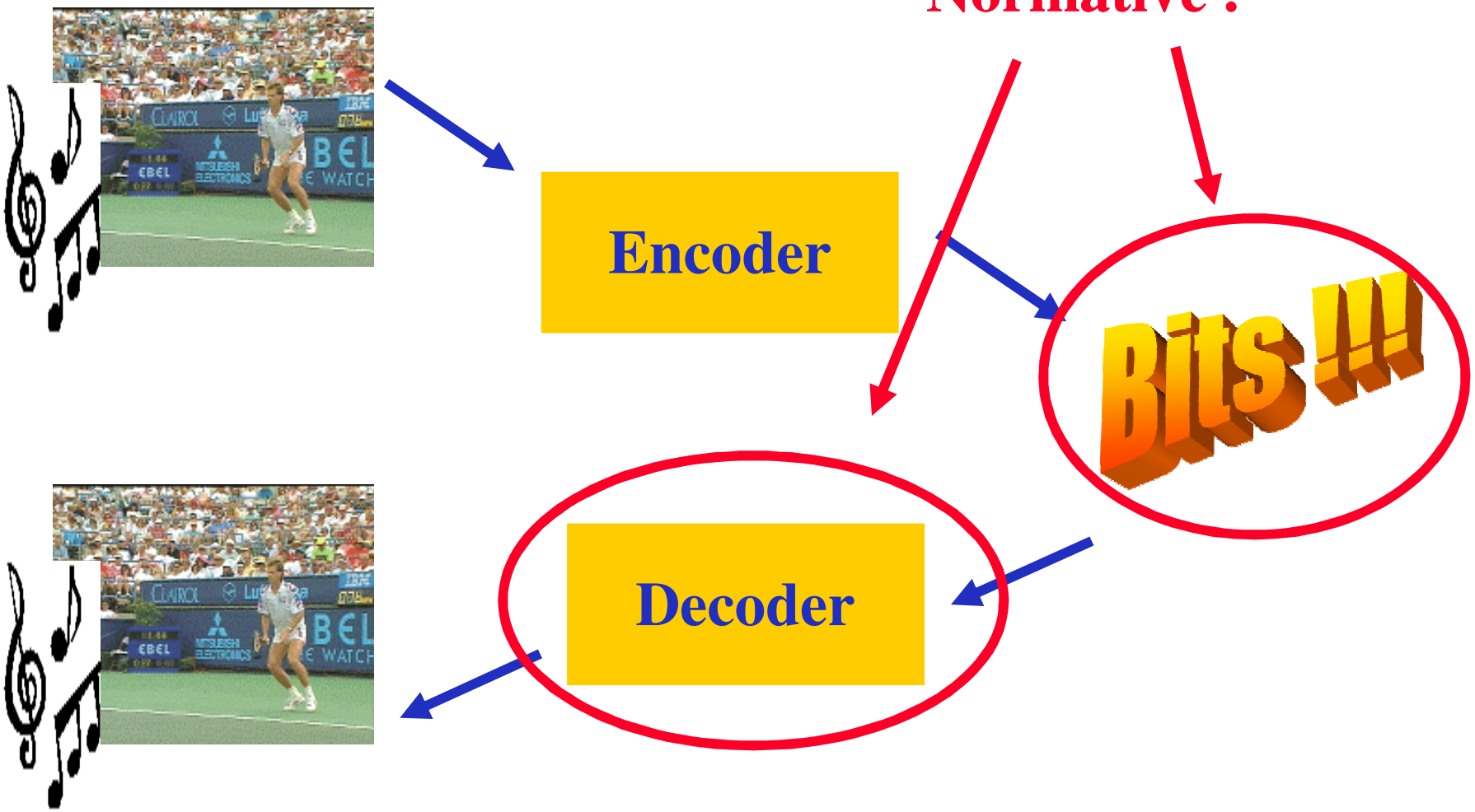
# Interoperability as a Major Requirement: Standards to Assure that More is not Less ...

- **Compression is essential for digital audiovisual services where interoperability is a major requirement.**

- **Interoperability requires the specification and adoption of standards, notably audiovisual coding standards.**

- **To allow some evolution of the standards and some competition in the market between compatible products from different companies, standards must specify the minimum set of technology possible, typically the bitstream syntax and the decoding process (not the encoding process).**

# Standards: a Trade-off between Fixing and Inovating

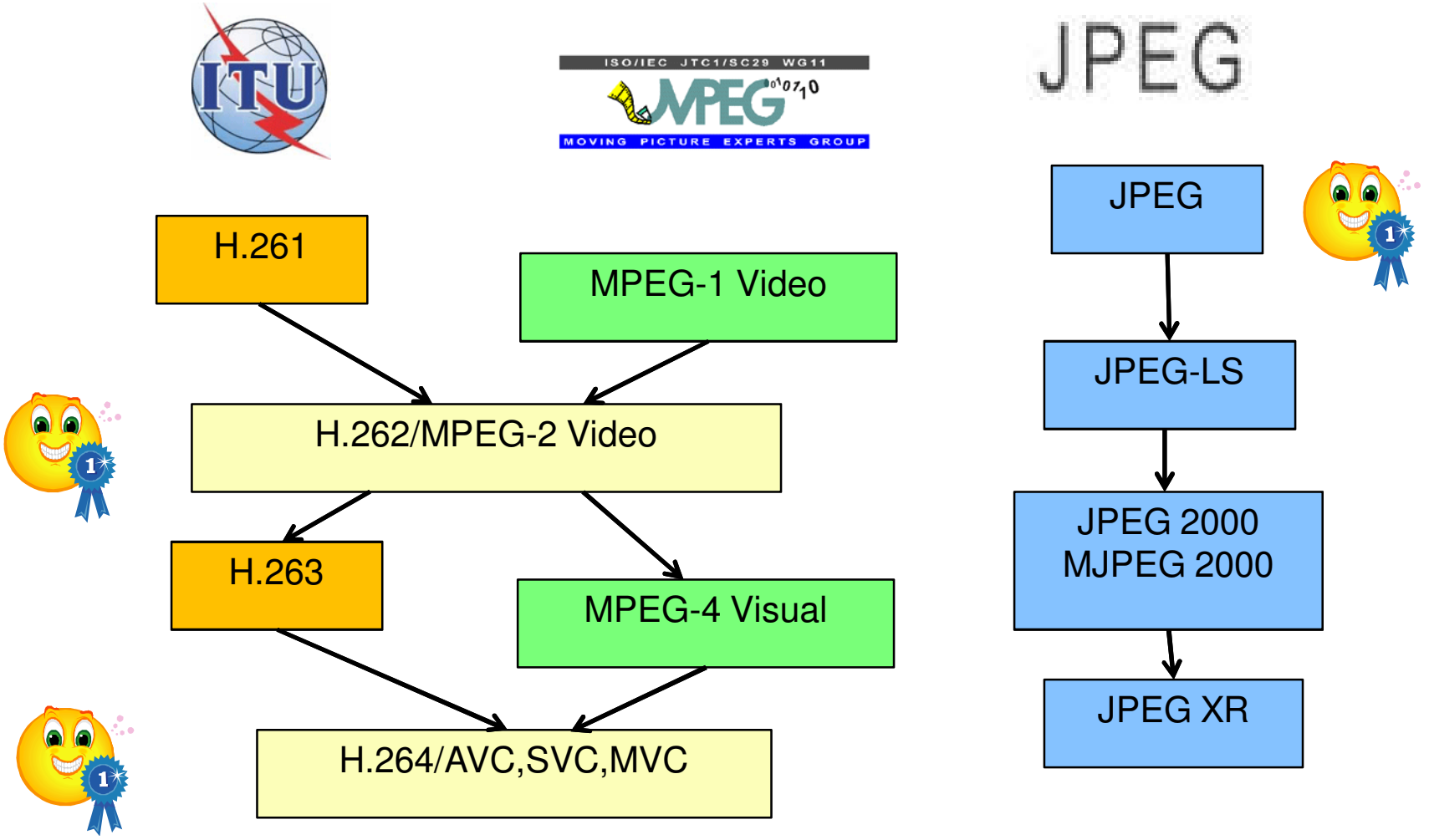



# Video Coding Standards …



- **ITU-T H.120 (1984) - Videoconference (1.5 - 2 Mbit/s)**
- **ITU-T H.261 (1988) – Audiovisual services (videotelephony and videoconference) at p×64kbit/s, p=1,…,30**
- **ISO/IEC MPEG-1 (1990)- CD-ROM Video**
- **ISO/IEC MPEG-2 also ITU-T H.262 (1993) – Digital TV**
- **ITU-T H.263 (1996) – PSTN and mobile video**
- **ISO/IEC MPEG-4 (1998) – Audiovisual objects, improved efficiency**
- **ISO/IEC MPEG-4 AVC also ITU-T H.264 (2003) – Improved efficiency**

# The Video Coding Standardization Path …




# ITU-T H.320 Terminals

# Videotelephony and Videoconference

# Videotelephony and Videoconference

## Personal (bidirectional) communications in real-time !


# ITU-T H.320 Recommendation: Motivation

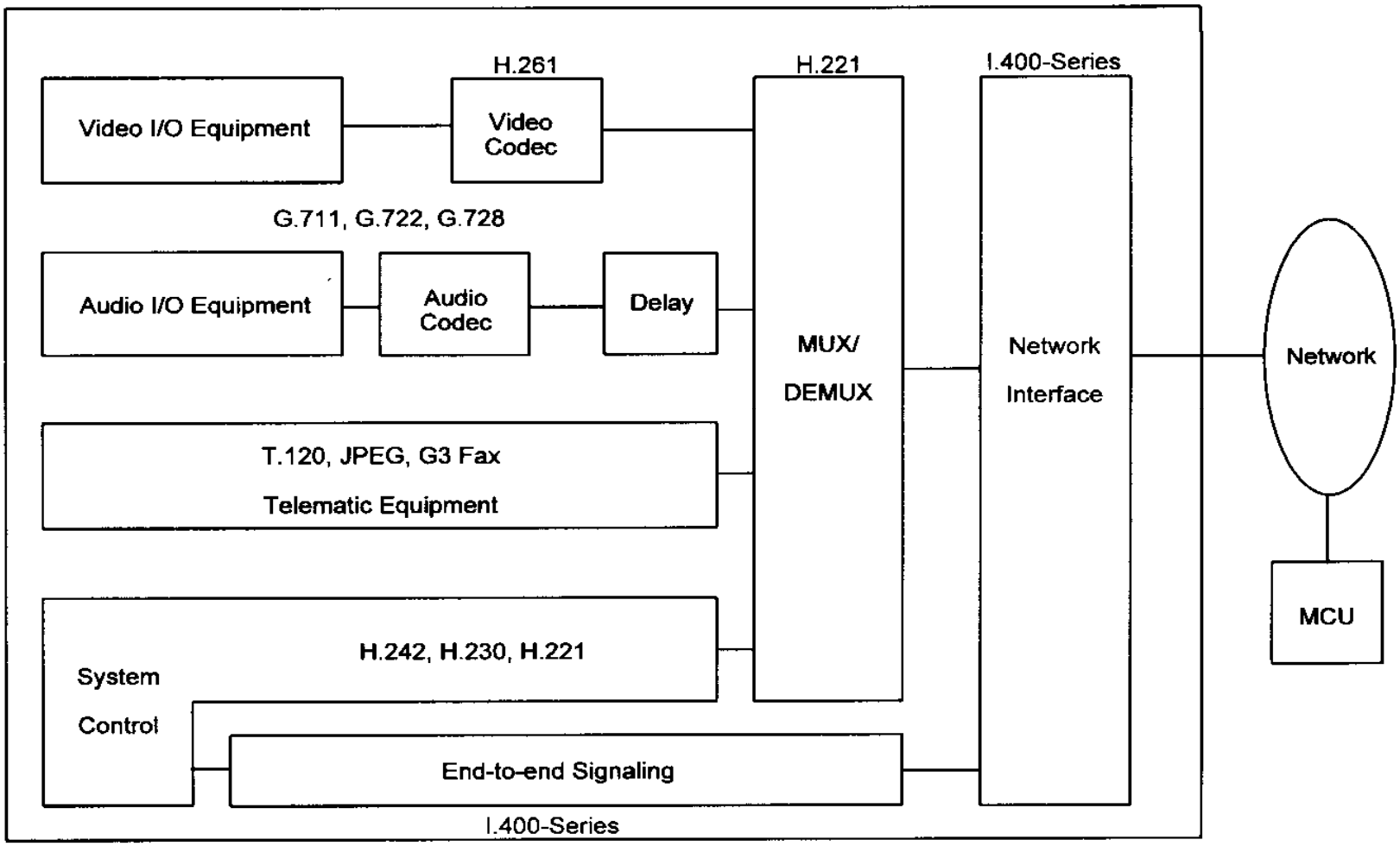
**The starting of the work towards Rec. H.320 and H.261 goes back to 1984 when it was acknowledged that:**

- **There was an increase in the demand for image-based services, notably videotelephony and videoconference.**
- **There was a growing availability of 64, 384 e 1536/1920 kbit/s digital lines as well as ISDN lines.**
- **There was a need to make available image-based services and terminals for the digital lines mentioned above.**
- **The acknowledgement that Rec. H.120, just issued at that time, for videoconference services, was already obsolete in terms of compression efficiency due to the fast development in the area of video compression.**

