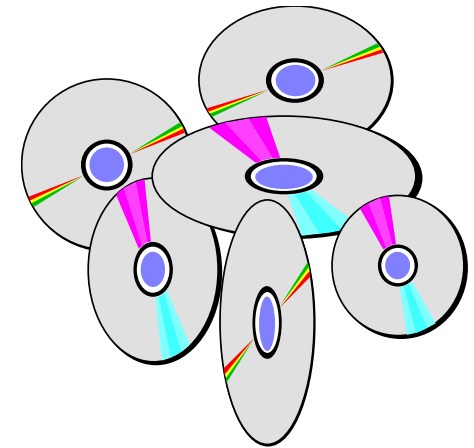
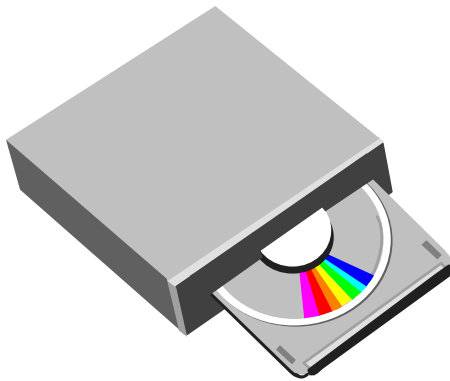


# DIGITAL VIDEO STORAGE



*Fernando Pereira*

*Instituto Superior Técnico*

# Digital Audio and Video Storage



**There are several technologies involved in digital audio and video storage, this means in the process of recording a physical support to store the audiovisual (AV) information at hand.**

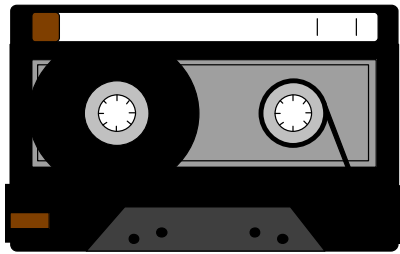
**One of the most important technologies for AV storage is audiovisual data coding which should provide the necessary compression efficiency and quality but also other storage functionalities such as random access, already provided in analogue recording.**



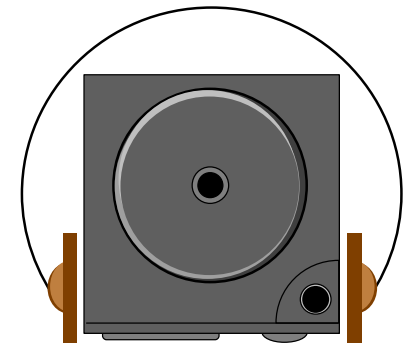
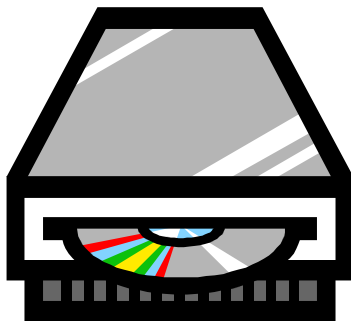
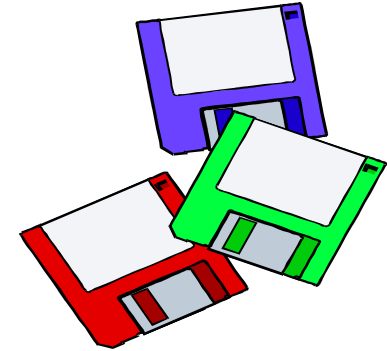
## Main Recording Functionalities ...

- **Normal video playback** – The usual play ...
- **Random access** – It shall be possible to access any part of the audiovisual data in a limited amount of time, e.g. 0.5 s.
- **Reverse playback** – Playing at regular speed, opposite to the usual temporal direction ...
- **Fast forward and Fast reverse** – Faster play (with time compression) in the usual and opposite time directions (more complex form of random access).
- **Edition** – Capability to edit the coded signal in a simple way.

# Main Storage Supports



- Magnetic tape
- Magnetic discs
- Optical discs
- ...





# Storage: Which Support ?

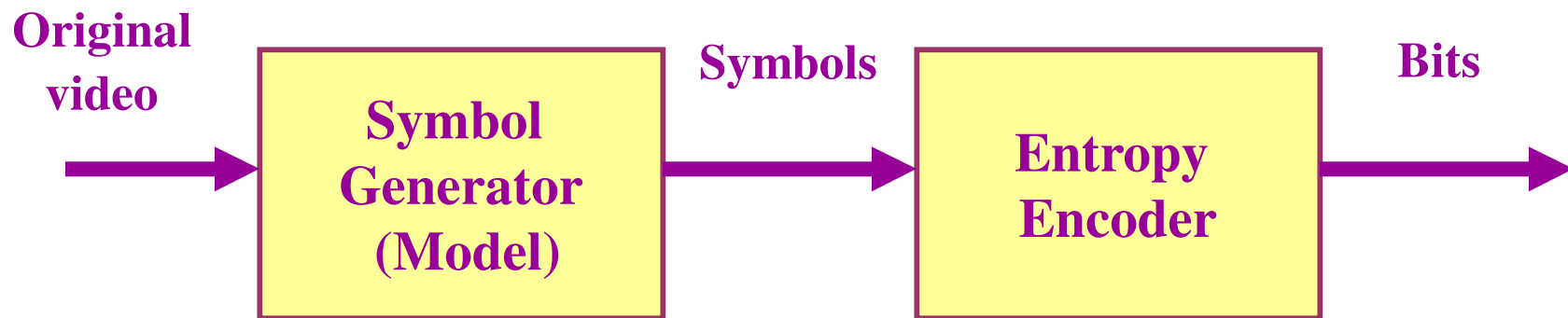
**Main factors to be taken into account to select an audiovisual storage support:**

- **Capacity (in MBytes)**
- **Reading speed (in Mbit/s)**
- **Time and form of access (e.g. sequential or random)**
- **Durability**
- **Mobility**
- **Cost**
- ...





# The Basic Coding Chain



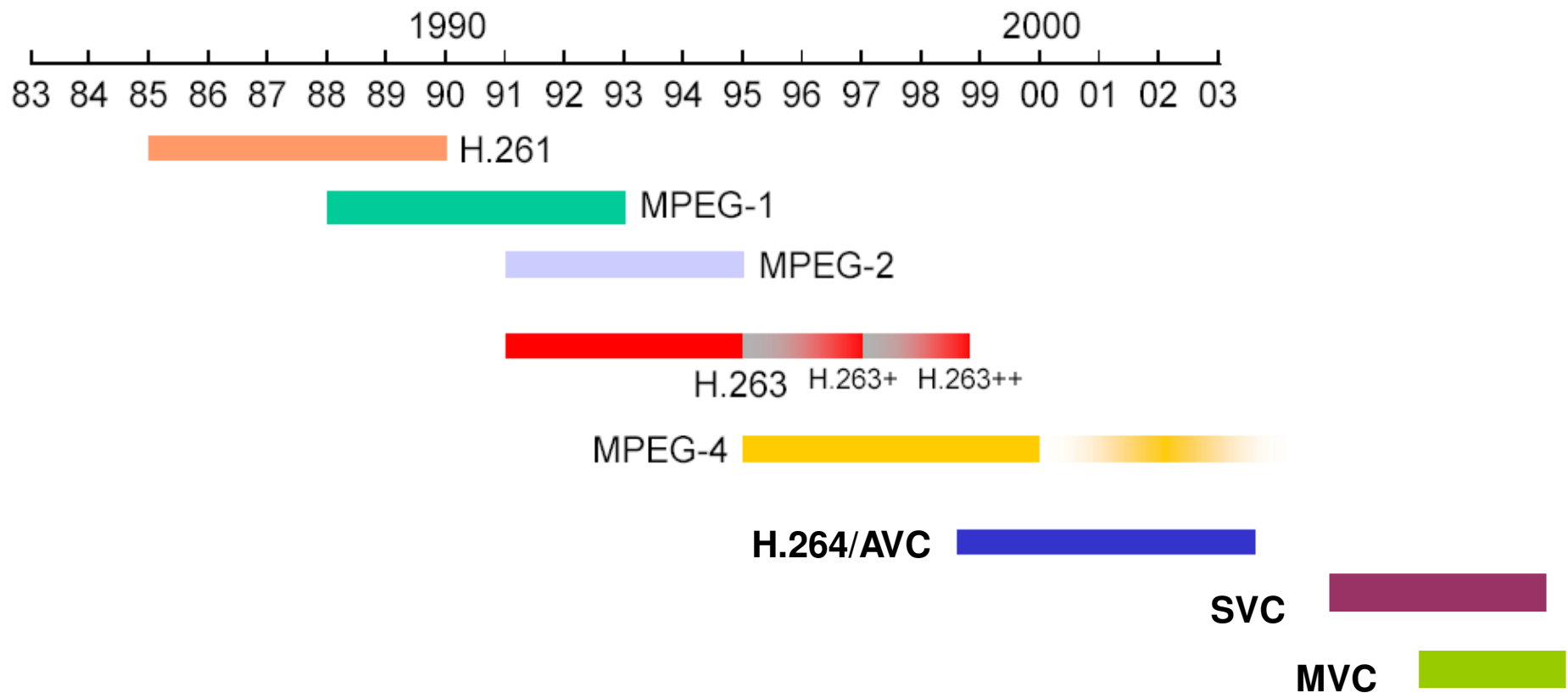
**The symbolic coding model depends on the type of data, e.g. audio, video, and on the application requirements, e.g. random access.**

# The MPEG Family for Source Coding

- **MPEG-1 (1988-1990):** Coding of video and associated audio for a target bitrate of 1.5 Mbit/s
  - CD Storage (initial target)
- **MPEG-2 (1990-1993):** Coding of video and associated audio (initially for bitrates up to 10 Mbit/s)
  - Digital TV (for any transmission channel) and DVD
- **MPEG-3 (X):** Coding of video and associated audio with bitrate up to 60 Mbit/s (finally not defined since MPEG-2 fulfils the needs)
  - High Definition TV (HDTV)
- **MPEG-4 (1994-2008):** Coding of video and associated audio (natural and synthetic) based on objects
  - Interactive multimedia applications and much more



# Standards Along the Years ...





# MPEG-1 Standard





# Motivation

- **The emergence of digital storage supports with large capacity and high reading speeds at increasingly lowers costs.**
- **The development of video coding algorithms reaching increasingly higher compression factors for a certain acceptable quality.**
- **The growing electronic integration capability of complex functions in reduced silicon areas (VLSI).**
- **The growing interaction between the telecommunications, computer and consumer electronics industries.**
- **The need to standardize in an area for which the technical development was ready to offer several *de facto* solutions, taking the opportunity to lower the costs and increase the production.**





## **MPEG-1: Storage Supports (~ 1990)**

- **CD-ROM (Compact-Disc Read Only Memory)**
  - Capacity between 600 MByte and 2 GByte (usually, 700 MB) with a reading speed of about 1.5 Mbit/s (and growing ...)
- **CD-WORM (CD-Write Once Read Many times)**
  - Reading speed of about 8 Mbit/s
- **Discos Winchester**
  - Capacity between 20 and 400 MByte (now above 1 GByte) with a reading speed of about 8 Mbit/s (now above 20 Mbit/s)
- **DAT (Digital Audio Tape)**
  - Reading speed of about 7.5 Mbit/s

# Storage: Which Support ?

**Main factors to be taken into account to select an audiovisual storage support:**

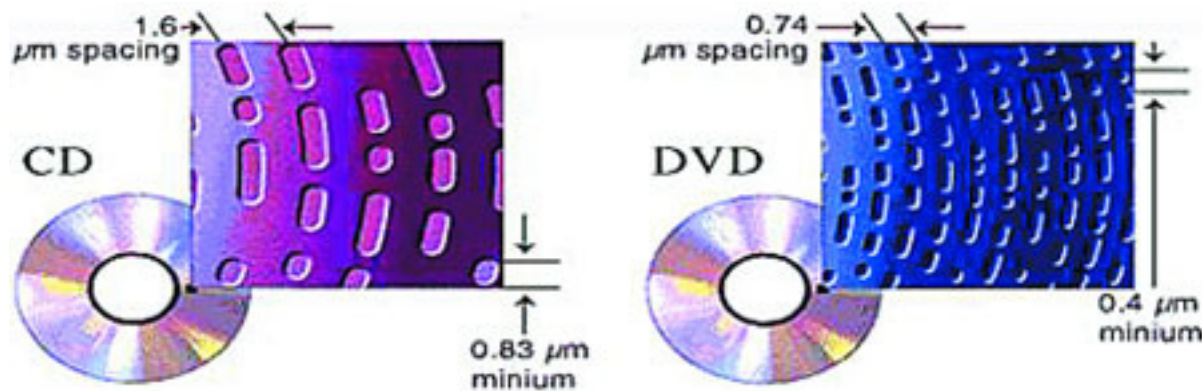
- **Capacity (in MBytes)**
- **Reading speed (in Mbit/s)**
- **Time and form of access (e.g. sequential or random)**
- **Durability**
- **Mobility**
- **Cost**
- **...**



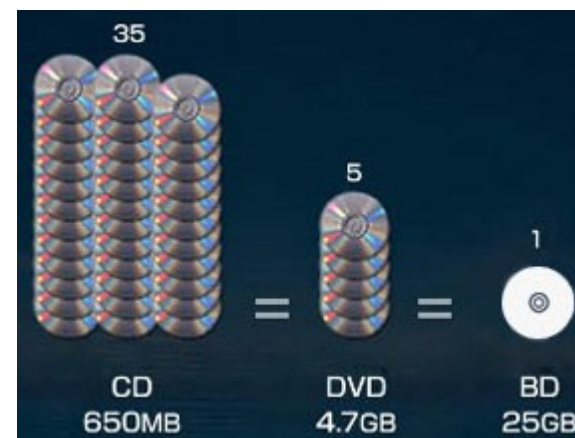
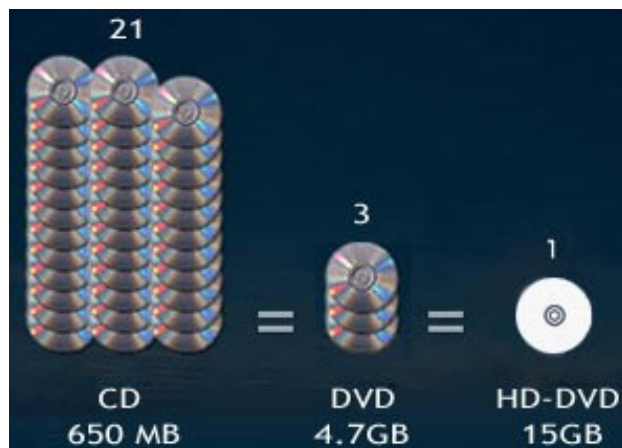
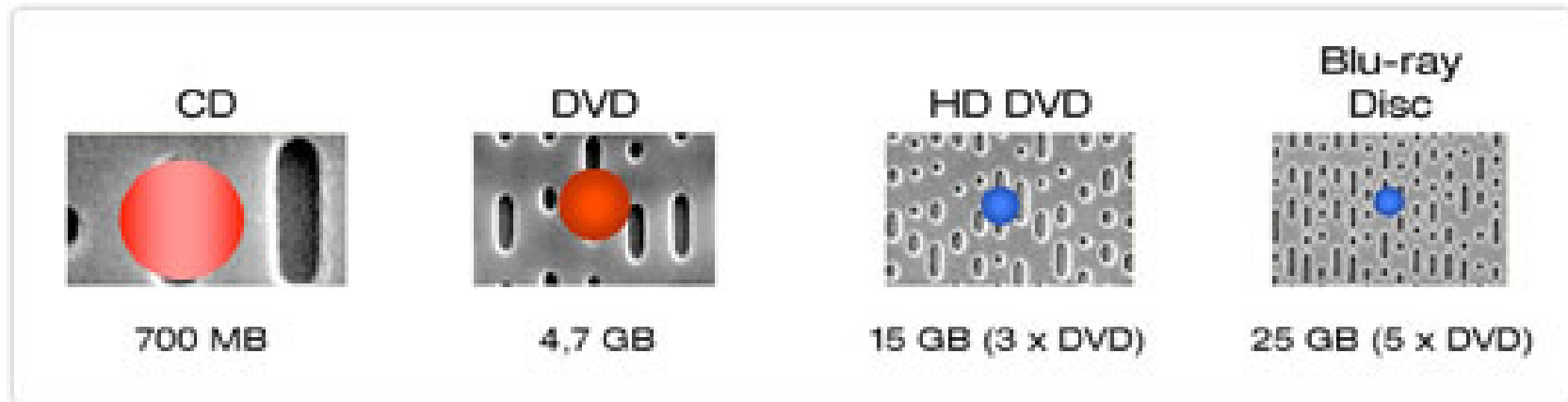
**The CD-ROM was selected as the most adequate storage support to offer, for the first time in large scale, interactive multimedia signals, mainly due to its large capacity and low cost.**

## Digital Versatile Disc (DVD): the Support Today !

- The DVD is essentially an evolution of the CD where it is possible to store more video, audio or any other type of digital data.
- There are 2 DVD sizes: 12 and 8 cm both 1.2 mm thick (two 0.6 mm glued substrates).
- DVDs may be single-sided or double-sided depending if only one or the two sides store data.
- Each side may have one or two storage layers.

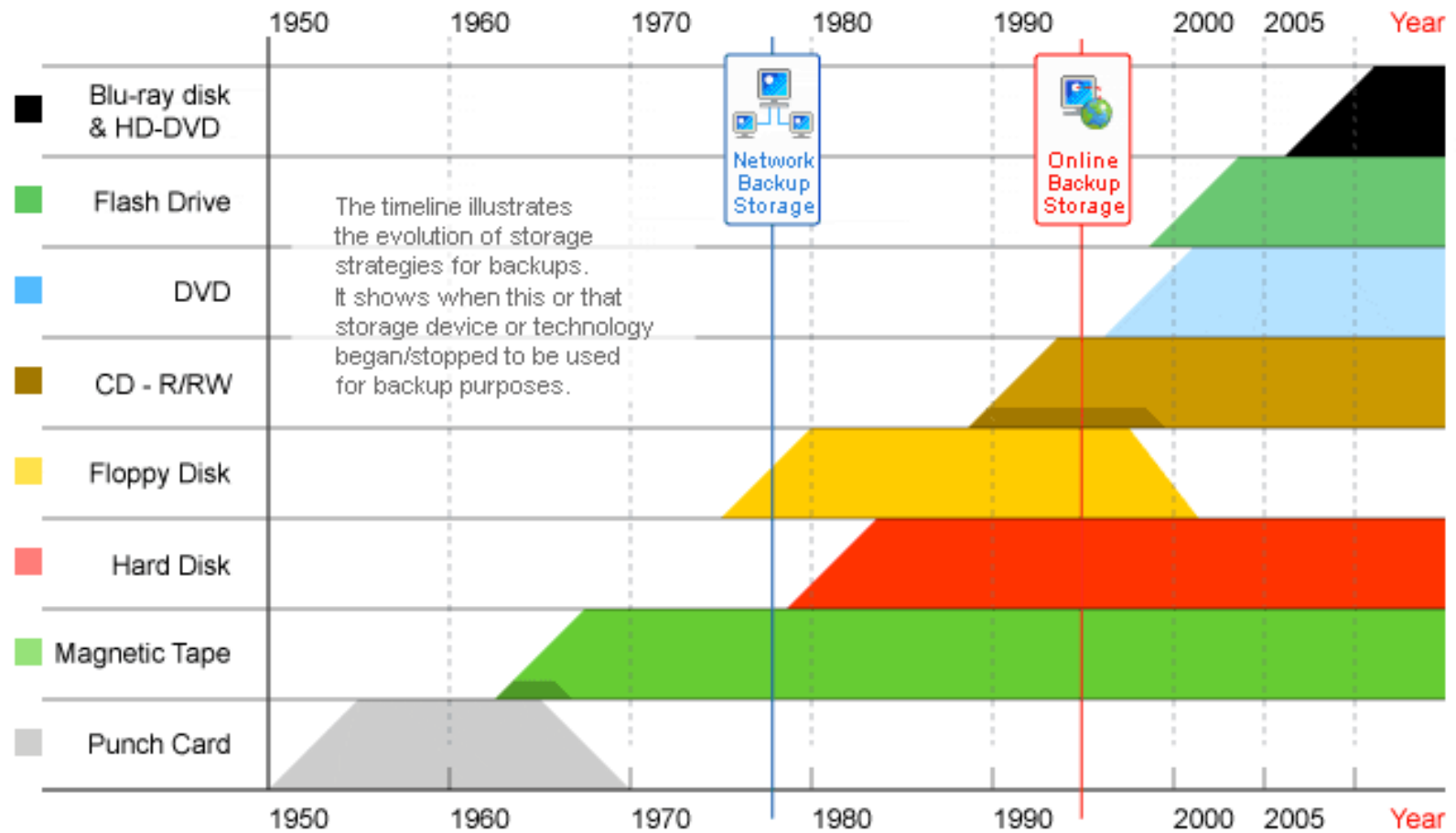


# Making Comparisons ...



# Storage along Times ...

**Timeline: Data Backup Storage**





# MPEG-1: Objectives

**Coding of video and associated audio with a total bitrate of about 1.5 Mbit/s with a minimum acceptable (subjective) quality.**

- **With MPEG-1, video signals just become another type of (digital) data that may be easily stored and processed, e.g. in a computer.**
- **MPEG-1 content in CD-ROM targeted ‘killing’ the (analogue) VHS business, dominant at that time.**
- **The video and associated audio information may be stored in any type of digital support or transmitted in any type of digital network.**

# MPEG-1: Example Applications

- **Asymmetric applications** – These applications involve the repeated usage of the decoding process after a single (or a limited number) of encodings

- Movies
- Games
- Education
- Tele-shopping
- Tourism

- **Symmetric applications** - These applications involve a similar usage of the encoding and decoding processes

- Videotelephony
- Videoconference
- Video-mail





# MPEG-1 Standard: Structure

## Part 1: Systems

Specifies the multiplexing of the several audio and video coded streams in a single stream with synchronization

## Part 2: Video

Specifies the video coding solution (bitstream and decoding) for bitrates of about 1.15 Mbit/s

## Part 3: Audio

Specifies the audio coding solution (bitstream and decoding) for bitrates of 32-448 kbit/s per channel (mono and stereo)

## Part 4: Conformance Testing

Specifies conformance tests for the streams and decoders

## Part 5: Reference Software

Software implementation of the parts 1, 2 and 3



# **MPEG-1 Standard**

## **Part 1: Systems**



# MPEG-1 Systems: Objectives

**The MPEG-1 Systems standard has the objective to combine one or more coded audio and video streams into a single binary stream, called MPEG-1 stream or ISO/IEC 11172 stream.**

**The MPEG-1 Systems standard defines:**

- **Syntax for the streams offering timing control**
- **Multiplexing and synchronization of the audio and video streams**

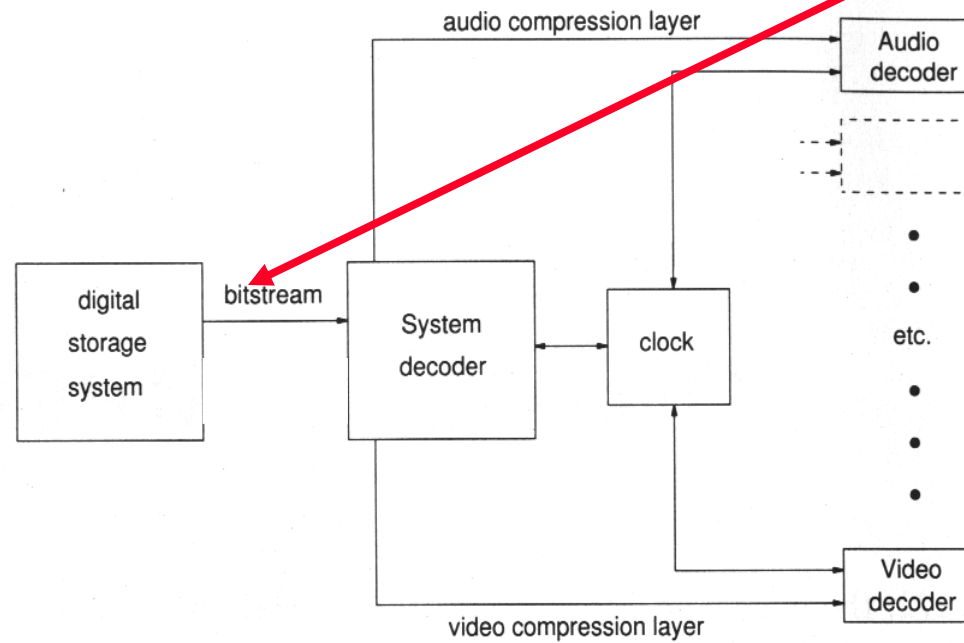
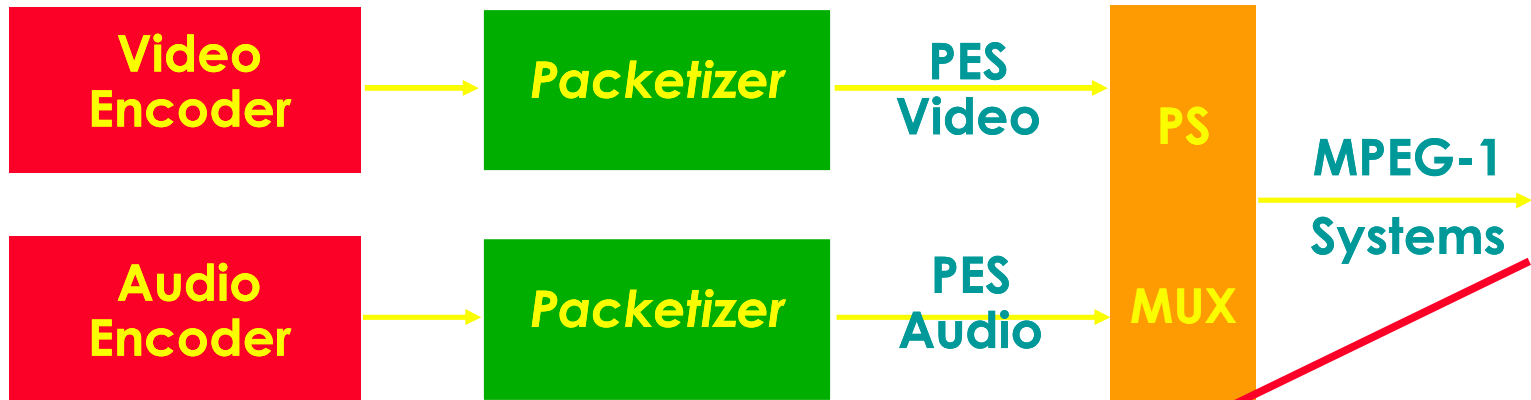


# MPEG-1 Streams

One MPEG-1 stream is formed by two layers:

- **SYSTEM** – Serves as envelope for the compression layers; offers the necessary information for the demultiplexing and timing of the compression layers.
- **COMPRESSION** – Includes the coded data that will be given to the audio and video decoders.

