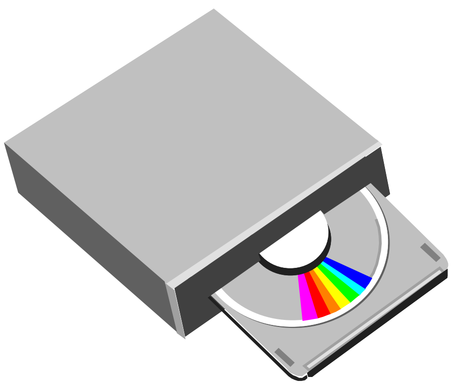
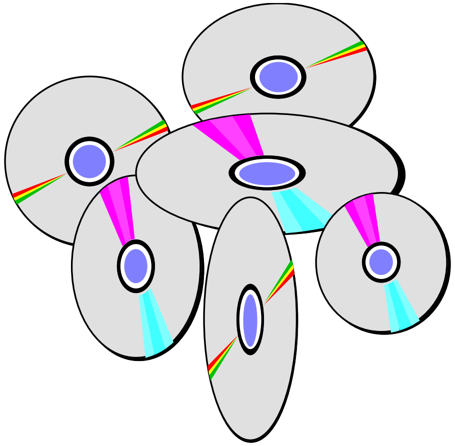
DIGITAL AUDIOVISUAL STORAGE: FIRST GENERATION





Fernando Pereira

Instituto Superior Técnico

New Application Domain ...

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.

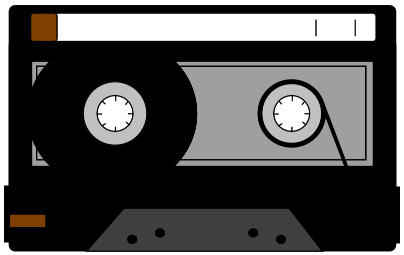
One of the most important technologies for AV storage is <u>audiovisual data coding</u> which should provide the necessary compression efficiency and quality but also other storage functionalities such as random access, already provided in analogue recording.

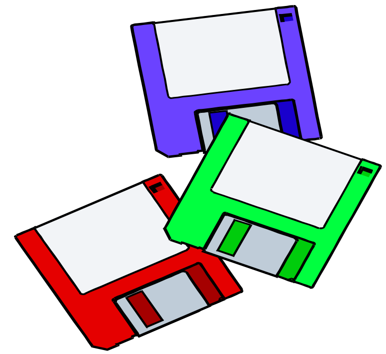
Main Recording Functionalities …

- **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.**

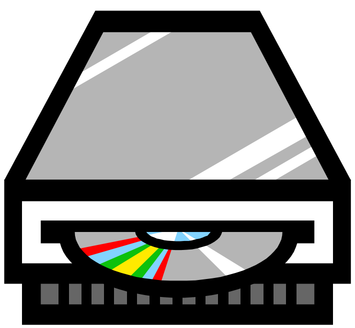
TÉCNICO LISBOA

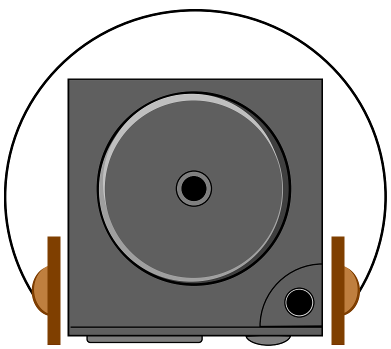
Main Storage Supports





- Magnetic tape
- Magnetic discs
- Optical discs
- …





Storage: Which Support ?

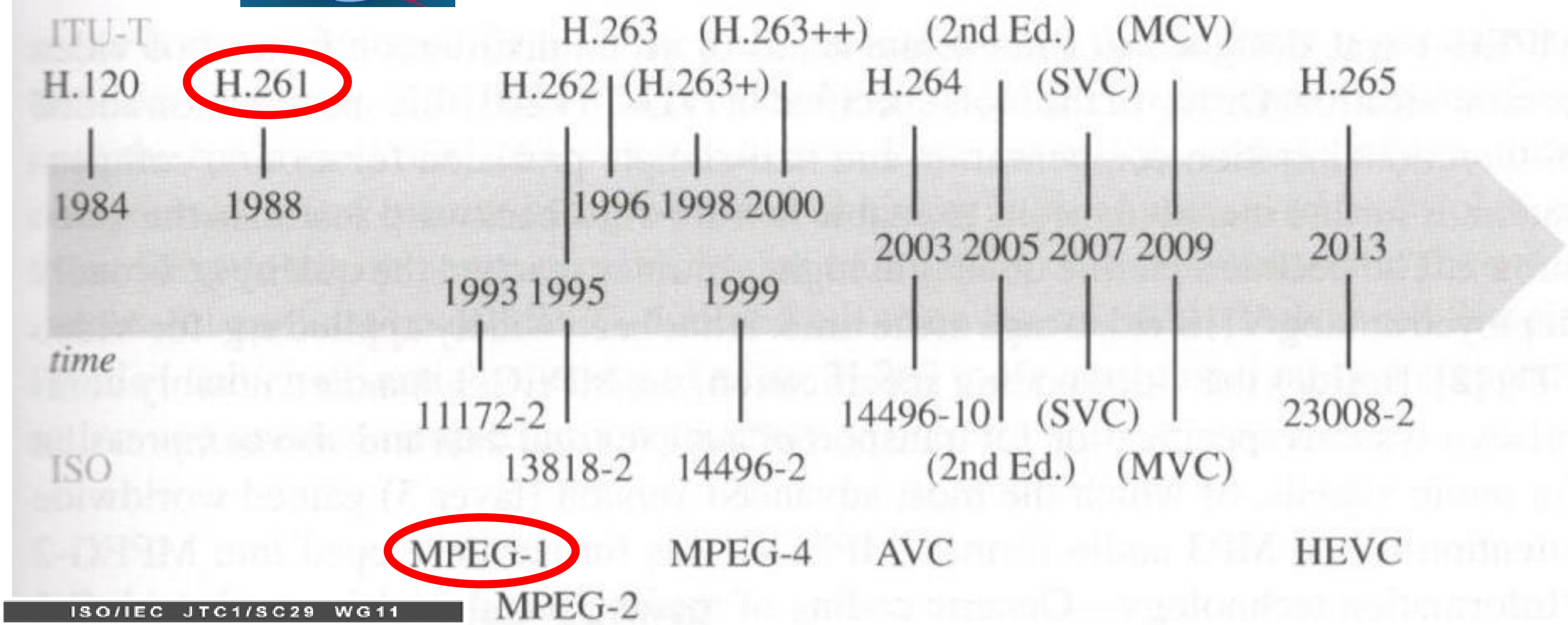
Main factors to be taken into account to select a multimedia storage support:

