H.261 VIDEO CODING

Laboratory session

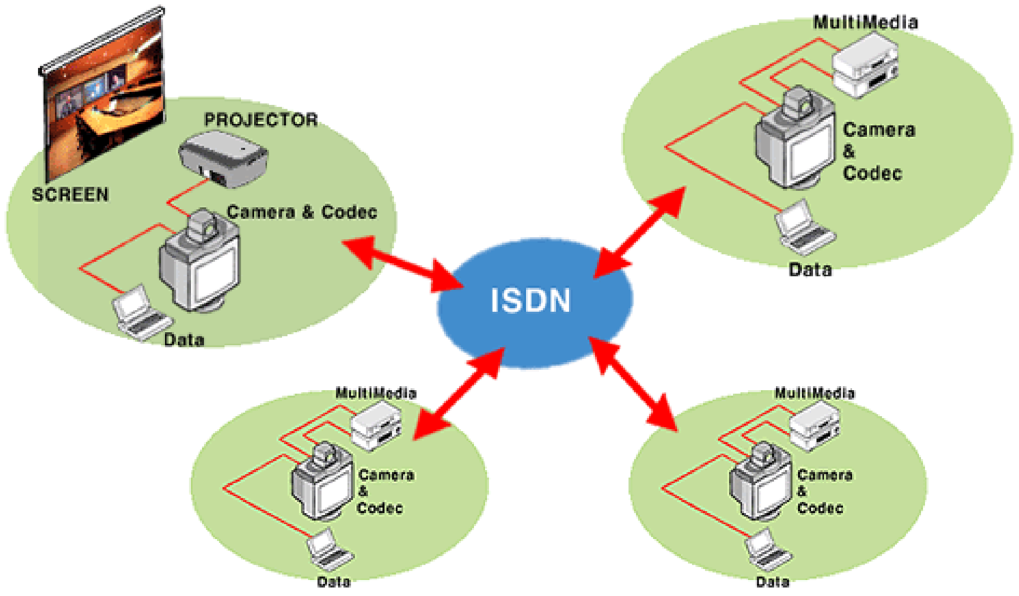
Fernando Pereira

Instituto Superior Técnico

Coding Changing Faces …

Recommendation ITU-T H.261: Objectives

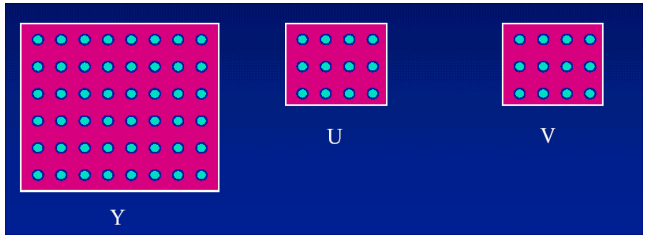
~1985



Efficient coding of videotelephony and videoconference visual data 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 market adoption, thus introducing the notion of backward compatibility in video coding standards.

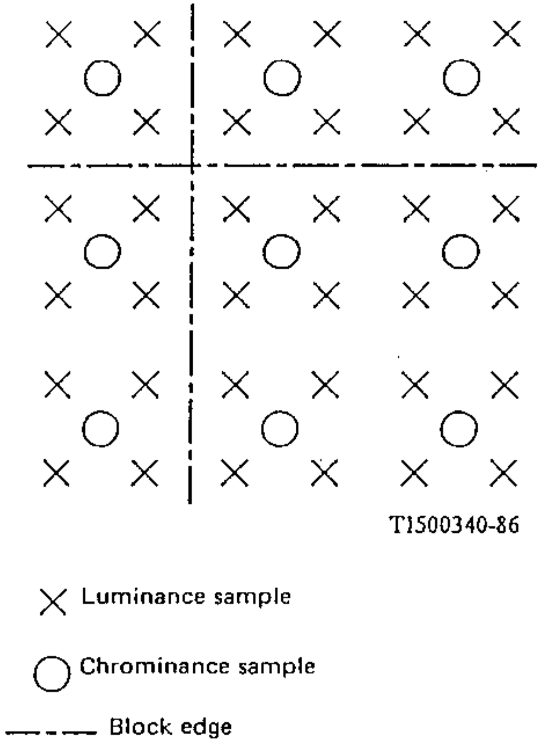
H.261: Image Resolution



Two spatial resolutions are possible:

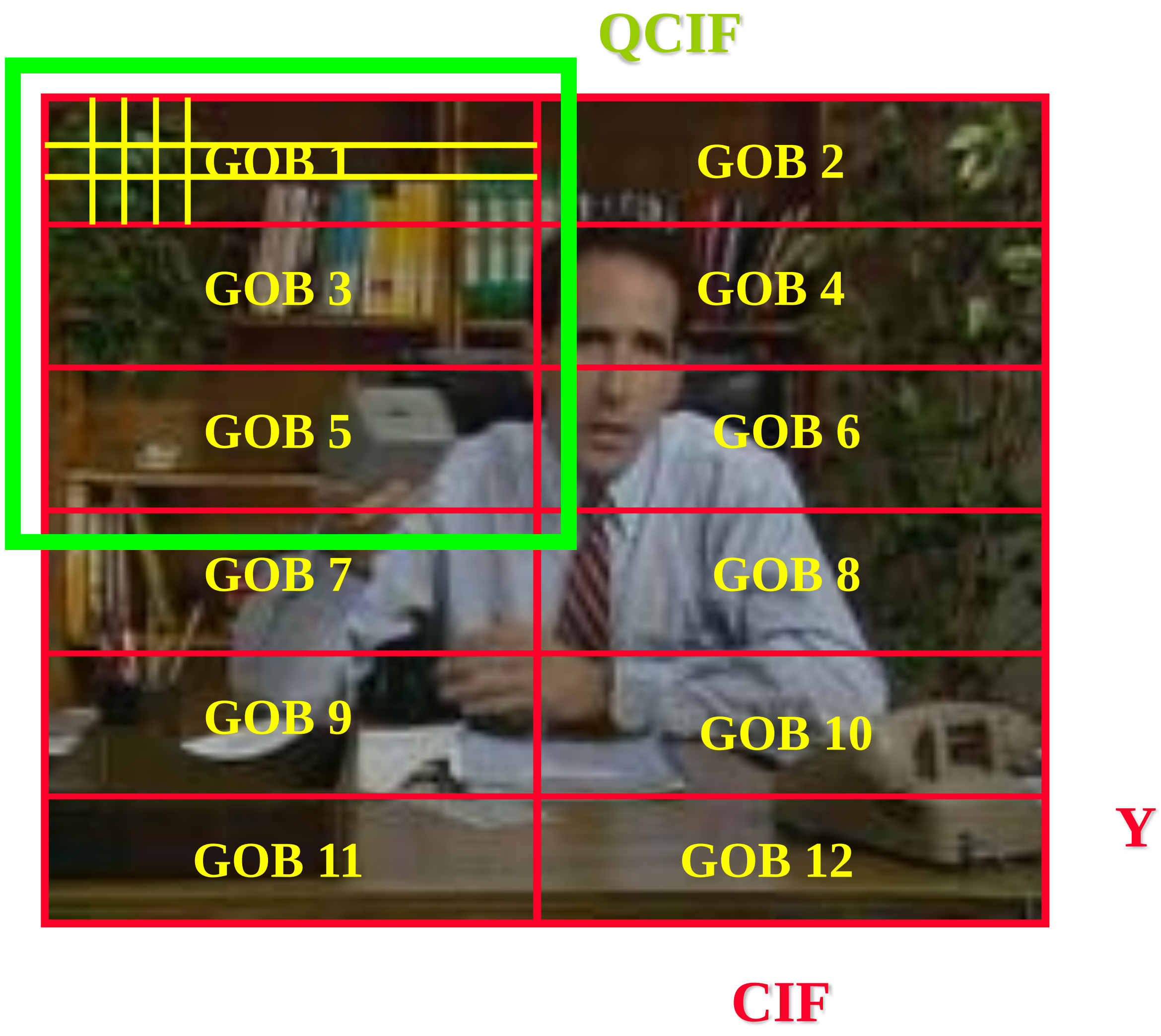
- CIF (*Common Intermediate Format*) - 288 × 352 samples for luminance (Y) and 144 × 176 samples for each chrominance (U,V) this means a 4:2:0 subsampling format, with 'quincux' positioning, progressive, 30 frame/s with a 4/3 aspect ratio.