# Videotelephony and Videoconference: Main Features

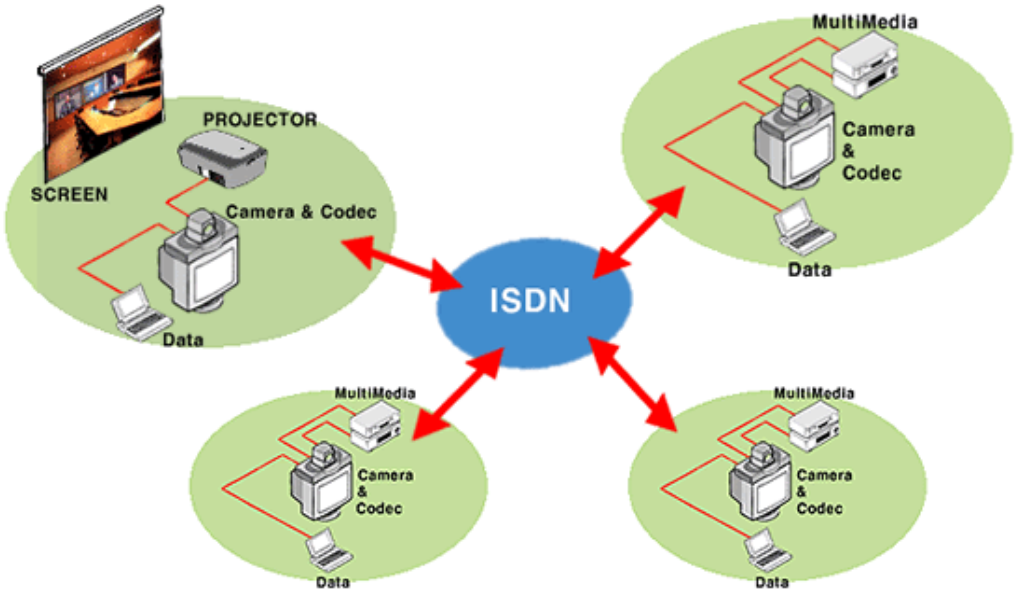
- Personal communications (point to point or multipoint to multipoint)
- Symmetric bidirectional communications (all nodes involved have the same similar features)
- Critical delay requirements
- Low or intermediate quality requirements
- Strong psychological and sociological impacts

# Rec. H.320 Terminal




# Video Coding: Rec. ITU-T H.261

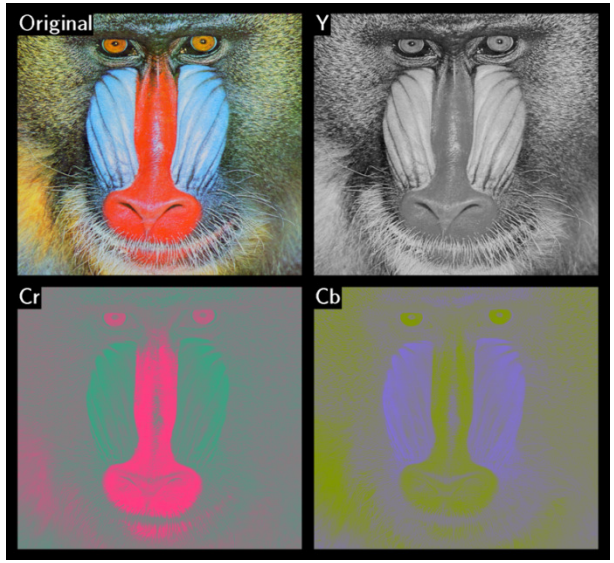
# Recommendation H.261: Objectives




**Efficient coding of videotelephony and videoconference sequences with a minimum acceptable quality using a bitrate from 40 kbit/s to 2 Mbit/s, targeting synchronous channels (ISDN) at p×64 kbit/s, with p=1,...,30.**

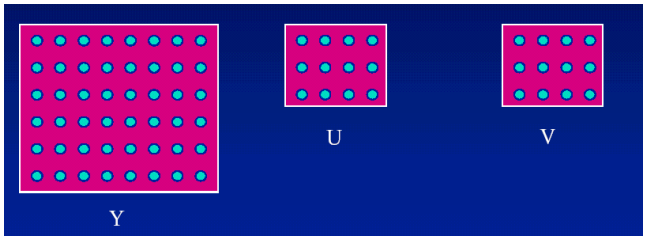
**This is the first international video coding standard with relevant adoption, thus introducing the notion of backward compatibility in video coding standards.**

# H.261: Signals to Code




- The signals to code for each image are the luminance (Y) and 2 chrominances, named $C_B$ and $C_R$ or U and V.
- The samples are quantized with 8 bits/sample according to Rec. ITU-R BT-601:
  - Black = 16; White = 235; Null colour difference = 128
  - Peak colour difference (U,V) = 16 and 240
- The coding algorithm operates over progressive (non-interlaced) content at 29.97 image/s.
- The frame rate (temporal resolution) may be decreased by skipping 1, 2 or 3 images between each transmitted image.

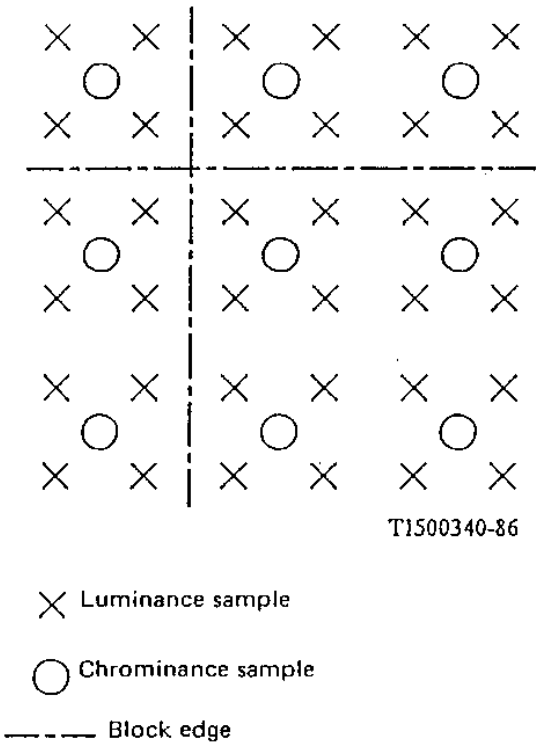
# H.261: Image Format




**Two spatial resolutions are possible:**

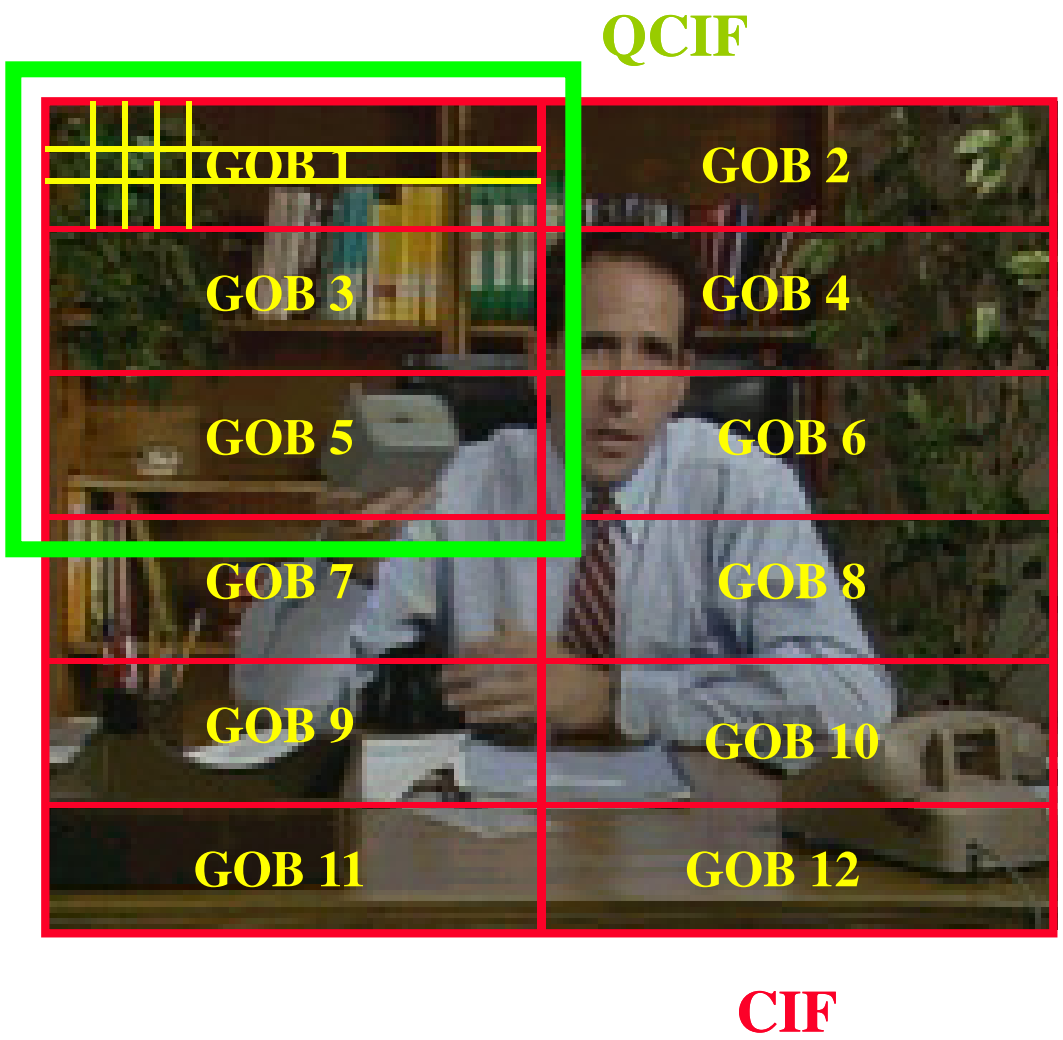
- **CIF (*Common Intermediate Format*) - 288 × 352 samples for luminance (Y) and 144 × 176 samples for each chrominance (U,V) this means a 4:2:0 subsampling format, with 'quincux' positioning, progressive, 30 frame/s with a 4/3 aspect ratio.**

- **QCIF (*Quarter CIF*) – Similar to CIF with half spatial resolution in both directions this means 144 × 176 samples for luminance and 72 × 88 samples for each chrominance.**



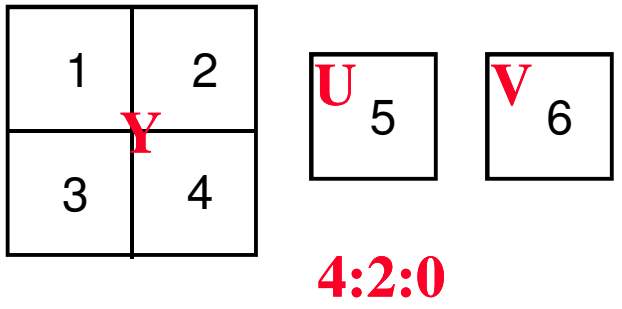

**All H.261 codecs must work with QCIF and some may be able to work also with CIF (resolution is determined after negotiation).**

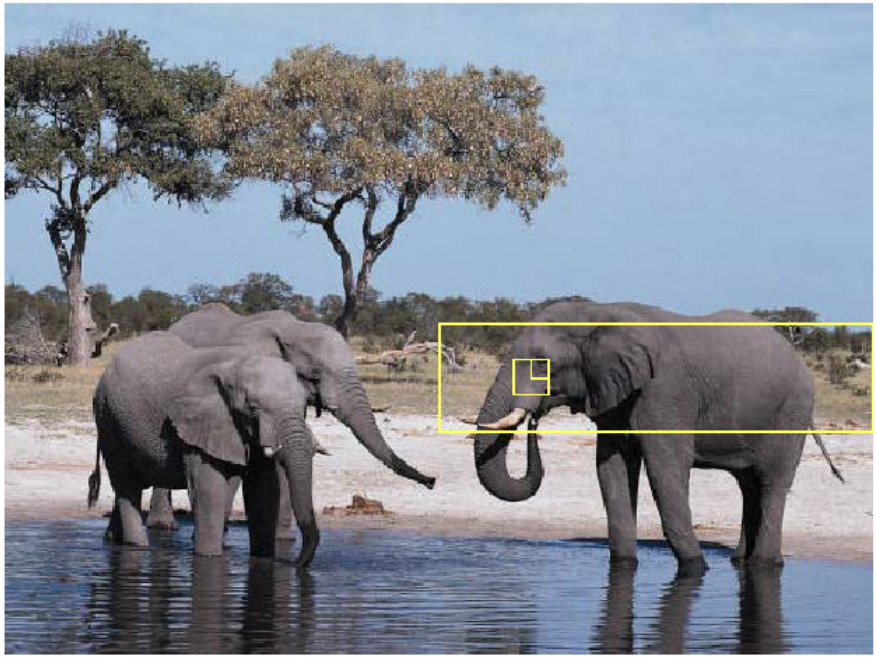
# Images, Groups Of Blocks (GOBs), Macroblocks and Blocks




The video sequence is spatially organized according to a hierarchical structure with 4 levels:

- Images
- Group of Blocks (GOB)
- Macroblocks (MB)
- Blocks



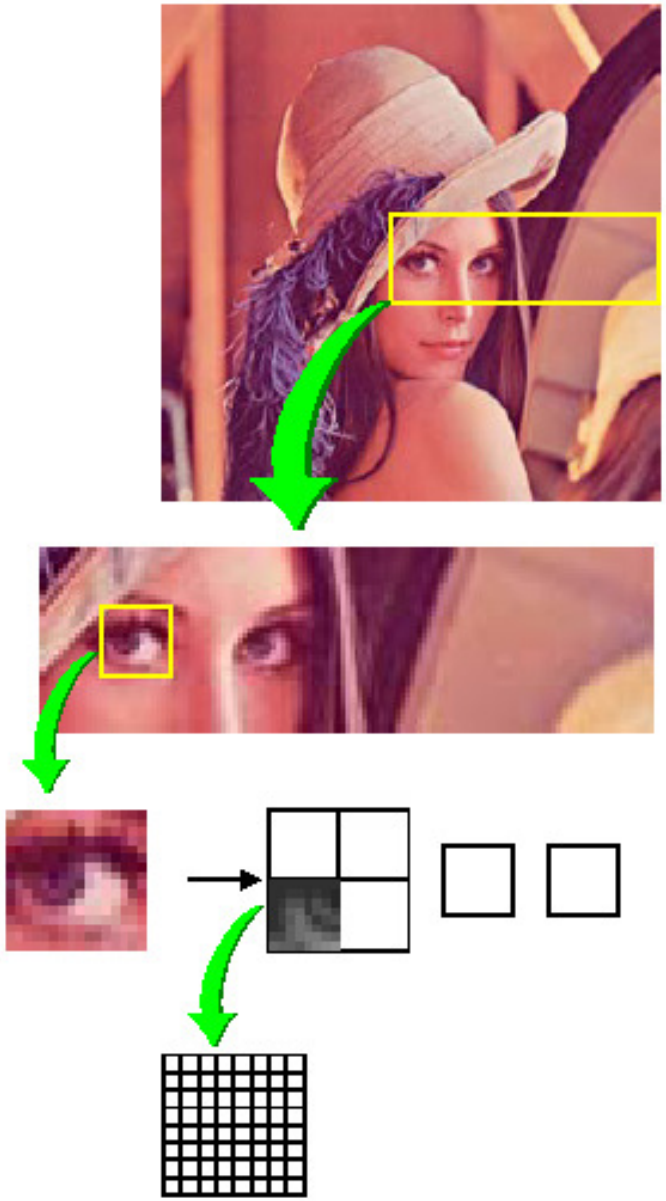


Picture

GOB

Macroblock

Block



# H.261: Coding Tools


- **Temporal Redundancy**

  Predictive coding: sending differences

  and motion compensation

- **Spatial Redundancy**

  Transform coding (Discrete Cosine Transform, DCT)

- **Statistical Redundancy**

  Huffman entropy coding


- **Irrelevancy**

  Quantization of DCT coefficients


# Exploiting

# Temporal Redundancy


# Temporal Prediction and Prediction Error

- **Temporal prediction is based on the principle that, locally, each image may be represented using as reference a part of some preceding image, typically the previous one.**
- **The prediction quality strongly determines the compression performance since it defines the amount of information to code and transmit, this means the energy of the error/difference signal called *prediction error*.**
- **The lower is the prediction error, the lower is the information/energy to transmit and thus**
  - Better quality may be achieved for a certain available bitrate
  - Lower bitrate is needed to achieve a certain video quality

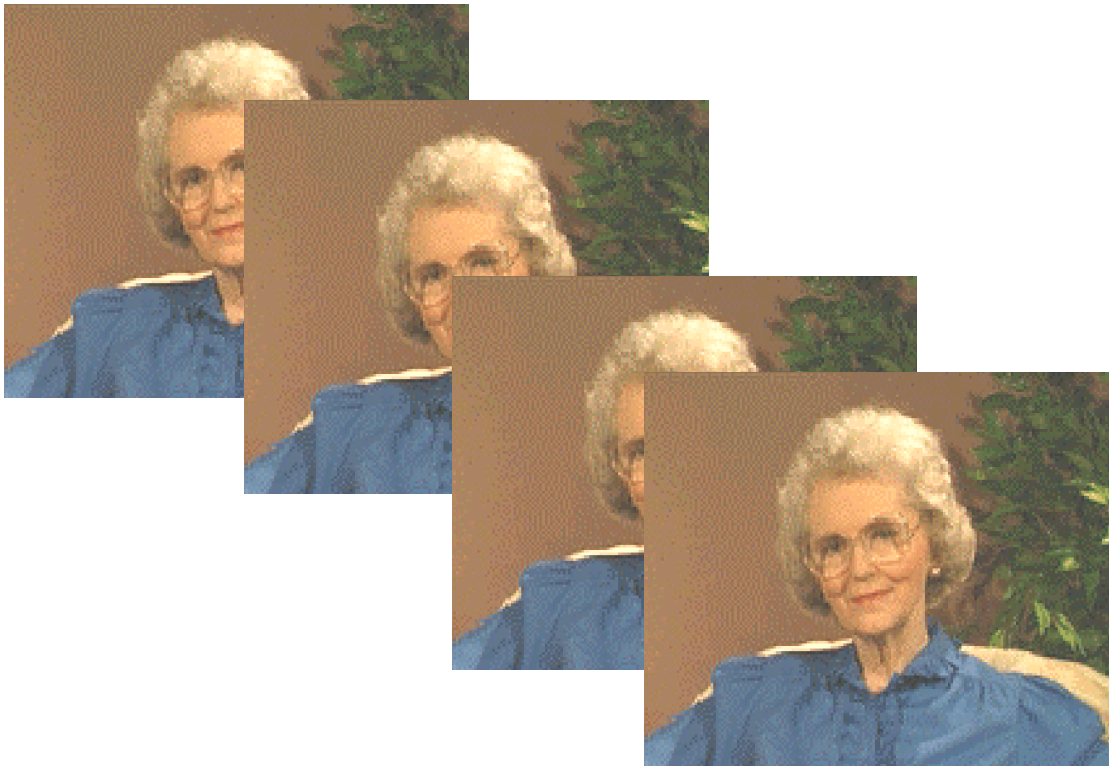
# H.261 Temporal Prediction

**Rec. H.261 includes 2 temporal prediction tools which have both the target to eliminate/reduce the temporal redundancy in the PCM video signal:**

## Sending the Differences

## Motion Compensation

# Temporal Redundancy: Sending the Differences



**The idea is that only the new information in the new image (this means what changes from the previous image) is sent; the previous image works as a simple prediction of the current image.**

●●●

**There are no losses !**

# Computing the Differences: an Example

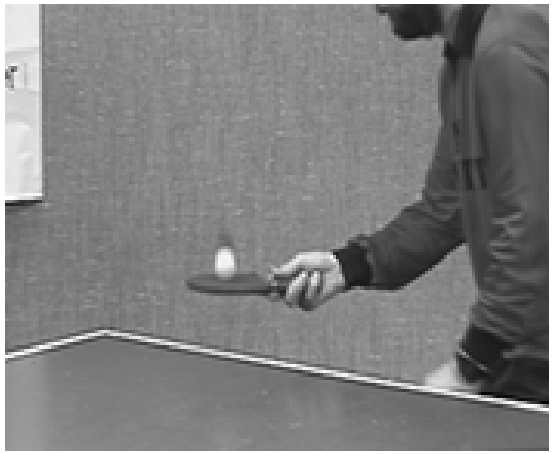


Image t

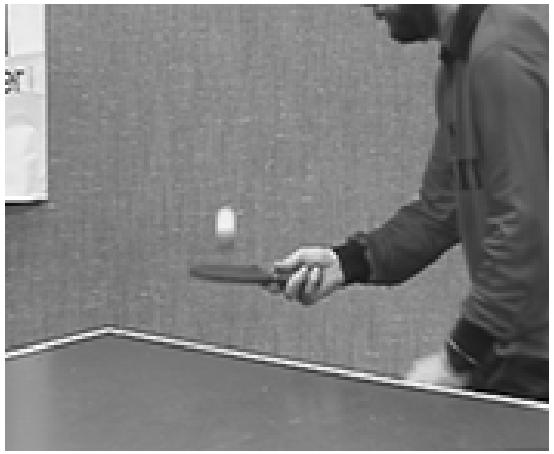
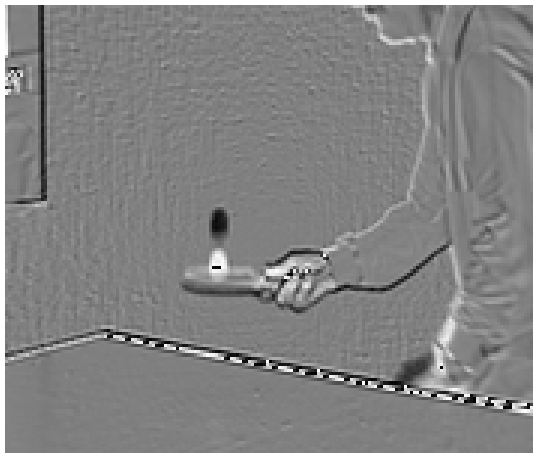
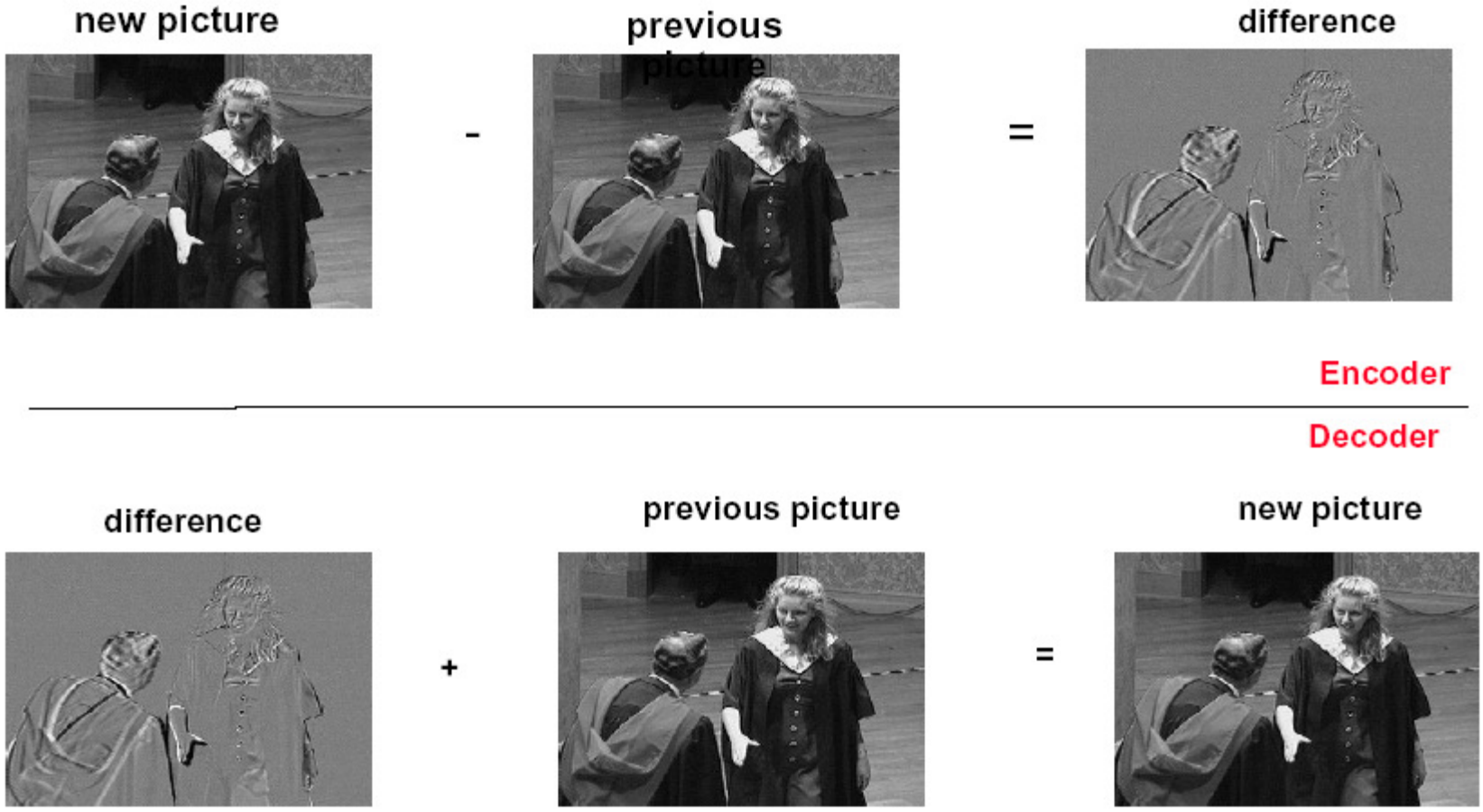