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

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 and also frames (Part 2); the Advanced Video Coding (AVC) standard, which has been jointly developed with ITU-T (H.264), is MPEG-4 Part 10**
  - **All types of applications**
- **MPEG-H (2012-?): High efficiency coding and media delivery in heterogeneous environments; the High Efficiency Video Coding (HEVC) standard, which has been recently jointly developed with ITU-T (H.265), is MPEG-H Part 2**
  - **All types of applications; newly targeting Ultra High Definition (UHD) content**

Standards Over Time …



MPEG-1 Standard

Motivation (~1990)

- 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 ?

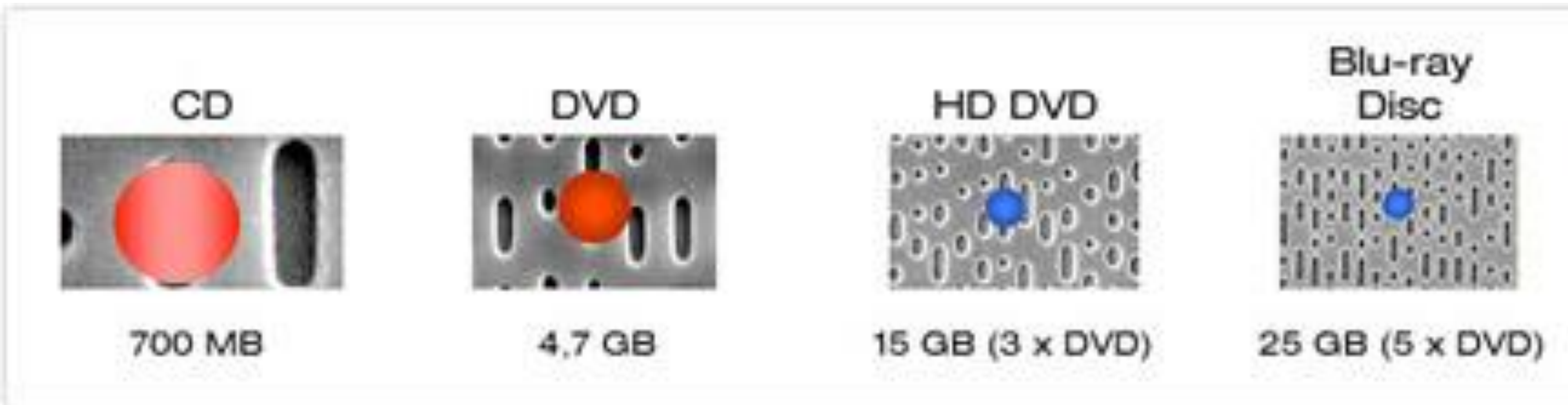
Main factors to be taken into account to select an audiovisual storage support:

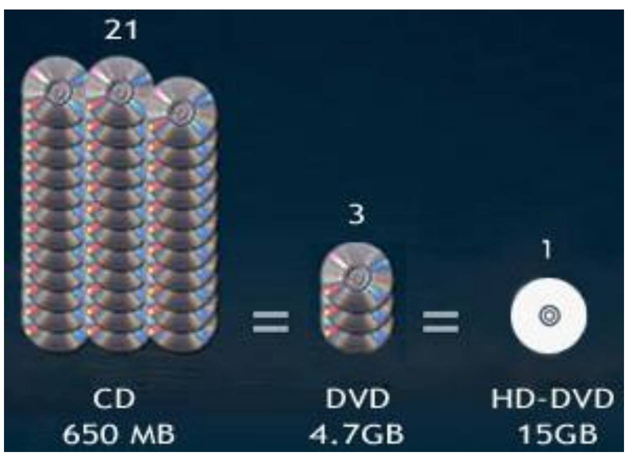
- **Capacity (in MBytes)**
- **Reading speed (in Mbit/s)**
- **Time and form of access (e.g. sequential or random)**
- **Durability**
- **Mobility**
- **Cost**
- **…**

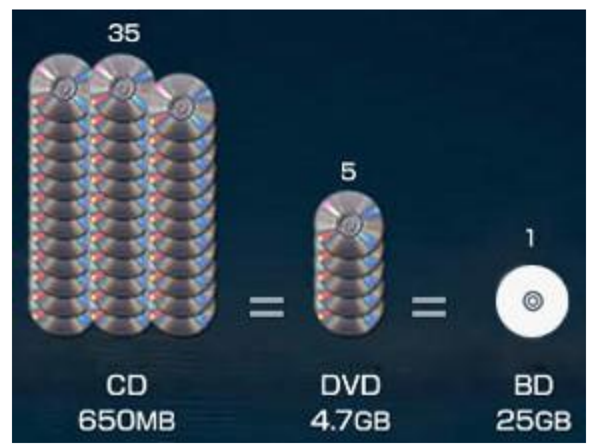
The CD-ROM was selected as the most adequate storage support to offer, for the first time in large scale, interactive multimedia signals, mainly due to its large capacity and low cost.