The elementary (coded) audio and video streams are divided into variable size packets – the packets – creating the so called *Packetized Elementary Streams (PESs)*.





# Packs and Packets

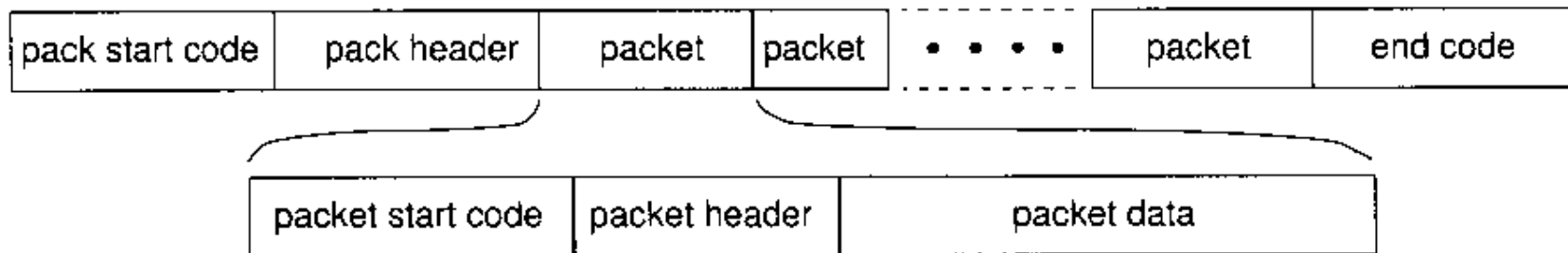
The operations to be performed by the Systems decoder regard the full MPEG-1 stream - *multiplex-wide*- or elementary streams, e.g. audio or video - *stream-specific*.

The MPEG-1 Systems stream is structured in two sub-layers:

- ***PACK sub-layer*** – Refers to the multiplex-wide operations such as the control of the reading of the stream from the storage support, if possible, the adjustment of the clocks, buffer management, and the definition of the resources needed for decoding.
- ***PACKET sub-layer*** – Refers to the stream-specific operations such as demultiplexing and synchronization of the various elementary streams; packets may have a fixed or variable length.

One *pack* corresponds to a collection of *packets* with additional *multiplex-wide* information.

# MPEG-1 Systems Stream Syntax



- **One MPEG-1 Systems stream consists in a sequence of packs, each one containing several packets (with coded audio OR video); one video (or audio) packet may start at any byte of the video (or audio) stream and may have a variable length.**
- **The Systems decoder parses the MPEG-1 stream, giving to the audio and video decoders their respective packets, after inspecting the *packet start codes*.**
- **At most, 32 audio streams, 16 video streams and 2 data streams may be multiplexed in a single MPEG-1 stream.**



# MPEG-1 Systems: Synchronization

MPEG-1 Systems synchronization relies on two basic elements:

- ***Systems Target Decoder (STD)*** – Hypothetical reference model used to define the ideal decoding process; in this ideal model, the transference, decoding and presentation of the information are instantaneous; in a real system, some delay is inevitable and thus should be accounted for.
- ***Master Time Base (MTB)*** – Reference timing information used to synchronize the presentation of the various elementary streams; it may correspond to the clock of one of the elementary decoders, the Digital Storage Media clock or to an external clock - *Systems Clock Reference (SCR)* derived from a *Systems Time Clock (STC)*.



# MPEG-1 Systems: Timing

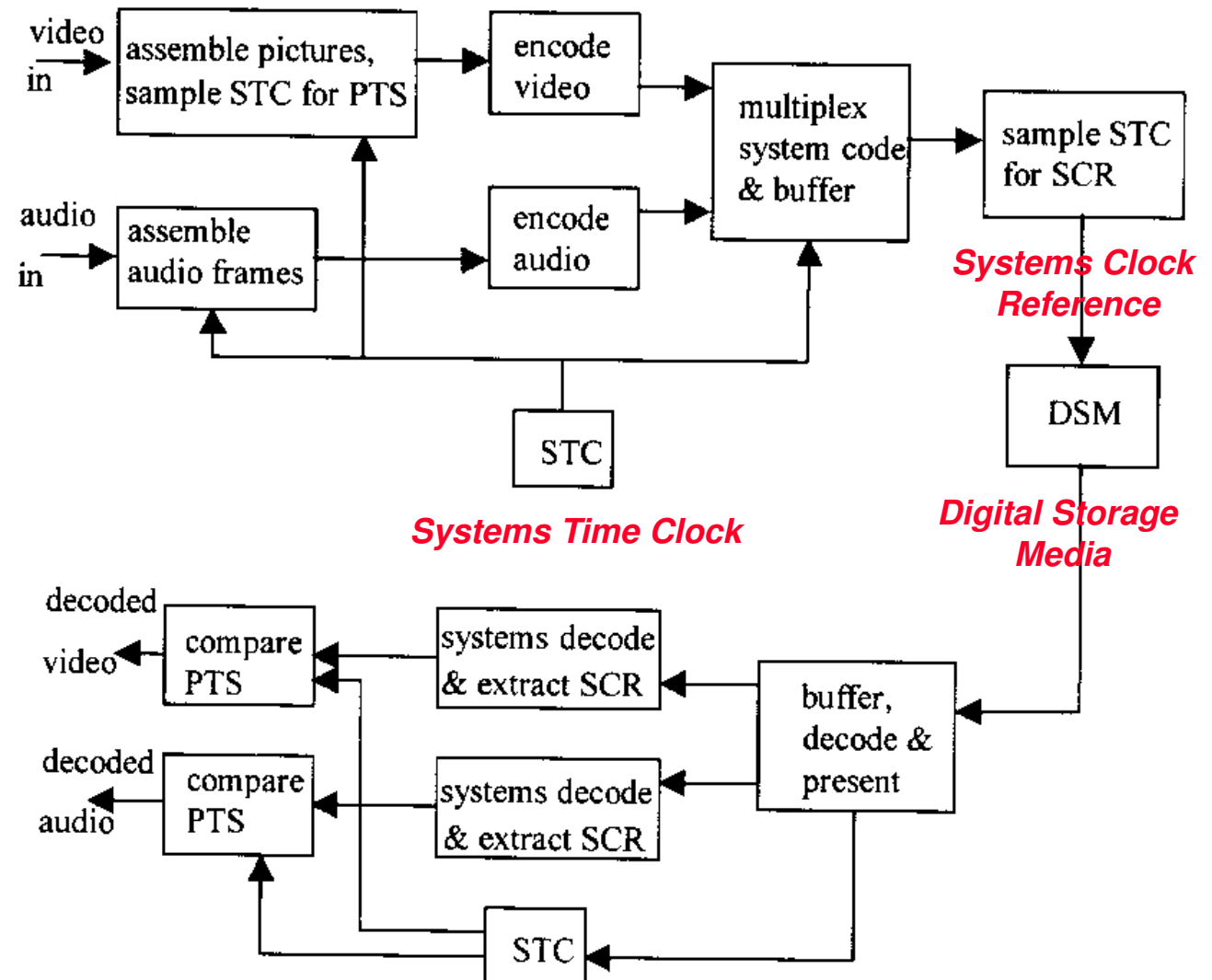


- ***Decoding Time Stamp (DTS)*** – Timing information that may be present in the packet header to indicate the moment when the corresponding coded information must be decoded in the *Systems Target Decoder (STD)*.
- ***Presentation Time Stamp (PTS)*** – Timing information that may be present in the packet header to indicate the moment when the corresponding decoded information must be presented in the *Systems Target Decoder (STD)*.

**MPEG-1 players use PTS to control the presentation of the decoded information regarding the reference clock.**

**PTS and DTS are different when the decoding and presentation orders are not the same, such as when using B video frames; the STD assumes instantaneous decoding.**

# MPEG-1 Chain Architecture





# **MPEG-1 Standard**

## **Part 2: Video**



# MPEG-1 Video: Requirements

- **Normal video playback** – The usual play ...
- **Random access** – It shall be possible to access any part of the audiovisual data in a limited amount of time, e.g. 0.5 s.
- **Reverse playback** – Playing at regular speed against the usual temporal direction ...
- **Fast forward and Fast reverse** – Faster play (with time compression) in the usual and opposite time directions (more complex form of random access).
- **Edition** – Capability to edit the coded signal in a simple way.
- **Audiovisual synchronization** – Need to guarantee synchronization between audio and video information.
- **Error resilience** – Need to provide some robustness to residual errors.
- **Total delay** – Depends on the applications and may be used to trade-off with quality.
- **Format flexibility** – E.g., it should be possible to use different spatial and temporal resolutions.
- **Cost** – Especially the decoders must have an acceptable (low) cost.

## MPEG-1 Video: Objective

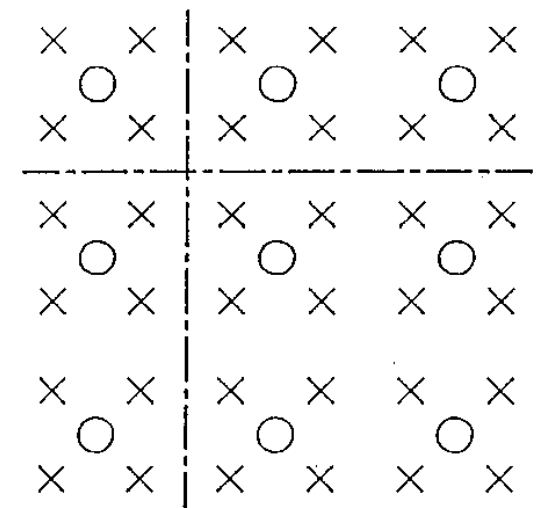


**Efficient coding of video information with a minimum acceptable quality with bitrates up to about 1.2 Mbit/s (only video); other rates may also be used.**

**The target quality for CD-ROM storage is the quality associated with VHS tapes, targeting the substitution of the popular analogue storage with digital storage.**

# MPEG-1 Video: Signals to Code

- **Signals** - The signals for each image are the luminance (Y) and two chrominances, designated as  $C_B$  and  $C_R$  or U and V. The R,G,B primary signals have a gamma correction around 2.2 - 2.8.
- **Resolution** - Luminance as twice the number of rows and columns of the chrominance; this means a 4:2:0 subsampling format is used considering the lower human visual sensibility to colour.
- **Bit depth** - Samples are quantized according to Recommendation ITU-R BT.601 this means with 8 bit/sample.



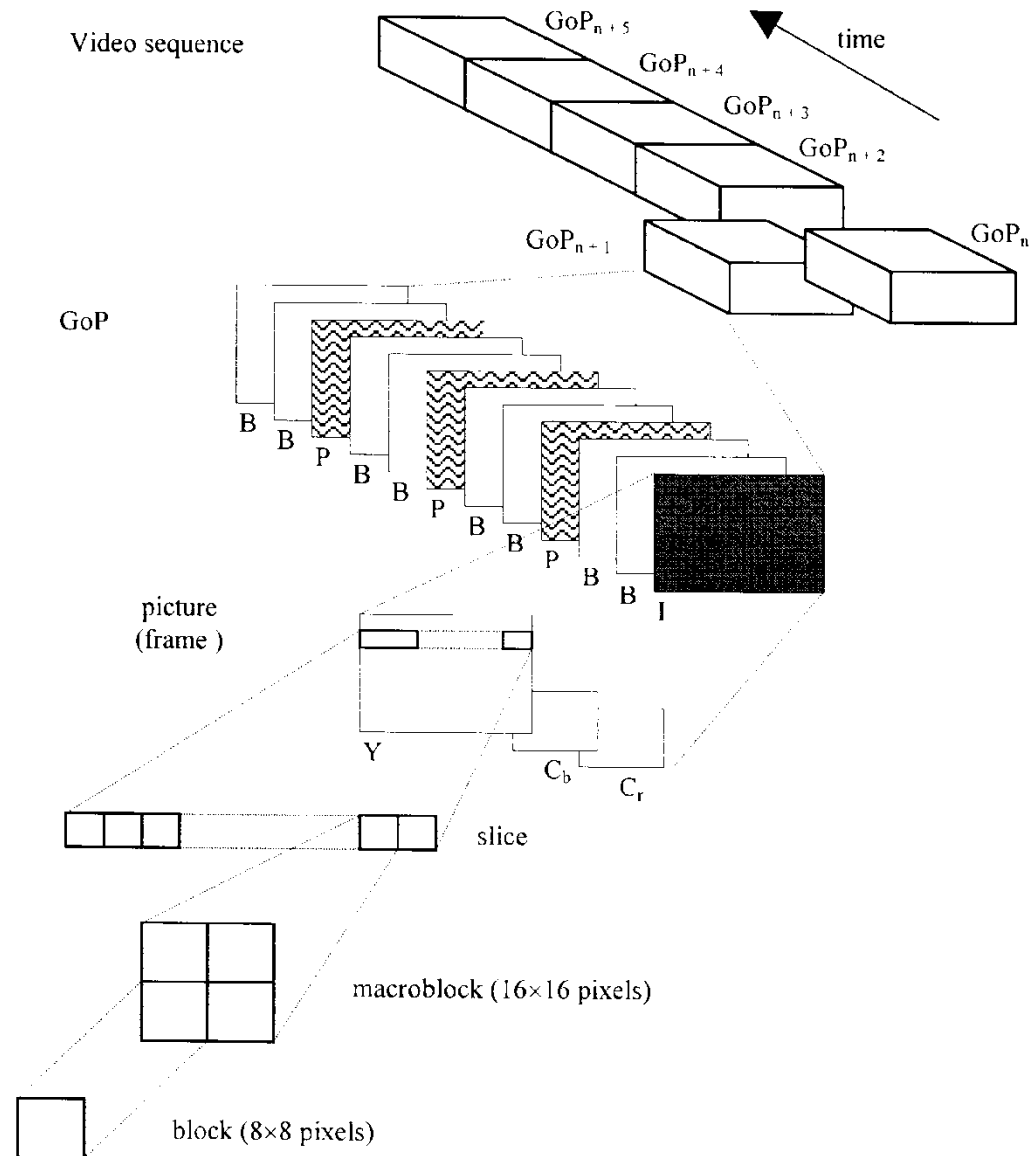
T1500340-86

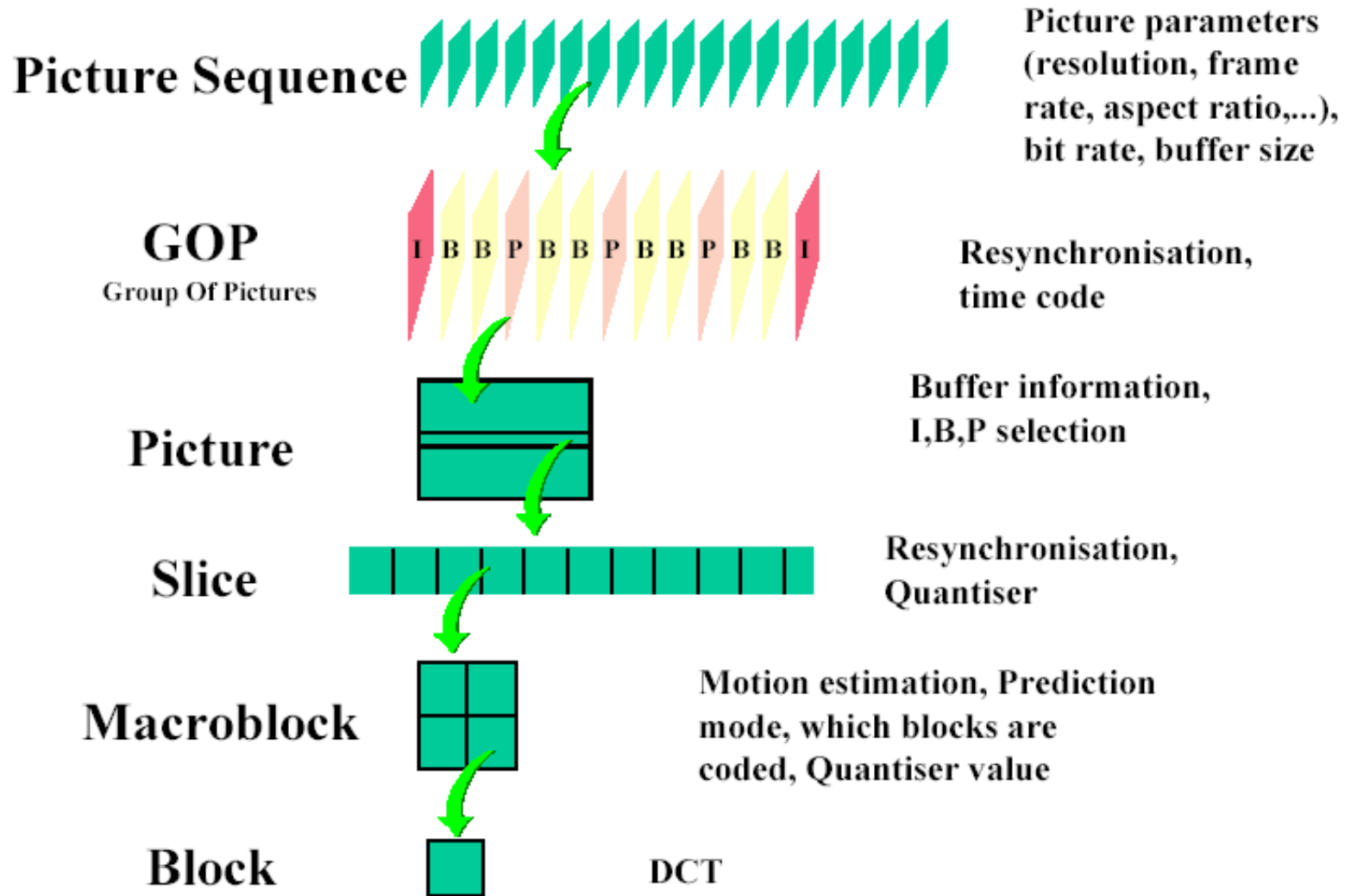
× Luminance sample  
 ○ Chrominance sample  
 - - - Block edge

# Video Structure

Spatially, the video data is organized in a hierarchical structure with 5 layers:

- Sequence
- Group of Pictures (GOP)
- Picture
- Slice
- Macroblock (MB)
- Block







# MPEG-1 Video: Coding Tools

**LOSSLESS**

- **Temporal Redundancy**

Predictive coding: temporal differences and motion compensation (uni and bidirectional;  $\frac{1}{2}$  pixel accuracy)

- **Spatial Redundancy**

Discrete Cosine Transform (DCT)

- **Statistical Redundancy**

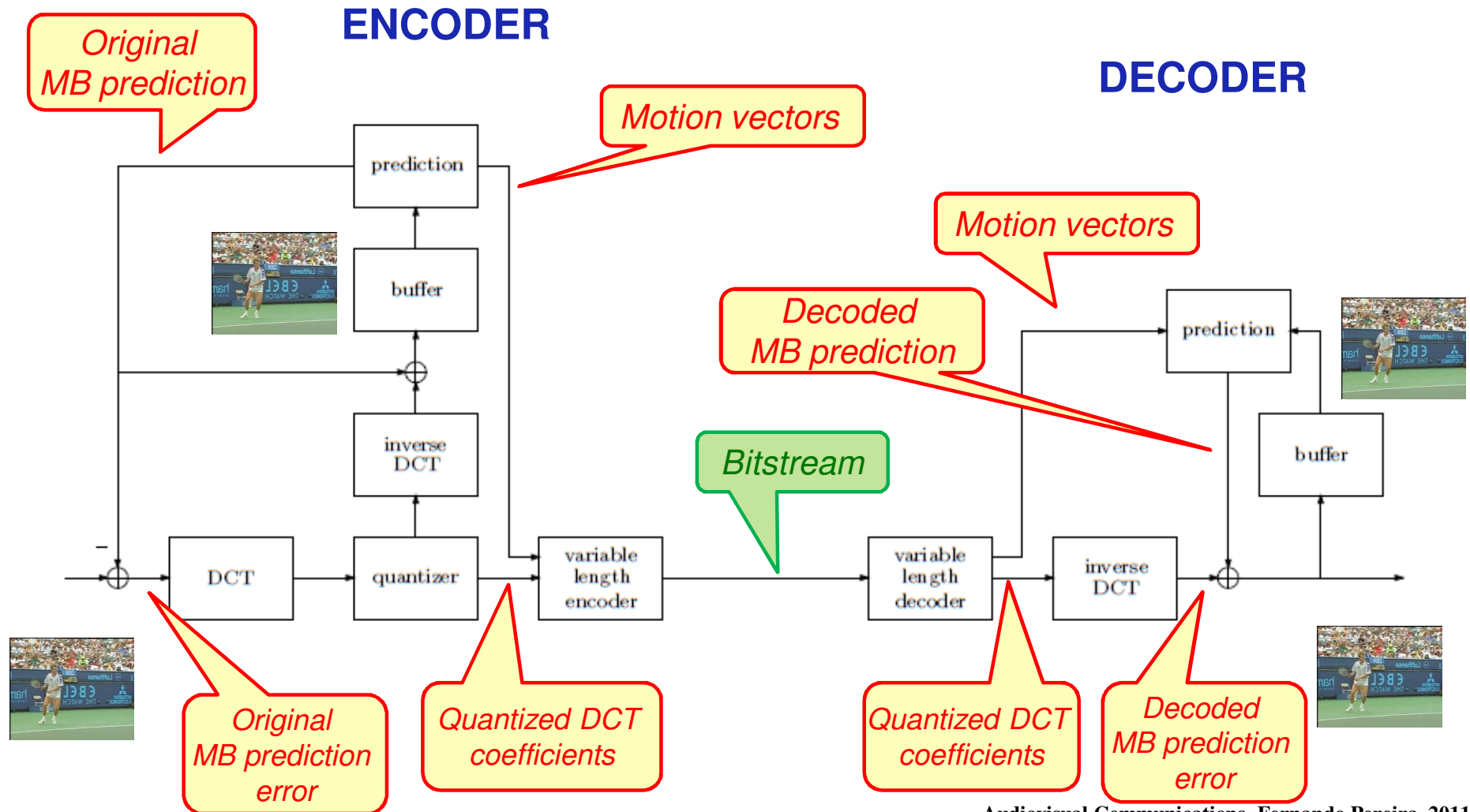
Huffman entropy coding

- **Irrelevancy**

DCT coefficients quantization

**LOSSY**

# Starting with the same Architecture ... Buying Quality with Computation, Memory and Delay ...

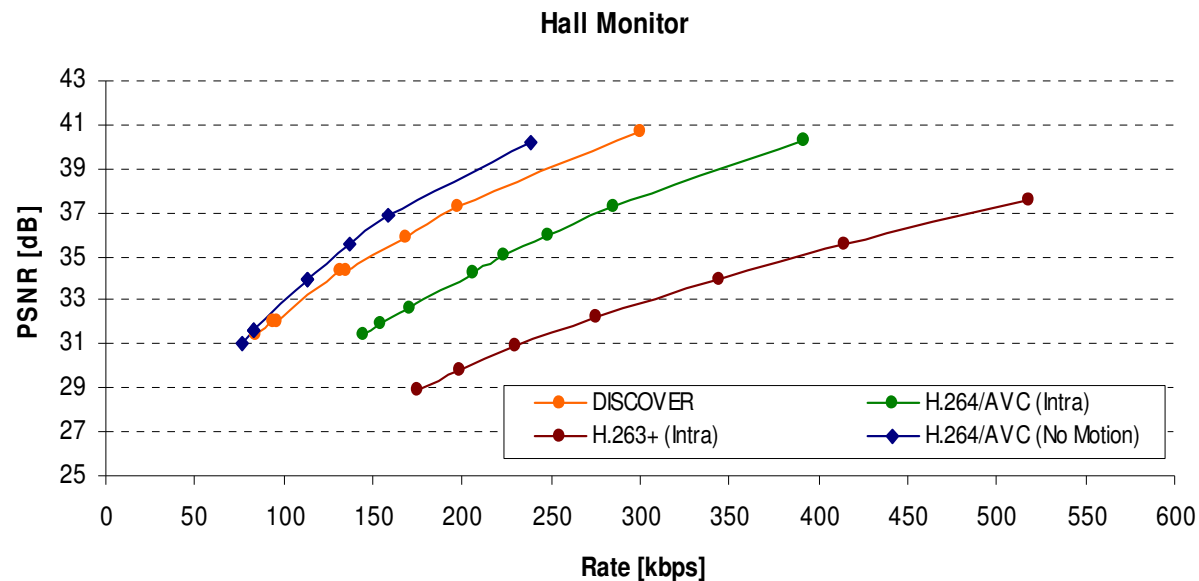




# Exploiting the Temporal Redundancy

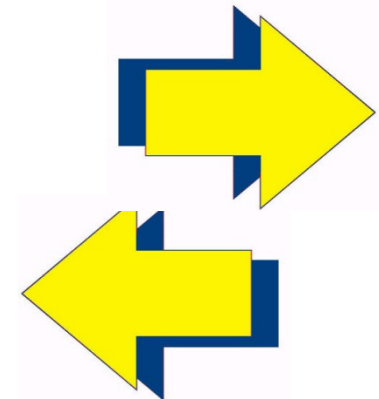
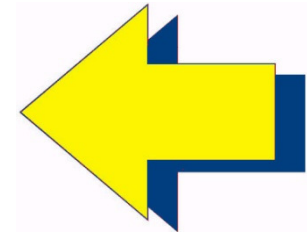
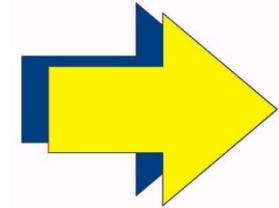
# The Key to Better RD Performance: Better Temporal Predictions ...