Image t-1

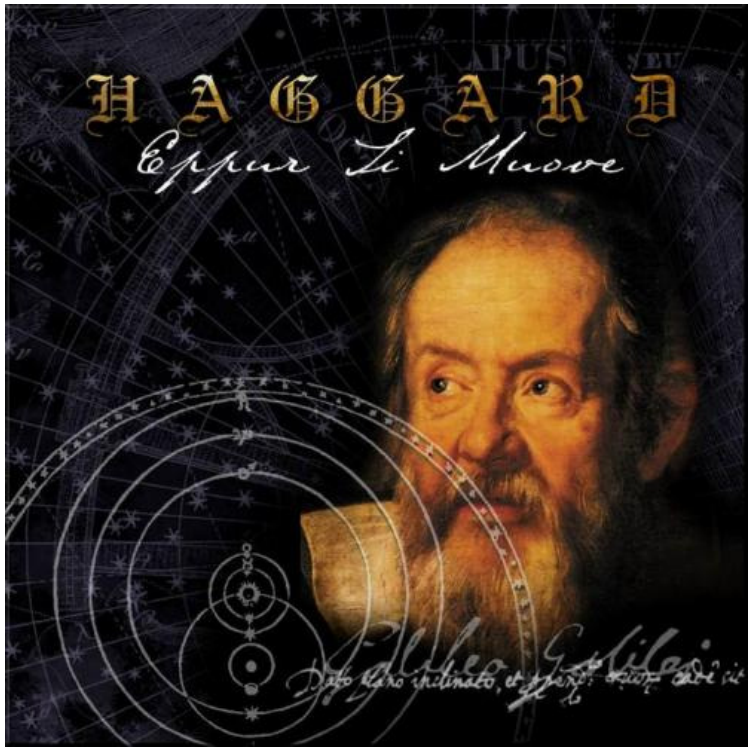


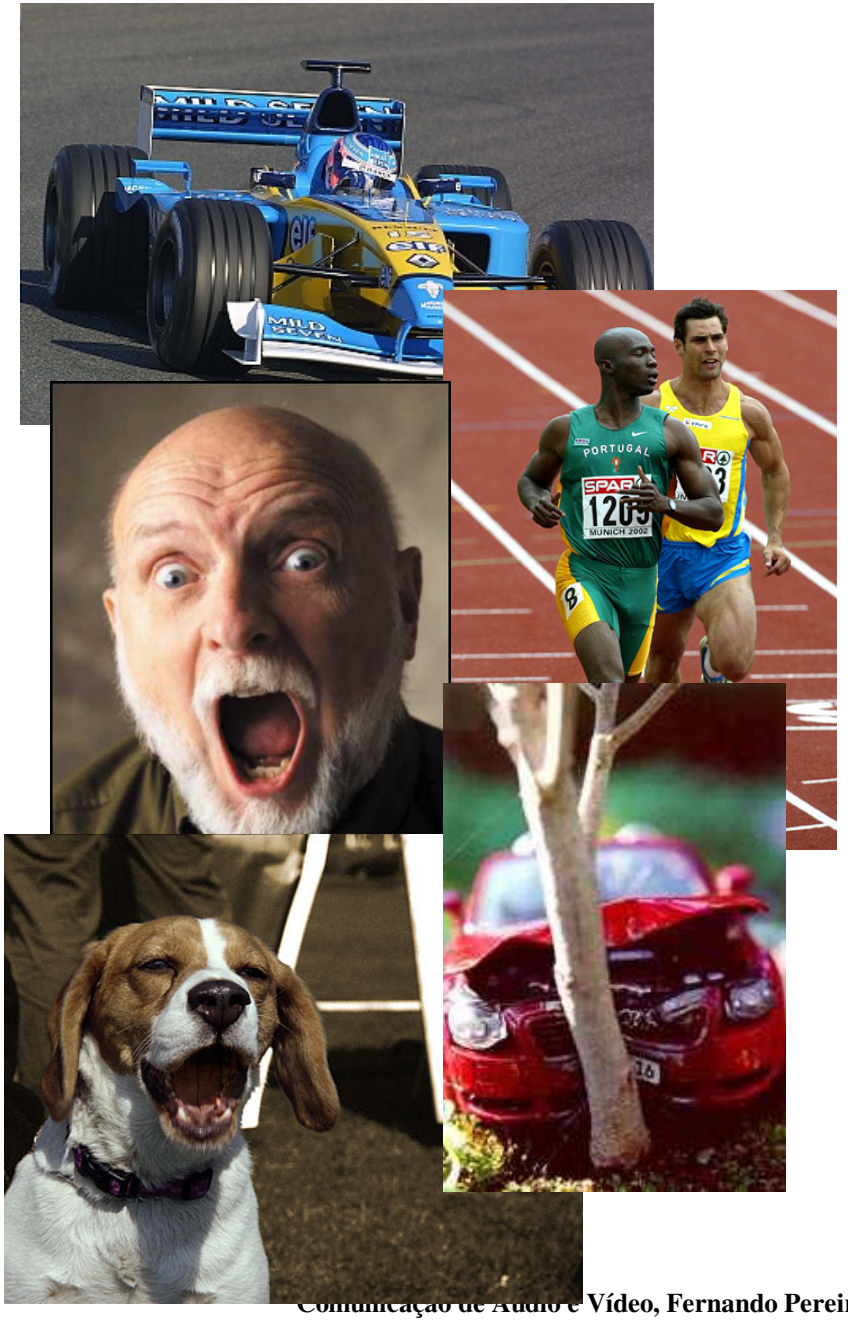
Differences

# Coding and Decoding ...



# *Eppur Si Muove …*



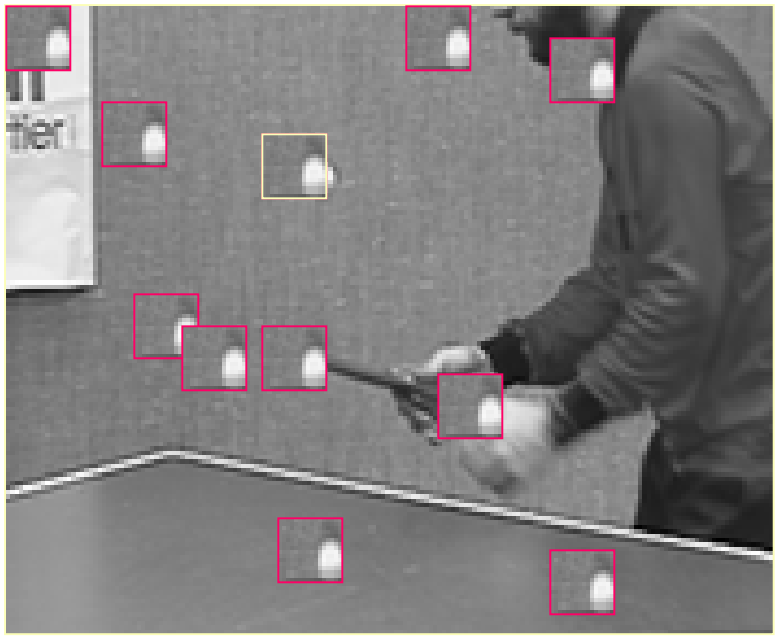

# Motion Estimation and Compensation

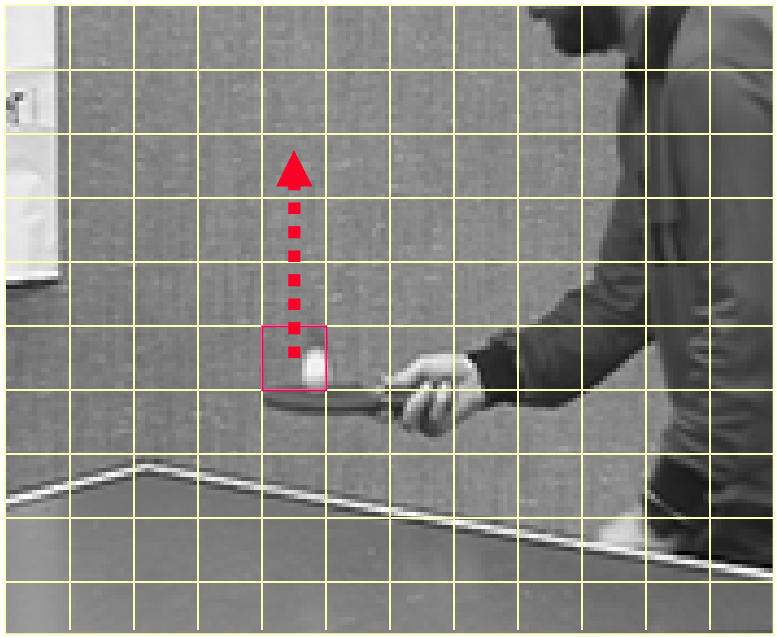
**Motion estimation and compensation have the target to improve the temporal predictions for each image zone by detecting, estimating and compensating the motion in the image.**

- **Motion estimation is not normative (it is part of the encoder) but the so-called *block matching* is the most used technique.**
- **In H.261, motion compensation is made at macroblock level. The usage of motion compensation for each MB is optional and decided by the encoder.**

**Motion estimation implies a very high computational effort. This justifies the usage of fast motion estimation methods which try to decrease the complexity compared to full search without much quality losses.**