This means putting generic audio and video coded data with acceptable quality in 1.5 Mbit/s.

Making Comparisons up to Blu-ray …







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, (generic) 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

Multimodalities and Multiplexing ...

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:

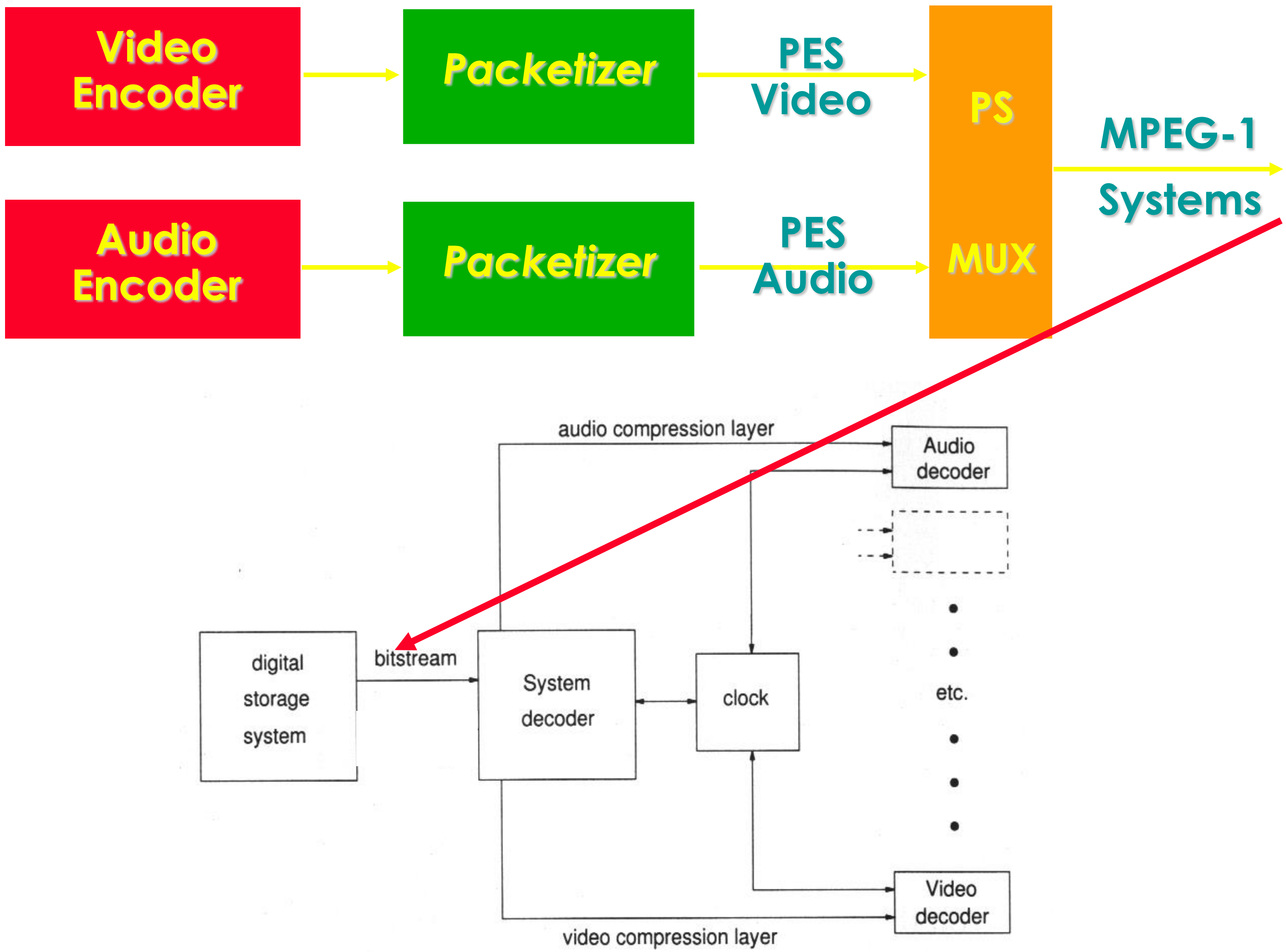
- **Syntax for the streams offering <u>timing control</u>**
- **<u>Multiplexing and synchronization</u> of the audio and video streams**

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).



Video Encoder
Packetizer
PES Video
Audio Encoder
Packetizer
PES Audio
PS MUX
MPEG-1 Systems
audio compression layer
Audio decoder
digital storage system
bitstream
System decoder
clock
etc.
Video decoder
video compression layer

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*.