- 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.



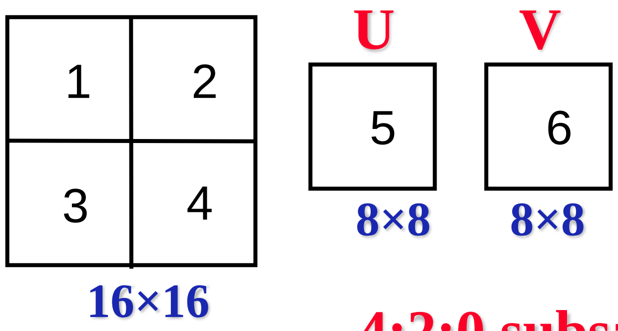
All H.261 codecs must work with QCIF and some may be able to work also with CIF (spatial resolution is set during initial negotiation).

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



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

- Images/Frames
- Group of Blocks (GOB)
- Macroblocks (MB) – 16×16 pixels
- Blocks - 8×8 samples



4:2:0 subsampling

TÉCNICO LISBOA

H.261: Coding Tools

Lossless ;-)

- **Temporal Redundancy**

  Predictive coding + residues: temporal prediction without and with motion compensation (only backward)

- **Spatial Redundancy**

  Transform coding (Discrete Cosine Transform, DCT)

- **Statistical Redundancy**

  Huffman entropy coding

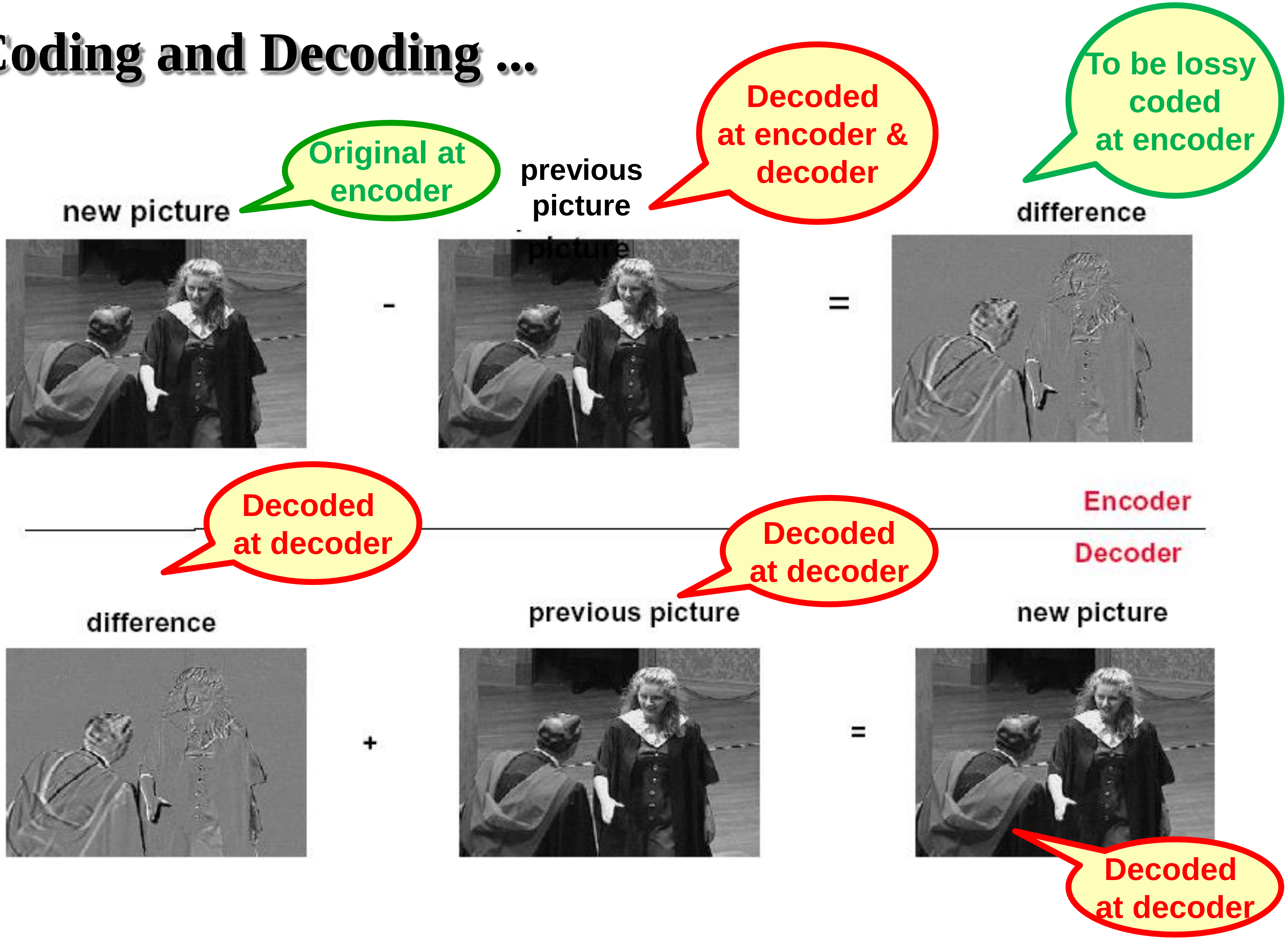
Lossy ;-(

- **Irrelevancy**

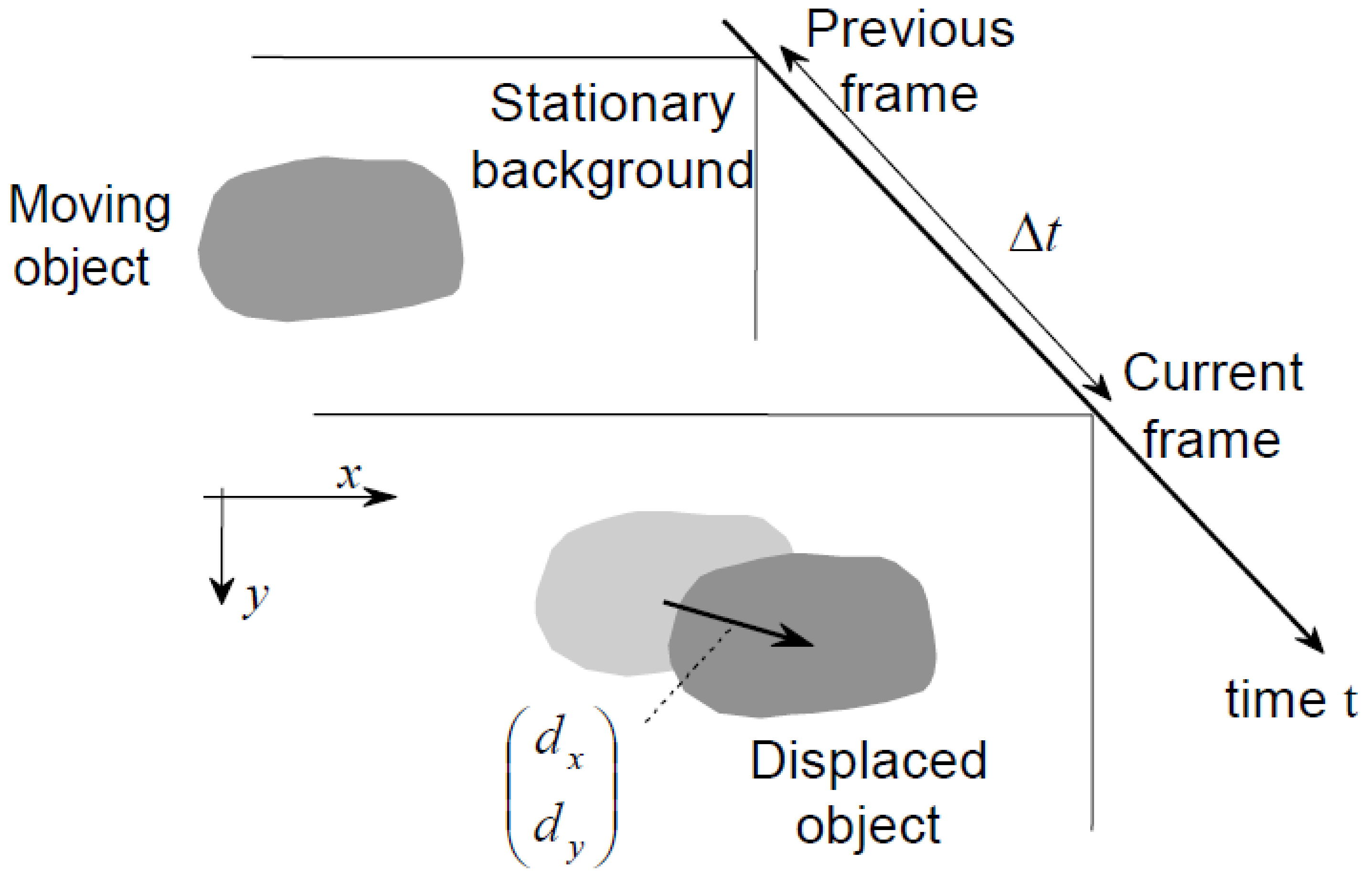
  Quantization of DCT coefficients

TÉCNICO LISBOA

Coding and Decoding ...