As for H.261, the quality of the temporal prediction strongly determines the video codec rate-distortion (RD) performance since it defines the energy of the difference signal, this means the prediction error.



# Better Predictions with more Motion Compensation

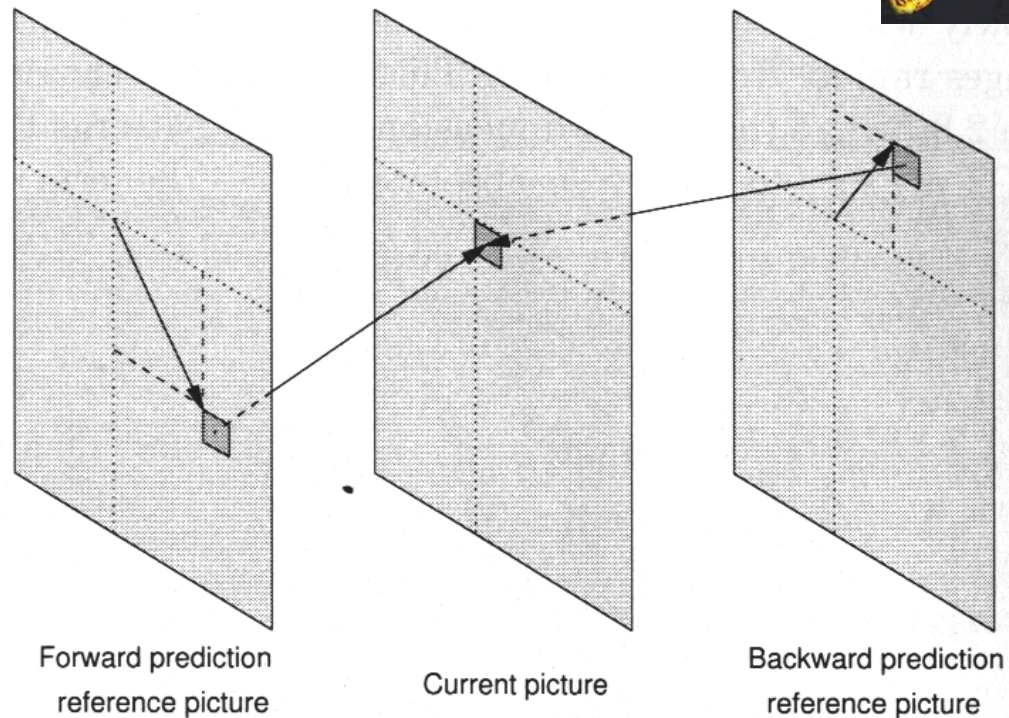
- **FORWARD PREDICTION** – It is based on the principle that, locally, each image (or part thereof) may be represented from one or more previous images after a translation.
- **BACKWARD PREDICTION** – It is based on the principle that, locally, each image (or part thereof) may be represented from one or more future images after a translation.
- **BIDIRECCIONAL PREDICTION** – It is based on the principle that, locally, each image (or part thereof) may be represented from a previous image (forward prediction), a future image (backward prediction), or a combination thereof, after corresponding translations.



## The Future is so Close ...



- Bidirectional predictions ‘buy’ quality with delay and complexity !
- This is possible especially if the application can accept the additional delay ...



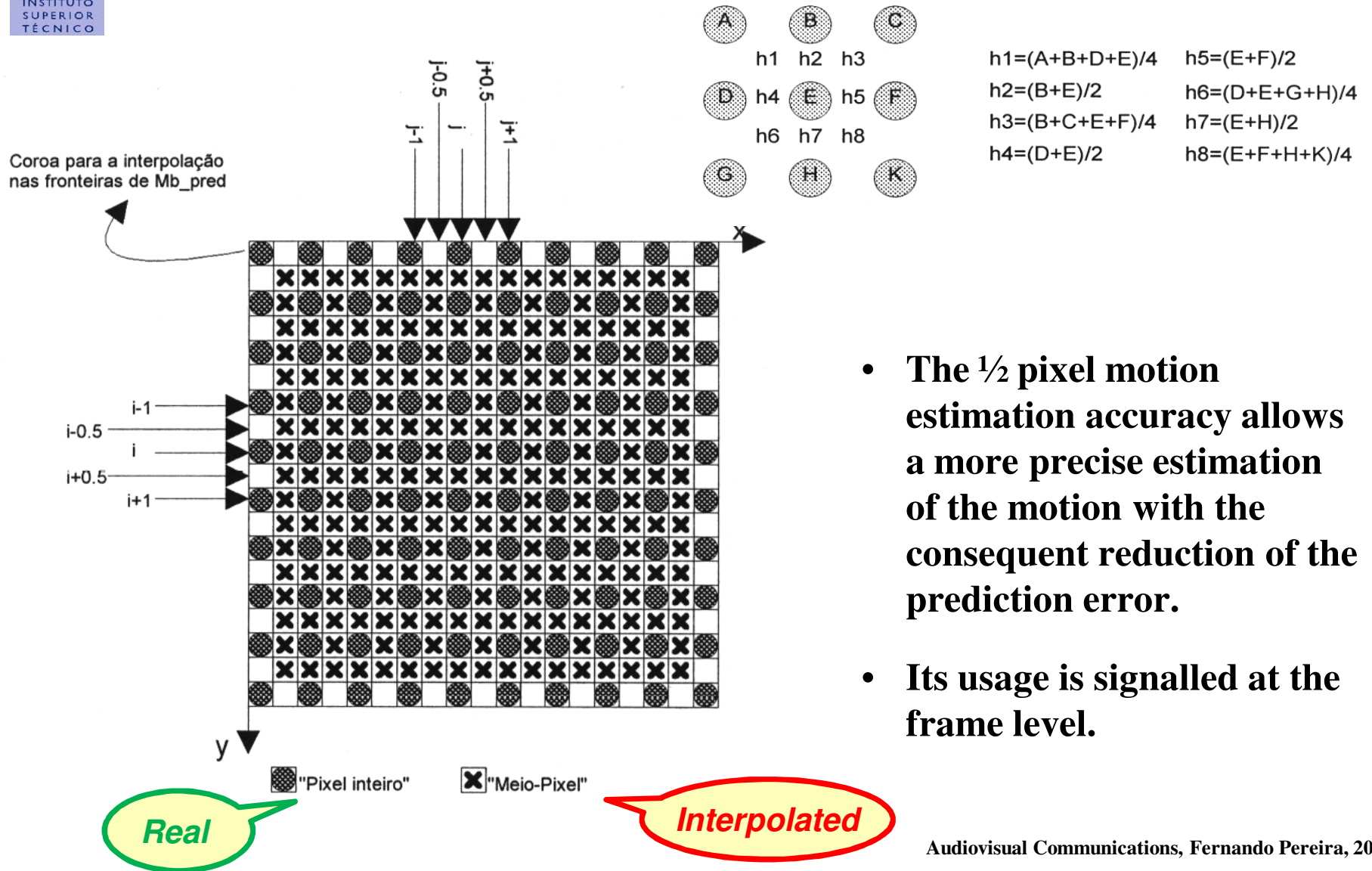
**Bidirectional prediction allows to reach better predictions (there is more information available to make the prediction) and, thus, to reach better RD performance in certain conditions; for example, it is useful to deal with uncovered backgrounds.**



# MPEG-1 Video: Motion Estimation

- **Spatial support** - Motion estimation and compensation is performed at the macroblock level.
- **Optional** - Motion estimation and compensation are always optional, meaning that the encoder may decide to use it or not (independently of this being a good or bad decision).
- **Non-normative** - Motion estimation is performed at the encoder and, thus, it is not normative ! There are hundreds of ways of doing motion estimation !
- **High complexity** - Motion estimation implies a high complexity, thus justifying the need for fast (non full search) motion estimation algorithms.
- **Block matching** - Since the bitstream syntax allows sending up to two motion vectors per MB, block matching motion estimation is the most used solution.
- **Prediction 'breaks'** - Bidirectional prediction cannot be applied to all frames of a sequence due to delay constraints; thus, there is a need to define relatively close prediction anchors (which do not predict from the future).

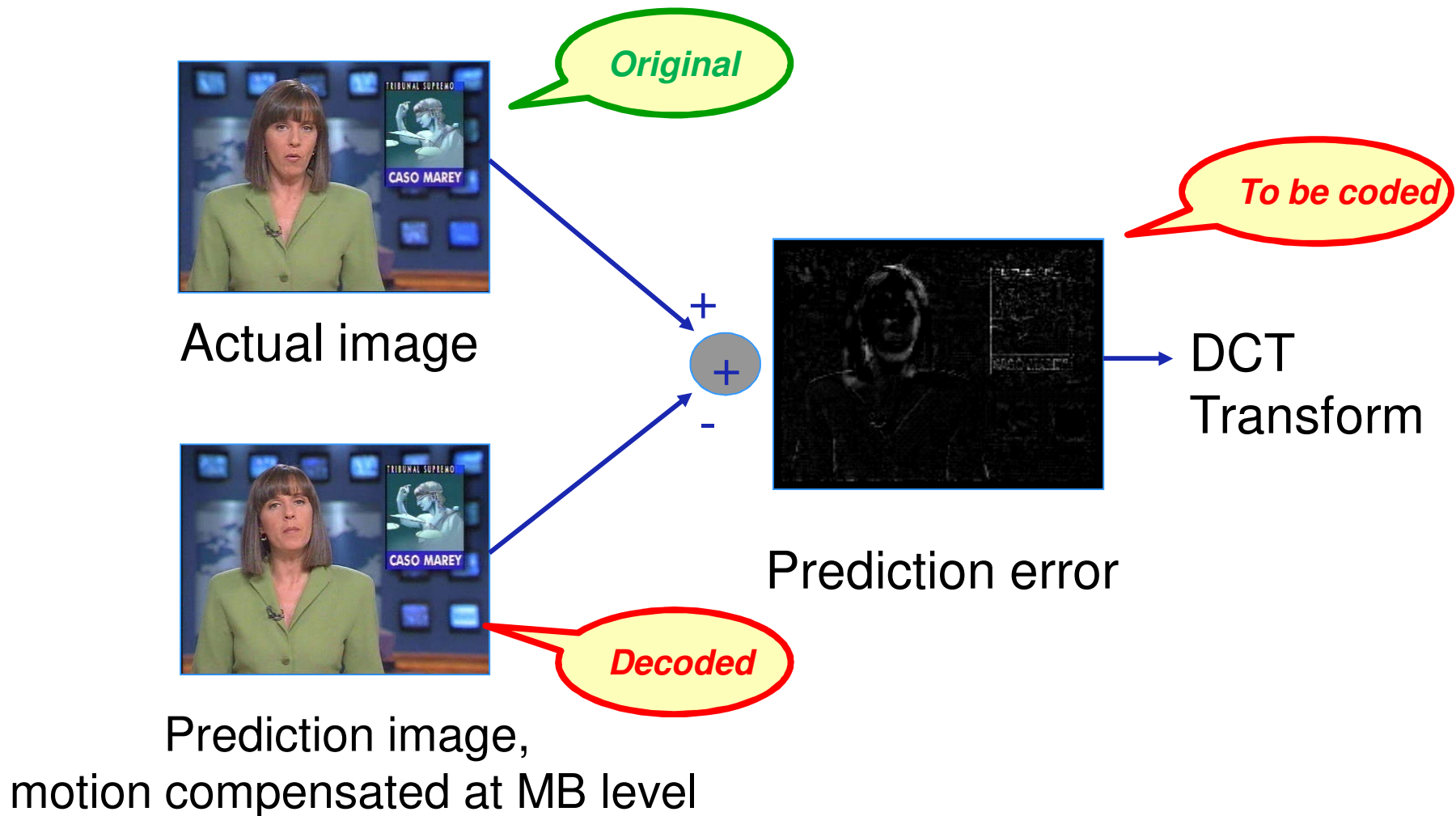
# 1/2 Pixel Motion Estimation and Compensation





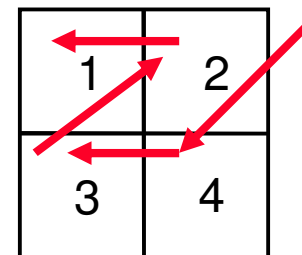
# **Exploiting the Spatial Redundancy and the Irrelevancy**

## After Time, the Space ...



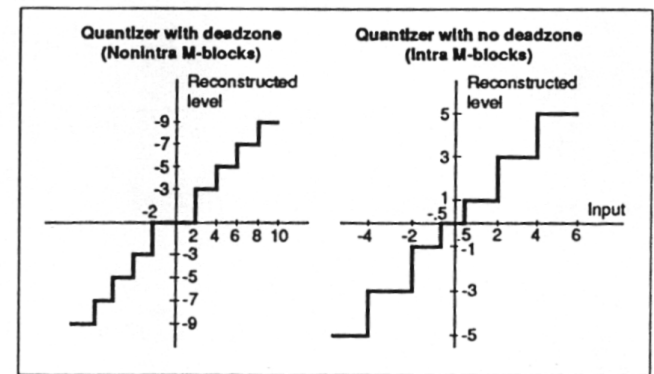
## MPEG-1 Video: How is the DCT Applied ...

- The DCT is applied to  $8 \times 8$  blocks of samples ( $N=8$ ).
- The inverse DCT (IDCT) precision is controlled as in H.261, e.g. the mismatch pixel error has to be always lower or equal to 1.
- The DCT coefficients to transmit are selected using non-normative thresholds, allowing the consideration of psychovisual criteria to optimize the final subjective impact for different types of content and applications.
- The quantized DCT coefficients in each block are zig-zag scanned to assure that they are transmitted according to their (decreasing) subjective relevance.
- The DC coefficients are differentially coded within each MB and between neighbour MBs (left to right and top-down).



# MPEG-1 Video: Quantization

- **MPEG-1 Video assumes the usage of uniform quantization with dead zone for the Inter MBs and without dead zone for the Intra MBs.**
- **The quantization step is determined through the quantization matrix and the quantization factor; it may be different for each DCT coefficient.**
- **The quantization steps may be changed at any MB.**
- **The standardized quantization matrix is different for Intra and Inter coded MBs. These matrices may be changed to more adequate matrices for the cases at hand, of course paying the necessary bitrate cost.**
- **For Intra coded MBs, the DC coefficient is always quantized with step 8.**





# Default Quantization Matrices

8	16	19	22	26	27	29	34	16	16	16	16	16	16	16	16	16
16	16	22	24	27	29	34	37	16	16	16	16	16	16	16	16	16
19	22	26	27	29	34	34	38	16	16	16	16	16	16	16	16	16
22	22	26	27	29	34	37	40	16	16	16	16	16	16	16	16	16
22	26	27	29	32	35	40	48	16	16	16	16	16	16	16	16	16
26	27	29	32	35	40	48	58	16	16	16	16	16	16	16	16	16
26	27	29	34	38	46	56	69	16	16	16	16	16	16	16	16	16
27	29	35	38	46	56	69	83	16	16	16	16	16	16	16	16	16

INTRA

INTER

**For Inter coding, the high frequency coefficients are not necessarily associated to high frequency image content since they can result from block artifacts in the reference image, poor motion compensation, or camera noise.**



# Exploiting the Statistical Redundancy



# MPEG-1 Video: Huffman Coding

- **Entropy coding is lossless !**
- **To any generated symbol, a codeword is attributed which length (in bits) is ‘inversely proportional’ to its probability. The more likely, the less bits since the more expectable ...**
- **The usage of variable length codes implies the need to use a output buffer to smooth the (asynchronous) data flow, notably if the output channel is synchronous (with a constant bitrate).**
- **The compression efficiency increase is obtained at the cost of a higher sensibility to the transmission errors, this means a lower error resilience.**
- **MPEG-1 Video uses Huffman codes for:**
  - **Differential motion vectors**
  - **DCT coefficients - (*run, level*) pairs**
  - **MB classes**
  - **MB addressing**
  - **...**



# Combining the Tools ...



## The Functional Sandwich ...

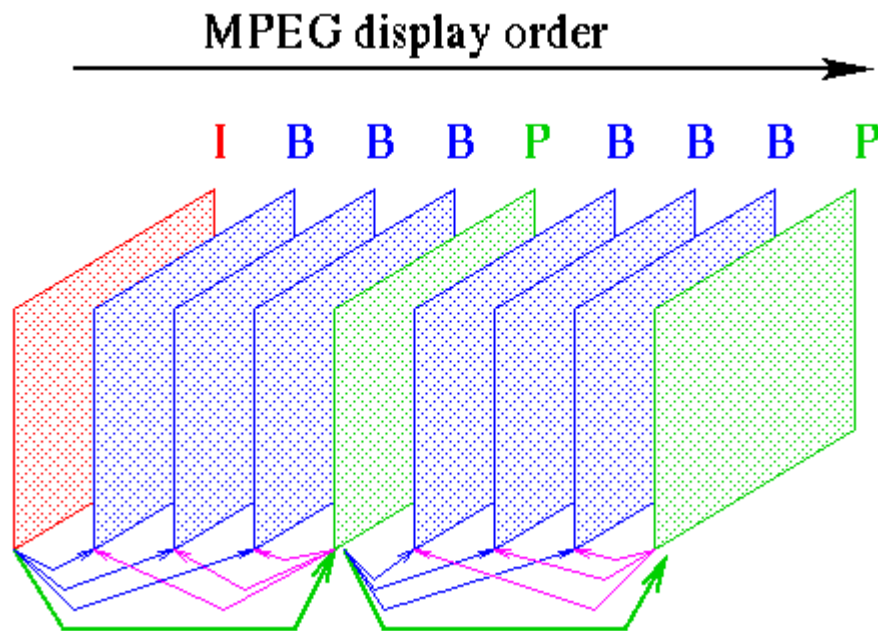
**Compression Efficiency**



**The CODEC !**

**Random Access**

# Temporal Prediction Structure



*Open GOPs use reference pictures from the previous GOP at the current GOP boundary.  
There is a flag to signal open and closed GOPs !*

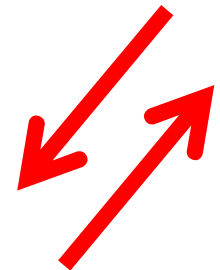
The “conflict” between coding efficiency and random access led to the definition of 3 frame types depending on the coding tools used:

- **Random access: Intra frames (I) – Don’t use temporal predictions**
- **Compression efficiency:**
  - **Predicted frames (P) – May only use *forward* prediction from previous I/P frame**
  - **Bidirectionally predicted frames (B) – May use both forward and backward prediction from previous and future I/P frame**

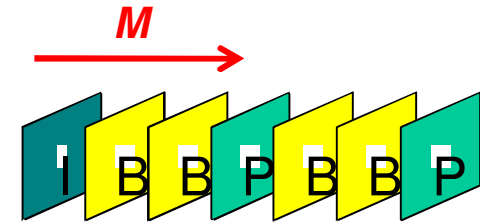
# The Frame Level Syntactical Restrictions

- **I (INTRA CODED) FRAMES** – All MBs in I frames are Intra codec blocks; no temporal predictions are allowed at all and, thus, no temporal redundancy is exploited making these frames rather expensive in rate to achieve a target quality.
- **P FRAMES** – MBs in P frames MAY use backward prediction, this means a prediction from a past frame, with or without a motion vector; for MBs in P frames, only the previous P/I frame may be used as forward prediction.
- **B FRAMES** – MBs in B frames MAY use backward, forward or bidirectional prediction (average prediction from a past and a future frame), with or without motion vector(s); for MBs in B frames, the previous P/I and next P/I frames may be used as forward and backward predictions.

**Motion**



# Temporal Prediction Structure



- The temporal prediction structure is rather flexible and may depend on the content or application.
- A good solution is to insert temporal anchors (I and P frames) about every 0.1 s using a combination like

... I B B P B B P B B P B B I B B P B ...

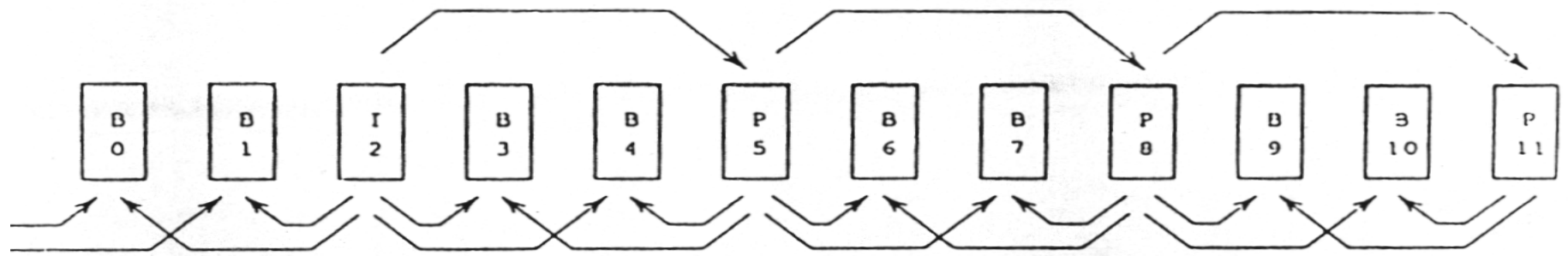
- If a regular prediction structure is used
  - $N$  (GOP size) – number of frames between two I frames + 1
  - $M$  – number of B frames between two anchors (I or P frames) + 1

this means that  $N$  is always a multiple of  $M$ ;  $M$  and  $N$  are not syntactic elements in the MPEG-1 Video bitstream.

# The Order is ... Out of Order ...

Since B frame decoding may only be made after receiving and decoding the corresponding anchor frames, the transmission of I and P frames out of the natural acquisition and visualization orders is inevitable !

This introduces an additional delay ...