# Temporal Redundancy: Motion Estimation

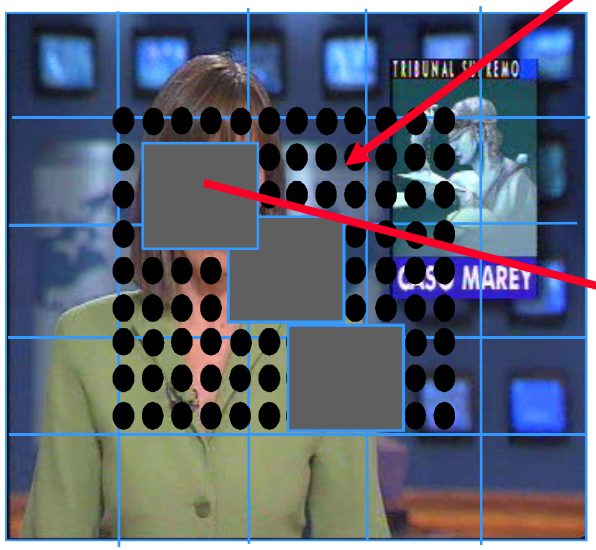




t

# Search, Where ?

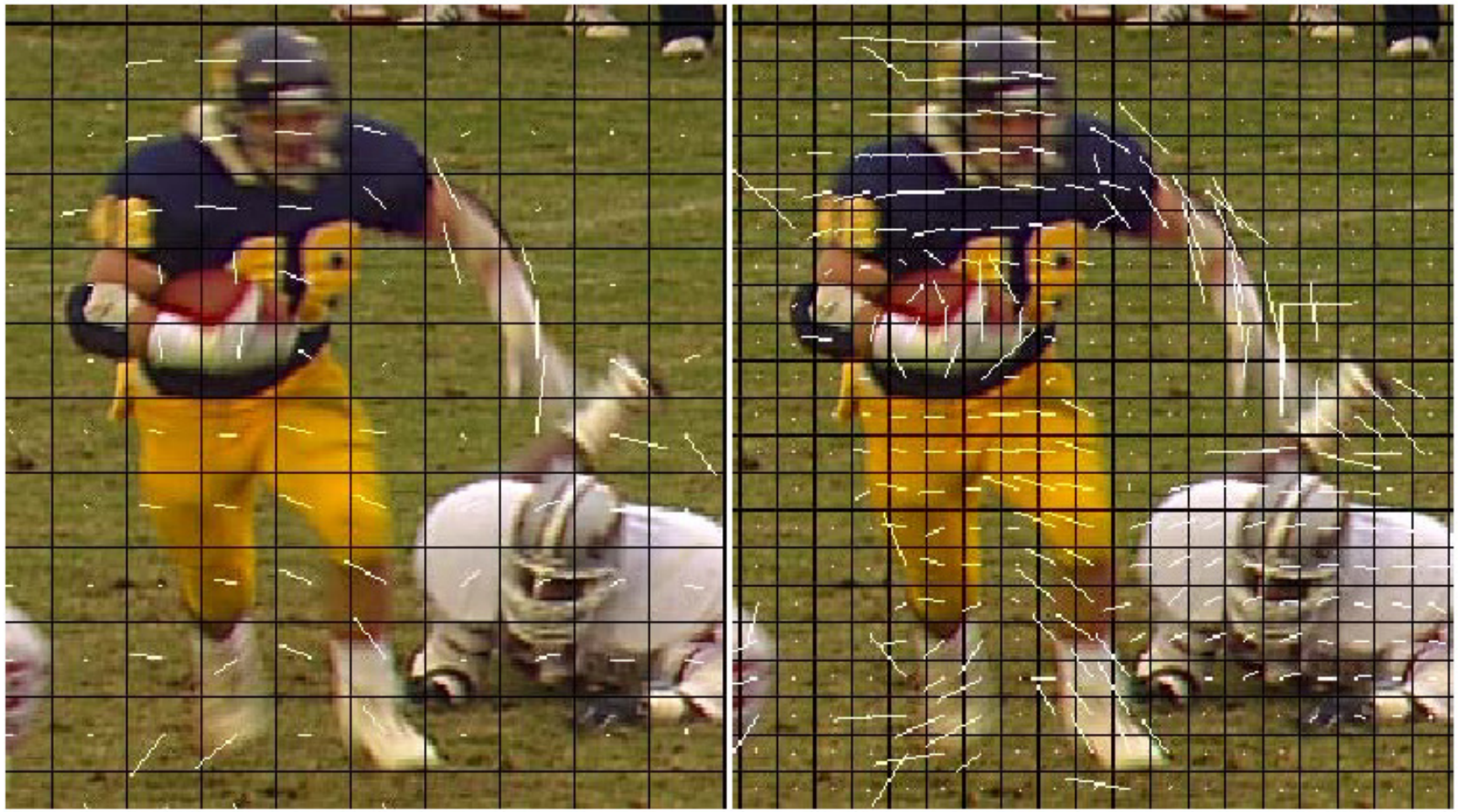
Searching area



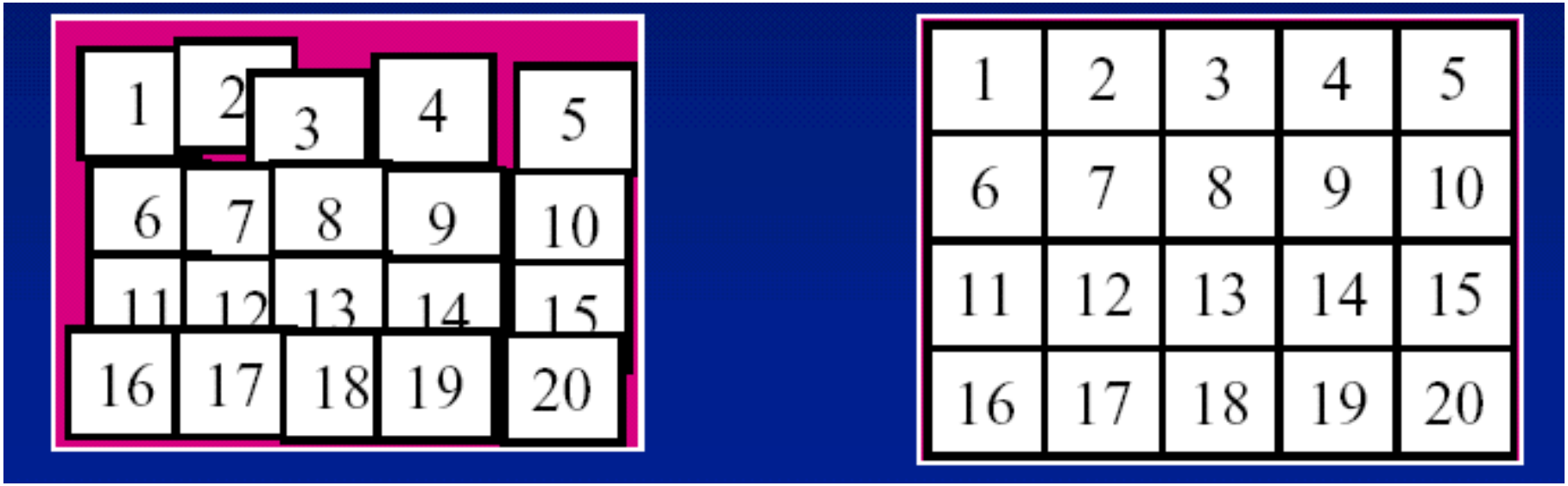
Reference image


Image to code

# Motion Vectors at Different Spatial Resolutions



# MBs to Code and Prediction MBs




**Reference content (coded macroblocks)**

**Current image under coding**

# Motion Compensation: an Example

Image t

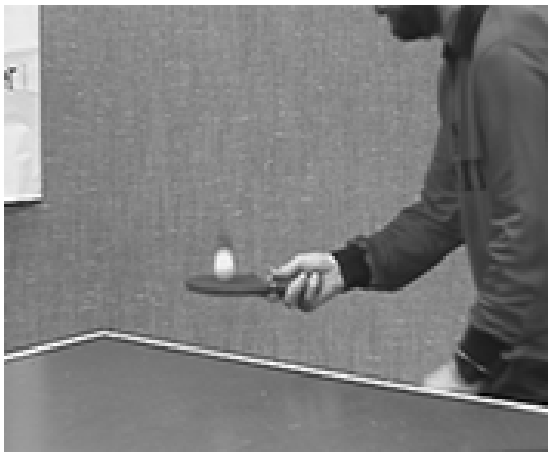
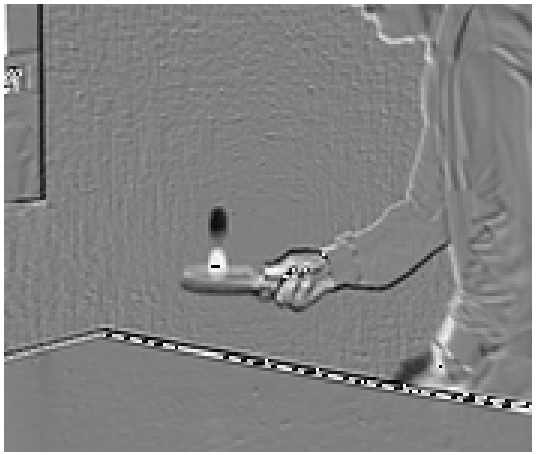
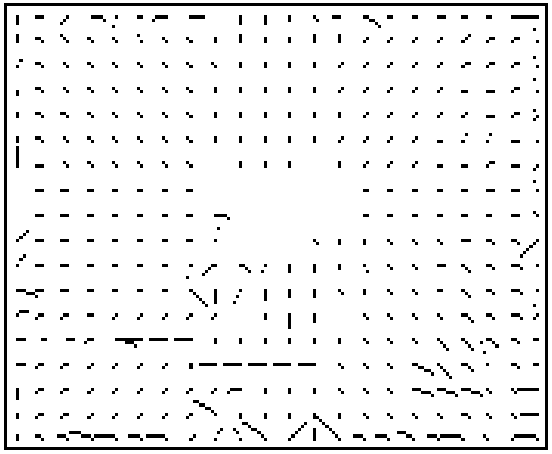


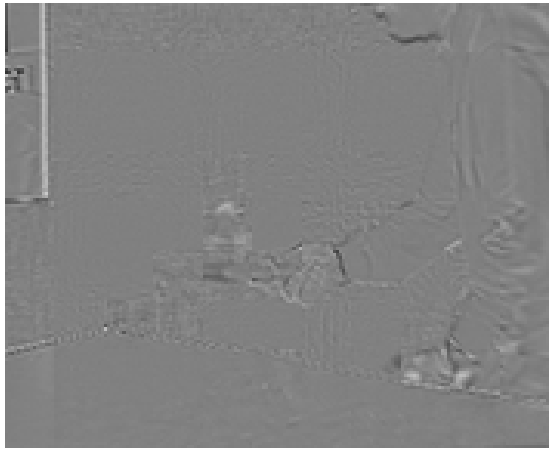
Image t-1



Diff. WITHOUT motion comp.



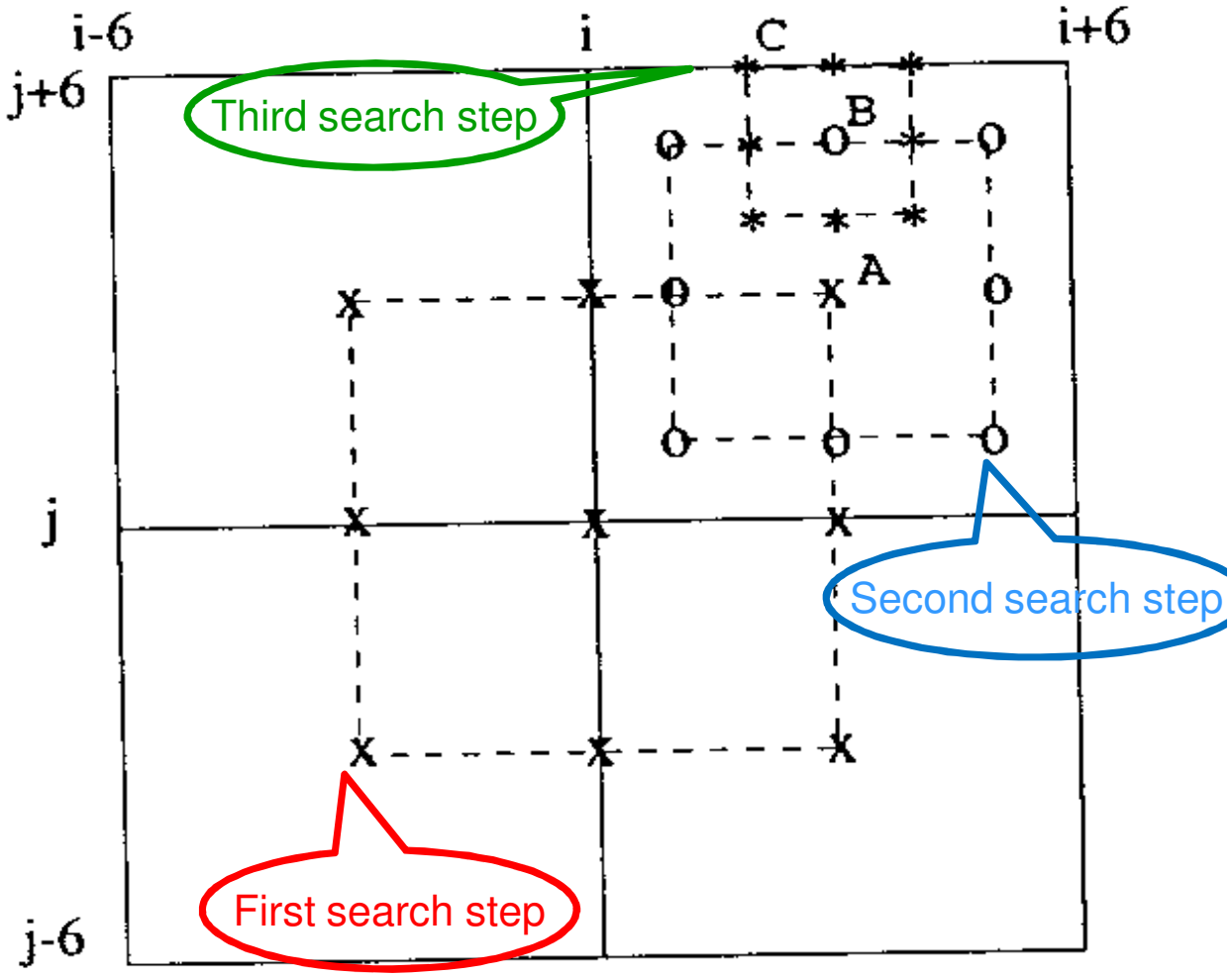
Motion vectors



Differences WITH motion comp.

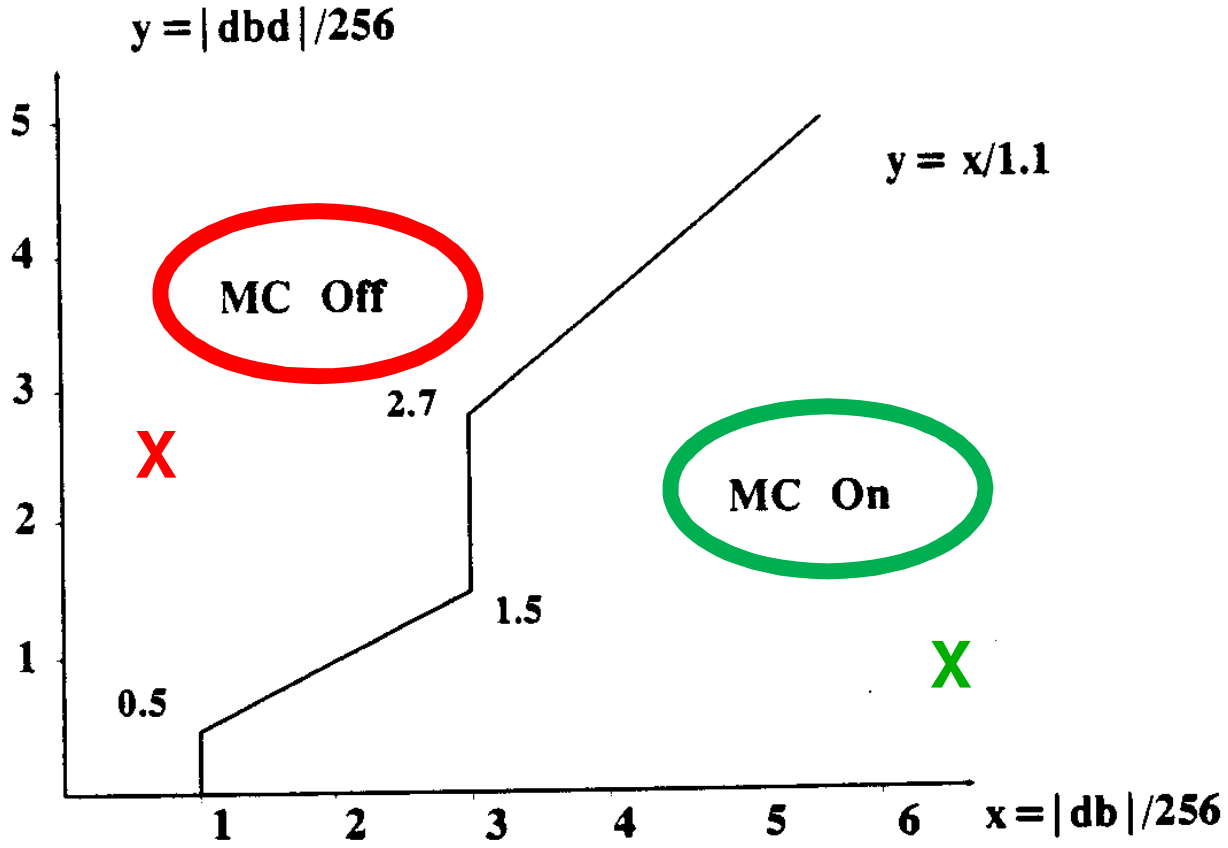
# Fast Motion Estimation: Three Steps Motion Estimation Algorithm

**Fast motion estimation algorithms offer much lower complexity than ful search at the cost o some small quality reduction since predictions are less optimal and thus the prediction erro is higher !**

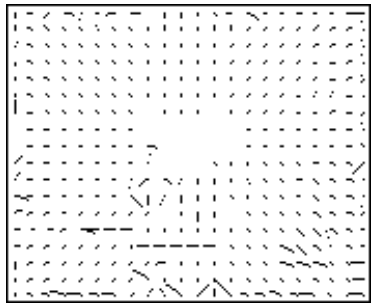


# Motion Compensation Decision Characteristic

**db** – *difference block*
**dbd** – *displaced block difference*



# H.261 Motion Estimation Rules …



- **Number of MVs** - One motion vector may be transmitted for each macroblock (if the encoder so desires).
- **Range of MVs** - Motion vector components (x and y) may take values from -15 to + 15 pels, in the vertical and horizontal directions, only the integer values.
- **Referenced area** - Only motion vectors referencing areas within the reference (previously coded) image are valid.
- **Chrominance MVs** - The motion vector transmitted for each MB is used for the 4 luminance blocks in the MB. The chrominance motion vector is computed by dividing by 2 and truncating the luminance motion vector.
- **Semantics** - A positive value for the horizontal or vertical motion vector components means the prediction must be made using the samples in the previous image, spatially located to the right and below the samples to be predicted.

# H.261 Motion Vectors Coding

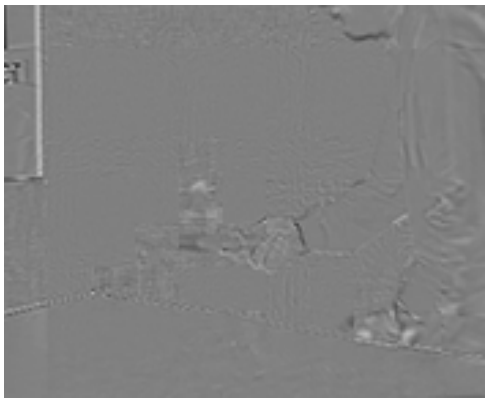
- To exploit the redundancy between the motion vectors of adjacent MBs, each motion vector is differentially coded as the difference between the motion vector of the actual MB and its prediction, this means the motion vector of the preceding MB.