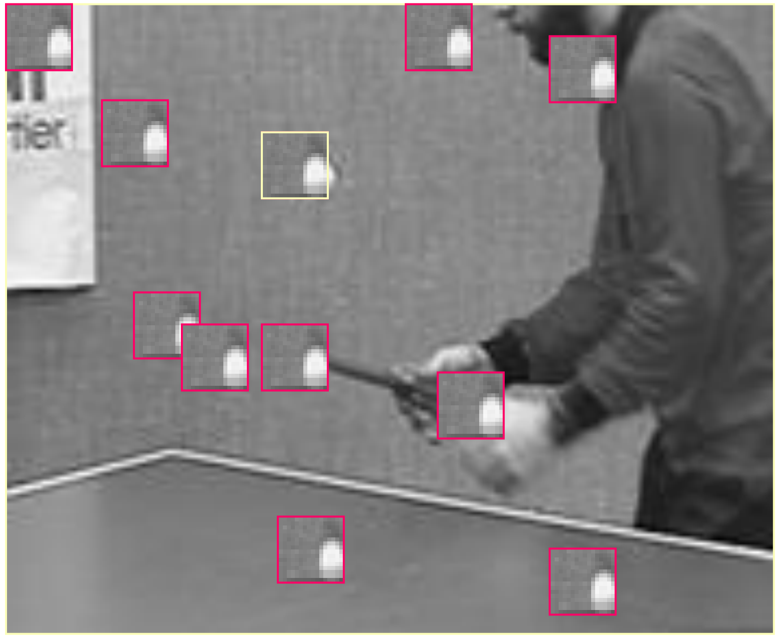
Motion in Action … Not New, Just Displaced !



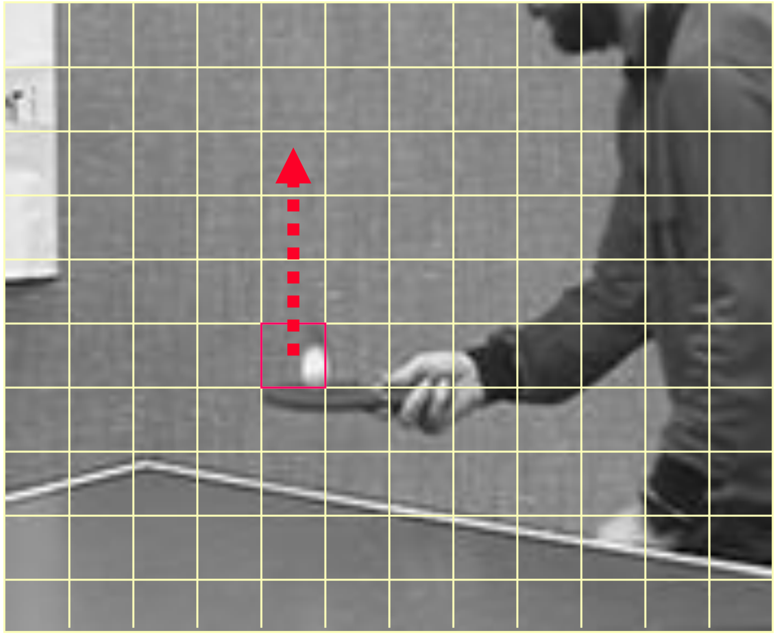
The key idea is to exploit as much as possible what the decoder already knows, i.e. the previously DECODED frames !

Motion Estimation by Macroblock Matching

Decoded Frame *i-1*
(available at encoder and decoder)



Original Frame *i*
(ONLY available at encoder)

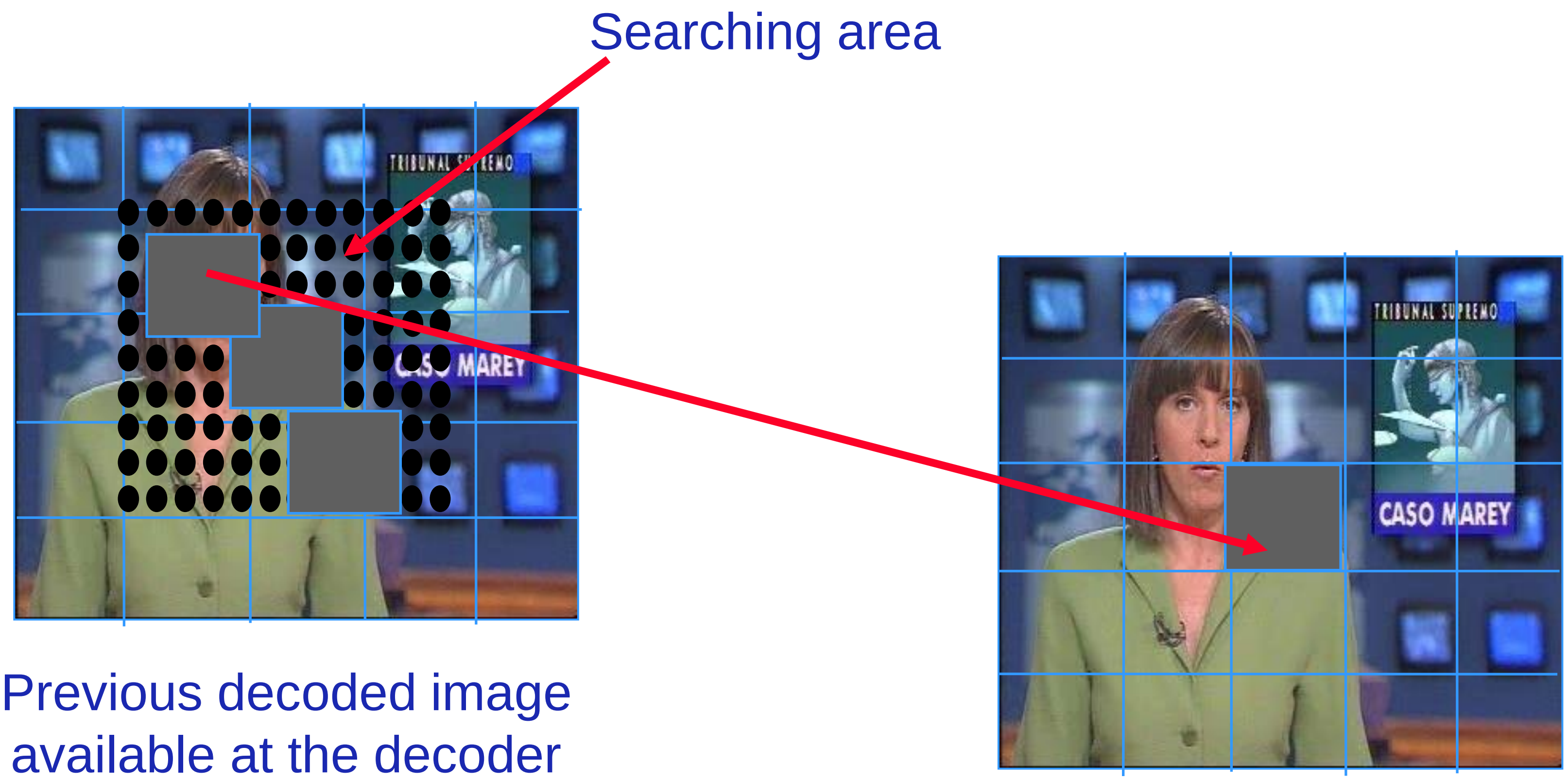


t

The best MB prediction from the previous decoded frame is the one minimizing an error metric, e.g. Mean Squared Error (MSE) or Mean Absolute Error (MAE), between the original MB in frame *i* and the candidate (displaced) MBs in frame *i-1*.

$$MSE = \frac{1}{n} \Sigma \underbrace{\left(y - \hat{y} \right)^2}_{\text{The square of the difference between actual and predicted}}$$

Motion Estimation: Where Worthwhile for Compression While Not Wasting Complexity ?