I B B P B B P B B P B B I B B P B B P  
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

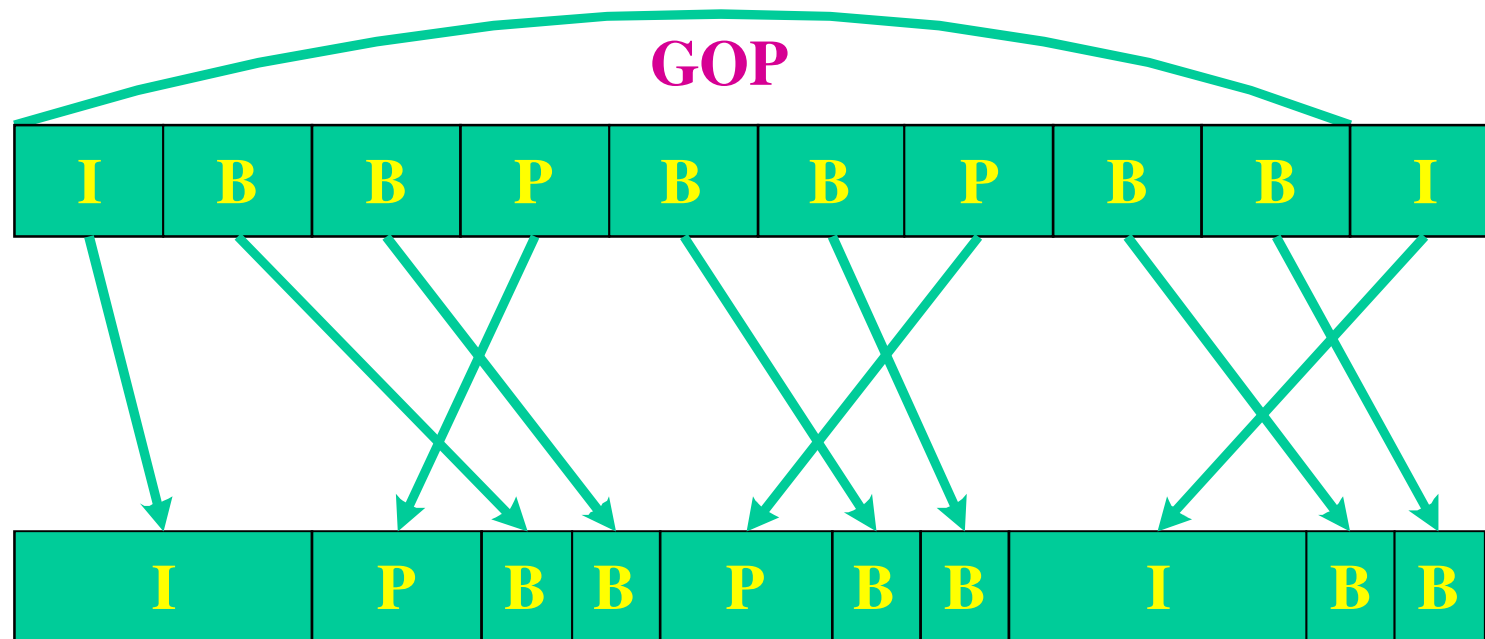
Acquisition  
and  
visualization

I P B B P B B P B B I B B P B B P B B  
0 3 1 2 6 4 5 9 7 8 12 10 11 15 13 14 18 16 17

Transmission

# Putting Order in the Orders ...

## Acquisition and Visualization Orders



## Coding, Transmission and Decoding Orders

## Constant Quality: How ?



- **Users enjoy content which is shown (and thus coded) with rather constant quality this means without noticeable quality variations along time and space.**
- **The need for random access led to the definition of 3 frame types depending on the used coding tools which have different compression powers.**
- **Since the uniform allocation of bitrate resources to the various frames would lead to noticeable quality variations in time, there is the need to non-uniformly allocate the bitrate resources depending on the compression power of the coding tools used for each frame.**
- **Experience has shown that for  $M=2-3$ , good results are achieved attributing similar quality to the I and P frames and a slightly lower quality to the B frames.**

## Who Does Take the Best Part ?

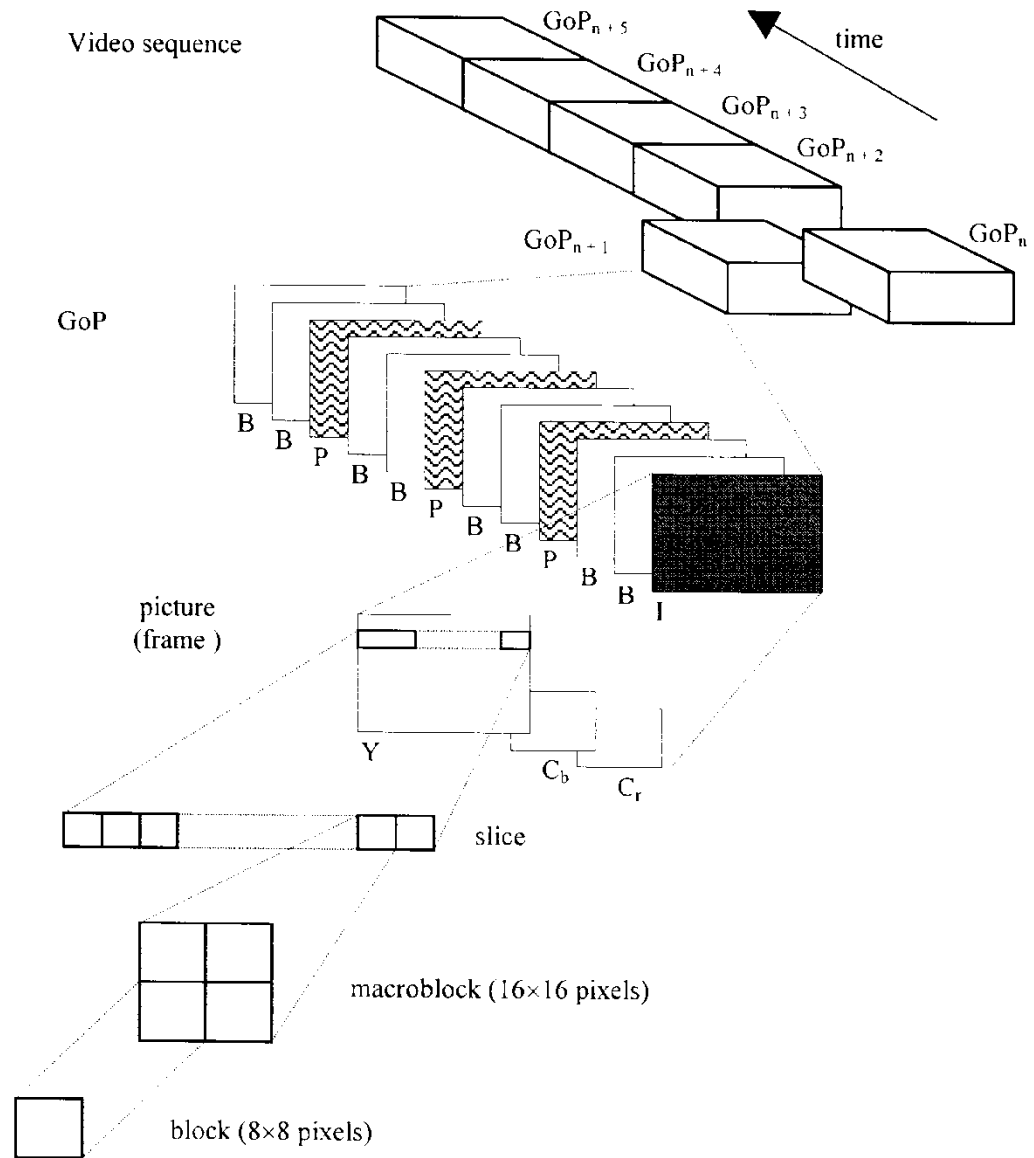


- **The ideal allocation of resources among the various frames depends on the specific video content; however, the following type of distribution typically leads to good quality results for natural images:**
  - **P frames with 2-5 times more bits than B frames**
  - **I frames with up to 3 times more bits than P frames**
  - **For low motion, more bits must be allocated to the I frames**
  - **For high motion, the proportion of I frames bits must be reduced passing these savings to the P frames**
- **These rules should only be taken as a starting point; the final bitrate allocation must be performed by the bitrate control method depending on the dynamic characteristics of the video frames.**

# Video Structure

The video data is organized in a hierarchical structure with 5 layers:

- Sequence
- Group of Pictures (GOP)
- Picture
- Slice (more flexible than H.261 GOBs)
- Macroblock (MB)
- Block





# Hierarchical Data Structure (1)

Each layer has a well defined target in terms of signal processing, e.g. DCT or motion compensation, or provided functionalities, e.g. resynchronization or random access.

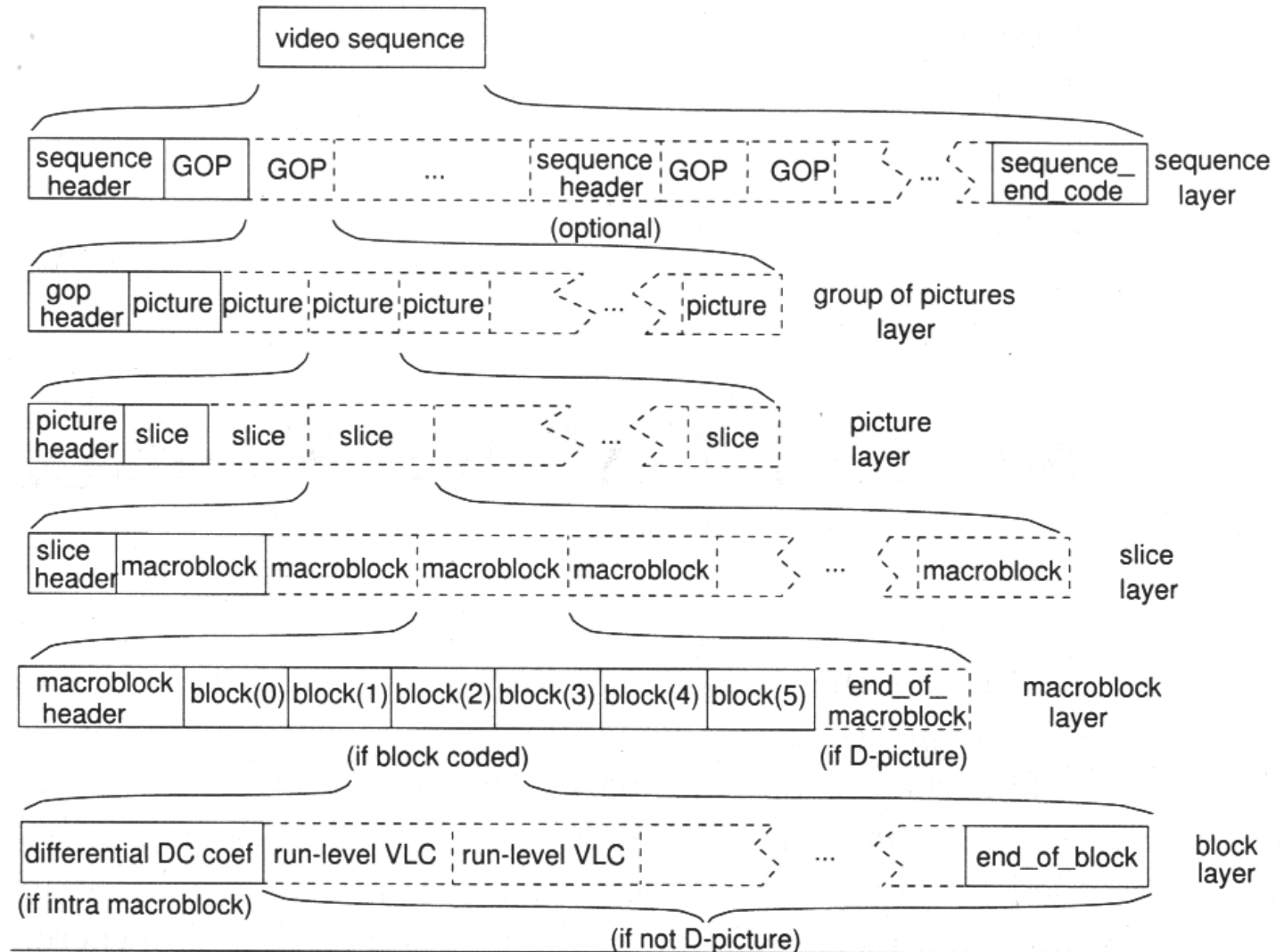
- **SEQUENCE** – One or more Group of Pictures; basic parameters, e.g. spatial format, temporal resolution, bitrate, etc.
- **GROUP OF PICTURES (GOP)** – Set of one or more frames directly accessible without reference to any other frames; essential for random access; may have one or more I frames.
- **PICTURE** – Basic coding unit which may be classified as I, P or B depending on the allowed coding tools; each frame may have one or more slices.



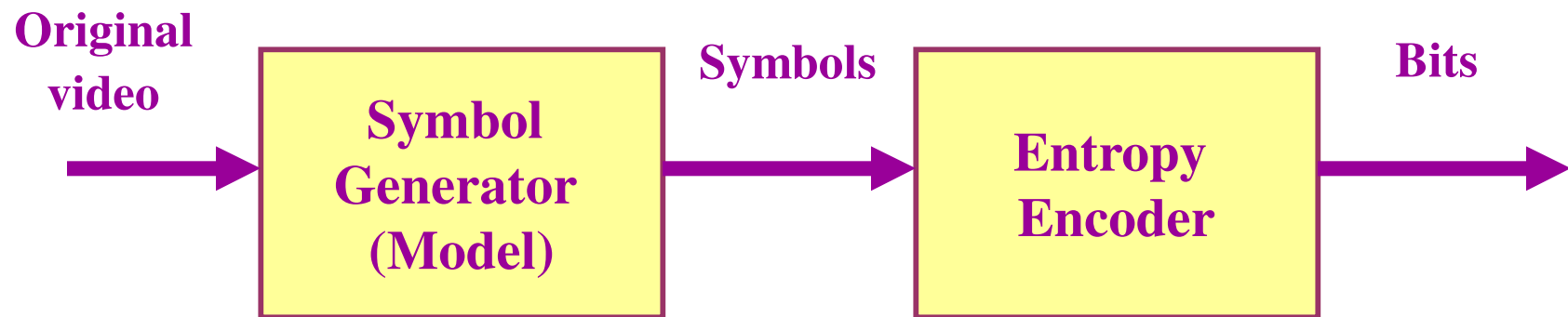
## Hierarchical Data Structure (2)

- **SLICE** – Set with any number ( $> 1$  and  $\leq$  number of MBs in frame) of macroblocks, ordered from left to right and top-down; it offers the most internal resynchronization layer; slice positions may change from frame to frame.
- **MACROBLOCK** – Set of  $16 \times 16$  luminance pixels and corresponding  $8 \times 8$  pixels for each of the two chrominances (for 4:2:0 subsampling); unit for motion estimation and compensation application.
- **BLOCK** – Set of  $8 \times 8$  luminance or chrominance pixels; unit for DCT application.

# MPEG-1 Video Syntax

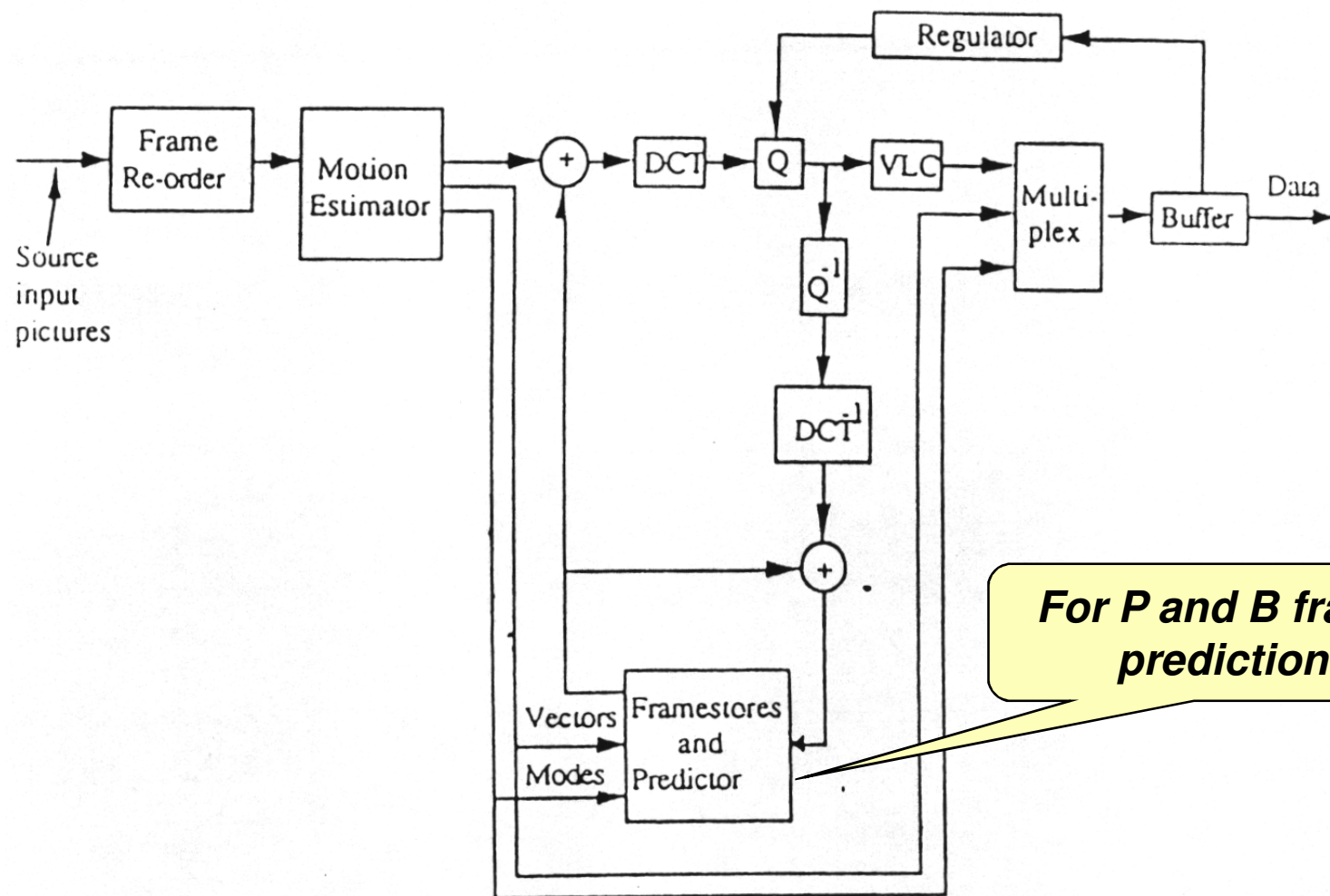


# The MPEG-1 Video Symbolic Model



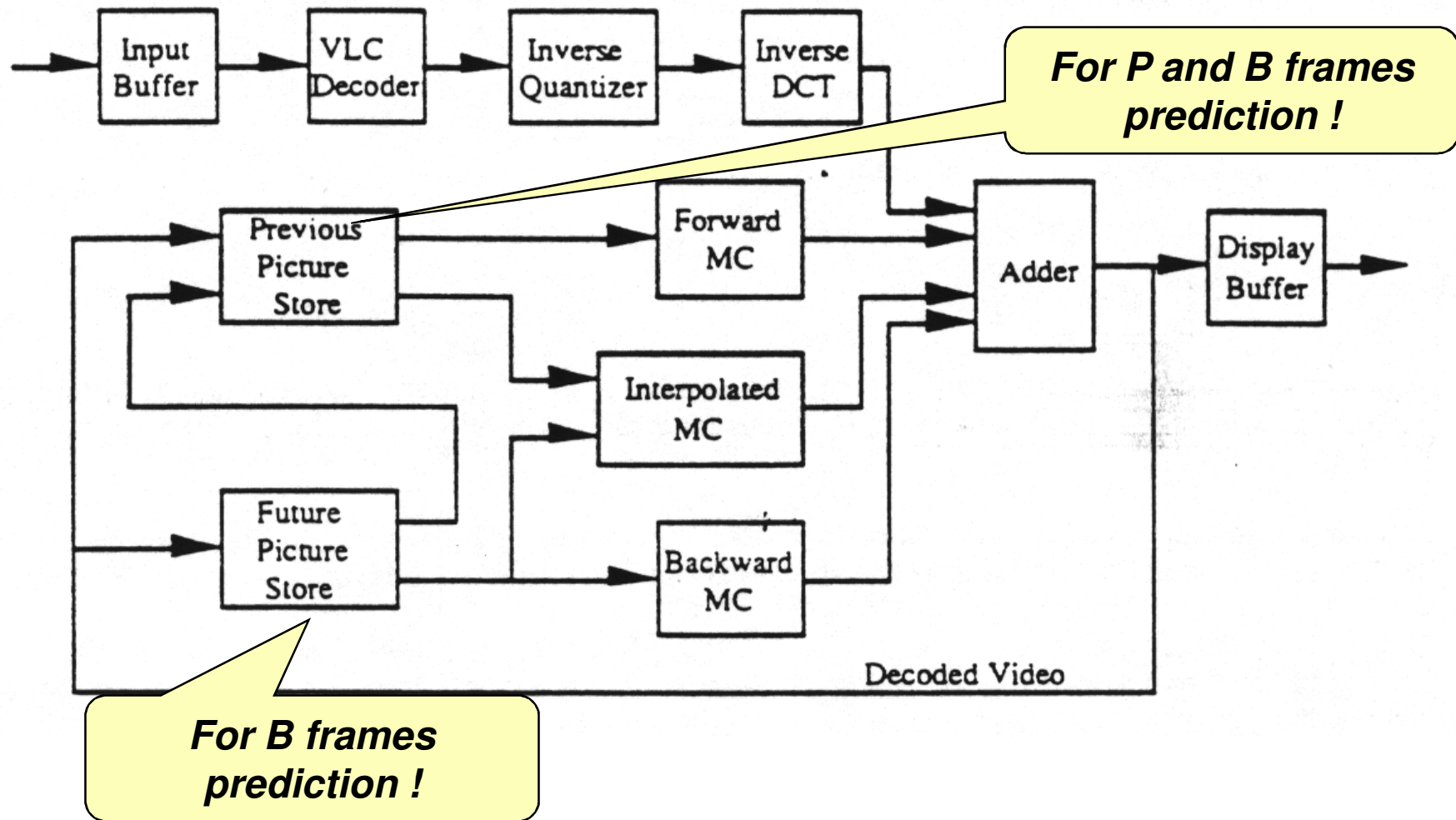
**A video sequence is represented as a succession of GOPs, including I, P and B coded frames, each structured in macroblocks, coded using motion vector(s) and/or DCT coefficients, following the constraints set by the frame coding type (I, P or B).**

# MPEG-1 Video: Encoder



**For P and B frames prediction !**

# MPEG-1 Video: Decoder





## **MPEG-1 Video and H.261: Which Relationship ?**

**VERY INTIMATE ... but ...**

- **H.261 targets real-time applications with a maximum delay around 150-200 ms.**
- **MPEG-1 Video does not have strong delay requirements since it mainly targets storage applications.**
- **MPEG-1 Video must offer all the typical functionalities already available in analogue video storage systems.**
- **MPEG-1 Video is optimized for higher bitrates.**

**There is the highest possible technical compatibility between MPEG-1 Video and H.261 to facilitate the simultaneous implementation of both codecs in certain systems.**



## The Output Buffer ...



**The bitrate production by the encoder is highly non-uniform in space and time, essentially due to:**

- The classification of the frames as I, P and B which implies the usage of coding tools with different compression power
- The allocation of different bitrate resources to every frame to compensate the usage of different coding tools
- The spatial variation of activity within the various image areas
- The temporal variation of activity along time
- The entropy coding of the generated symbols

**To make the variable bitrate produced by the encoder ‘compatible’ with the constant bitrate drained by the channel, an output buffer must be used !**

## Rate Control ... and Offline Encoding ...

- The encoder must control the produced bitrate along time and within each image in order to reach the best overall subjective quality with the available resources.
- While the encoder has the mission to take important decisions, the decoder is a ‘slave’ limiting itself to follow the ‘orders sent by the encoder – the ‘boss’.

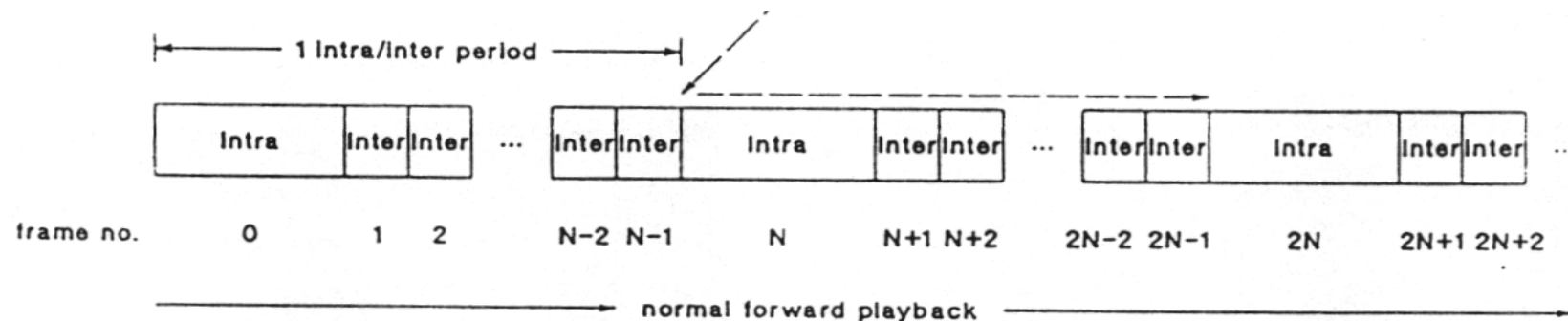


**For the most important MPEG-1 applications, encoding may be performed off-line (taking whatever time, iterative encoding, multiple passes, etc), thus achieving much higher quality than real-time encoding for similar bitrate resources.**



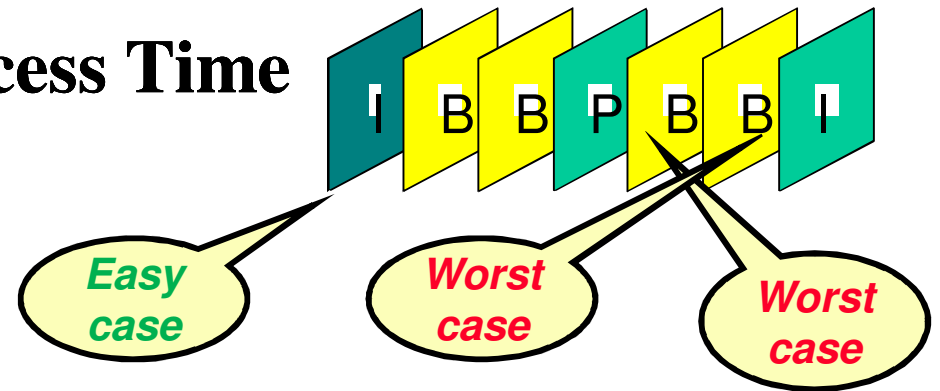
# Especial Access Modes

## Random Access ...



- The random access facility allows to access – read the bits, decode, and visualize - any video frame within a small, limited time, typically around 0.5 s; this imposes the usage of anchor frames like I frames.
- The storage device has an address table allowing the fast access to all reference frames, this means I frames; from those I frames, reading proceeds towards the target frame.

## Maximum Random Access Time



For the CD-ROM, the Maximum Random Access Time (MRAT) depends on the allocation of bits among the various types of frames, the frame rate and the time between I frames:

$$\text{MRAT} = T_{\text{DSM}} + [ 2 \text{BF}_I + (N/M-1) \text{BF}_P + 1 \text{BF}_B ] \times (1/f_s) \text{ (s)}$$

- $\text{BF}_X$  is the maximum frame peak factor for the bits spent in all X type frames
- $T_{\text{DSM}}$  is the sum of the various access times due to the need to jump in the CD-ROM to read only the strictly necessary bits.