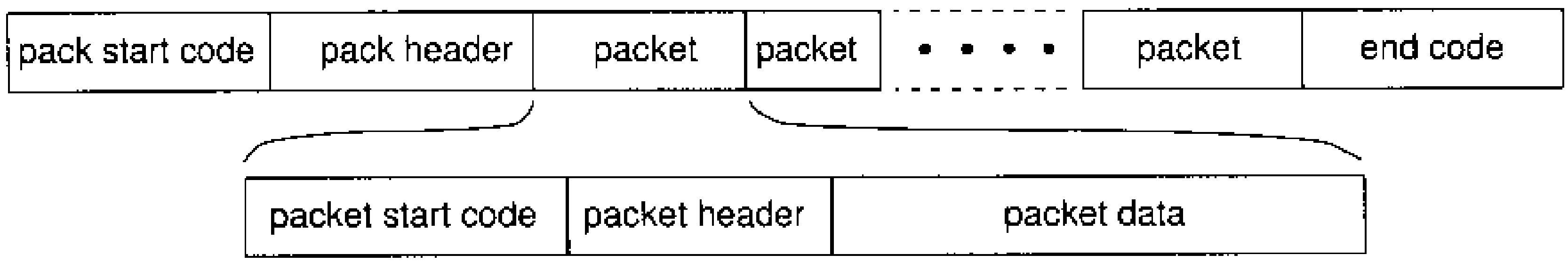
The MPEG-1 Systems stream is structured in two sub-layers:

- ***PACK sub-layer*** **– Refers to 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; <u>packets may have a fixed or variable length</u>.**

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

MPEG-1 Systems Stream Syntax



- **One <u>MPEG-1 Systems stream</u> consists in a <u>sequence of packs</u>, each one containing <u>several packets</u> (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 <u>variable length</u>.**
- **One pack corresponds to the audiovisual data for a certain period of time.**
- **The Systems decoder parses the MPEG-1 stream, giving to the audio and video decoders their respective packets, after inspecting the *packet headers*.**
- **At most, 32 audio streams, 16 video streams and 2 data streams may be multiplexed in a single MPEG-1 stream.**

MPEG-1 Systems: Timing

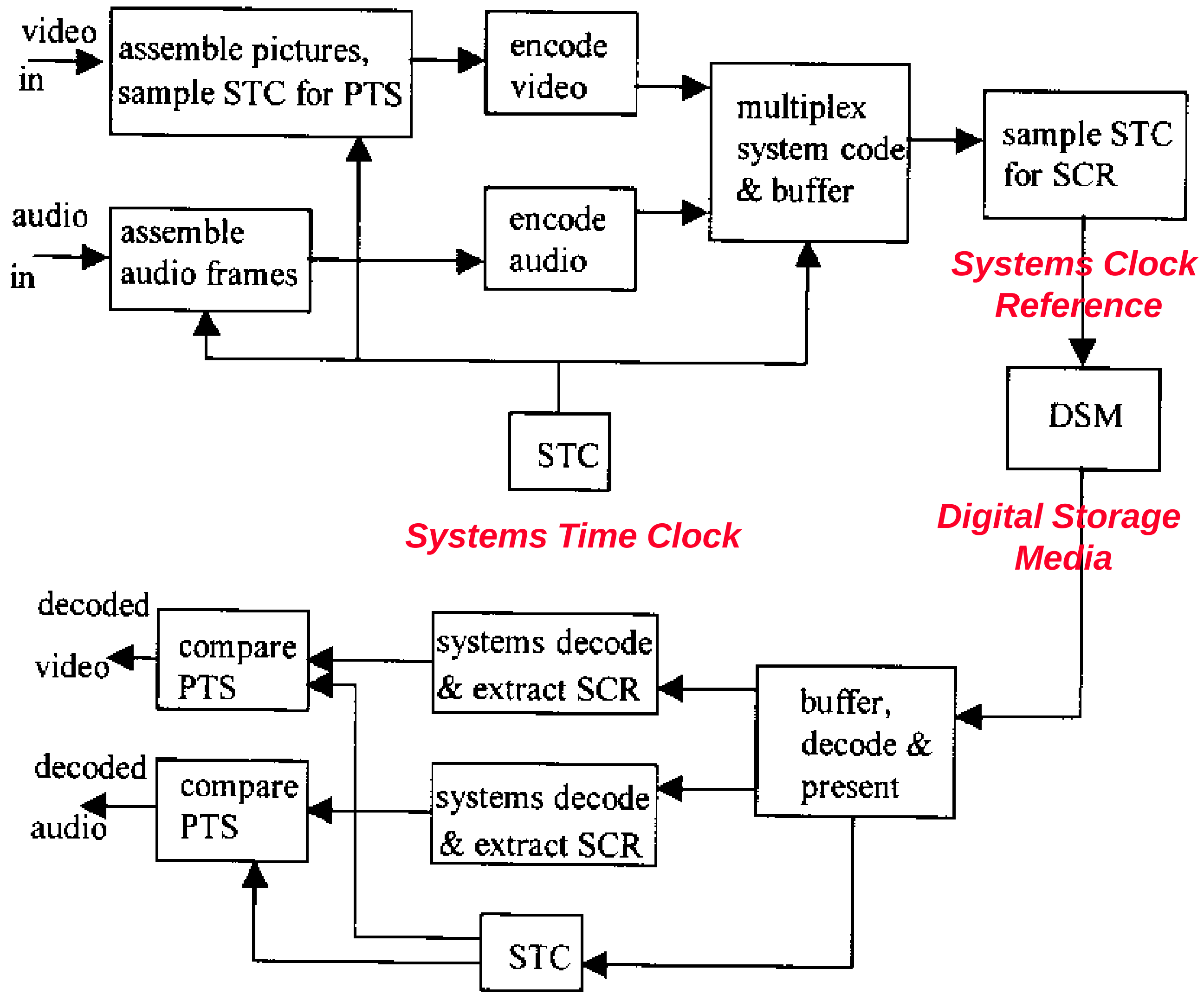
- ***Decoding Time Stamp* (DTS) – Timing information that may be present in the packet header to indicate the moment when the corresponding coded information <u>must be decoded</u> in the *Systems Target Decoder* (STD).**

- ***Presentation Time Stamp* (PTS) – Timing information that may be present in the packet header to indicate the moment when the corresponding decoded information <u>must be presented</u> 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 frames in video; the STD assumes instantaneous decoding.