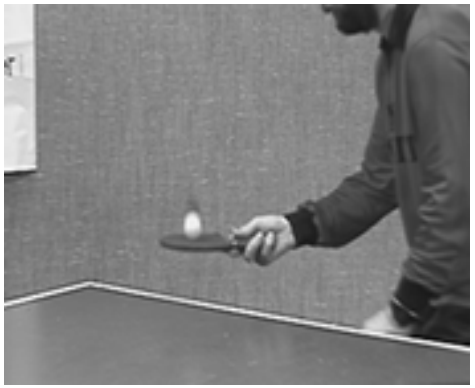
- The motion vector prediction is null when no redundancy is likely to be present, notably when:
  - The actual MB is number 1, 12 or 23
  - The last transmitted MB is not adjacent to the actual MB
  - The preceding and contiguous MB did not use motion compensation

# Inter Versus Intra Coding

**In H.261, the MBs are coded either in Inter or Intra coding modes:**

- **INTER CODING – To be used at block level when there is substantial temporal redundancy; may imply the usage or not of motion estimation.**
- **INTRA CODING – To be used when there is NO substantial temporal redundancy; no temporal predictive coding is used in this case ('absolute' coding like in JPEG is used).**






# Exploiting Spatial Redundancy and Irrelevancy

# After Time, the Space


Actual image


Prediction image,
motion compensated

+

+

-

Prediction error

DCT
Transform

# Transform Coding

**Transform coding involves the division of the image in N×N blocks to which the transform is applied, producing blocks with N×N coefficients.**

**A transform is formally defined through the formulas with the direct and inverse transforms:**

$$F(u,v) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j)\ A(i,j,u,v)$$

$$f(i,j) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u,v)\ B(i,j,u,v)$$

*where*

*f(i,j) – Input signal (in space)*

*A (i,j,u,v) – Direct transform basis functions*

*F(u,v) – Transform coefficients*

*B (i,j,u,v) – Inverse transform basis functions*

# Discrete Cosine Transform (DCT)

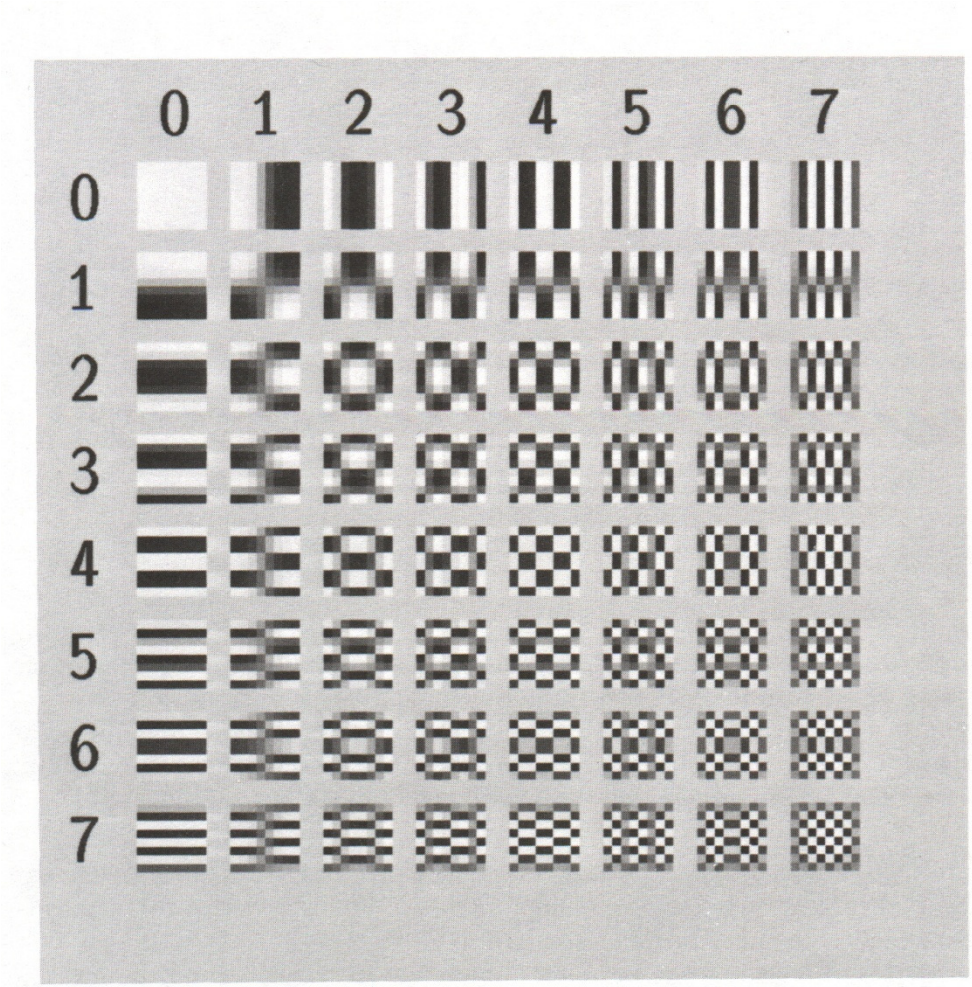
**The DCT is one of the sinusoidal transforms which means its basis functions are sampled sinusoidal functions.**

$$F(u,v) = \frac{2}{N} C(u)C(v) \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} f(j,k) \cos\left(\pi \frac{u(2j+1)}{2N}\right) \cos\left(\pi \frac{v(2k+1)}{2N}\right)$$

$$f(j,k) = \frac{2}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)C(v)F(u,v) \cos\left(\frac{u(2j+1)}{2N}\pi\right) \cos\left(\frac{v(2k+1)}{2N}\pi\right)$$

**The DCT is undoubtedly the most used transform in image and video coding since its performance is close to the KLT compacting performance for signals with high correlation and there are fast implementation solutions available.**

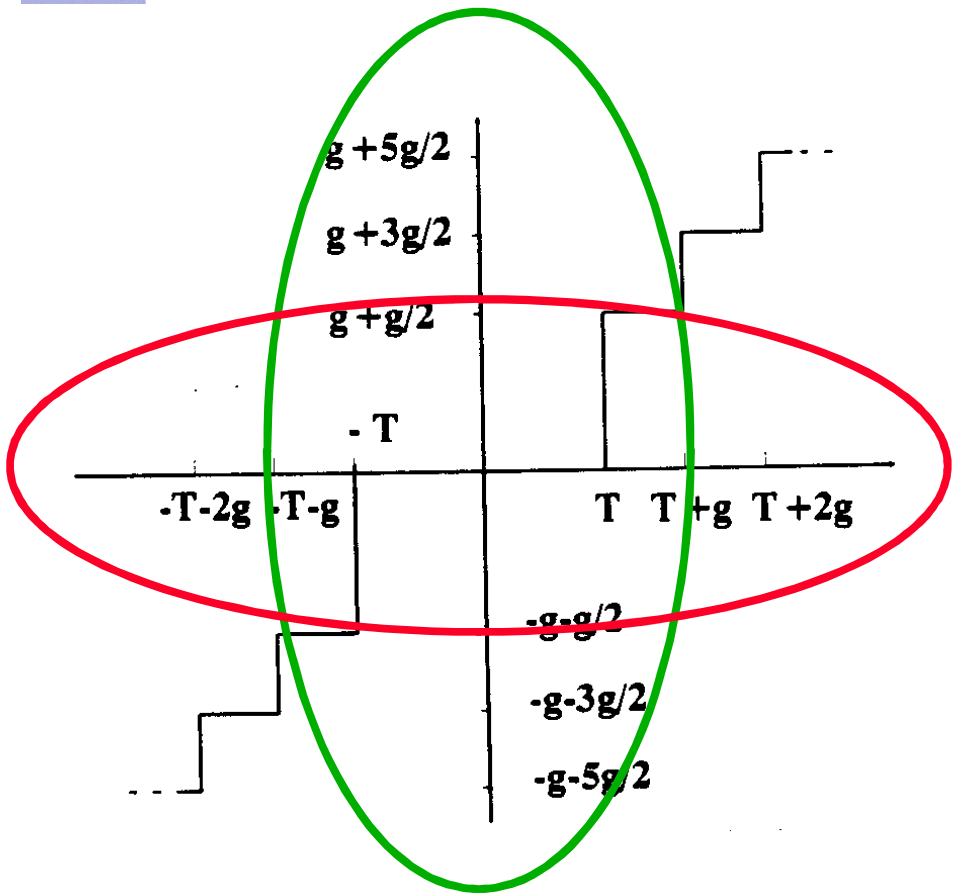
# Bidimensional DCT Basis Functions (N=8)




# The DCT Transform in H.261

- **Block size** - In H.261, the DCT is applied to blocks with 8×8 samples. This value results from a trade-off between the exploitation of the spatial redundancy and the computational complexity.
- **Coefficients selection** - The DCT coefficients to transmit are selected using non-normative thresholds allowing the consideration of psycho visual criteria in the coding process, targeting the maximization of the subjective quality.
- **Quantization** - To exploit the irrelevancy in the original signal, the DCT coefficients to transmit for each block are quantized.
- **Zig-Zag scanning** - Since the signal energy is compacted in the upper, left corner of the coefficients' matrix and the human visual system sensibility is different for the various frequencies, the quantized coefficients are zig-zag scanned to assure that more important coefficients are always transmitted before less important ones.

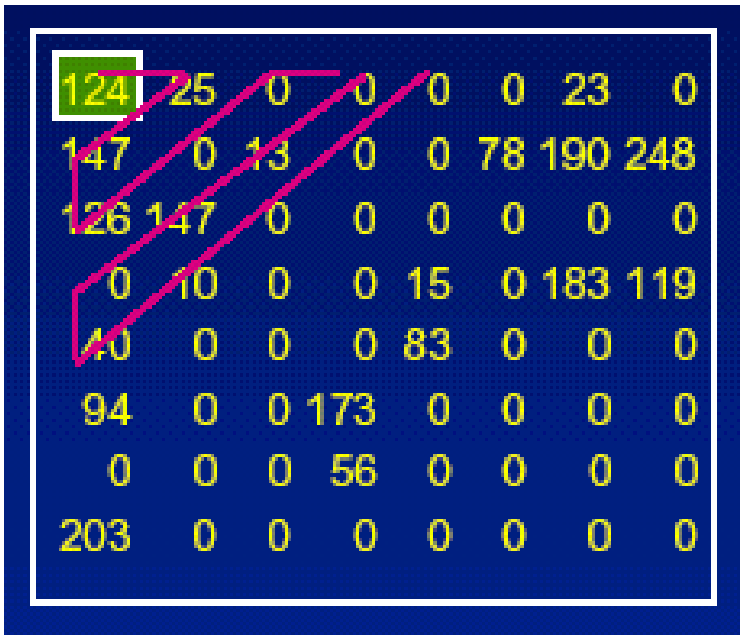
# H.261 Quantization




Example quantization characteristic

- Rec. H.261 uses as quantization steps all even values between 2 and 62 (31 quantizers available).
- Within each MB, all DCT coefficients are quantized with the same quantization step with the exception of the DC coefficient for Intra MBs which are always quantized with step 8.
- Rec. H.261 normatively defines the regeneration values for the quantized coefficients but not the decision values which may be selected to implement different quantization characteristics, uniform or not.

# Serializing the DCT Coefficients




- **The transmission of the quantized DCT coefficients requires to send the decoder two types of information about the coefficients: their position and quantization level (for the selected quantization step).**

- **For each DCT coefficient to transmit, its position and quantization level are represented using a bidimensional symbol**

**(*run, level*)**

**where the *run* indicates the number of null coefficients before the coefficient under coding, and the *level* indicates the quantized level of the coefficient.**


# Exploiting Statistical Redundancy

# Statistical Redundancy: Entropy Coding

**Entropy coding**

**CONVERTS SYMBOLS IN BITS !**

**Using the statistics of the symbols to transmit to achieve additional (lossless) compression by allocating in a clever way bits to the input symbol stream.**


- **A, B, C, D -> 00, 01, 10, 11**
- **A, B, C, D -> 0, 10, 110, 111**

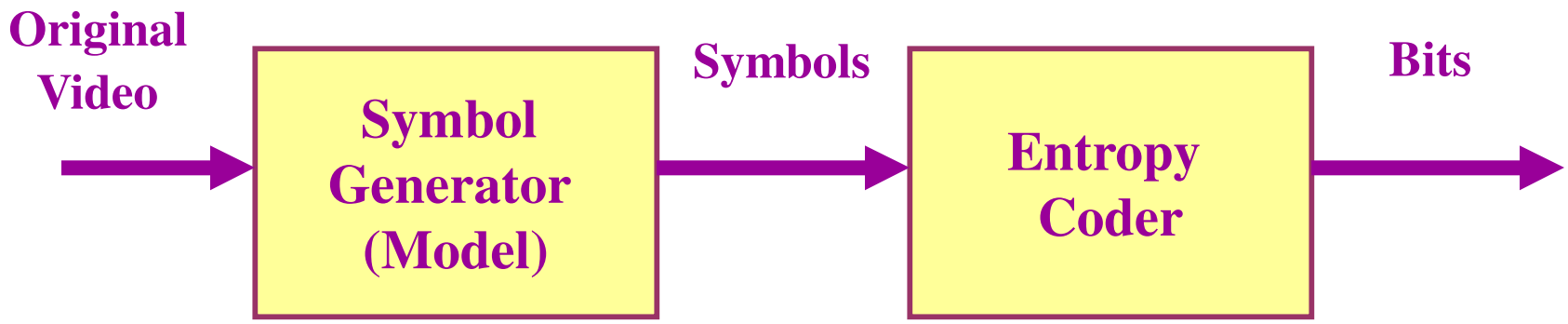

# Huffman Coding

**Huffman coding is one of the entropy coding tools which allows to exploit the fact that the symbols produced by the encoder do not have equal probability.**

- **To each generated symbol is attributed a codeword which size (in bits) is 'inversely' proportional to its probability.**
- **The usage of variable length codes implies the usage of an output buffer to 'smooth' the bitrate flow, if a synchronous channel is available.**
- **The increase in coding efficiency is 'paid' with an increase in the sensibility to channel errors.**