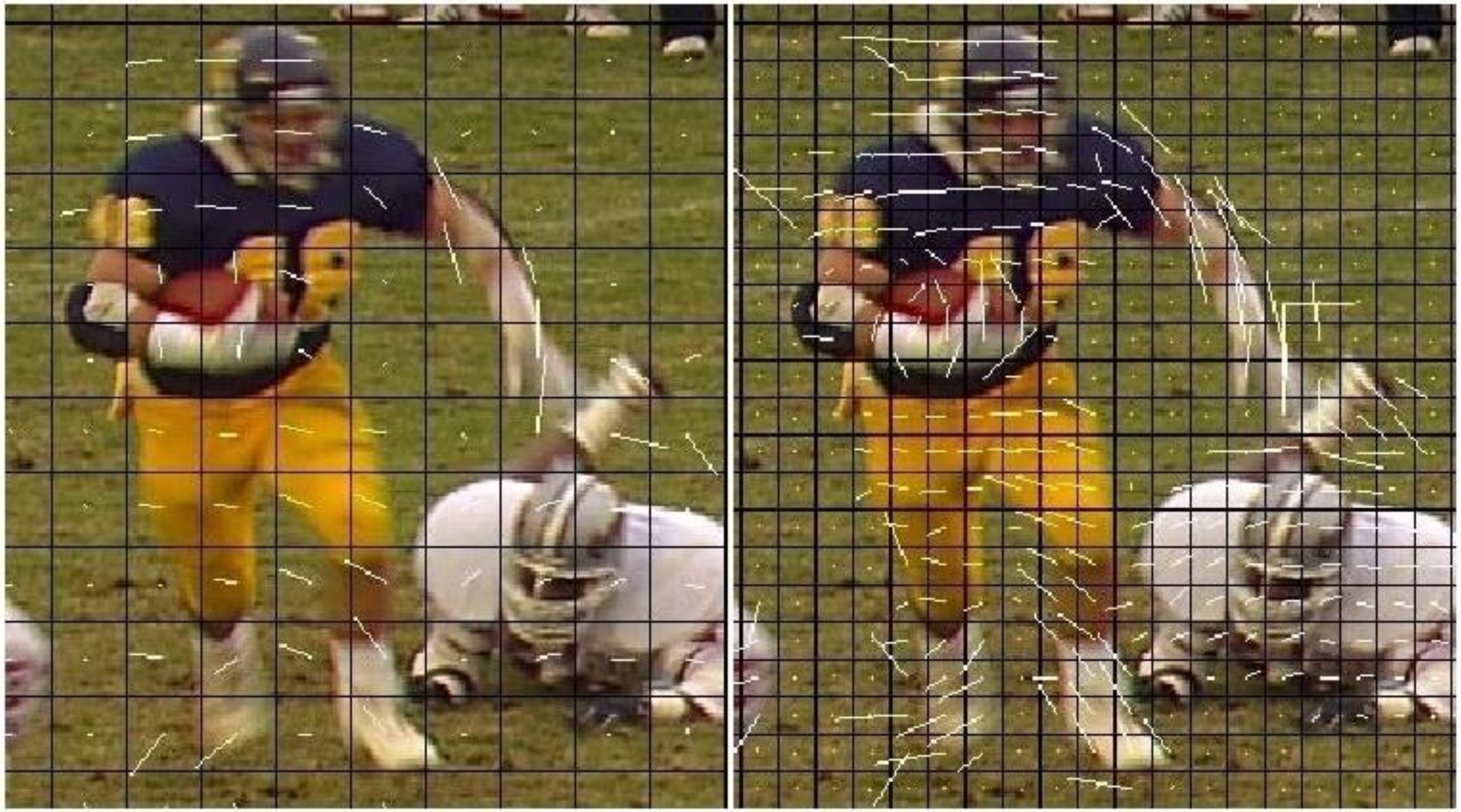


Previous decoded image available at the decoder AND encoder

Original image to encode ONLY available at encoder

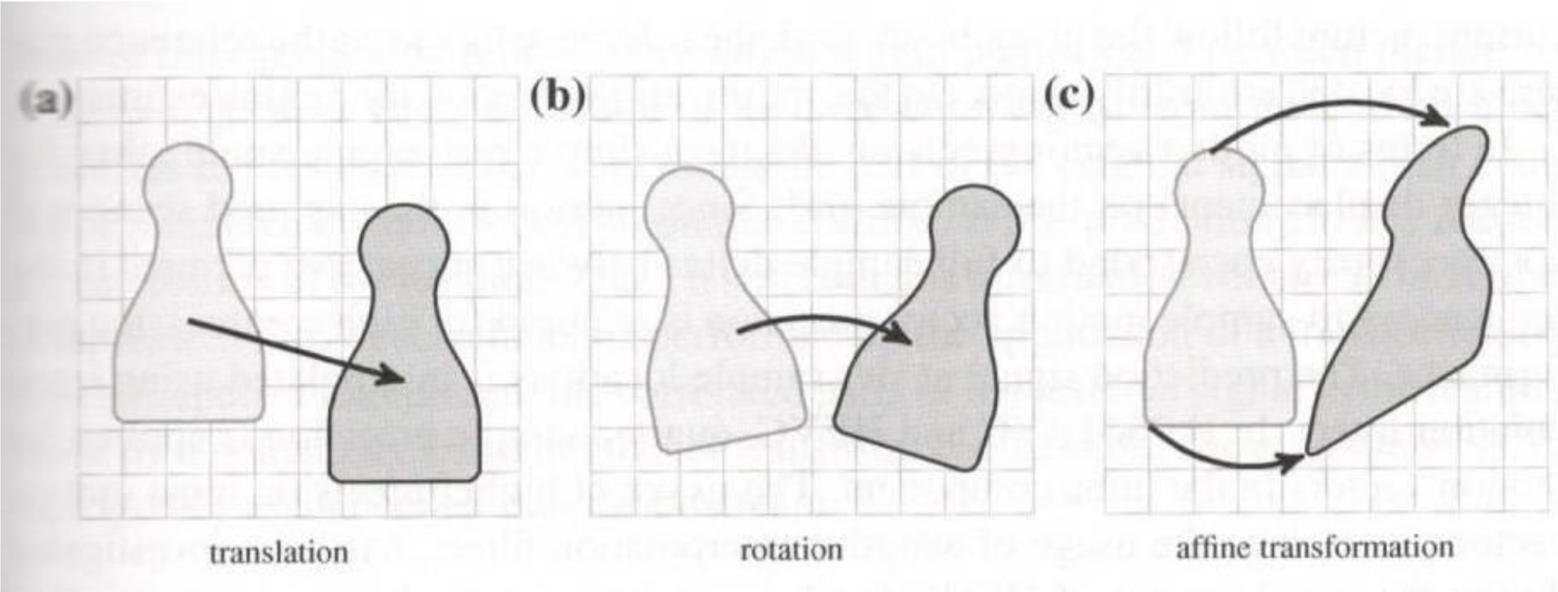
Motion Vectors at Different Spatial Resolutions

Low resolution | **Higher resolution**



MBs have always the same (16×16) number of pixels

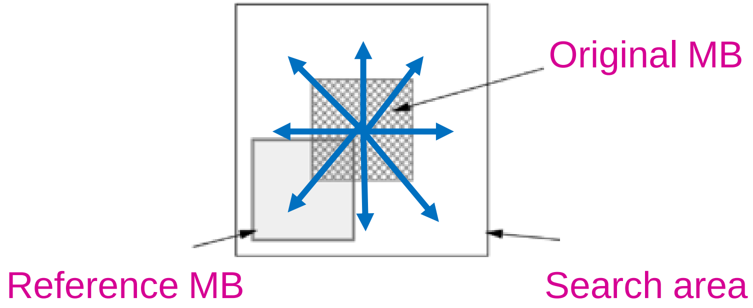
Motion is More than Translations !