MPEG-1 Systems 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

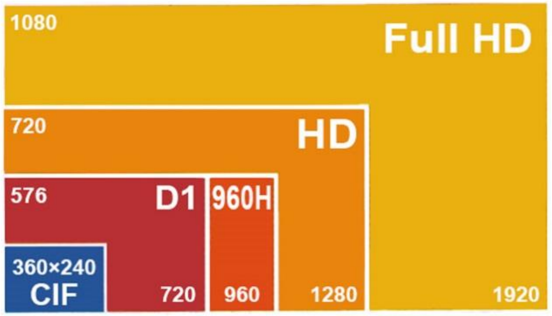
- **Signals -** The signals for each image are the **luminance (Y) and two chrominances**, designated as C_B and C_R or U and V. The R,G,B primary signals have a gamma correction around 2.2 - 2.8.

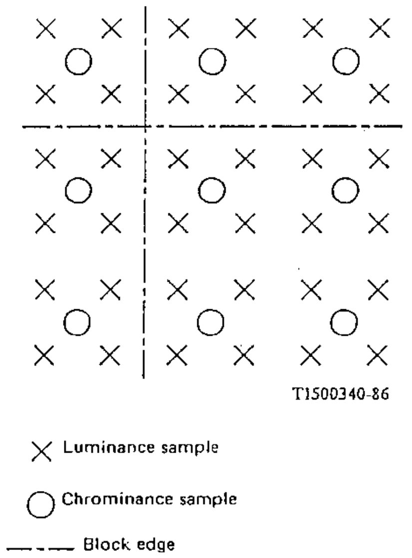
- **Spatial resolution –** Typical spatial resolution is CIF (Common Intermediate Format) with **288×352 luminance samples.**

- **Frame rate –** Typical frame rate is **25/30 Hz**.

- **Colour subsampling -** 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**.