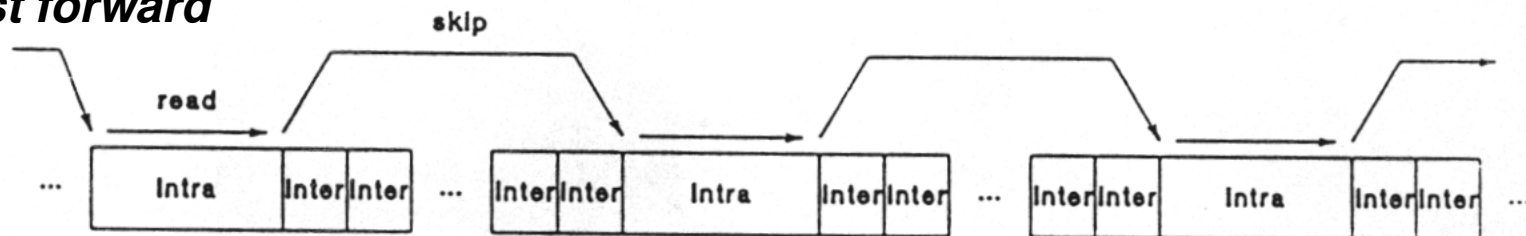
## **Fast Forward and Fast Reverse**



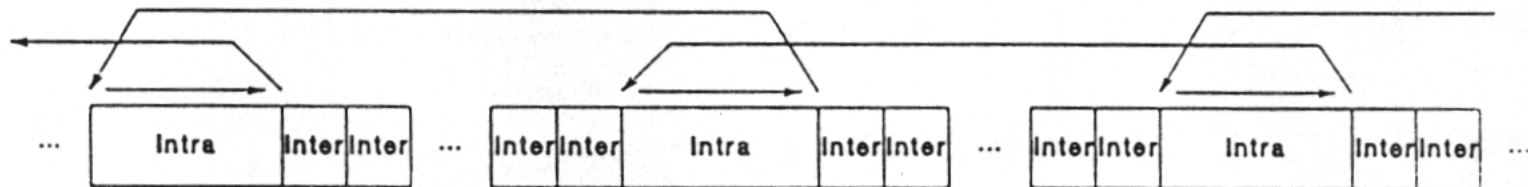
- **MPEG-1 offers fast modes with a speed up factor between 8 and 10 which means the video information corresponding to 1 s must be, on average, visualized in 0.1 to 0.125 s.**
- **These fast modes have to be based on I and/or P frames depending on the temporal prediction structure, notably the M and N values for regular structures.**
- **The most basic limitation is imposed by the reading speed which limits the number of frames that may be read in a certain time, especially those with the lowest compression factors, this means I and P frames.**

# Fast Forward and Fast Reverse: Examples

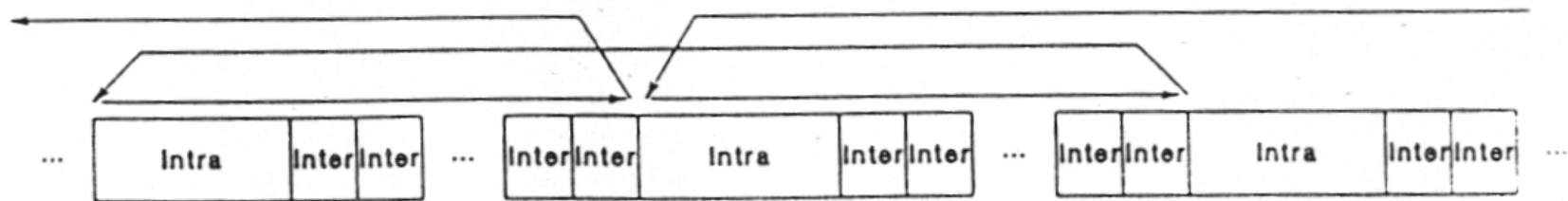
## Fast forward



## Fast reverse

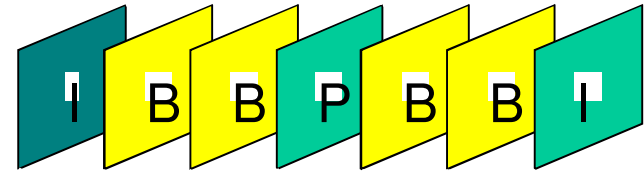


# Normal Reverse: Example





## Speed Up Factor



The *Speed Up Factor* (SUF) is computed as the ratio between the ‘real’ time corresponding to the frames read and their corresponding ‘visualization’ time.

- If only I and P frames are read:

$$\text{SUF} = \langle \text{real time} \rangle / \langle \text{visualization time} \rangle = [(K+1) M] / V$$

- $K$  is the number of I and P frames skipped between each one read (if P frames are read, no P frames may be skipped)
  - $V$  is the number of frame periods for the visualization of each frame read
- When only I frames are used, it comes  $K = N/M - 1$ .



## Fast Forward and Reverse: Selecting the Frames

Since the reading speed (constant or peak, if they are different) limits the number of frames that may be read during the time each frame is visualized, there is a minimum value for this visualization time in order the next frame to be visualized may be ready.

- As the time to visualize a frame has to be higher than the time to access the next frame to be visualized, which is a I frame in the worst case, it comes:

$$V/f_s \geq BF_I/f_s + T_{DSM} \Rightarrow V \geq f_d \times (BF_I/f_s + T_{DSM})$$

- $V$  is the number of frame periods (integer and positive) each frame is visualized
- $T_{DSM}$  corresponds to the total device access time (for the CD-ROM it includes the time to jump between access points which depends on their number, latency, angular speed, and protocol)

If  $M= 4$ ,  $N= 24$  and  $BF_I = 3$ , it comes  $V \geq 6$  and  $SUF= 8$ , if only I frames are used.



# MPEG-1 Video: Constrained Parameters (1)

- MPEG-1 Video offers great flexibility for the video sequence parameters (included in the bitstream), accepting a large range of spatial and temporal resolutions, aspect ratios and bitrates.
- Since it is important to avoid forcing the manufacturers to produce equipment which is unnecessarily complex to guarantee interoperability, a set of values for the basic coding parameters has been defined in the standard, allowing to create the so-called *Constrained Parameters Bitstreams*.
- The *constrained parameters* guide (and constraint) the product manufacturers and content producers since all MPEG-1 Video decoders must be able to decode *Constrained Parameters Bitstreams*.
- However, bitstreams using other parameters may be created: a flag in the bitstream signals if the bitstream follows or not the limitations imposed by the *constrained parameters*.

## MPEG-1 Video: Constrained Parameters (2)

Horizontal picture size	Less than or equal to 768 pels
Vertical picture size	Less than or equal to 576 lines
Picture area	Less than or equal to 396 macroblocks
Pel rate	Less than or equal to 396x25 macroblocks per second
Picture rate	Less than or equal to 30 Hz
Motion vector range	Less than -64 to +63.5 pels (using half-pel vectors) [backward_f_code and forward_f_code <= 4
Input buffer size (in VBV model)	Less than or equal to 327 680 bits
Bitrate	Less than or equal to 1 856 000 bits/second (constant bitrate)

**VBV in bytes not bits !**





# Error Concealment



**If MPEG-1 Video coded content is used for transmission over error prone channels, the decoder should:**

- **Detect the residual errors at syntactic and semantic levels**
- **Minimize the negative subjective effect of the transmission errors by applying (non-normative) error concealment methods such as:**
  - **Substituting the corrupted image areas with the co-located areas from a previous frame, ☹**
  - **Substituting the corrupted areas with the motion compensated areas from a previous frame, ☺**

## MPEG-1 Video: Final Remarks

- **MPEG-1 Video is one of the most common formats for digital video in PCs, e.g. Windows has a MPEG-1 Video player (software).**
- **A large share of the digital video in the Internet is in the MPEG-1 Video format.**
- **There are many products and services based on the MPEG-1 Video standard, notably video cameras; however, MPEG-1 is not anymore the state-of-the-art in terms of video coding for entertainment content ...**
- ***Video CD* is based on MPEG-1 and sold hundreds of millions of players in China.**





# MPEG-1 Standard: Structure

## **Part 1: Systems**

**Specifies the multiplexing of the several audio and video coded streams in a single stream with synchronization**

## **Part 2: Video**

**Specifies the video coding solution (bitstream and decoding) for bitrates of about 1.15 Mbit/s**

## **Part 3: Audio**

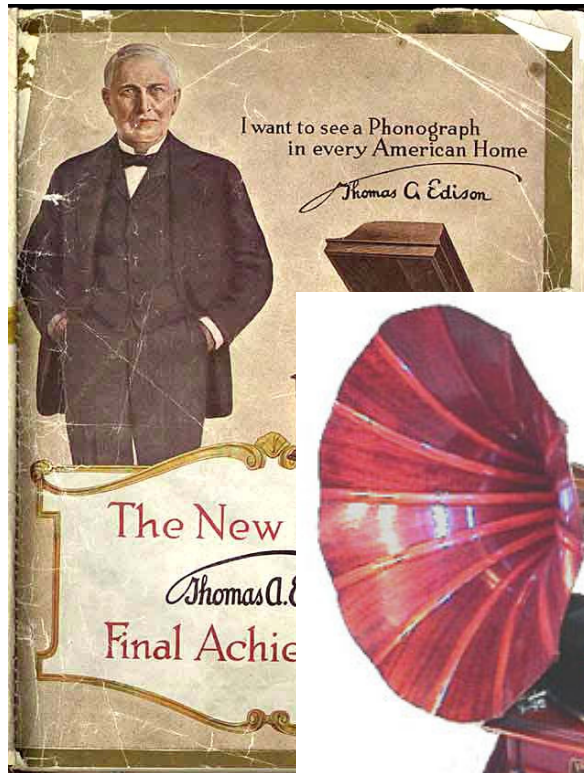
**Specifies the audio coding solution (bitstream and decoding) for bitrates of 32-448 kbit/s per channel (mono and stereo)**

## **Part 4: Conformance Testing**

**Specifies conformance tests for the streams and decoders**

## **Part 5: Reference Software**

**Software implementation of the parts 1, 2 and 3**



VICTOR IV

# MPEG-1 Standard

## Part 3: Audio

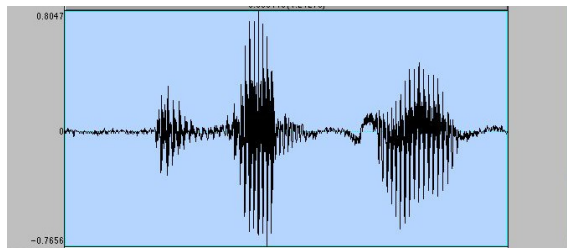
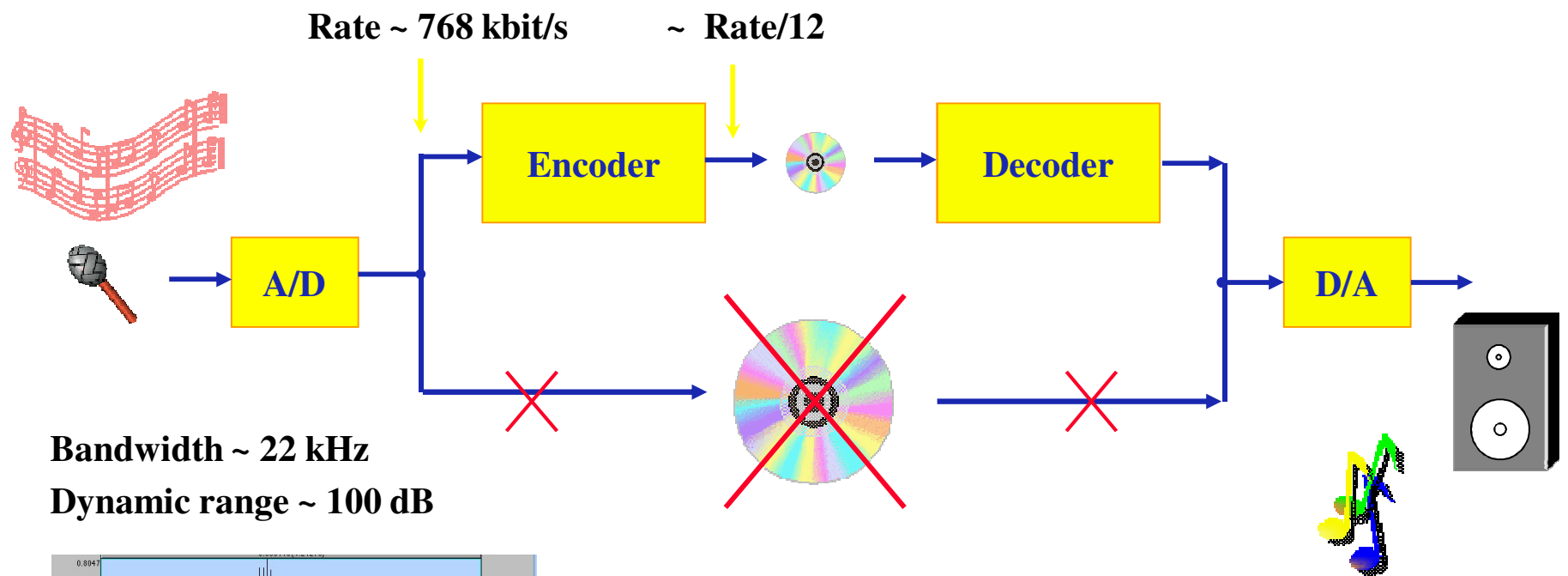


## **MPEG-1 Audio: Objective**

**Efficient audio coding, mono and stereo, with high quality, at 32-448 kbit/s per channel, using sampling rates of 32, 44.1 and 48 kHz, and targeting digital audiovisual storage at an overall rate of 1.5 Mbit/s.**

**The target audio RD performance is the CD-ROM (PCM) quality at 256 kbit/s, for stereo content.**

# The Audio Compression Chain ...





# **MPEG-1 Audio: Applications**

- **Audio production**
- **Audio distribution and sharing**
- **Internet streaming**
- **Portable audio**
- **Audio archival**
- **Digital radio and television (DAB and DVB)**
- **Digital audio storage**
- **Multiple multimedia applications**
- **....**



## PCM for Various Sound Signals ...

	<b>Frequency (Hz)</b>	<b>Sampling rate (kHz)</b>	<b>bit/sample (PCM)</b>	<b>PCM bitrate (kbit/s)</b>
<b>Speech (telephone)</b>	<b>300-3400</b>	<b>8</b>	<b>8</b>	<b>64</b>
<b>Speech (wideband)</b>	<b>50-7000</b>	<b>16</b>	<b>8</b>	<b>128</b>
<b>Audio (medium band)</b>	<b>10-11000</b>	<b>24</b>	<b>16</b>	<b>384</b>
<b>Audio (wideband)</b>	<b>10-22000</b>	<b>48</b>	<b>16</b>	<b>768</b>



## Compact Disc (CD) and Digital Audio Tape (DAT)

Storage Support	Audio rate	Overhead	Total rate
CD	1.41 Mbit/s	2.91 Mbit/s	4.32 Mbit/s
DAT	1.41 Mbit/s	1.67 Mbit/s	3.08 Mbit/s

The CD audio rate,  $44100 \times 2 \times 16 = 1.41$  Mbit/s, may increase up to  $1.41 \times 49/16 = 4.32$  Mbit/s due to the usage of a synchronization and correction code which spends 49 bits per each 16 audio bits.

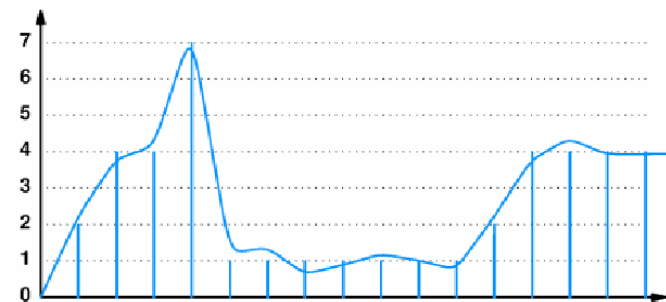
Typically, the audio samples are structured in frames with 2352 bytes at a rate of 75 frames/s which corresponds to  $75 \times 2352 \times 8 = 1.41$  Mbit/s.

One CD with 780 Mbyte corresponds to 74 minutes of music.



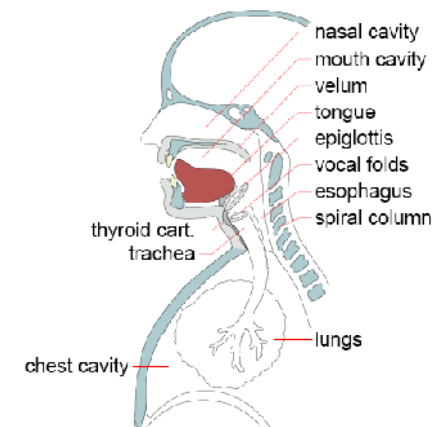
# MPEG-1 Audio: Requirements

- **High (decoded) signal quality independently of the spectral and amplitude characteristics of the coded signal**
- **Low encoding and decoding delays**
- **Spatial integrity for stereo and multichannel signals**
- **Error resilience to uniform and burst errors and packet losses**
- **Graceful degradation for higher error probabilities and loss rates**
- **Resilience to cascading, i.e. successive coding and decoding processes**
- **Capability to edit, mix, etc.**
- **Low implementation complexity**
- **Low energy consumption**



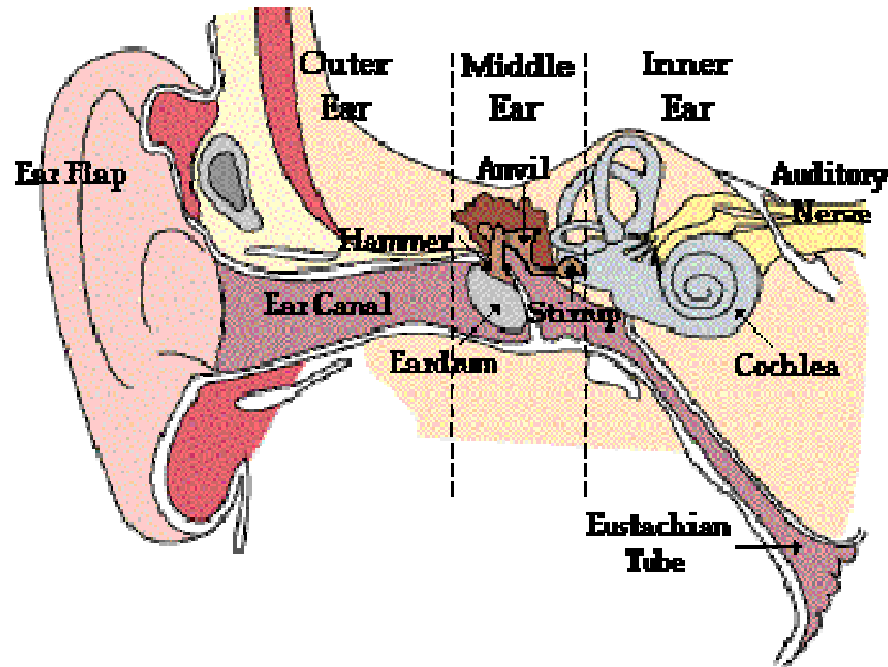
## Audio Coding Peculiarities ...

- Very high dynamic range (ratio between the maximum and minimum amplitudes) and larger bandwidth in comparison with speech.
- Absence of a universal source production model; the existence for speech of this model allows reaching much higher compression factors.
- Certain simplifying assumptions usually adopted for speech coding are not valid anymore such as:
  - Gaussianity
  - Stationarity
  - Spectral smoothness



**Compression gains for high quality audio coding mainly result from irrelevancy reduction (as redundancy is short ...).**

# Human Auditory System



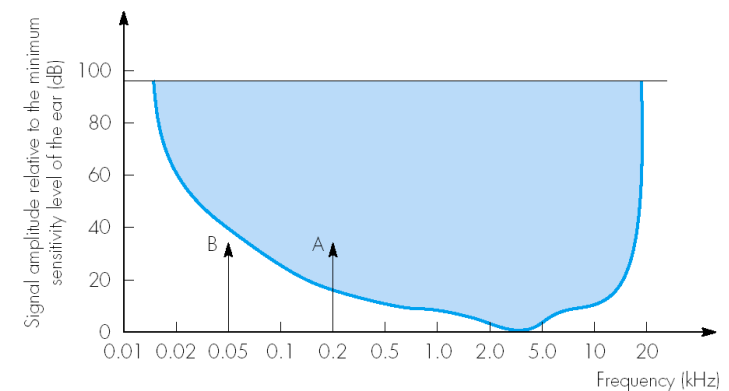
The ear has three main sections:

- 1) Outer ear – Directs the sound to the eardrum.
- 2) Middle ear – Transforms the sound pressure into mechanical vibration.
- 3) Inner ear – Converts these mechanical vibrations into excitations of the auditory nerves which send electrical signals to the brain.

- The perception of audio quality depends on the Human Auditory System (HAS).
- The Human Auditory System processing includes physiological and psychological effects.

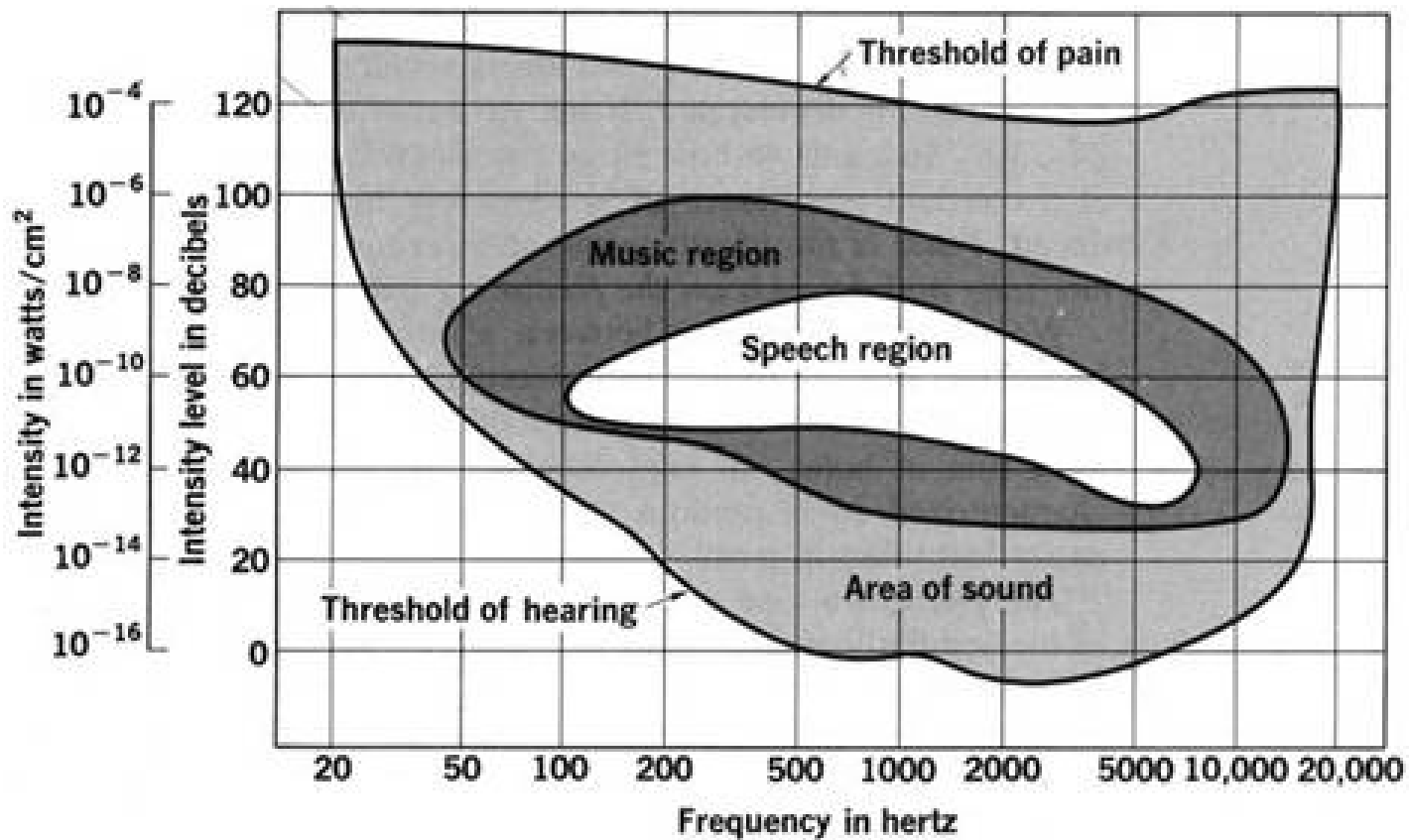
# Physiological Effects: the Thresholds

- **Threshold of Hearing** – Defines the minimum sound intensity which may be perceived; this threshold varies along the audio band.
- **Threshold of Feeling or Pain** – Defines the sound intensity above which the sounds may cause pain and provoke hearing damages.



Typically, the threshold of pain is about 120 to 140 dB; sound intensity is measured in terms of Sound Pressure Level relatively to a reference intensity with  $10^{-16}$  W/cm<sup>2</sup> at 1 kHz.

## Sound Sensibility ...



**The human hearing dynamic range is about 100 dB.**

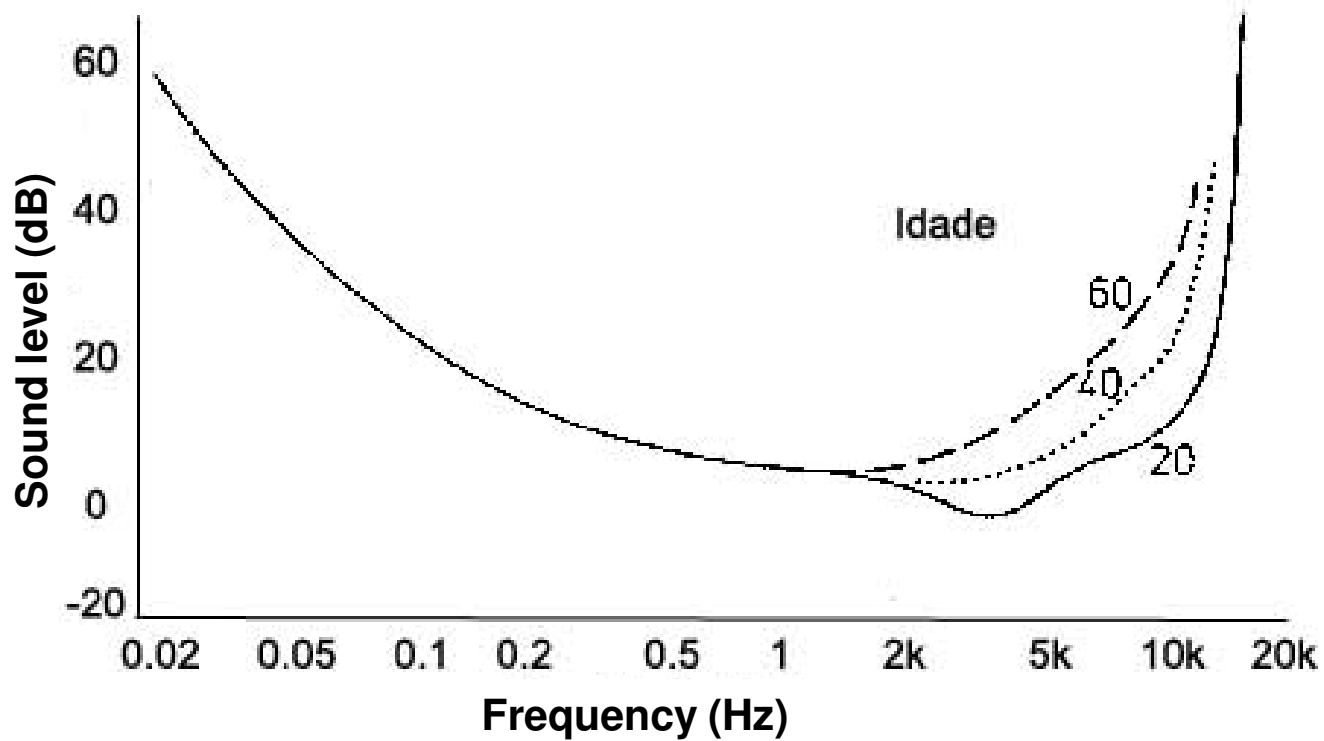
## The Attacks ...