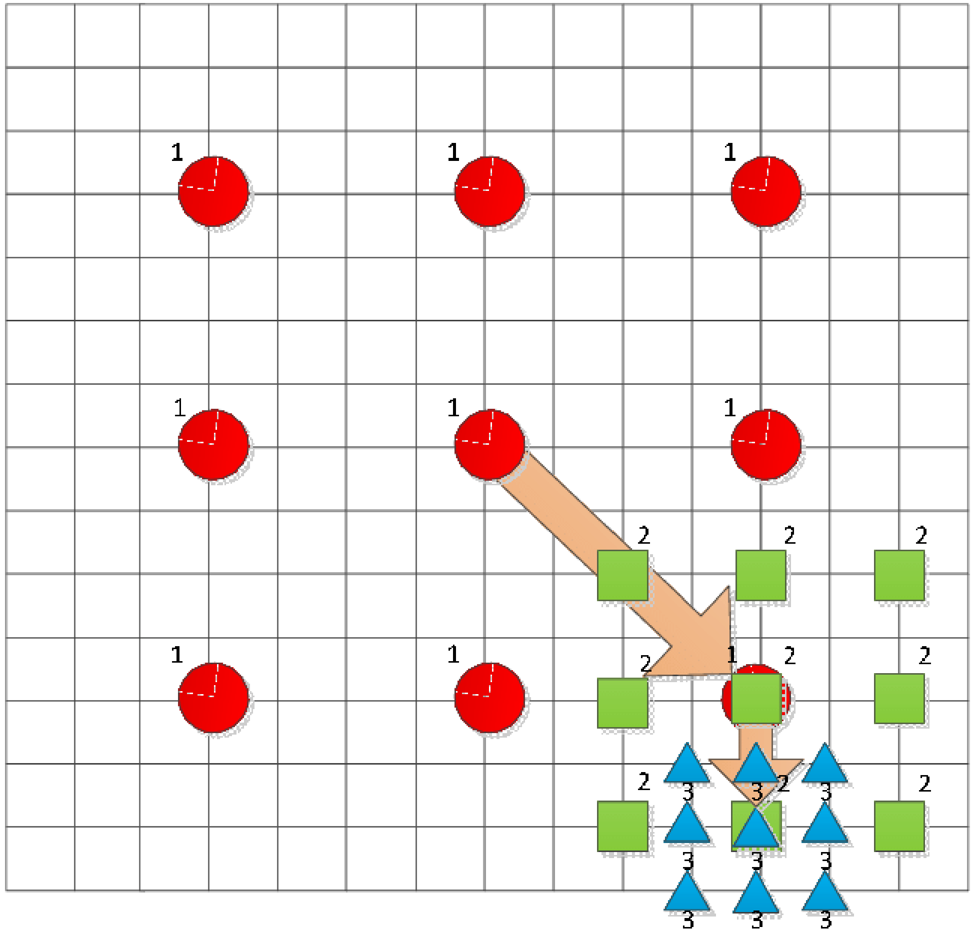


Clearly, a (translational) motion model (with 2D motion vectors) cannot represent well many types of motion …

But it is still very much worthwhile for efficient coding!

Fast Motion Estimation: Three-Steps Motion Estimation Algorithm





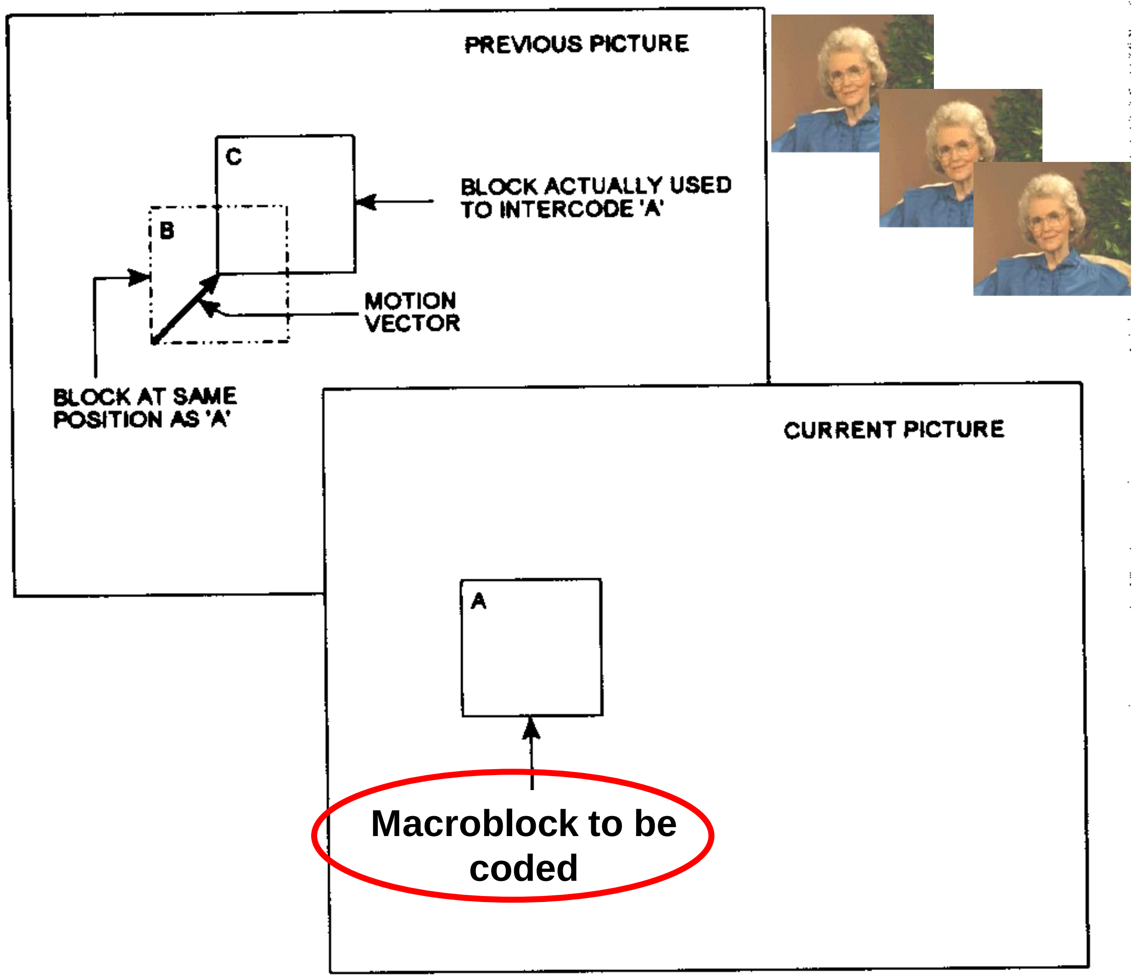
Fast motion estimation algorithms offer much lower computationl complexity than <u>full, exaustive search</u> at the cost of some small prediction residue increase/penalty since the motion-based prediction is less optimal and thus the prediction error is (slightly) higher !

Predicting in Time … With or Without Motion

Two main temporal prediction coding modes (Inter modes) are available for each MB:

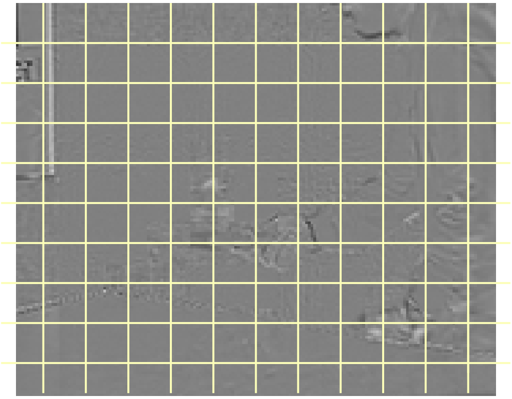
- Inter with No motion vector: Prediction from the same position in the previous frame
- Inter With a motion vector: Prediction from a displaced position in the previous frame

The encoder has to choose the *best coding deal* using some (non-normative) criteria !