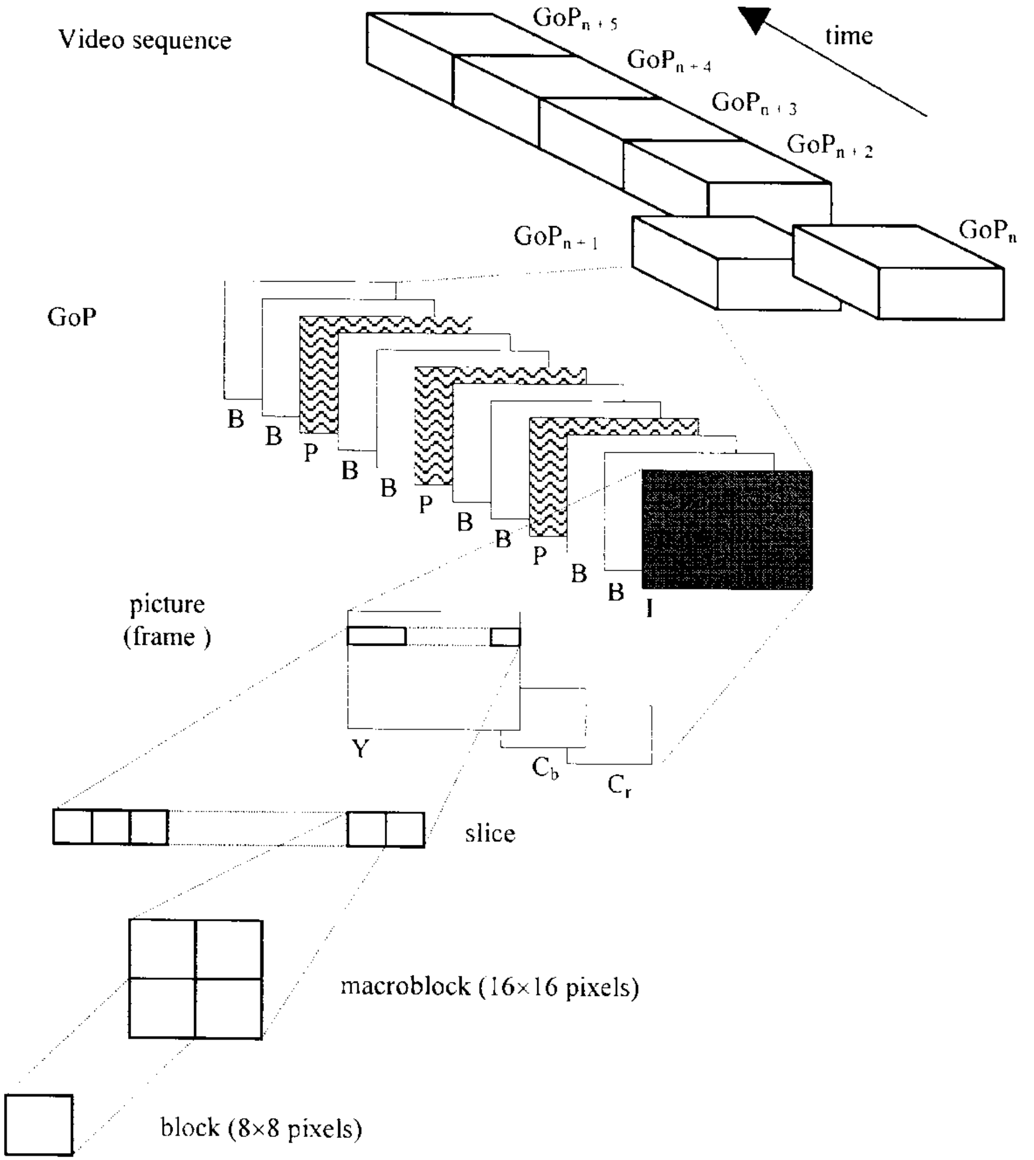




Video Structure

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; ½ pixel accuracy)

- **Spatial Redundancy**

Discrete Cosine Transform (DCT)

- **Statistical Redundancy**

Huffman entropy coding

Lossy :-(

- **Irrelevancy**

DCT coefficients quantization

Squeezed between Rate and Quality …

The MPEG-1 Video standard had to fulfil:

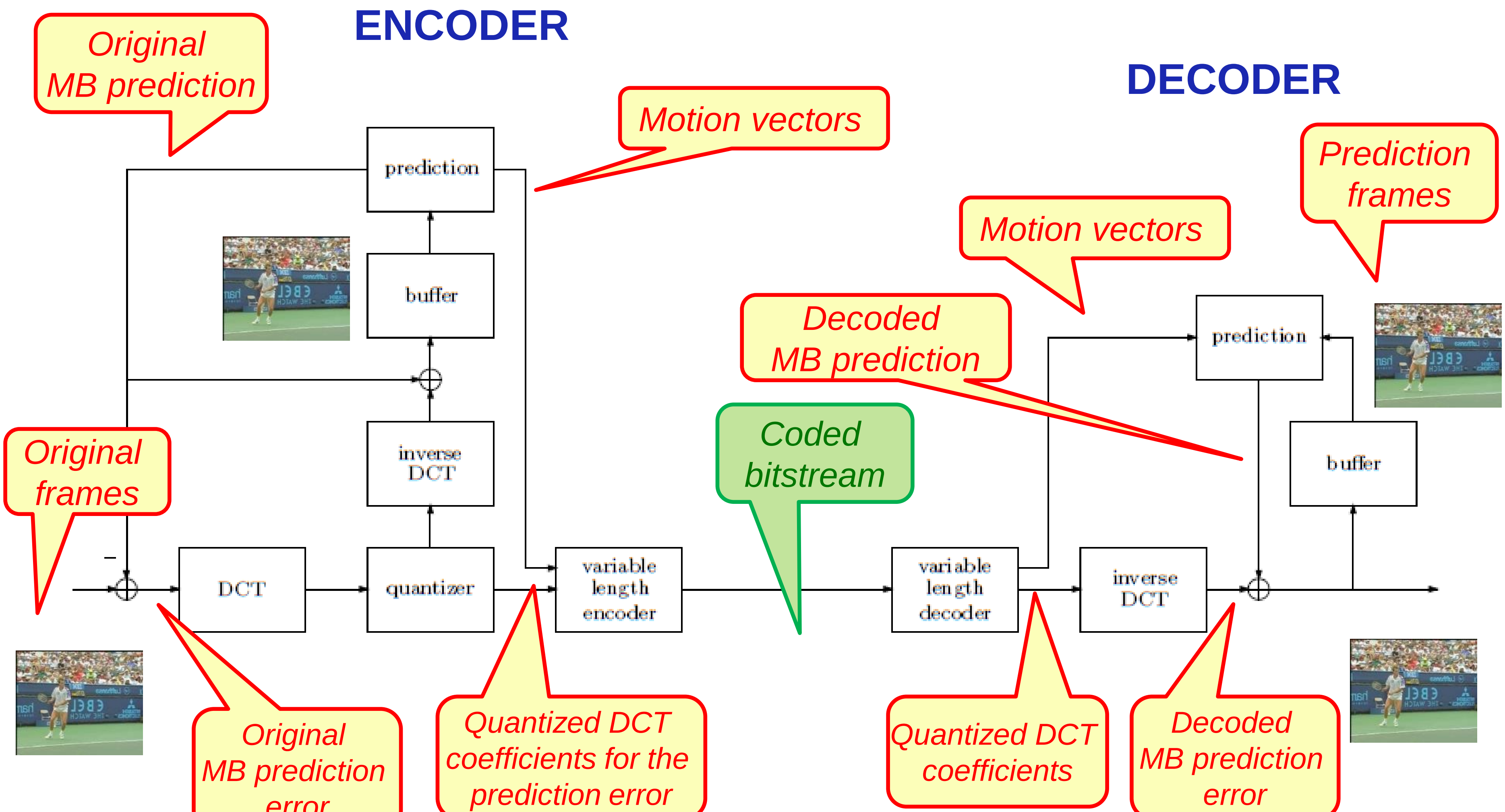
1. **Quality Target -** Better video quality than the VHS tapes that MPEG-1 was targeting to substitute (while offering random access)
2. **Rate Target -** At most 1.2 Mbit/s since the CD-ROM rate was limited to around 1.5 Mbit/s (the remaining 300 kbit/s should be for stereo audio and multiplexing/synchronization)