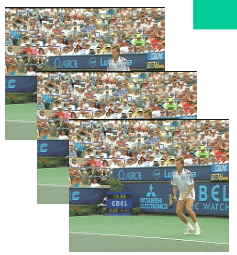

# Combining the Tools ...

# The H.261 Symbolic Model

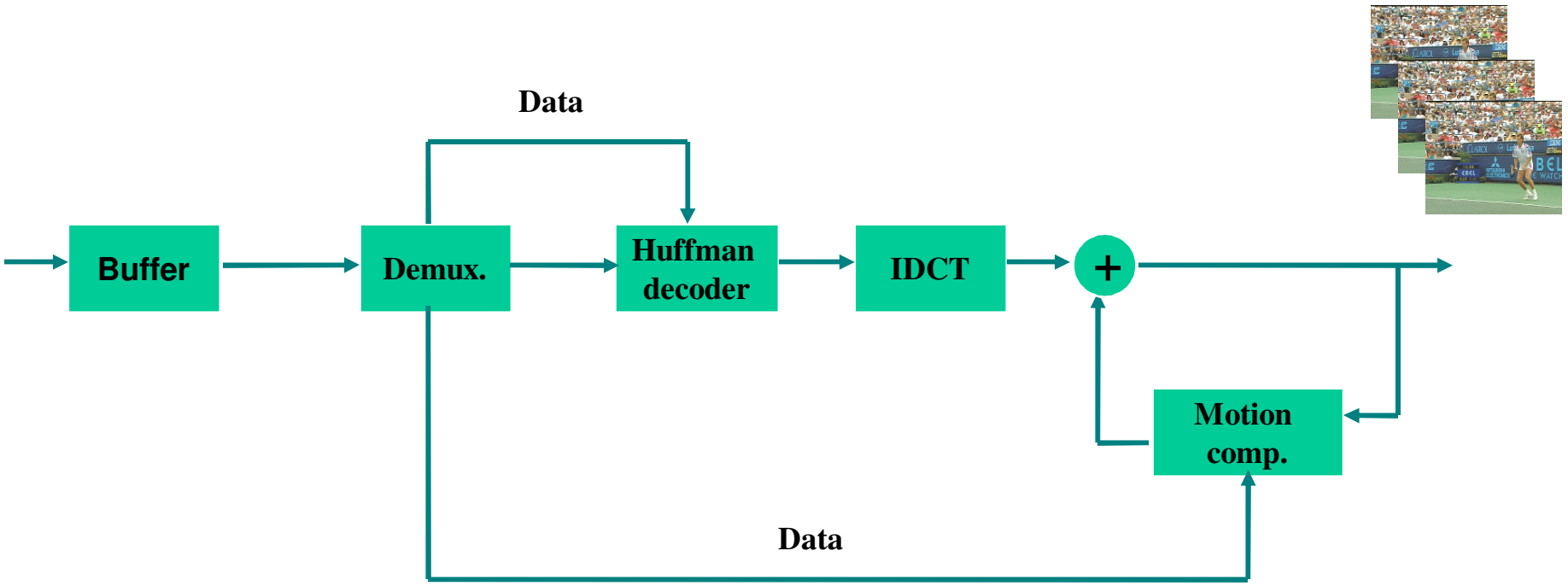



**A video sequence is represented as a sequence of images structured in macroblocks, each of them represented with motion vectors and/or (Intra or Inter coded) DCT coefficients for 8×8 blocks.**

# Encoder: the Winning Cocktail !



# Decoder: the Slave !



# Output Buffer

**The production of bits by the encoder is highly non-uniform in time, essentially because of:**

- **Variations in spatial detail for the various parts of each image**
- **Variations of temporal activity along time**
- **Entropy coding of the coded symbols**

**To adapt the variable bitrate flow produced by the encoder to the constant bitrate flow transmitted by the channel, an output buffer is used, which adds some delay.**

# Bitrate Control

**The encoder must efficiently control the way the available bits are spent in order to maximize the decoded quality for the synchronous bitrate/channel available.**

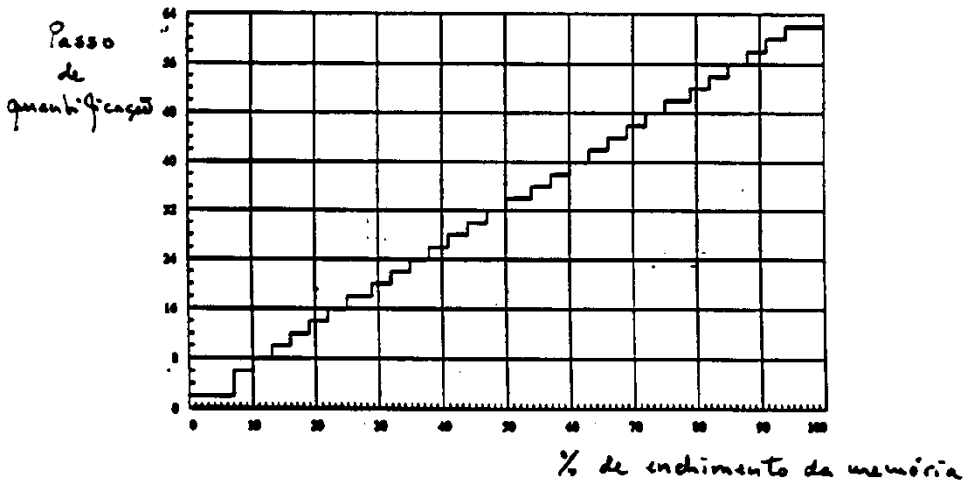
**Rec. H.261 does not specify what type of bitrate control must be used; various tools are available:**

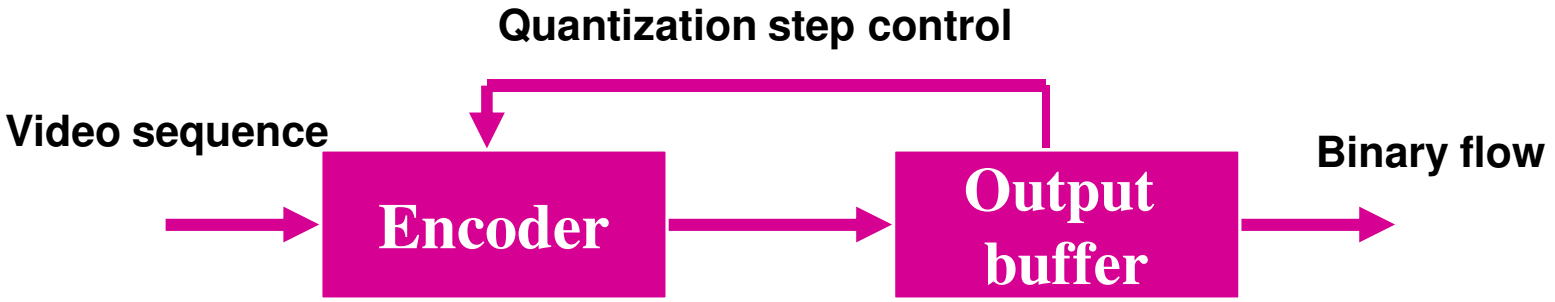
- **Changing the temporal resolution/frame rate**
- **Changing the spatial resolution, e.g. CIF to QCIF and vice-versa**
- **Controlling the macroblock classification**
- **CHANGING THE QUANTIZATION STEP VALUE**

**The bitrate control strategy has a huge impact on the video quality that may be achieved with a certain bitrate (and is not normative) !**

# Quantization Step versus Buffer Fullness

**The bitrate control solution recognized as most efficient, notably in terms of the granularity and frequency of the control, controls the quantization step as a function of the output buffer fullness.**

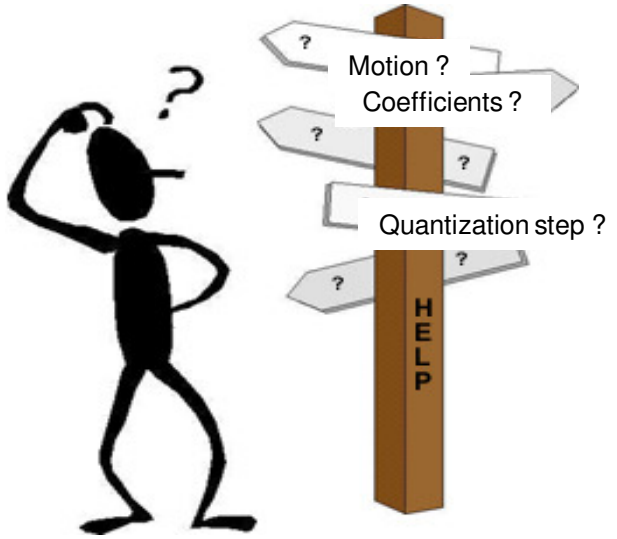


# The Importance of Well Choosing !

**To well exploit the redundancy and irrelevancy the video sequence, the encoder has to select:**

- **Which coding tools are used for each MB, depending of its characteristics;**
- **Which set of symbols is the best to represent MB, e.g. motion vector and DCT coefficients.**




**While the encoder has the mission to take important decisions and make critical choices, the decoder is a 'slave', limited to follow the 'orders' sent by the encoder; decoder intelligence is only shown for error concealment.**

# A Tool Box for Macroblock Classification

- **Macroblocks are the basic coding unit since it is at the macroblock level that the encoder selects the coding tools to use.**
- **Each coding tool is more or less adequate to a certain type of content and thus MB; it is important that, for each MB, the right coding tools are selected.**
- **Since Rec. H.261 includes several coding tools, it is the task of the encoder to select the best tools for each MB; MBs are thus classified following the tools used for their coding.**
- **When only spatial redundancy is exploited, MBs are INTRA coded; if also temporal redundancy is exploited, MBs are INTER coded.**

# Macroblock Classification Table

**VLC table for MTYPE**

| Prediction | MQUANT | MVD | CBP | TCOEFF | VLC |
|---|---|---|---|---|---|
| Intra | | | | x | 0001 |
| Intra | x | | | x | 0000 001 |
| Inter | | | x | x | 1 |
| Inter | x | | x | x | 0000 1 |
| Inter + MC | | x | | | 0000 0000 1 |
| Inter + MC | | x | x | x | 0000 0001 |
| Inter + MC | x | x | x | x | 0000 0000 01 |
| Inter + MC + FIL | | x | | | 001 |
| Inter + MC + FIL | | x | x | x | 01 |
| Inter + MC + FIL | x | x | x | x | 0000 01 |

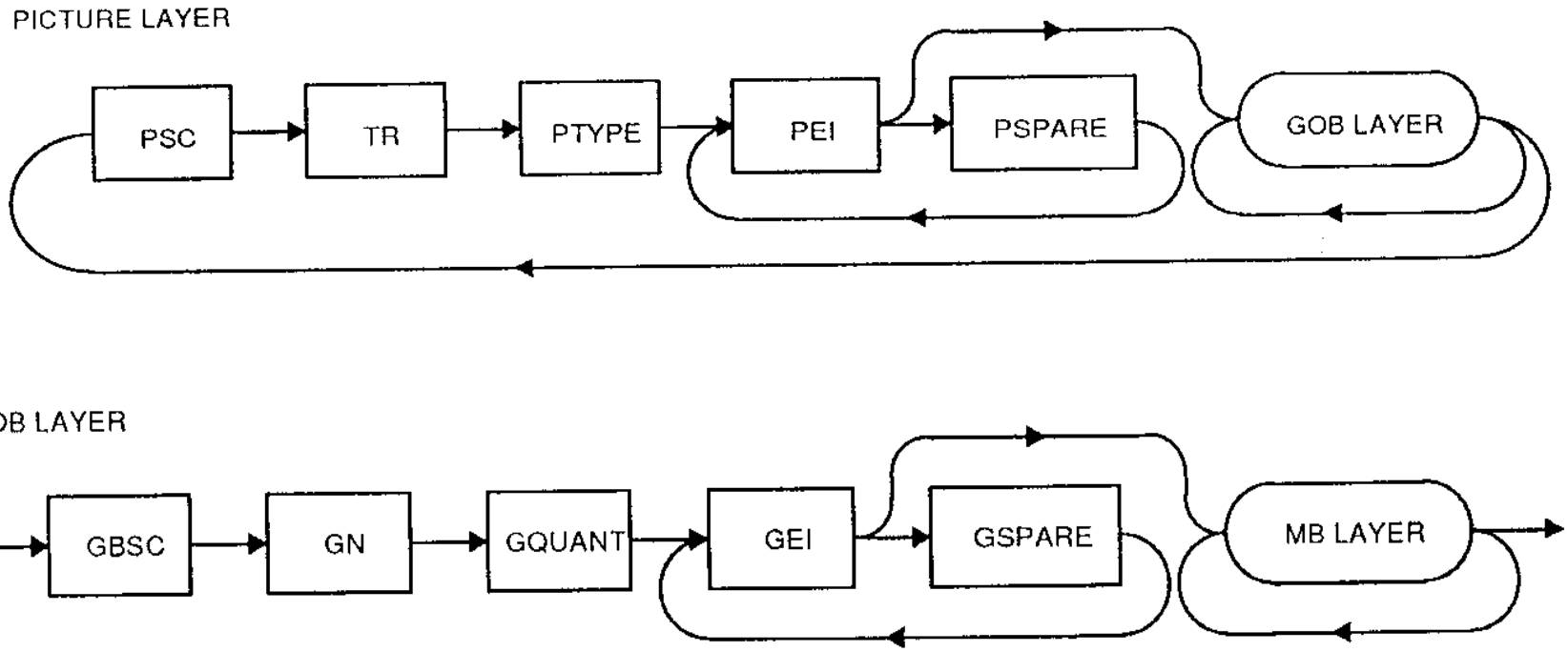
*Note 1* – "x" means that the item is present in the macroblock.