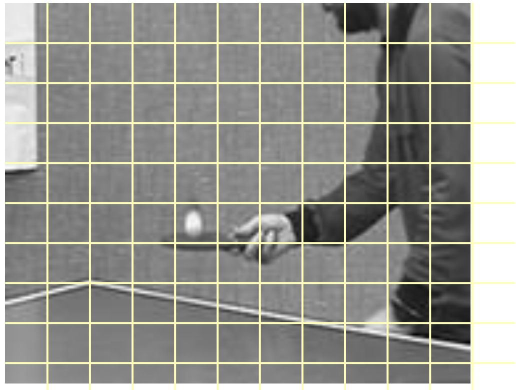
Inter Versus Intra Coding

Typically, the MBs are coded either in Inter or Intra coding mode:



- **INTER CODING MODE** – To be used when there is substantial temporal redundancy; may imply the usage or not of motion compensation, i.e. *Inter+MC* and *Inter(+noMC)*.

- **INTRA CODING MODE** – 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 to exploit the spatial redundancy).



After Temporal Redundancy, Spatial Redundancy

Original

Actual original image

+

+

-

Prediction image,
motion compensated at MB level
from previous decoded image

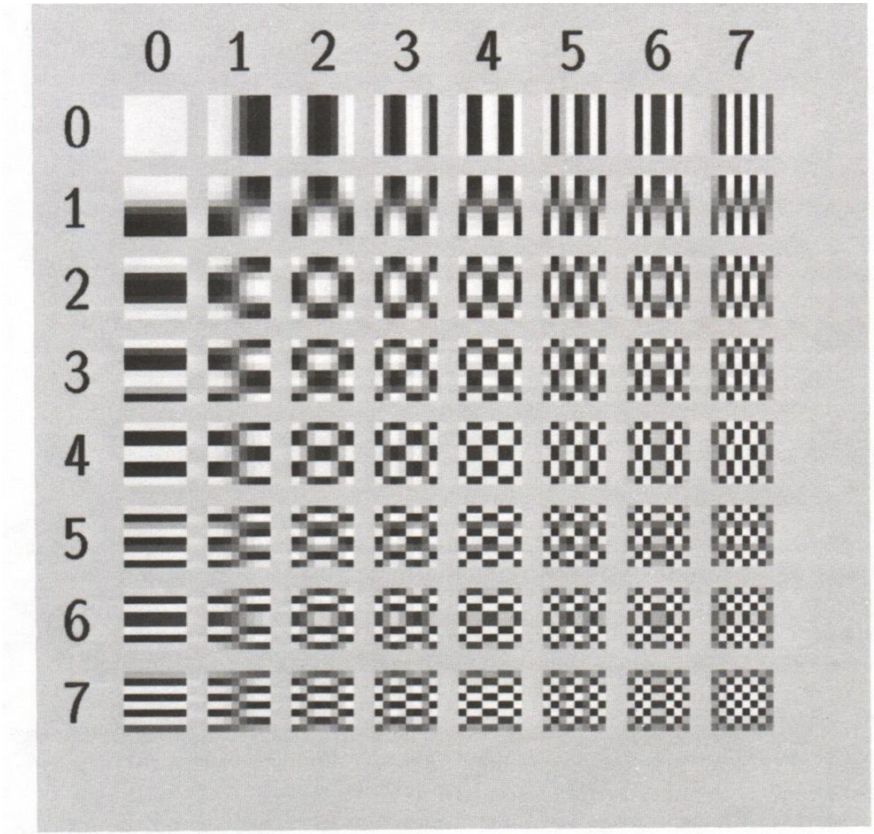
Created from previous decoded image with motion compensation

Residual to be coded with quantization

Prediction error

DCT Transform and quantization

Bidimensional DCT Basis Functions (N=8)



Exploiting Spatial Redundancy ...

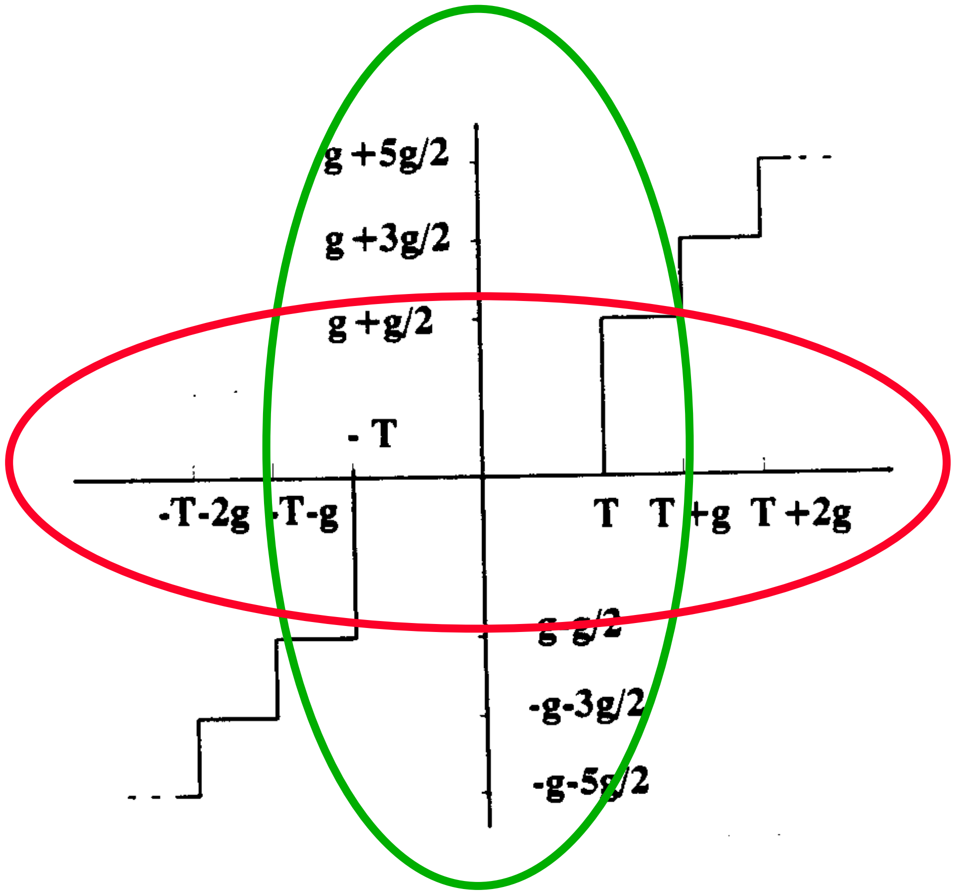
in the usual way ...

now also for the MB prediction error ...

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

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

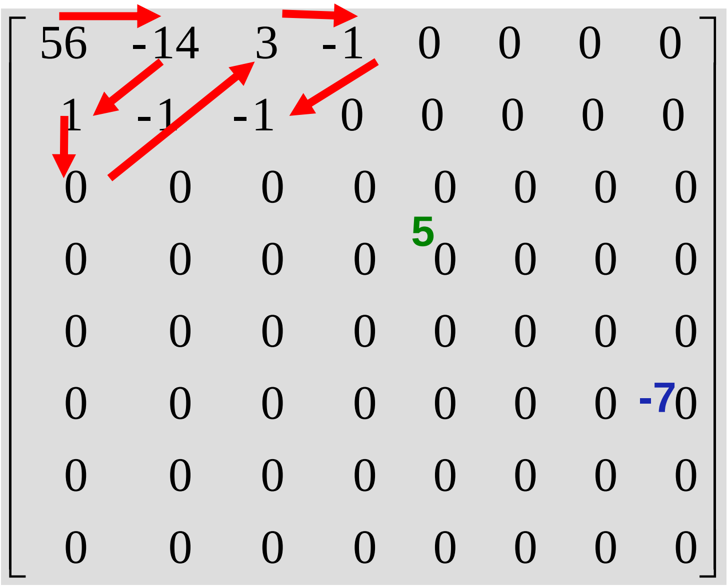
H.261 Quantization



Example quantization function

- 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.
- The usage of a same constant quantization step for all the AC DCT coefficients is motivated by the fact that a prediction error (and not absolute sample values) is being coded.
- H.261 normatively defines the reconstruction values for the quantized coefficients but not the decision values which may be selected to implement different quantization characteristics (uniform or not).

Serializing the Residual DCT Coefficients



(0, 56), (0, -14), (0, 1),
(1, -1), (0, 3), (0, -1),
(0, -1),
(Y, 5)
(X, -7)
EOB

- **The transmission of the quantized DCT coefficients requires to send the decoder two types of information about the coefficients: their <u>position</u> and <u>quantization level</u> (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.