- Rocket launching 180 dB
- Jet plane flying 140 dB
- Jet plane taking off 120 dB
- Disco 110 dB
- Underground 100 dB
- Urban traffic 90 dB
- Alarm-clock (at 1 meter) 80 dB
- Restaurant 70 dB
- Ar conditioning 60 dB
- Road (at 50 meters) 50 dB
- Living room 40 dB
- Library 30 dB
- Studio 20 dB
- Threshold of hearing 0 dB



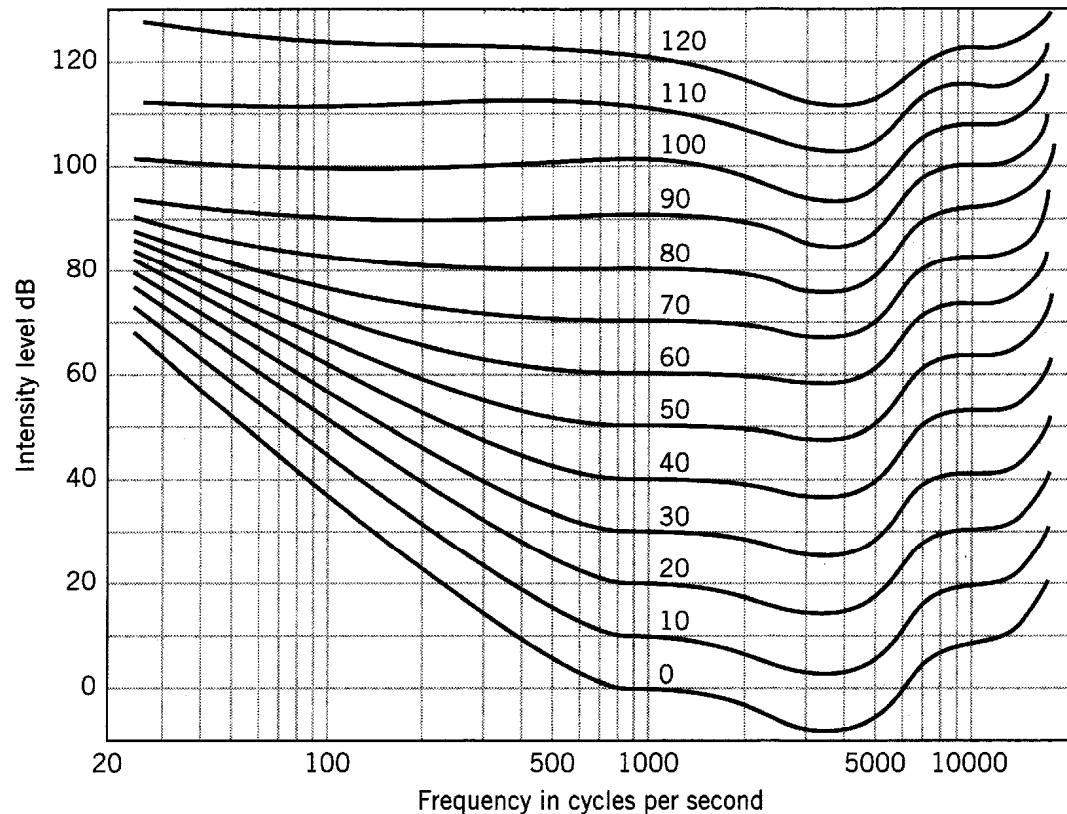
**The reference  
sound intensity (0  
dB) is  $10^{-16}$  W/cm<sup>2</sup>,  
for a sound at 1  
kHz.**

## Hearing Threshold Variation with Age ...

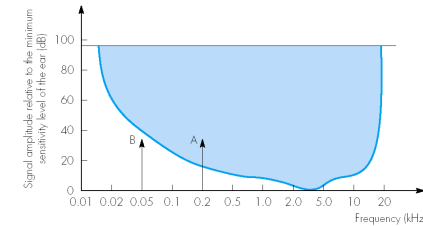


# Equal Loudness Curves

- The concept of a **threshold of hearing** is often extended to produce curves of equal perceived intensity for sounds.
- These 'equal loudness' curves describe the perceived loudness of a sound relative to its actual intensity.
- A similar sensation at the lower frequencies requires a higher intensity since there is less sensibility.

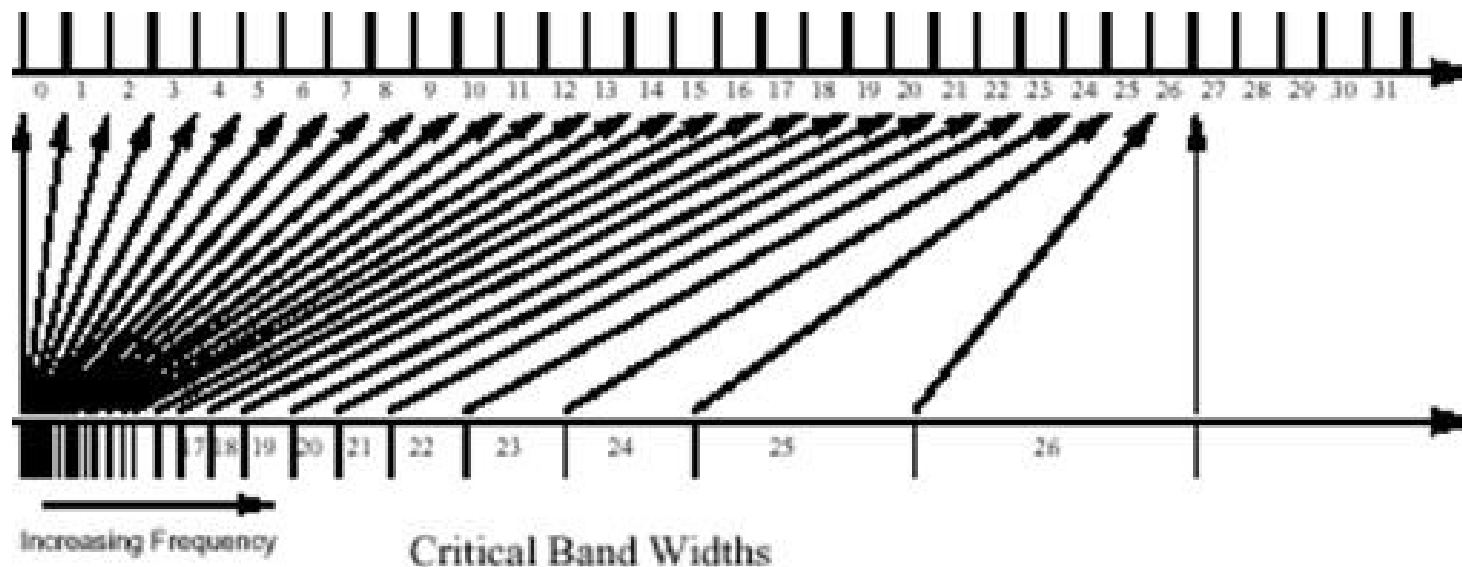


# Critical Bands



**Critical bands are the bands for which the auditory features are reasonably constant.**

- For  $f \approx 500$  Hz                       $LB_c = 100$  Hz
- For  $f \approx 1$  kHz ( $2 \times 500$  Hz)               $LB_c = 200$  Hz ( $2 \times 100$  Hz)
- For  $f \approx 5$  kHz ( $10 \times 500$  Hz)               $LB_c = 1000$  Hz ( $10 \times 100$  Hz)





# MPEG-1 Audio: Coding Tools

- **Redundancy**

Frequency coding

Window switching

- **Statistical Redundancy**

Huffman entropy coding

- **Irrelevancy**

Perceptive coding, masking and quantization

Dynamic allocation of bits

**LOSSLESS**

**LOSSY**



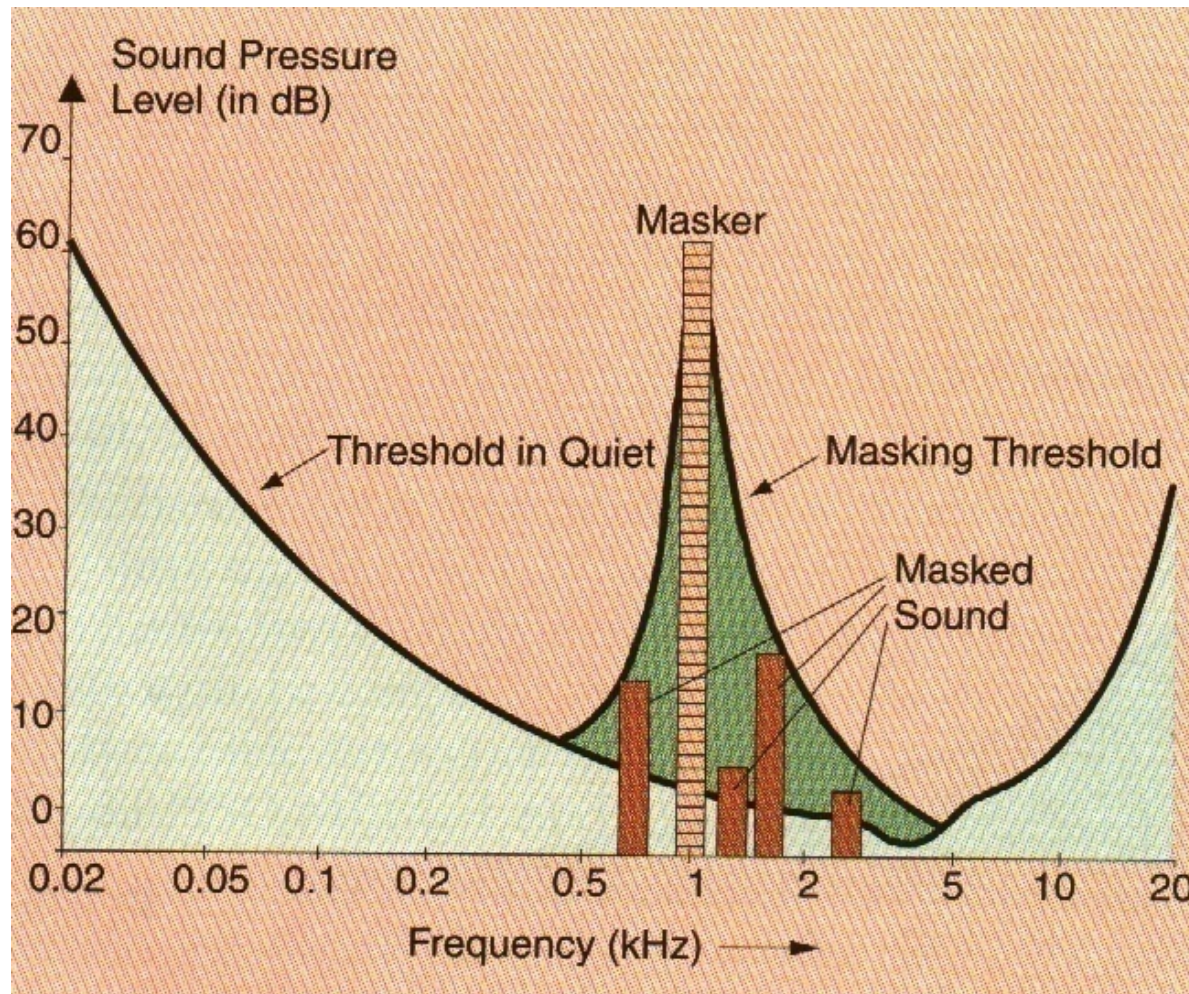
# Defining Audio Masking

**Auditory masking is the hearing behavior when the perception of one sound is affected by the presence of another sound; in this case, certain sound components may not be partially or totally perceived due to the prominence of other sound components.**

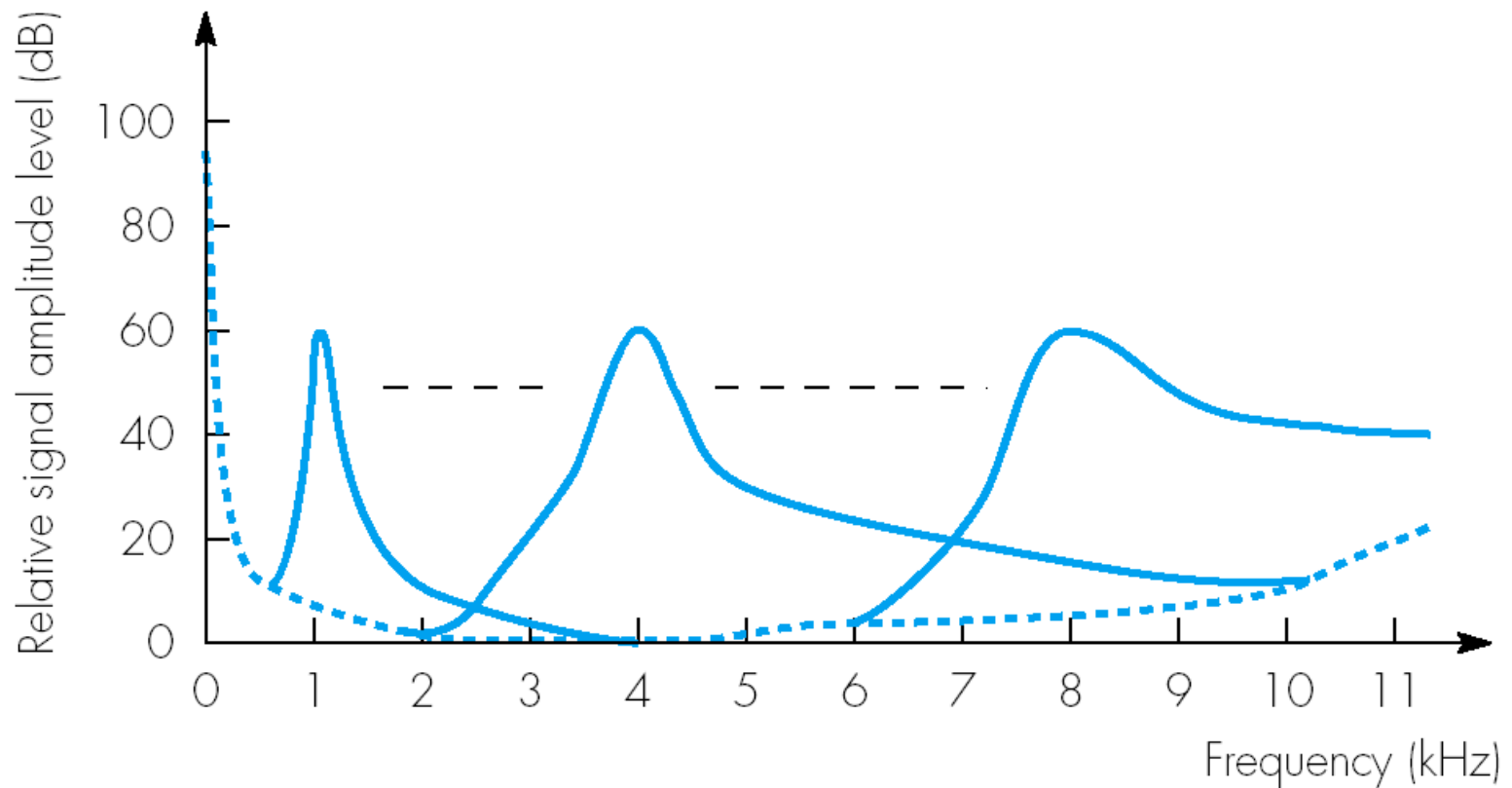
- **One sound may simply obscure another sound or increase its hearing threshold.**
- **The masking sound depends on the circumstances: for example, although it may be possible to speak ‘normally’ with someone at a party, any distraction may result in the background noise masking the voice of the other person.**

**The masking effect is highly non-linear and its effects are very diverse.**

# Frequency Masking



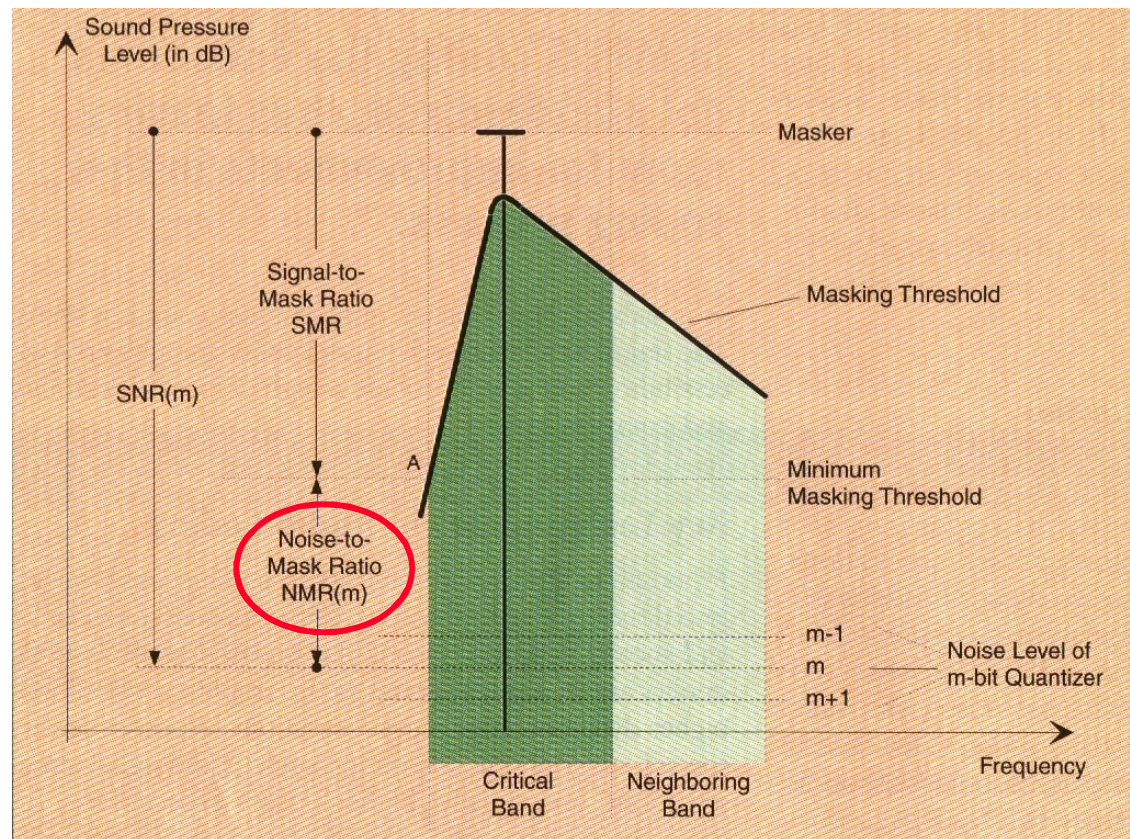
# Masking Width Variation with Frequency



# Frequency Masking

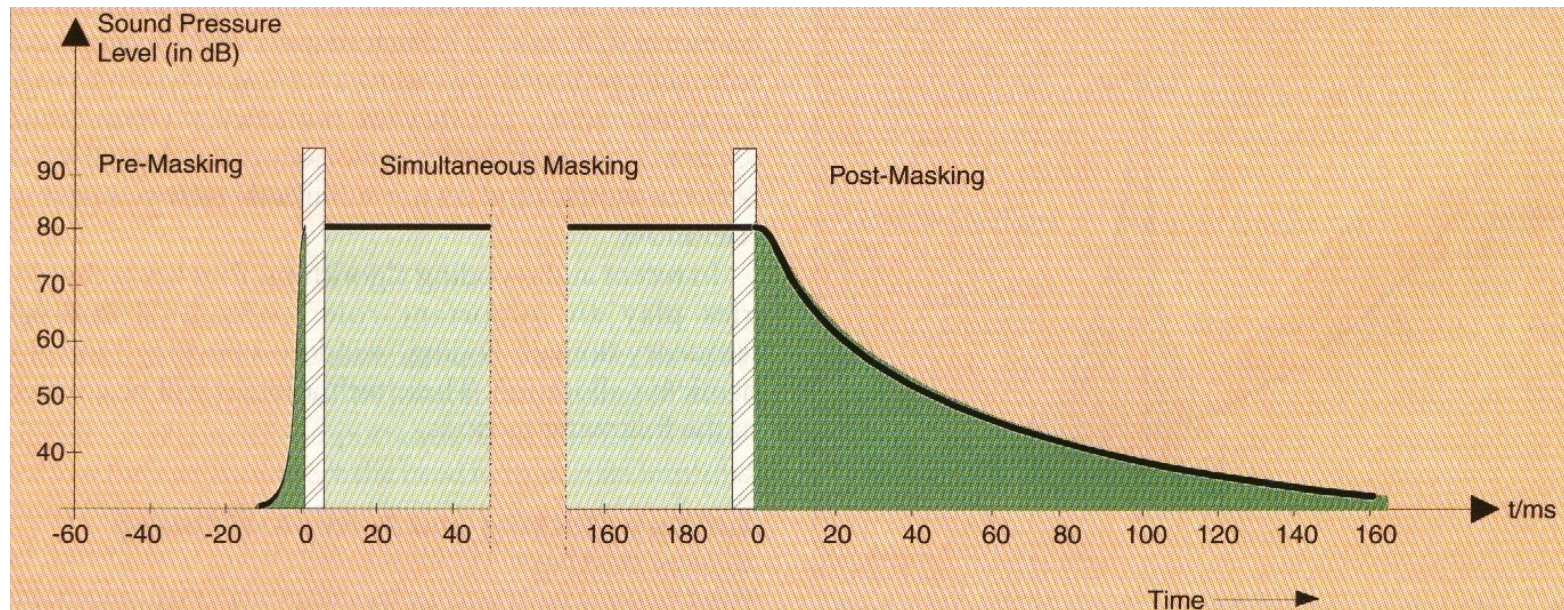
**NMR** measures the difference between the quantization noise level and the level at which the distortion becomes audible for a certain band.

The coding noise is not relevant while NMR is negative (NMR=SMR-SNR).



**Masking and the critical bands shape have been much studied to model the behaviour of the human auditory system.**

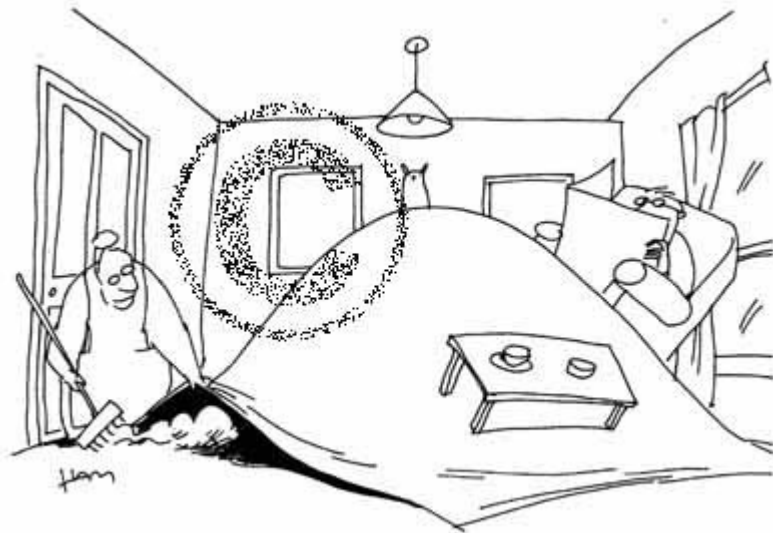
# Temporal Masking



- Temporal masking occurs when a sudden stimulus sound makes inaudible other sounds which are present immediately preceding or following the stimulus.
- Masking obscuring a sound immediately preceding the masker is called backwards masking or pre-masking ( $< 5$  ms) and masking obscuring a sound immediately following the masker is called forwards masking or post-masking ( $\approx 20$  ms).