*Note 2* – It is possible to apply the filter in a non-motion compensated macroblock by declaring it as MC + FIL but with a zero vector.

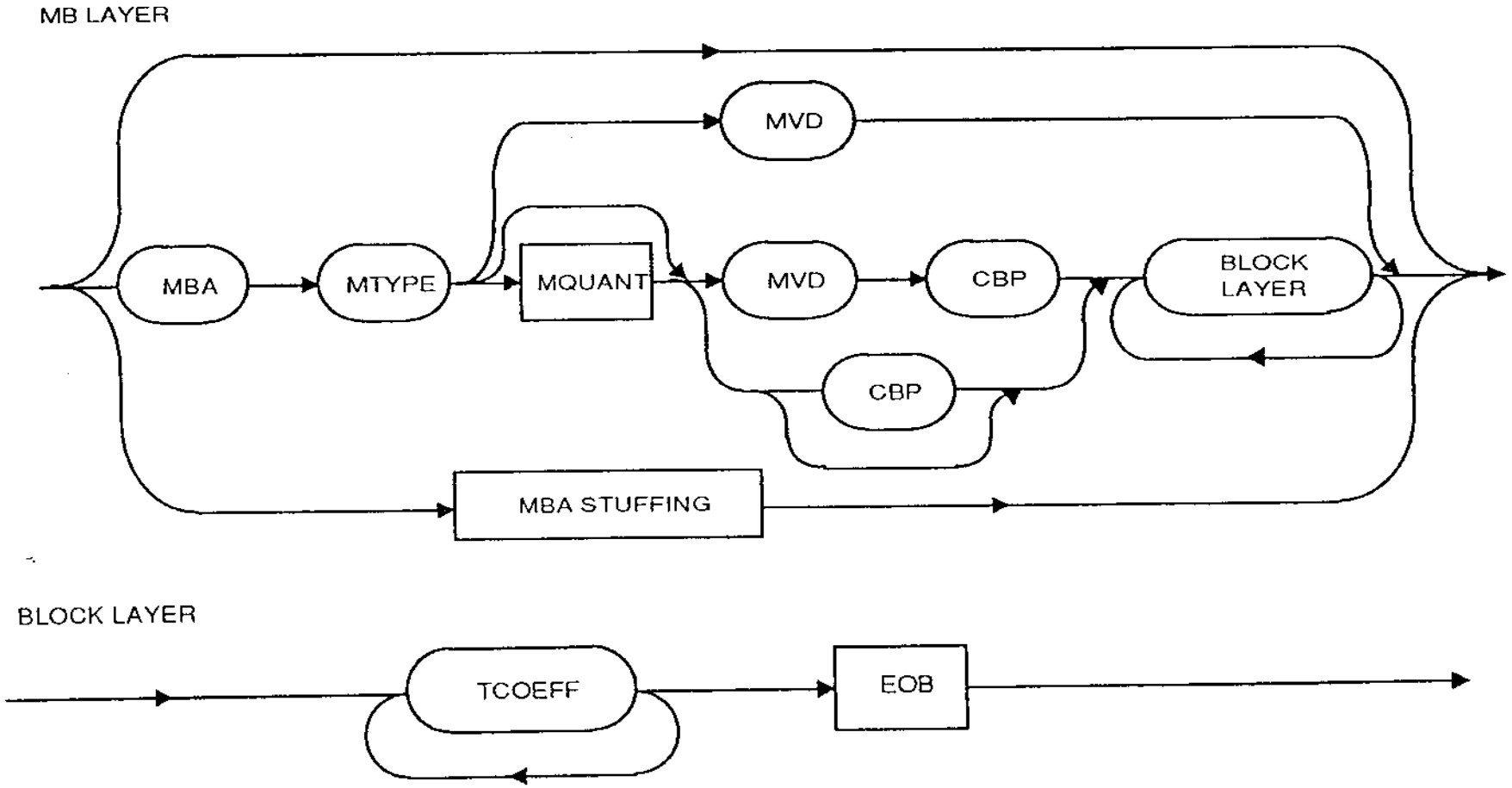

# Hierarchical Information Structure

- **Image**
  - **Resynchronization (*Picture header*)**
  - **Temporal resolution control**
  - **Spatial resolution control**
- **Group of Blocks (GOB)**
  - **Resynchronization (*GOB header*)**
  - **Quantization step control (mandatory)**
- **Macroblock**
  - **Motion estimation and compensation**
  - **Quantization step control (optional)**
  - **Selection of coding tools (MB classification)**
- **Block**
  - **DCT**

# Coding Syntax: Image and GOB Levels



# Coding Syntax: MB and Block Levels



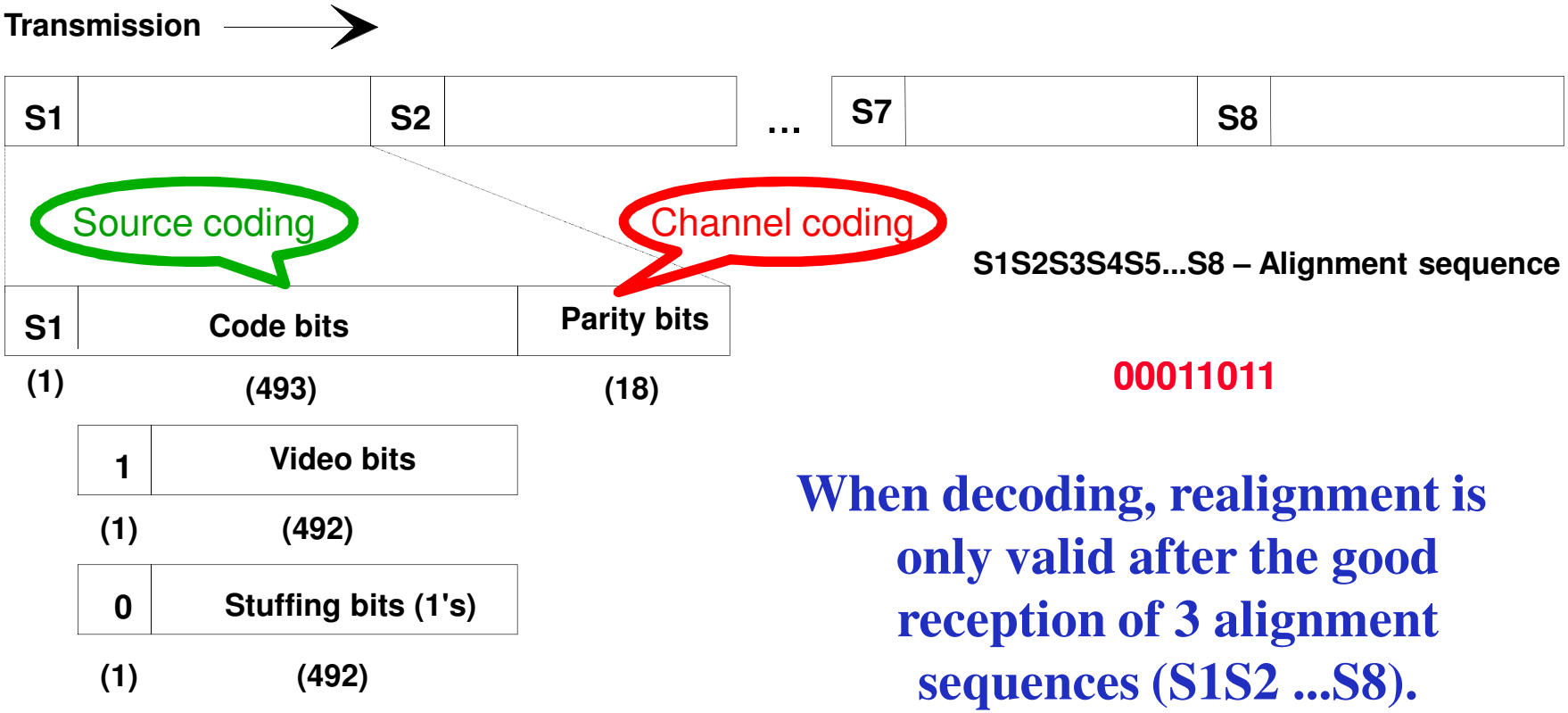

# Error Protection for the H.261 Binary Flow

- **Error protection for the H.261 binary flow is implemented by using a BCH (511,493) - *Bose-Chaudhuri-Hocquenghem* – block coding.**
- **The usage of the channel coding bits (also parity bits) at the decoder is optional.**
- **The syndrome polynomial to generate the parity bits is**

$$g(x) = (x^9 + x^4 + x)(x^9 + x^6 + x^4 + x^3 + 1)$$

# Error Protection for the H.261 Binary Flow

## The final video signal stream structure (multiframe with 512×8 = 4096 bits):

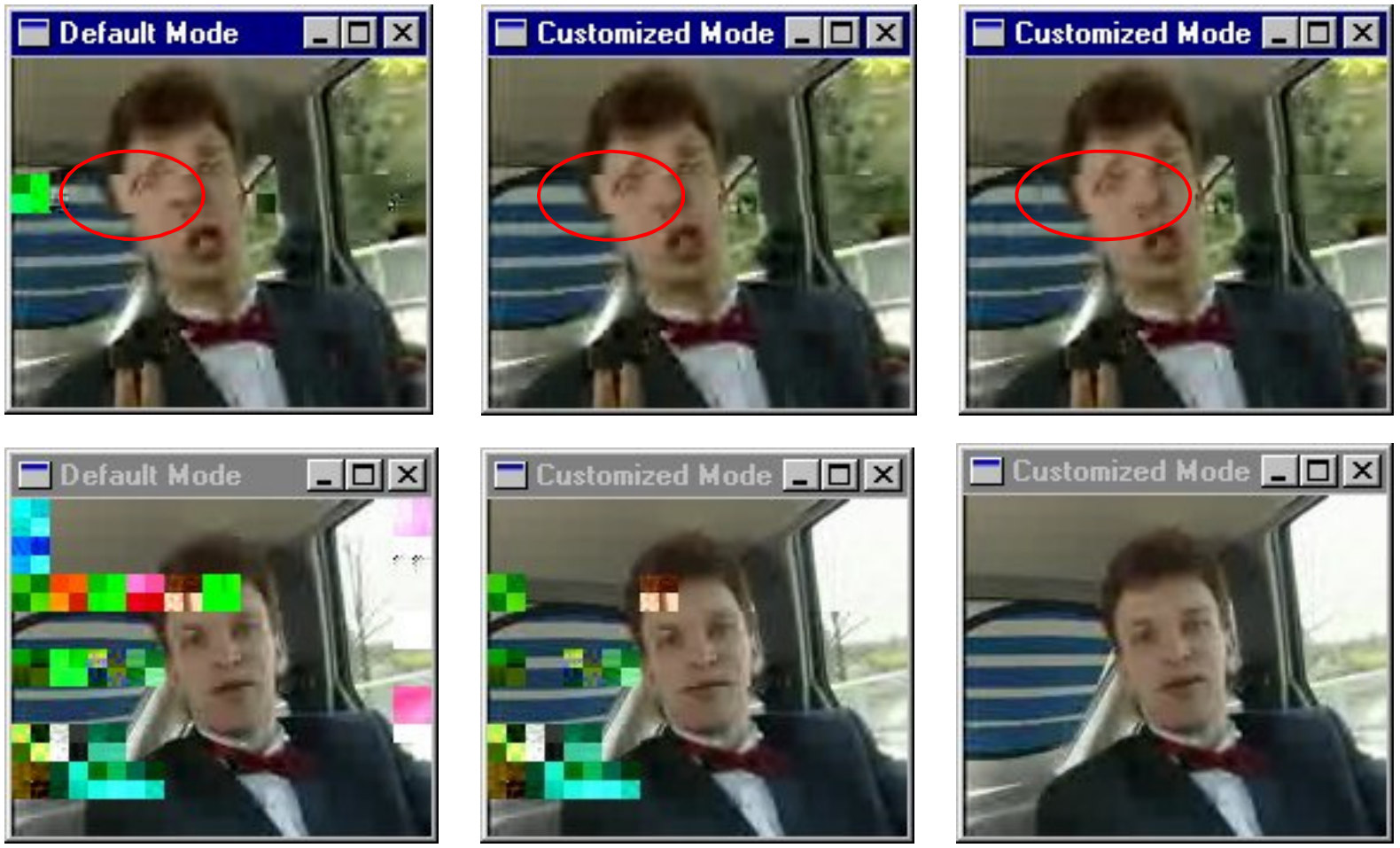



**When decoding, realignment is only valid after the good reception of 3 alignment sequences (S1S2 ...S8).**


# Error Concealment

- **Even when channel coding is used, some residual (transmission) errors may end at the source decoder.**
- **Residual errors may be detected at the source decoder due to syntactical and semantic inconsistencies.**
- **For digital video, the most basic error concealment techniques imply:**
  - **Repeating the co-located data from previous frame**
  - **Repeating data from previous frame after motion compensation**
- **Error concealment for non-detected errors may be performed through post-processing.**

# Error Concealment and Post-Processing Examples




# Final Comments

- **Rec. H.261 has been the first video coding international standard with relevant adoption.**
- **As the first relevant video coding standard, Rec. H.261 has established legacy and backward compatibility requirements which have influenced the standards to come after, notably in terms of technology selected.**
- **Many products and services have been available based on Rec. H.261.**
- **However, Rec. H.261 does not represent anymore the state-of-the-art on video coding (remind this standard is from 1990).**


# Bibliography


- **Videoconferencing and Videotelephony,** R. Schaphorst, Artech House, 1996
- **Image and Video Compression Standards: Algorithms and Architectures,** V. Bhaskaran and K. Konstantinides, Kluwer Academic Publishers, 1995
- **Multimedia Communications,** F. Halsall, Addison-Wesley, 2001
- **Multimedia Systems, Standards, and Networks,** A. Puri & T. Chen, Marcel Dekker, Inc., 2000