Small runs and small absolute levels are more probable and entropy coding will exploit that.

Channel and Coding Bitrates ...

Channel bitrate

The channel limits the coding bitrate

1st

Motion vectors

Lossless

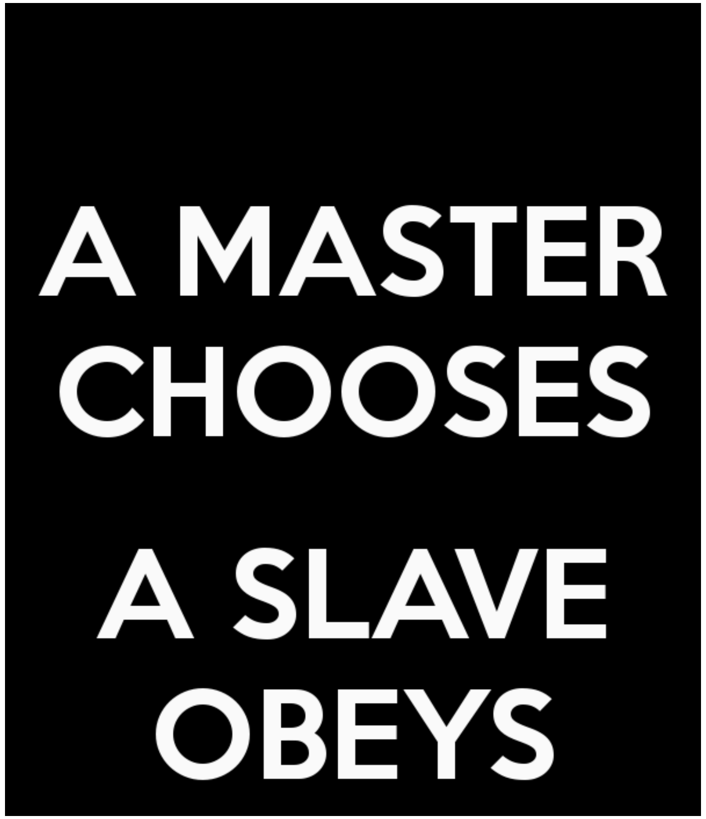
2nd

Residual and Intra DCT Coefficients (with quantization)

Lossy

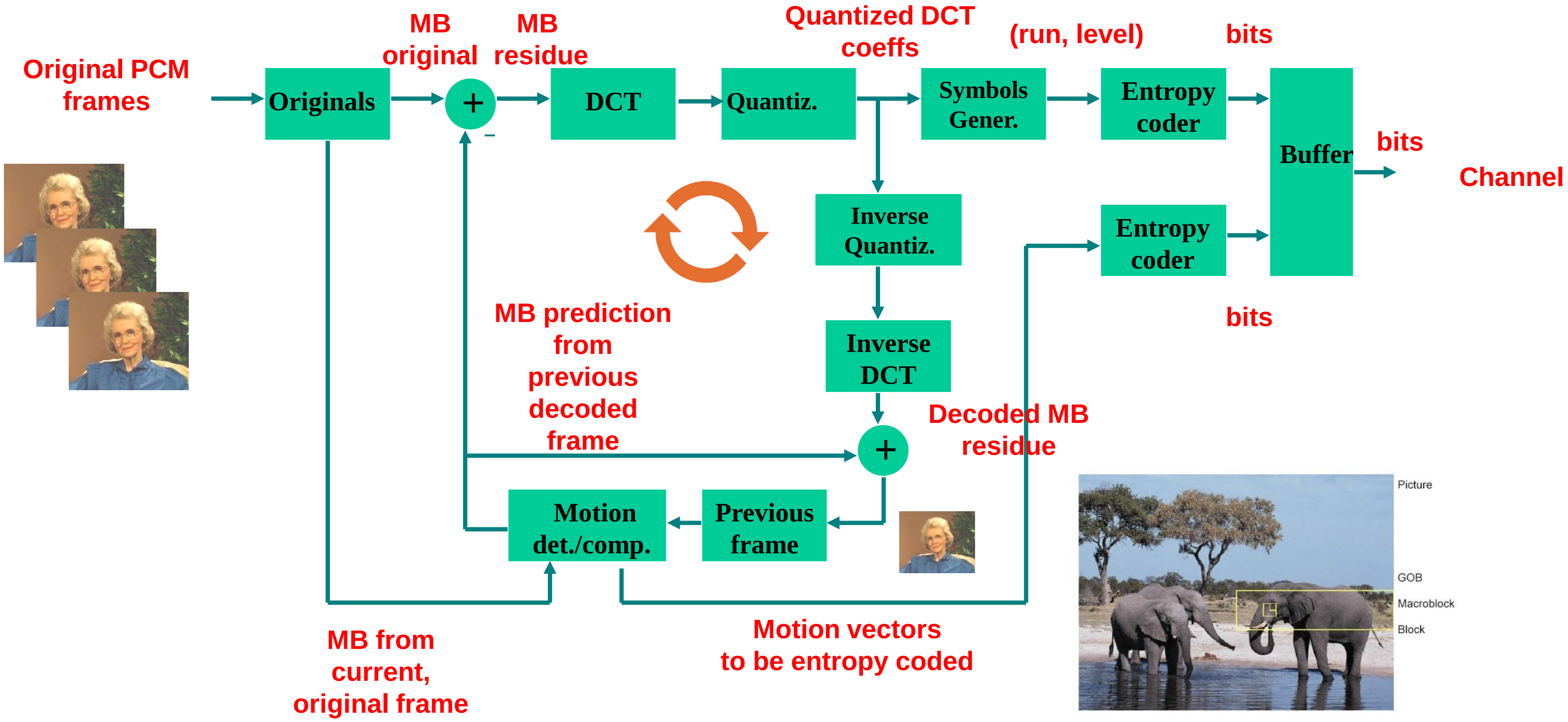
Encoder-Decoder or Master-Slave ?

Master
Complex
Intelligent
Non-normative
Defines performance
...