# Perceptive Coding

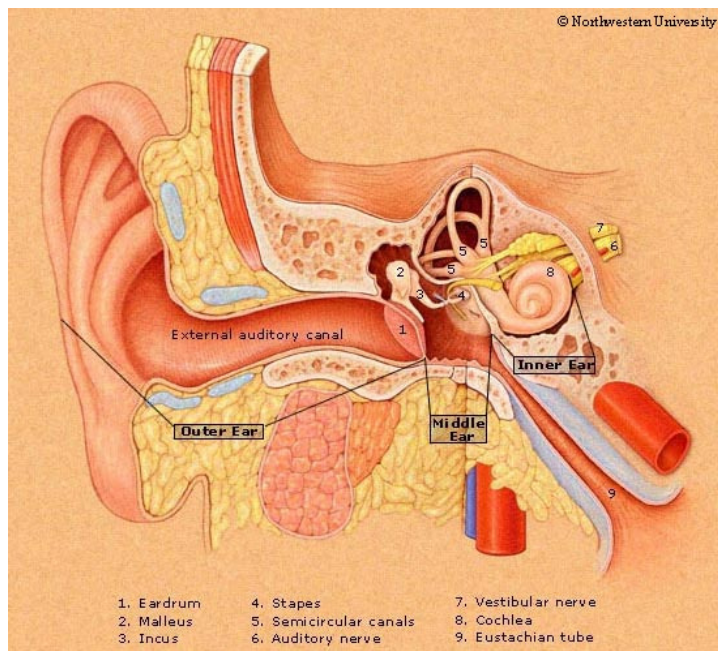
**Irrelevancy manifests itself as amplitude or frequency information (resolution, detail) which cannot be perceived by humans. All masked signal components do NOT need to be coded/transmitted.**



**Perceptive coding is based on the idea of ‘hiding’ more noise (coding error) in the frequency zones where that noise is better tolerated, e.g. due to masking, using a psychoacoustic model.**

**Perceptive coding exploits the characteristics of the receiver and not of the source as in speech coding.**

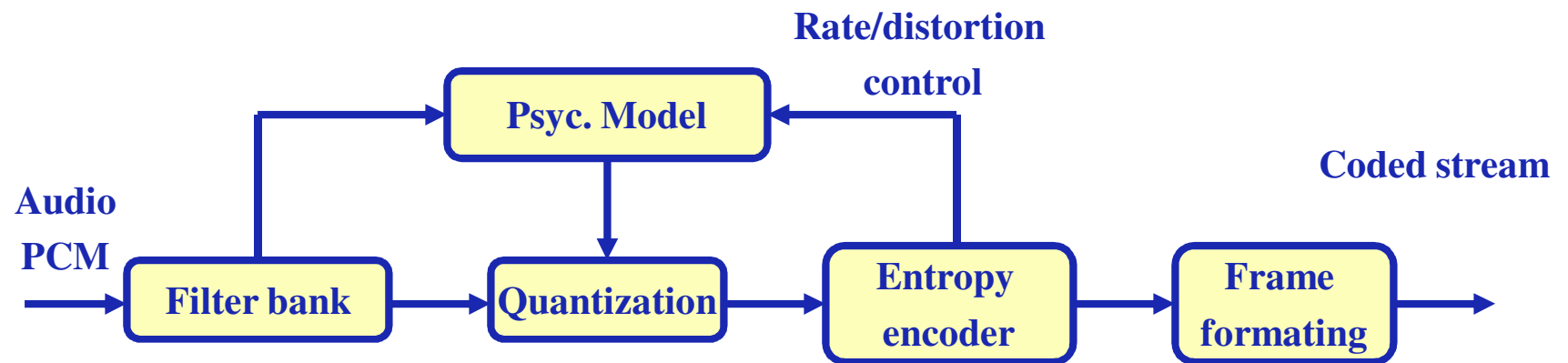
# Psychoacoustic Model: the Secret !



A psychoacoustic model is a mathematical model which defines, in a simplified way, the main properties and tolerances of the human auditory model, notably its sound intensity perception, its spectral selectivity and, especially, the masking effect.

It is very useful to dynamically and adaptively estimate the amount and shape of the coding noise that may be injected in the audio signal without becoming perceptible, in order to reduce the final coding rate.

# Perceptive Encoder Architecture



**The psychoacoustic model controls the quantization noise/error to introduce in each audio band.**



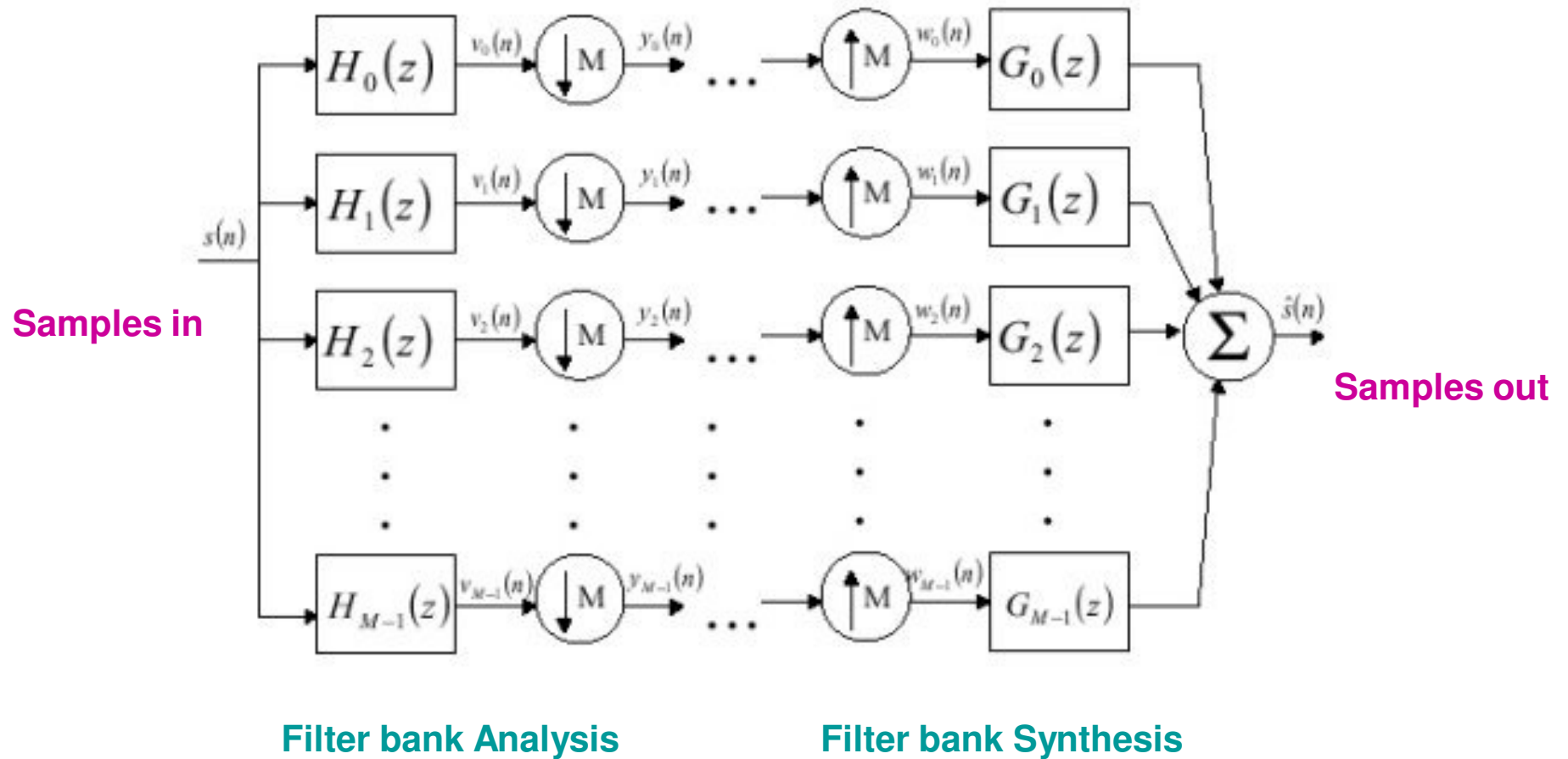
# Frequency Coding

**Coding in the frequency domain divides the audio signal spectrum in frequency bands; a filter bank is used to generate uncorrelated spectral components and independently quantize those components.**

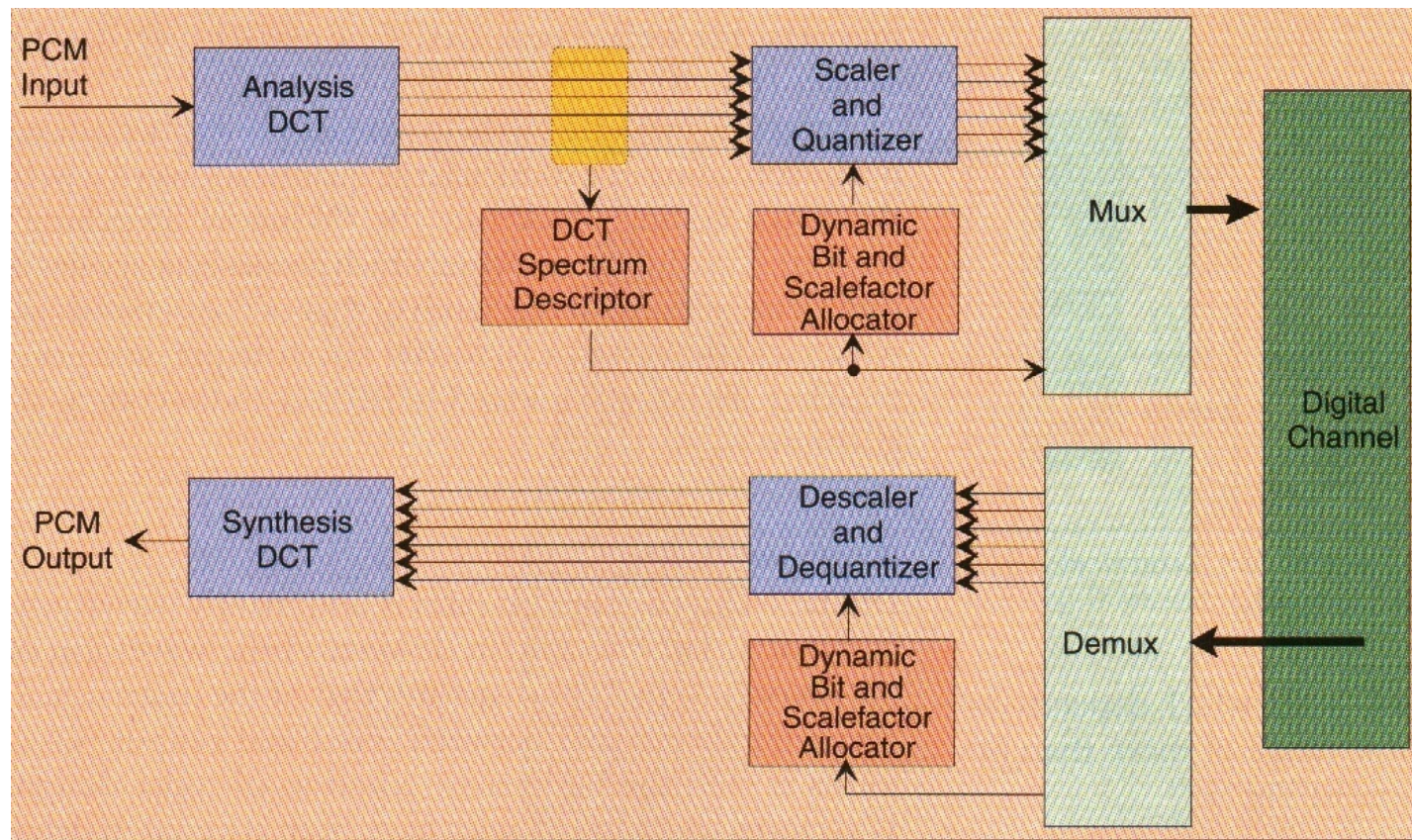
**There are two main ways to perform frequency coding:**

- **TRANSFORM CODING** – A samples block is linearly transformed using a discrete transform into a set of quasi-uncorrelated coefficients.
- **SUBBAND CODING** – A samples block is decomposed into several samples subsets using  $M$  band pass filters, contiguous in frequency, in order the set of generated subbands may be additively recombined to synthesise the original signal.

# The Subband Frequency Decomposition



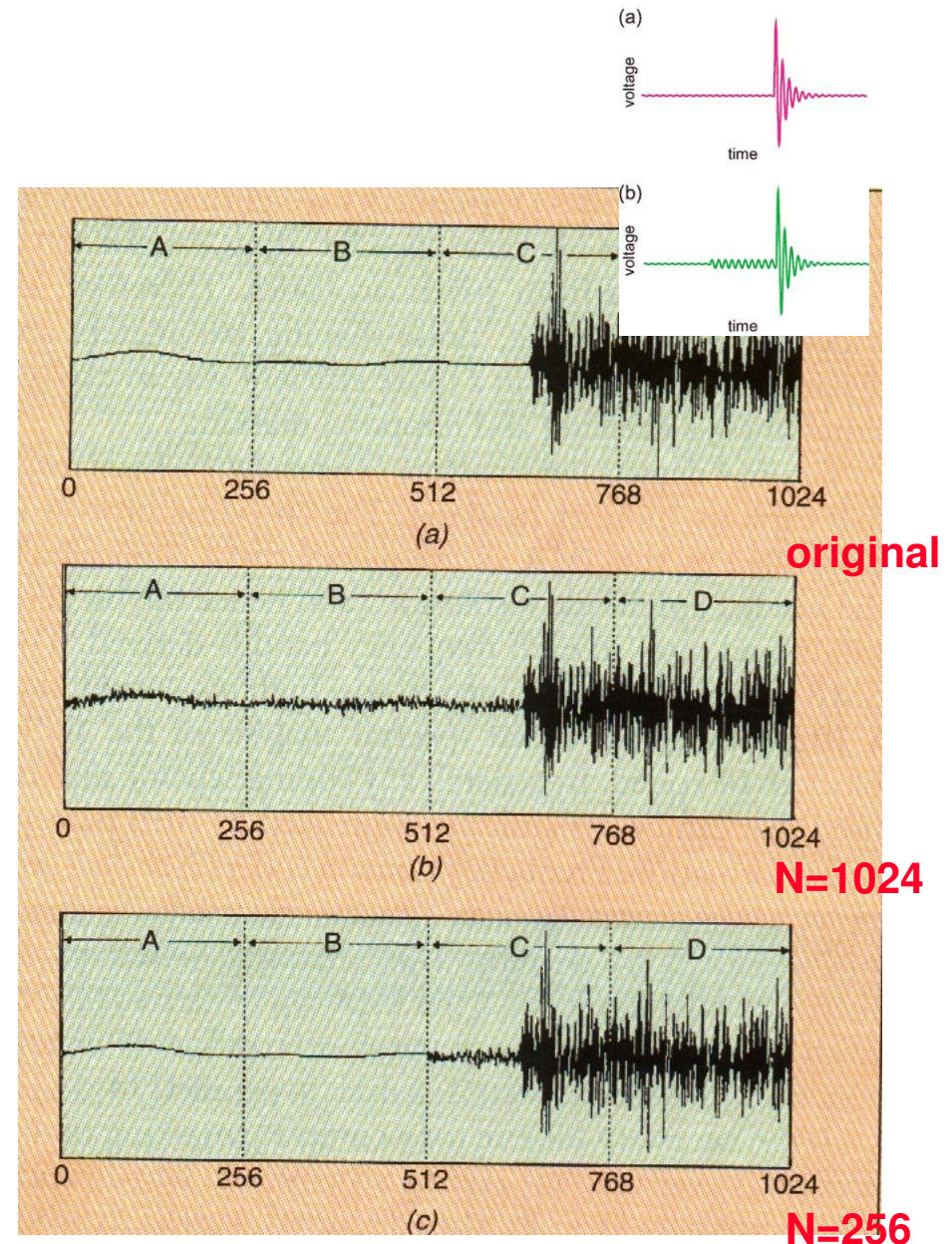
# Transform Coding Architecture



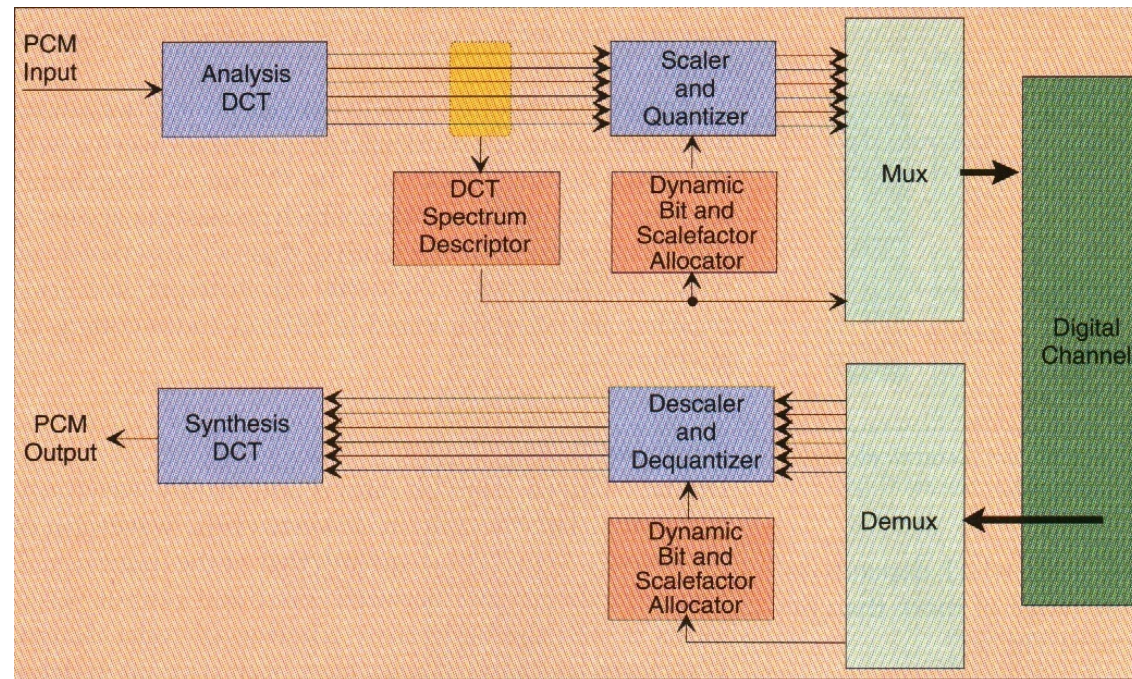
**Depending on their frequency band, the various frequency coefficients are differently quantized.**

# Window Switching

- The usage of frequency coding for blocks of samples where silence is followed by a strong signal creates the so-called pre-echoes since the signal synthesis may significantly change the silent part of the signal (in a more or less stronger way depending on the quantization).
- To limit this phenomenon, variable size transform windows may be used with the encoder selecting the adequate window size depending on the signal characteristics.

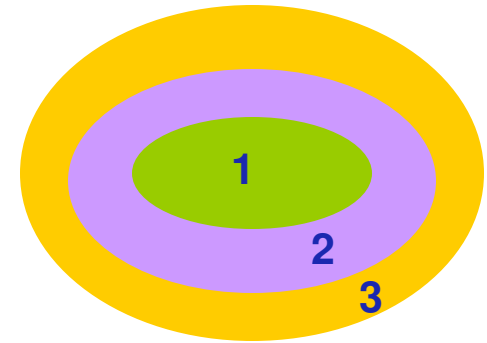


# Dynamic Allocation of Bits



- **The quantization of the coefficients (after bit allocation) is performed depending on the spectral characteristics of the samples block to be coded.**
- **It may originate block effects and pre-echoes.**

# MPEG-1 Audio: the 3 Layers



**MPEG-1 Audio specifies the coded representation and decoding process of audio (mono or stereo pair) signals in three layers where:**

- **Each layer offers a rate/quality/complexity trade-off**
- **Higher layers have higher complexity, delay and coding efficiency**
- **N layer decoders are able to decode N-1 layers coded streams defining a hierarchy of decoders and bitstream syntaxes**

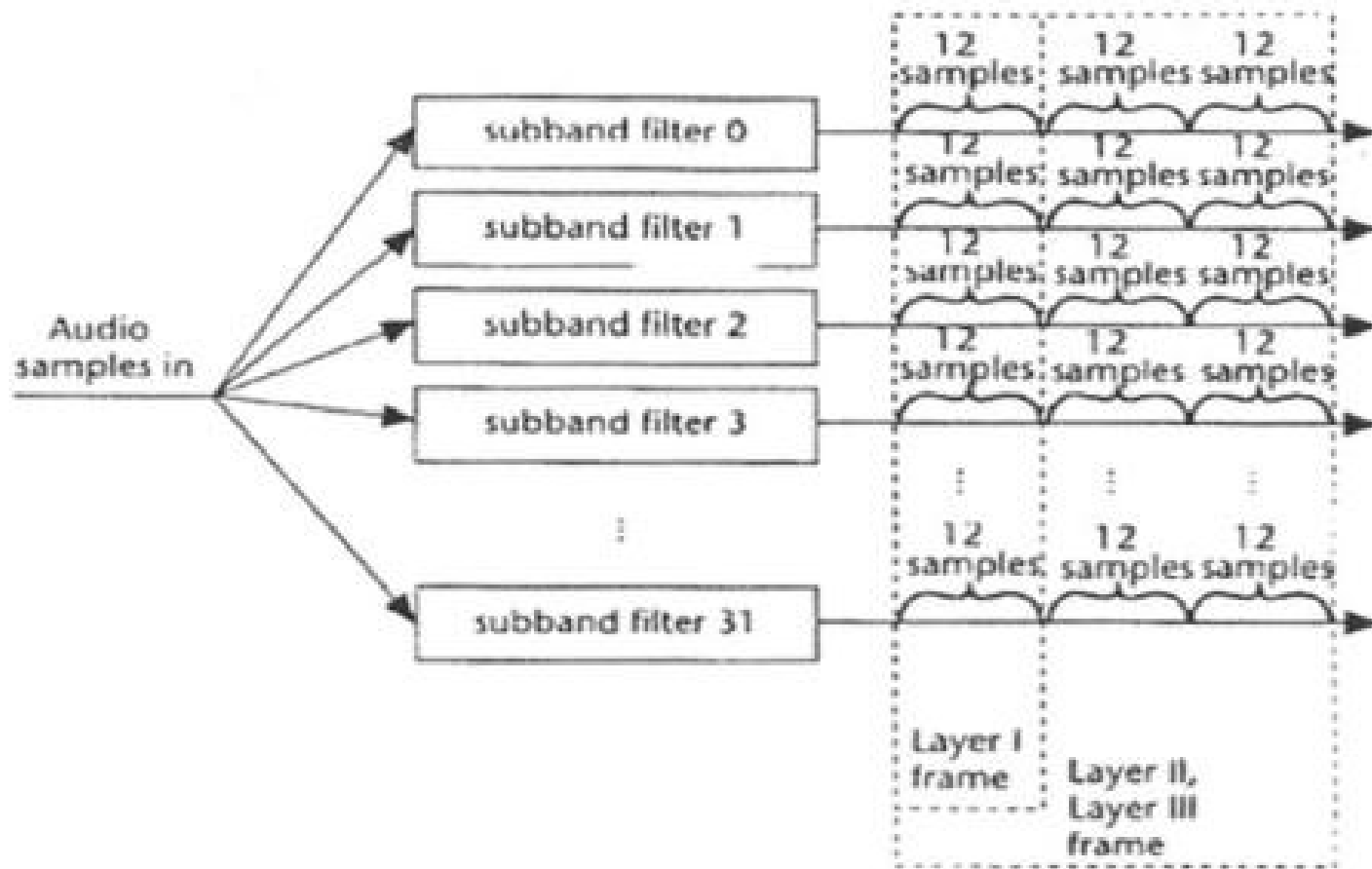
Layer	Typical rate	Minimum coding delay
1	32-448 kbit/s	$(256+256+12 \times 32)/48k \approx 19$ ms
2	32-384 kbit/s	$(256+256+12 \times 32 \times 3)/48k \approx 35$ ms
3	32-320 kbit/s	$(256+256+18 \times 2 \times 32 \times 2)/48k \approx 59$ ms



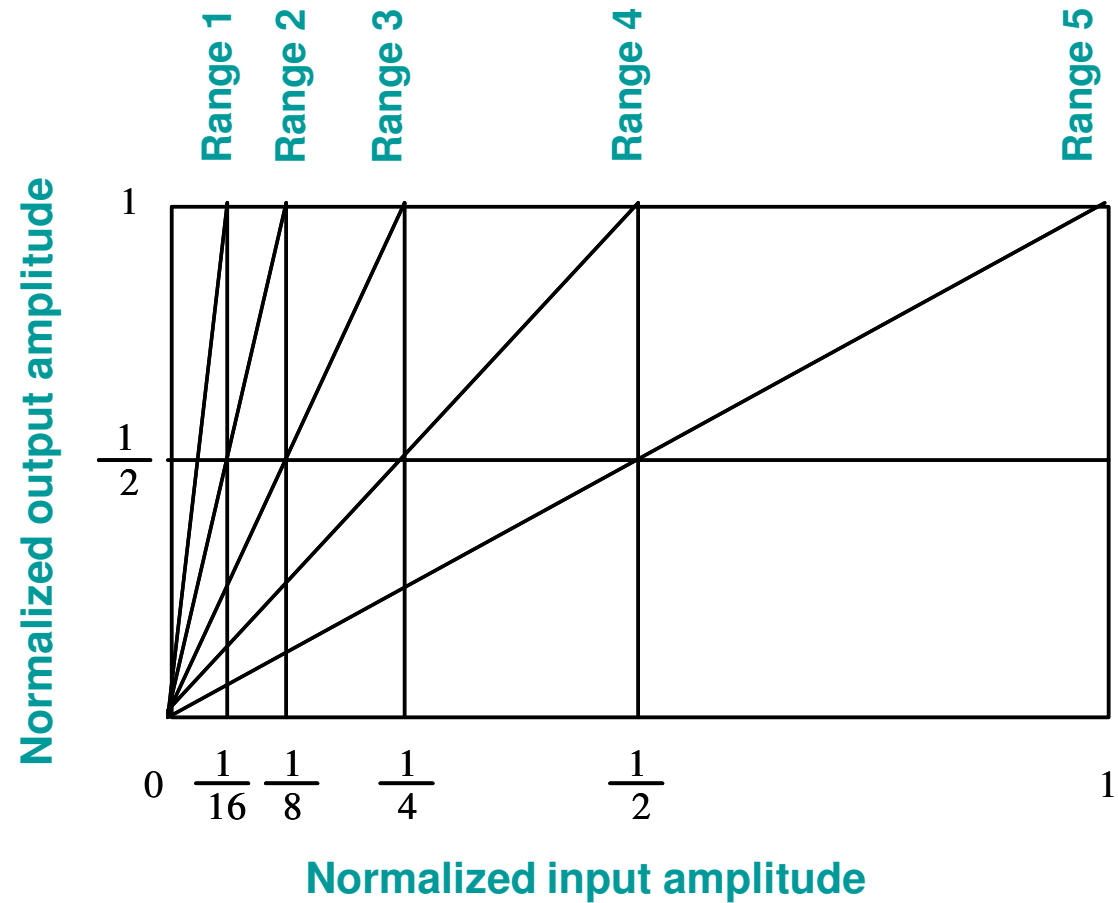
# MPEG-1 Audio: Layer 1

- **Blocks with 384 audio samples are coded (corresponding to 8 ms at 48 kHz)**
- **Signal is decomposed into 32 uniform subbands**
- **Fixed segmentation of 12 samples per subband ( $12 \times 32 = 384$  samples)**
- **APCM type quantization (block companding using a scale factor) for each subband with 0-15 bit/sample; this value may change for each subband; each scale factor ‘costs’ 6 bits (maximum  $6 \times 32 = 192$  bits/frame)**
- **Psychoacoustic models 1 or 2 suggested in the standard (there are 2 in the standard without normative value)**
- **Iterative rate/distortion adjustment to minimize the NMR (*Noise-to-Mask Ratio*) ratio for each subband**
- **Transparent quality regarding the CD quality (PCM) at 384 kbit/s which a typical compression factor of 4**