To solve this dilemma, MPEG-1 Video had to 'buy' quality not with rate but with

1. **Additional Encoder complexity (computational and memory)**
2. **Additional Delay**

Starting with the Same Architecture …

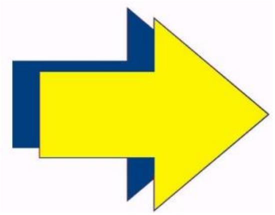
Buying Quality with Computation, Memory and Delay …



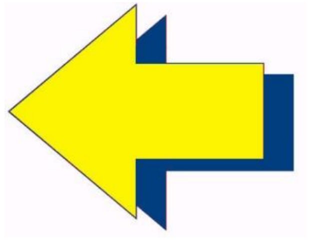
Exploiting the Temporal Redundancy

Better Predictions with more Motion Data ...

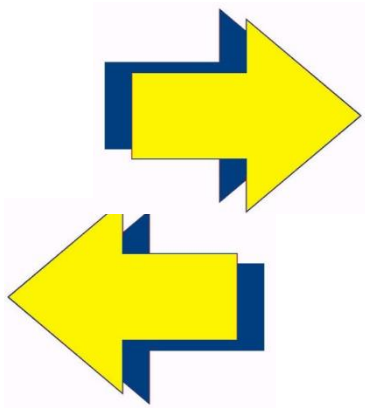
- **FORWARD PREDICTION – It is based on the principle that, locally, each image (or part thereof) may be represented from one or more previous images after a translation.**



- **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.**

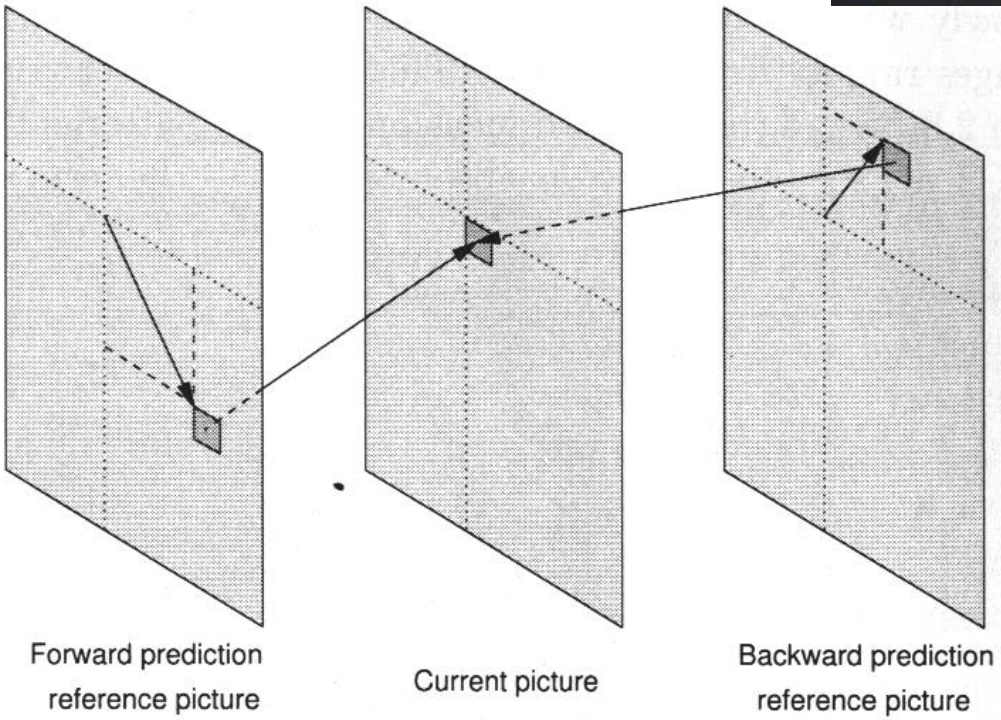


- **BIDIRECCIONAL PREDICTION – It is based on the principle that, locally, each image (or part thereof) may be represented from a previous image (forward prediction), a future image (backward prediction), or a combination thereof, after corresponding translations.**