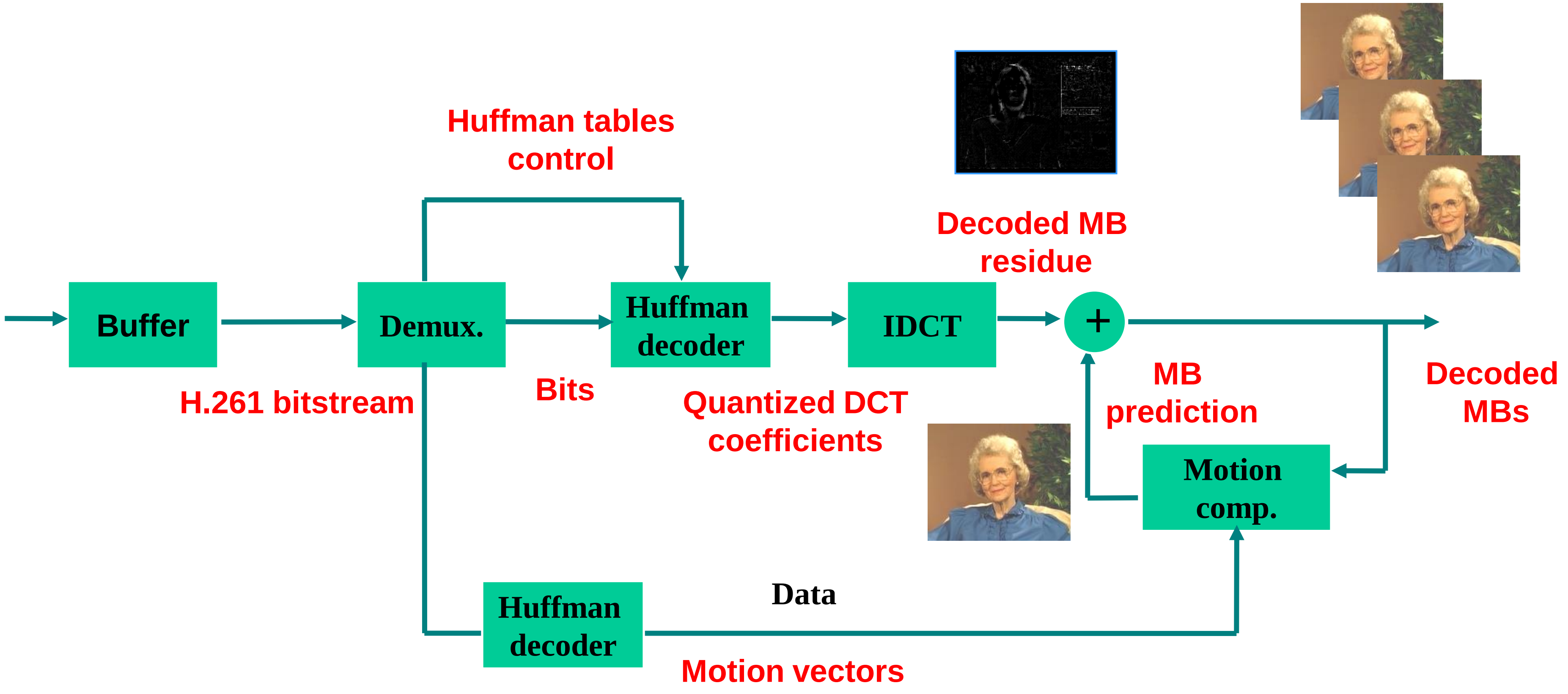


Slave
Simple
No room for intelligence
Normative
Does not define performance
...

Encoder: the Winning Cocktail !



Decoder: the Slave !

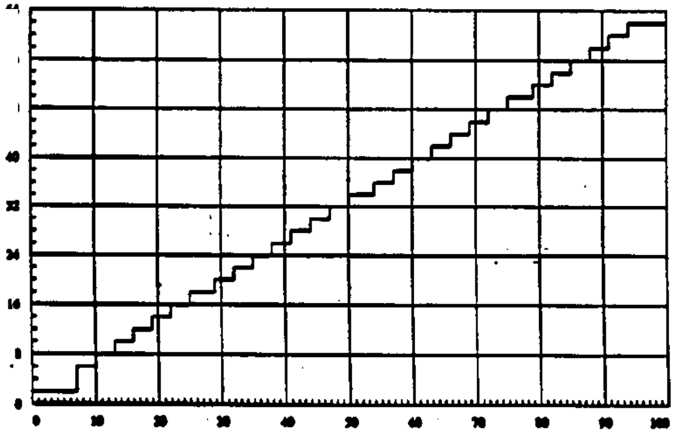


Quantization Step versus Buffer Fullness

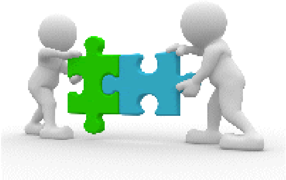
The bitrate control solution recognized as most efficient controls the MB quantization step, e.g. as a function of the output buffer fullness.

The granularity of this control is associated to the *31 quantization step values available;* the frequency of change may be at *MB level.*

Quantization step



Buffer fullness (%)



Quantization step control

Video sequence

Encoder

Output buffer

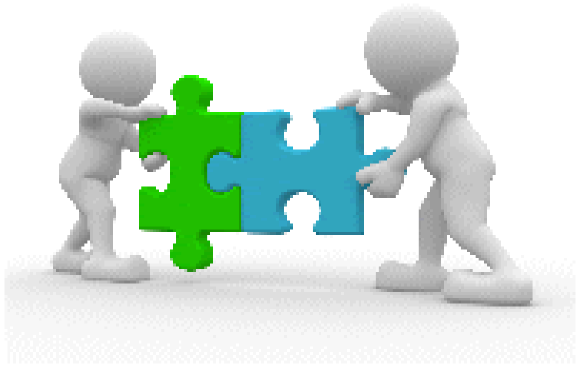
Coding rate fitting the channel rate

Variable encoder rate with an average below the channel rate

Constant channel rate

Matching the Encoder Variable Rate with the Channel Constant Rate

Variable content
Variable rate
Need to adapt

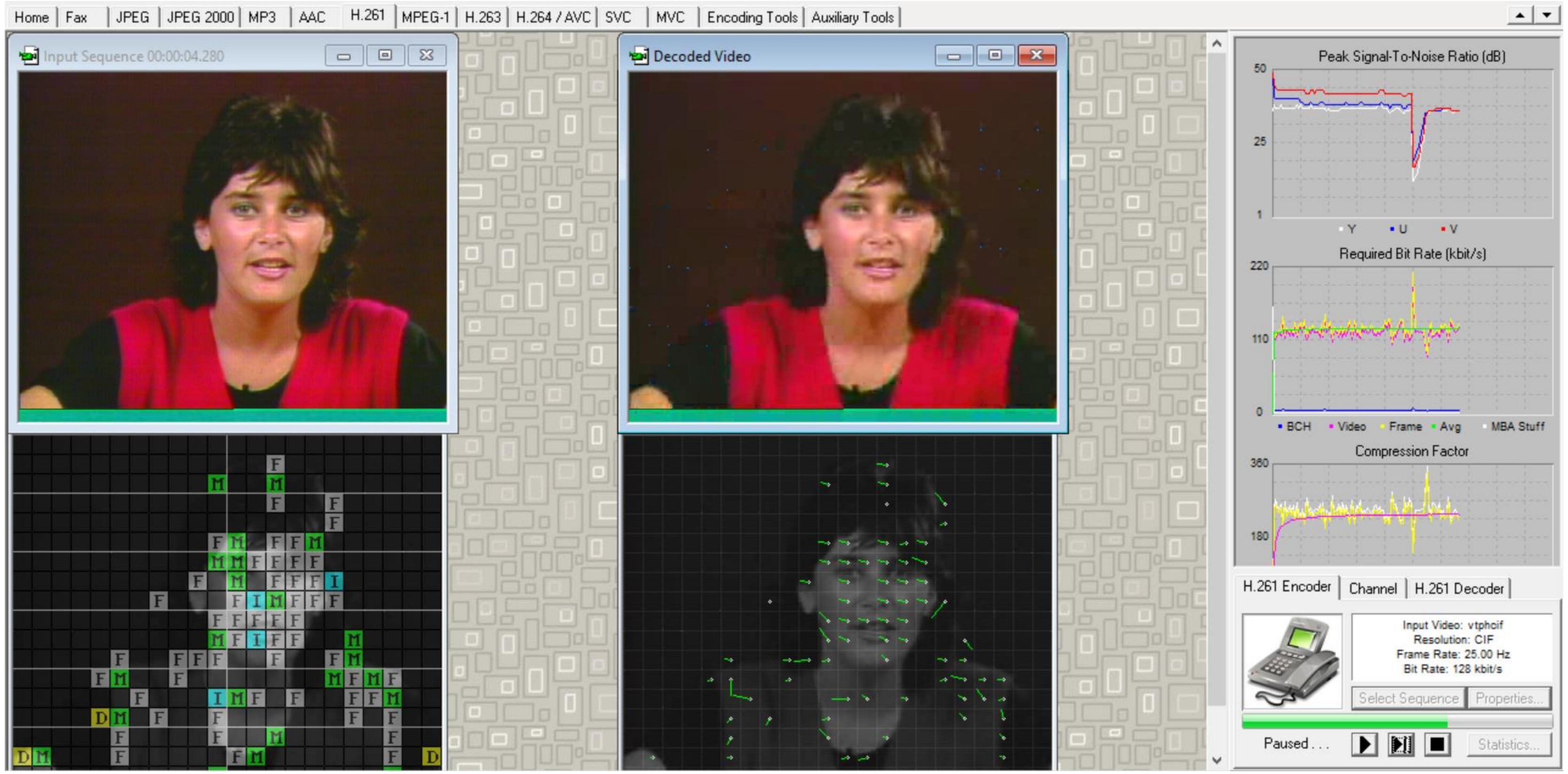


Constant rate channel
Uniform resources

The final video quality depends on:

- **The (constant) channel rate**
- **The (spatially and temporally varying) content**
- **The quantization selection algorithm**
- **The (limited) buffer size (for the critical moments)**

Let's Play and See …



I – Intra coding;
D – Inter coding without motion compensation (only error differences);
M – Inter coding with motion compensation;
F – Inter coding with motion compensation and the in-loop filter;
O - Overflow (indicating that the output buffer is too full and no bits may be sent).


LET'S
HAVE
FUN