# Samples, Frames and Subbands ...



# Block Compressing or Scale Factors ...

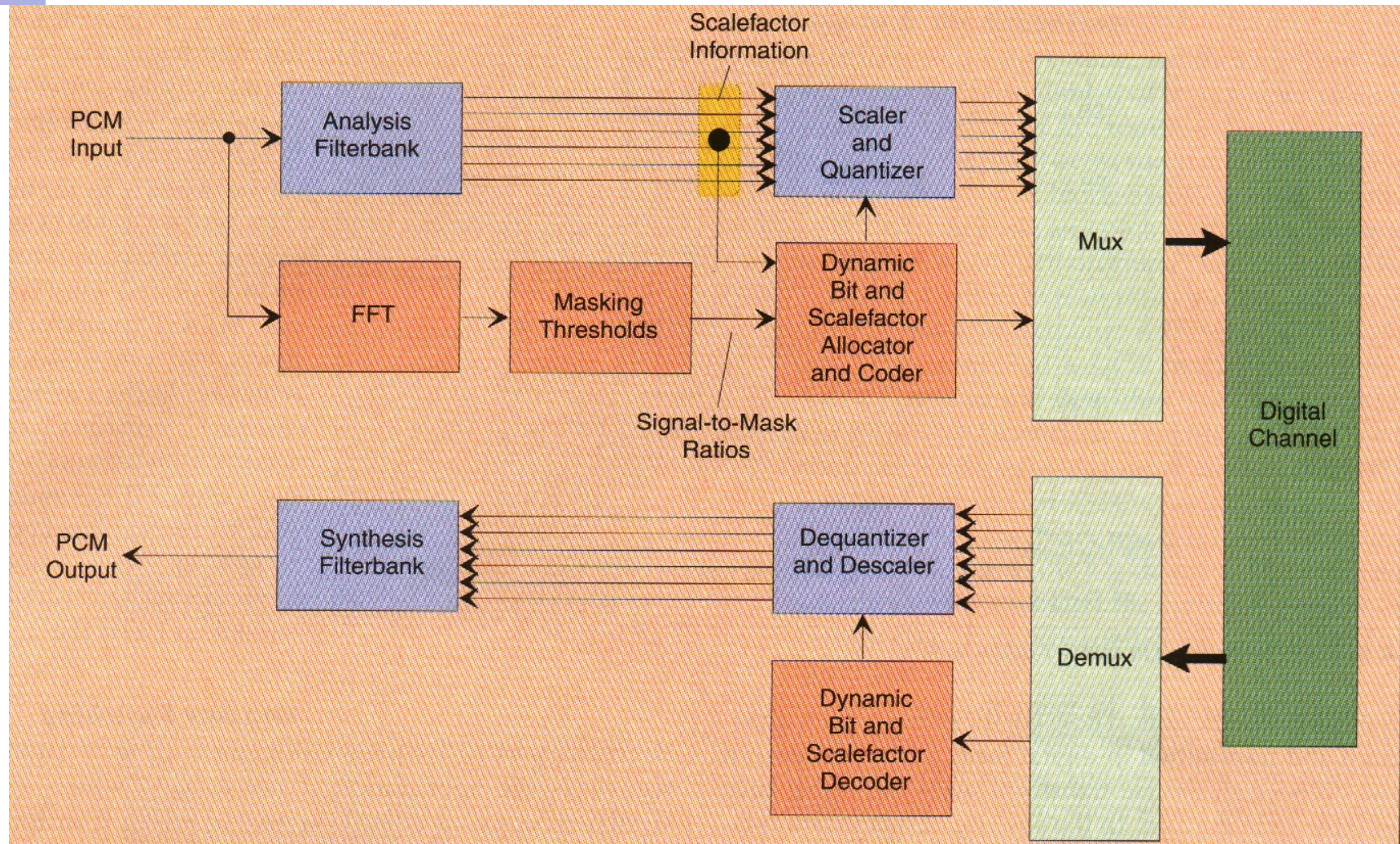




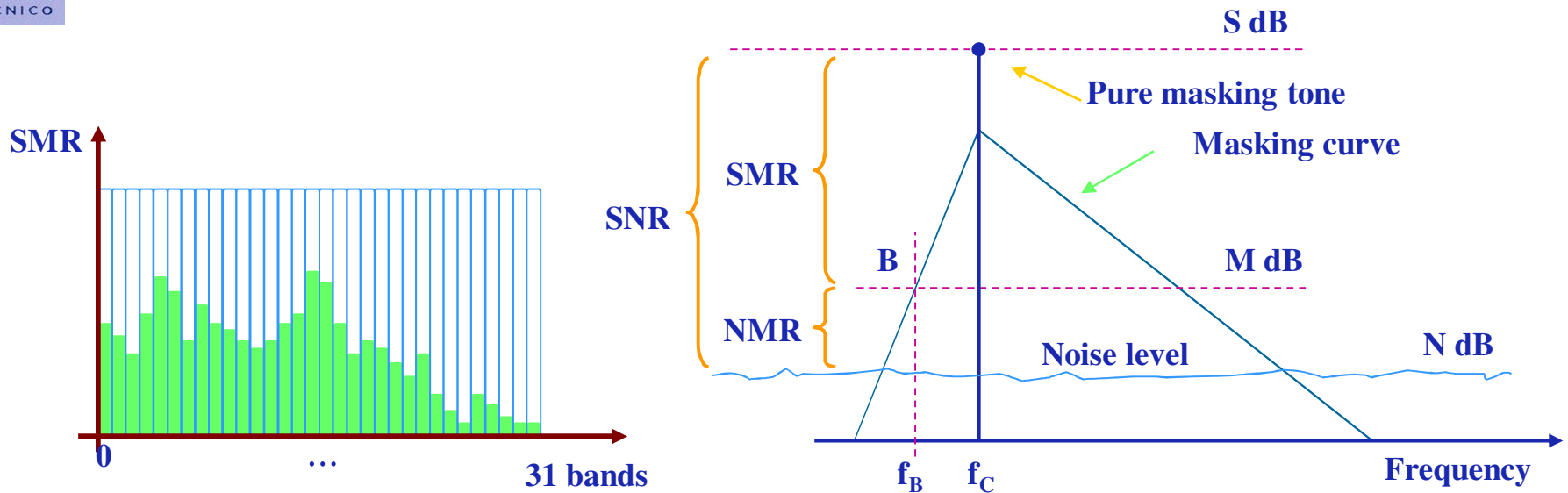
## MPEG-1 Audio: Layer 2

- **Blocks with  $3 \times 384 = 1152$  audio samples are coded (corresponding to 24 ms at 48 kHz)**
- **Fixed segmentation with  $3 \times 12 = 36$  samples per subband**
- **Coding algorithm as for Layer 1 with the exception of using more efficient methods to code the quantization scale factors by exploiting the redundancy between the adjacent scale factors within the 3 sub-blocks of 12 samples.**
  - Scale factors are shared among the 3 consecutive ‘granules’ for each sub-band.
  - When they are similar or when temporal post-masking can hide the distortion, only one or two scale factors need to be coded.
- **Transparent quality regarding the CD quality (PCM) at 192 kbit/s which a typical compression factor of 8**

# Layers 1 and 2 Encoder Architecture



# MPEG-1 Audio: Quantization



The psychoacoustic model is used to define the masking curve for each band, determining the noise allowed for each band and, thus, the number of quantization levels to use for the signal components above the masking threshold.

If the quantization noise remains below the masking threshold, the coded signal is subjectively indistinguishable from the original signal.



# MPEG-1 Audio, Layers 1 and 2: Quantization

- **The number of quantization levels for each subband is obtained through the dynamic allocation of bits controlled by the psychoacoustic model.**
- **The number of quantization levels for each subband is the one minimizing the NMR for that band, this means the one allowing ‘hiding’ more quantization noise and thus spending the minimum rate.**
- **The adoption of block companding determines the coding of samples normalized to the maximum value per subband (using a scale factor by subband), thus allowing to reduce the quantization error for the smaller samples.**
- **The normalized samples per subband (using the scale factors) are PCM coded with a varying number of bits/sample.**
- **When decoding, all samples of all bands without allocated bits are set to zero.**



# MPEG-1 Audio: Layer 3 (the Famous MP3 !)



- **Blocks with 1152 audio samples are coded (2 groups with 576 samples each)**
- **Hybrid time/frequency coding structure** - The filter bank (subbands) is followed by transform coding (Modified DCT).
- **Dynamic window switching** – To increase the frequency resolution, the 32 subbands are frequency subdivided by applying to each of them a transform with 6 or 18 coefficients; this results into a maximum number of frequency components of  $32 \times 18 (6) = 576$  (or 192). The smallest window allows to control the temporal resolution and thus to reduce the pre-echoe effect.
- **Overlapping windows** - The MDCT is applied with 50% window overlapping to reduce the block artefacts which means that the MDCT is applied to sets of 12 or 36 subband samples.

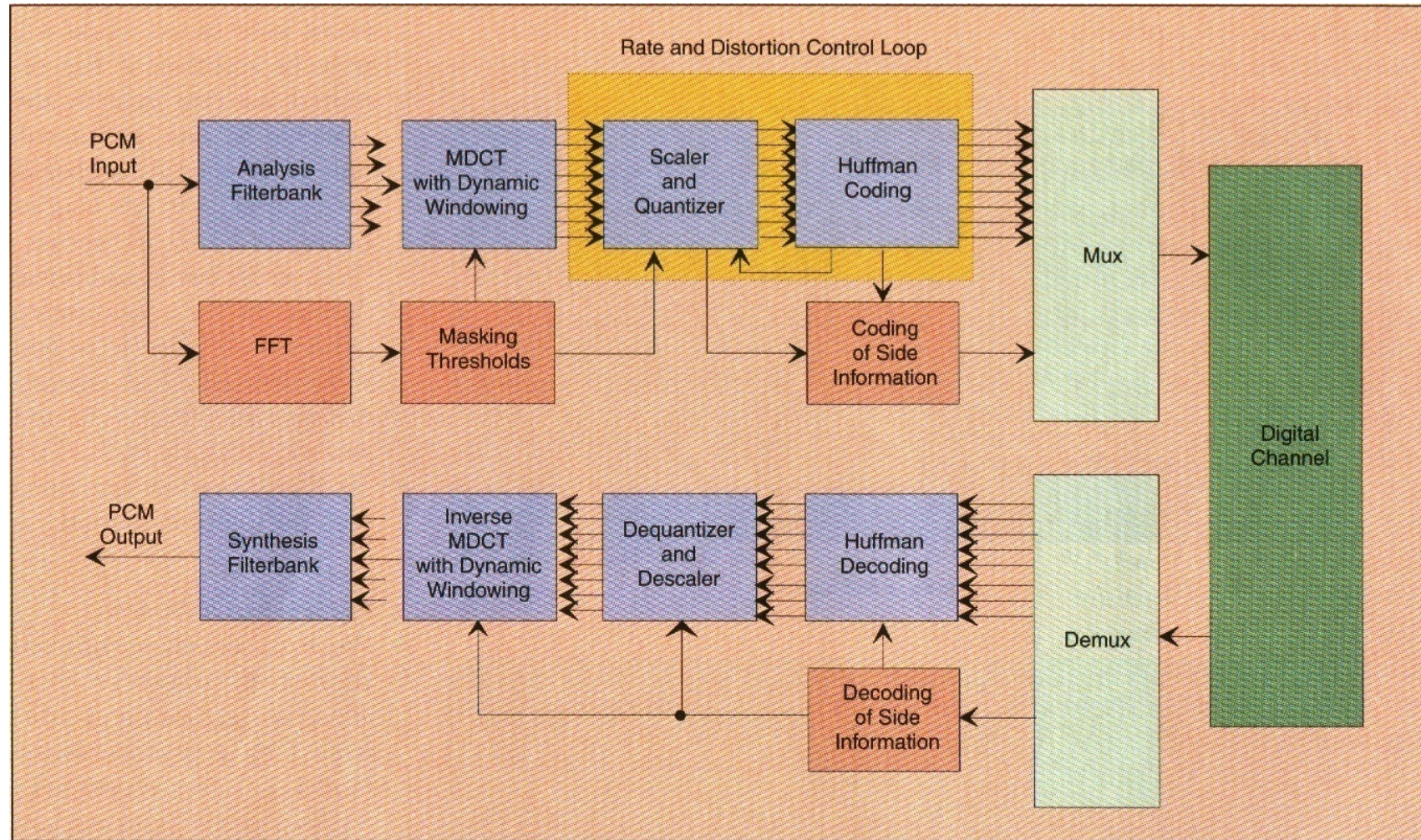


## MPEG-1 Audio: Layer 3 (the Famous MP3 !)



- **Quantization** - Non-uniform quantization of the MDCT coefficients (exponential like) introducing higher quantization error for the higher amplitude coefficients (where there is lower sensibility to errors); a mechanism with two nested cycles is typically used to control the quantization and coding.
- **Entropy Coding** - Huffman entropy coding of the quantized MDCT coefficients and the scale factors.
- **Psychoacoustic model** - Psychoacoustic model 2 suggested in the standard (more complex than model 1).
- **VBR** - More targeted to variable rate coding (useful for some applications)
- **Target** - Transparent quality regarding the CD quality (PCM) at 128 kbit/s which a typical compression factor of 12

# MP3 Encoder Architecture





# MP3 Encoding Walkthrough ...

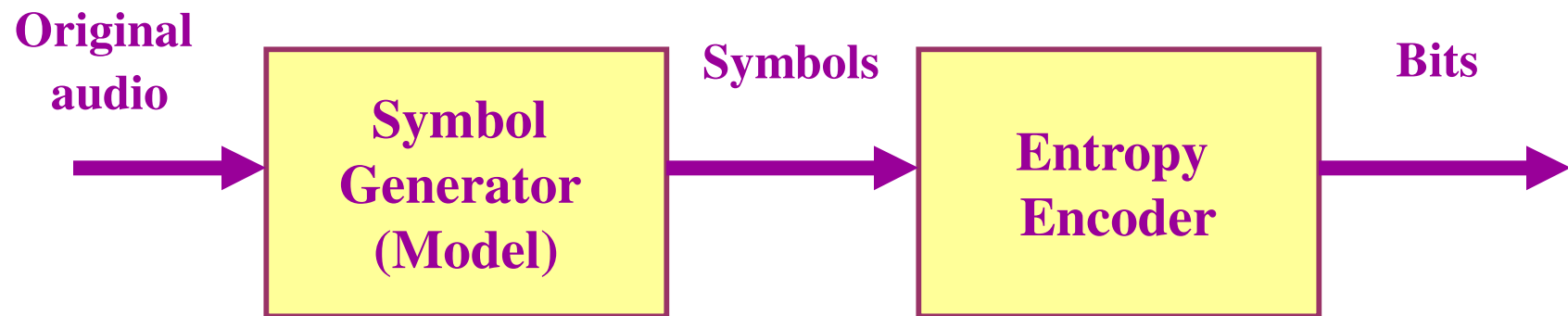
- **FREQUENCY DECOMPOSITION** - Divide the audio spectrum into 32 frequency bands (known as sub-bands) using a filter bank.
- **DCT TRANSFORM** - Calculate 6 or 18 DCT frequency components for each sub-band; use 6 frequency components when there is a need to control time artifacts (pre-echo and post-echo).
- **MASKING THRESHOLDS COMPUTATION** - Use the psychoacoustic model to compute the masking thresholds for the audio (or the allowed noise) for each spectrum partition.
- **QUANTIZATION** - Quantize the DCT components for each band using the defined quantization step and scale factor. If the quantization noise can be kept below the masking threshold, then the compression results should be indistinguishable from the original signal.
- **ENTROPY CODING** – The quantized DCT components for each band are entropy coded.



# MP3 Performance ...

Sound quality	Bandwidth	Mode	Bitrate	Compression factor
telephone sound	2.5 kHz	mono	8 kbps *	96:1
better than shortwave	4.5 kHz	mono	16 kbps	48:1
better than AM radio	7.5 kHz	mono	32 kbps	24:1
similar to FM radio	11 kHz	stereo	56...64 kbps	26...24:1
near-CD	15 kHz	stereo	96 kbps	16:1
CD	>15 kHz	stereo	112..128kbps	14..12:1

# The MP3 Symbolic Model



**An audio sequence is represented as a succession of frames, each with a certain number of audio samples, represented using MDCT coefficients for each subband, quantized based on a psychoacoustic model.**



# Stereo ... and More ....



# MPEG-1 Audio: Stereo Coding



iPod Not Included

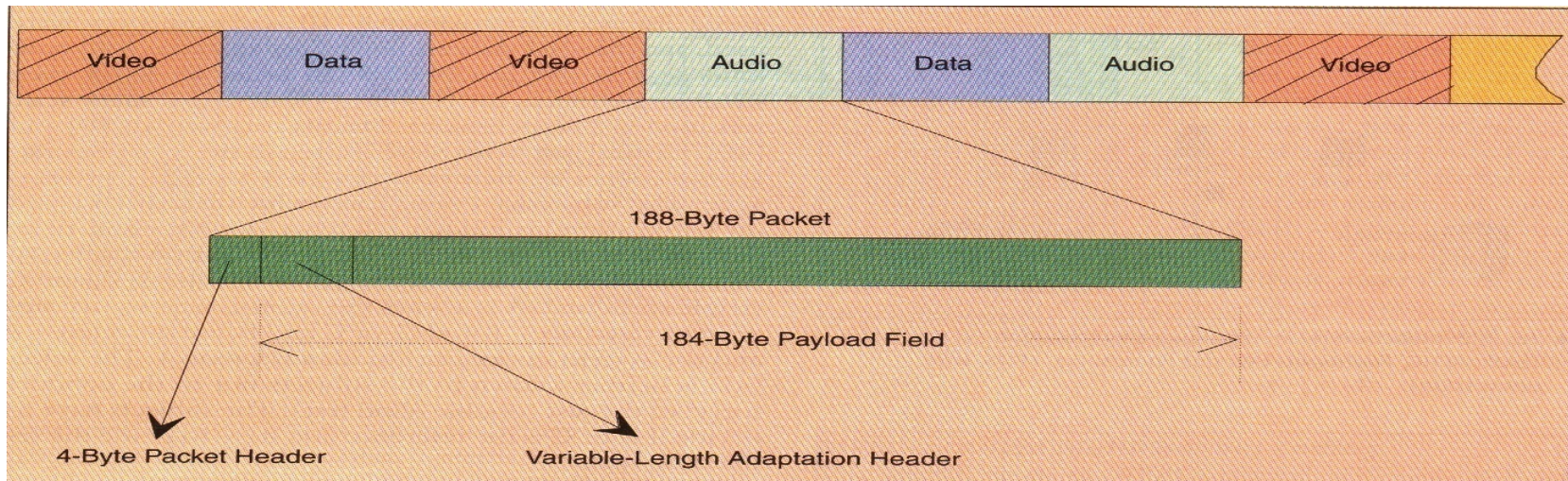
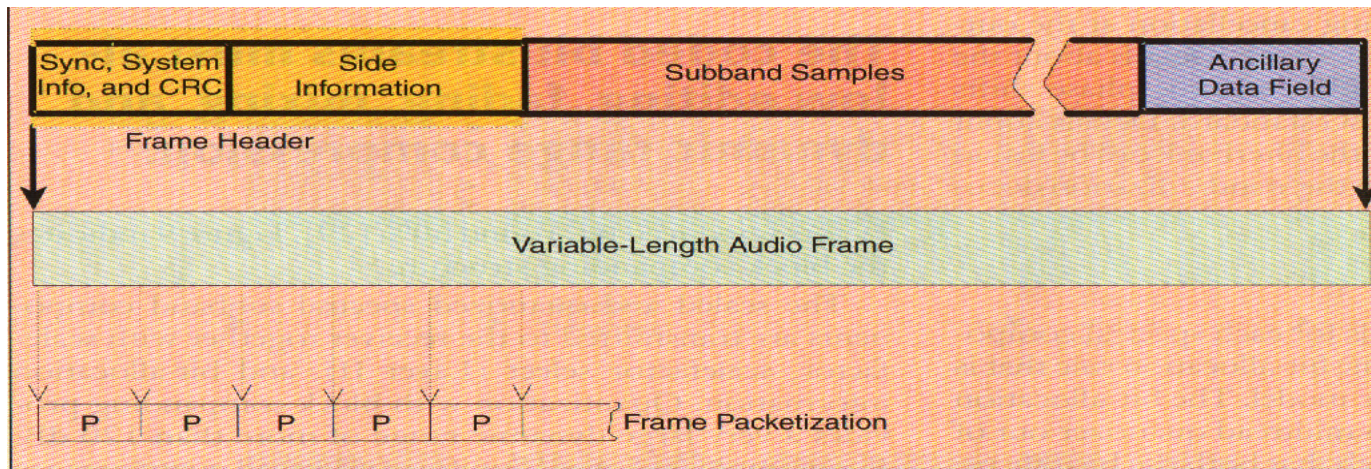
**Joint stereo coding takes advantage of the fact that the two channels of a stereo pair contain redundant information. These stereophonic irrelevancies and redundancies are exploited to reduce the total bitrate. Joint stereo is used in cases where only low bitrates are available but stereo signals are desired.**

**There are 5 audio coding modes:**

- **Mono**
- **Dual Stereo** – Channels are independently coded, e.g. 2 different languages.
- **Stereo** – Independent coding but sharing certain fields in the coded frame.
- **Joint Stereo** – Channel dependency is exploited through the so-called *intensity stereo technique*; above 2 kHz, the L+R signal is coded together with scale factors for the 2 channels (L and R) since there is lower hearing sensibility.
- **Mono/Stereo (MS)** (only layer 3) – The 2 channels are coded as L+R (middle) and (side) L-R allowing to better control the spatial location of the quantization noise.

# Audio Syntax and Systems Packet Level

CRC is optional !





# MP3 Licensing

## PC Software Applications

mp3	Decoder	• US\$ 0.75 per unit or US\$ 50 000.00 - US\$ 60 000.00 one-time paid-up
	Codec	• US\$ 2.50 - US\$ 5.00 per unit
mp3PRO	Decoder	• US\$ 1.25 per unit or US\$ 90 000.00 one-time paid-up
	Codec	• US\$ 5.00 per unit

## Hardware Products

mp3	Decoder	• US\$ 0.75 per unit
	Codec	• US\$ 1.25 per unit
mp3PRO	Decoder	• US\$ 1.25 per unit
	Codec	• US\$ 5.00 per unit

## ICs / DSPs

For available software, supported platforms, porting and licensing options, please [contact](mailto:info@mp3licensing.com) us at [info@mp3licensing.com](mailto:info@mp3licensing.com).

## Games

mp3	• US\$ 2 500.00 per title
mp3PRO	• US\$ 3 750.00 per title

## Electronic Music Distribution / Broadcasting / Streaming

mp3	• 2.0 % of related revenue
mp3PRO	• 3.0 % of related revenue

© Original Artist

Reproduction rights obtainable from  
www.CartoonStock.com

© Mike Baldwin / Comed



"This next block of silence is for all you folks who download music for free, eliminating my incentive to create."

**With MP3, it is effectively easier to 'pirate' music**

...

**Which does not mean one should do it**

...

**Or even that it is advantageous to do it, at least in the long term ...**

## Final Remarks

- **There is a growing number of devices, applications, etc, notably portable, including MP3 players.**
- **MPEG-1 Audio Layer (MP3) is commonly used for music in the Web.**
- **Digital Audio Broadcasting (DAB) and Digital Video Broadcasting (DVB) used for a long time only MPEG-1 Audio Layer 2.**
- **MP3 provoked the explosion of one of the biggest current multimedia problem this means digital rights management ... and Napster ... and peer-to-peer ...**





# Bibliography

- *MPEG Video Compression Standard*, J. Mitchell, W. Pennebaker, C. Fogg, D. LeGall, Chapman & Hall, 1996
- *Video Coding: an Introduction to Standard Codecs*, M. Ghanbari, IEE Press, 1999
- *Multimedia Communications*, F. Halsall, Addison-Wesley, 2001
- *Introduction to Digital Audio Coding and Standards*, M. Bosi, R. E. Goldberg, The Springer International Series in Engineering and Computer Science, 2003