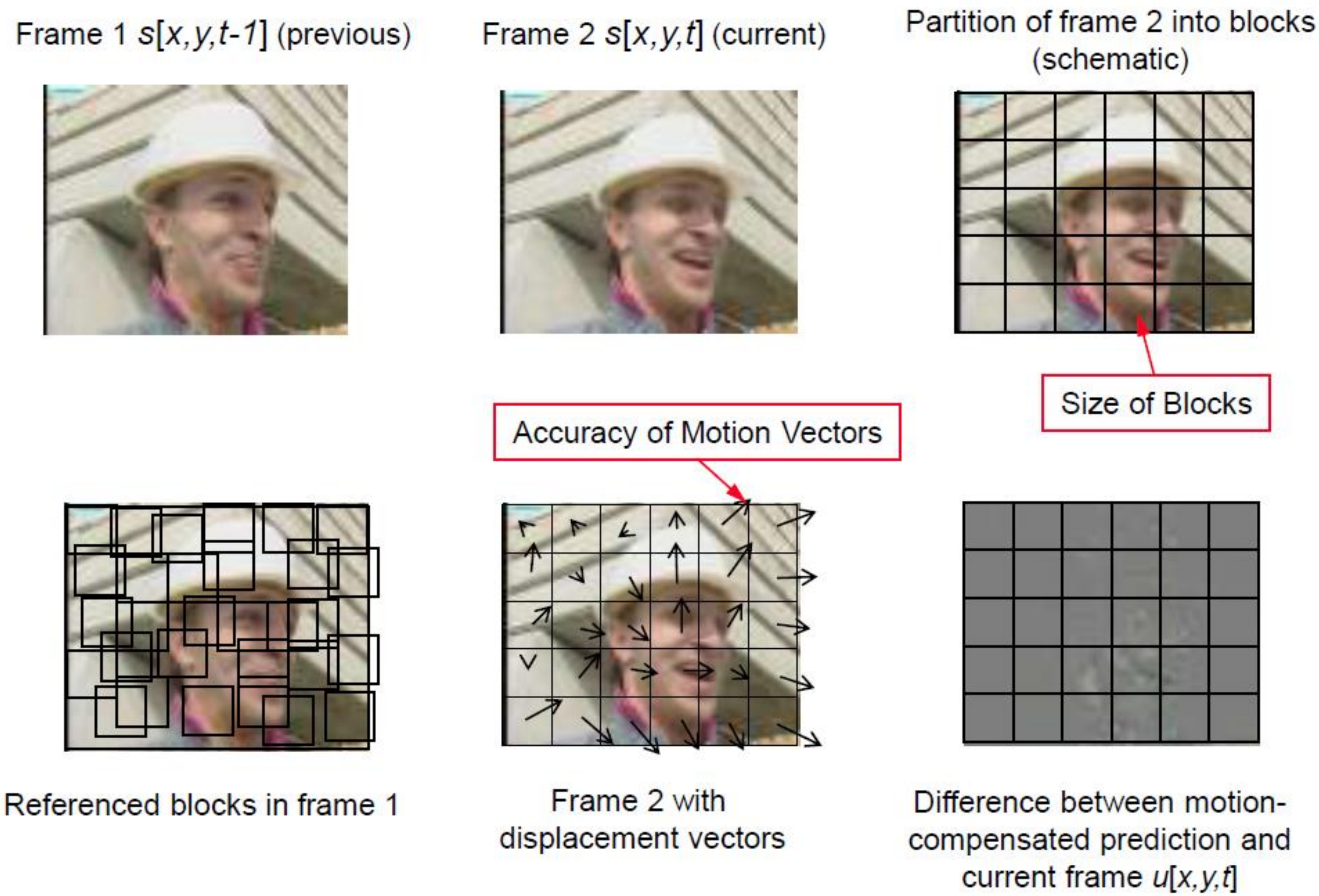
Predicting also from the Future …

- **Bidirectional predictions ‘buy’ quality with computational complexity, memory and DELAY !**
- **This is possible especially if the application can accept the additional delay …**

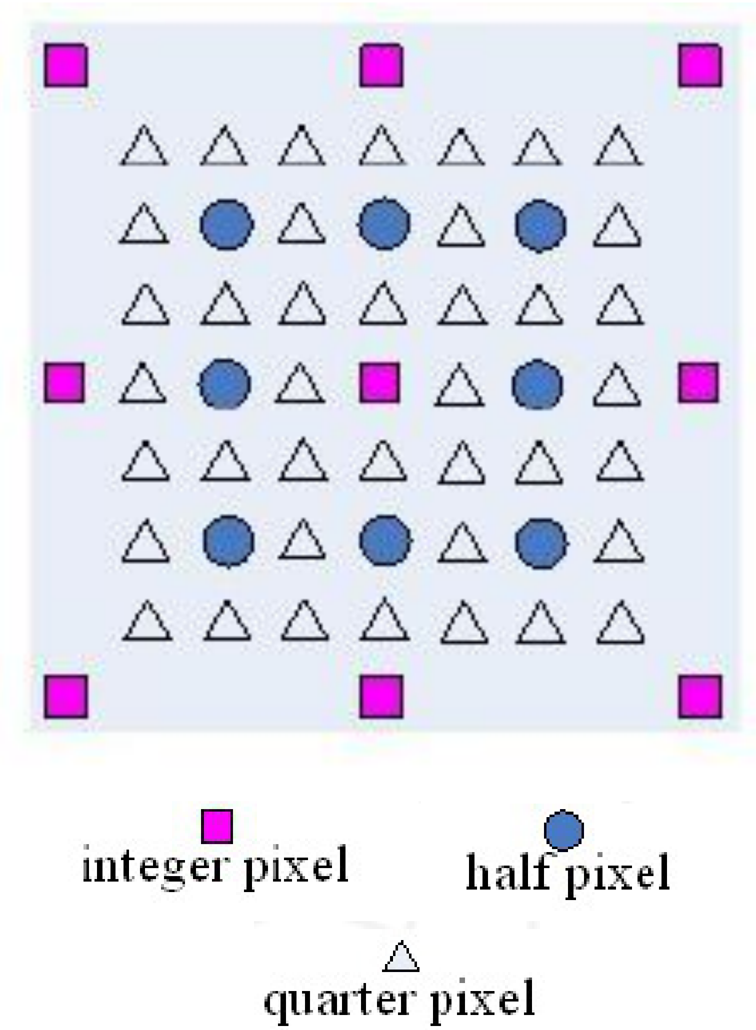


Bidirectional prediction allows to reach better predictions (there is more information available to make the prediction) and, thus, to reach better RD performance in certain conditions; for example, it is useful to deal with uncovered backgrounds.

Coding with Motion Compensation …

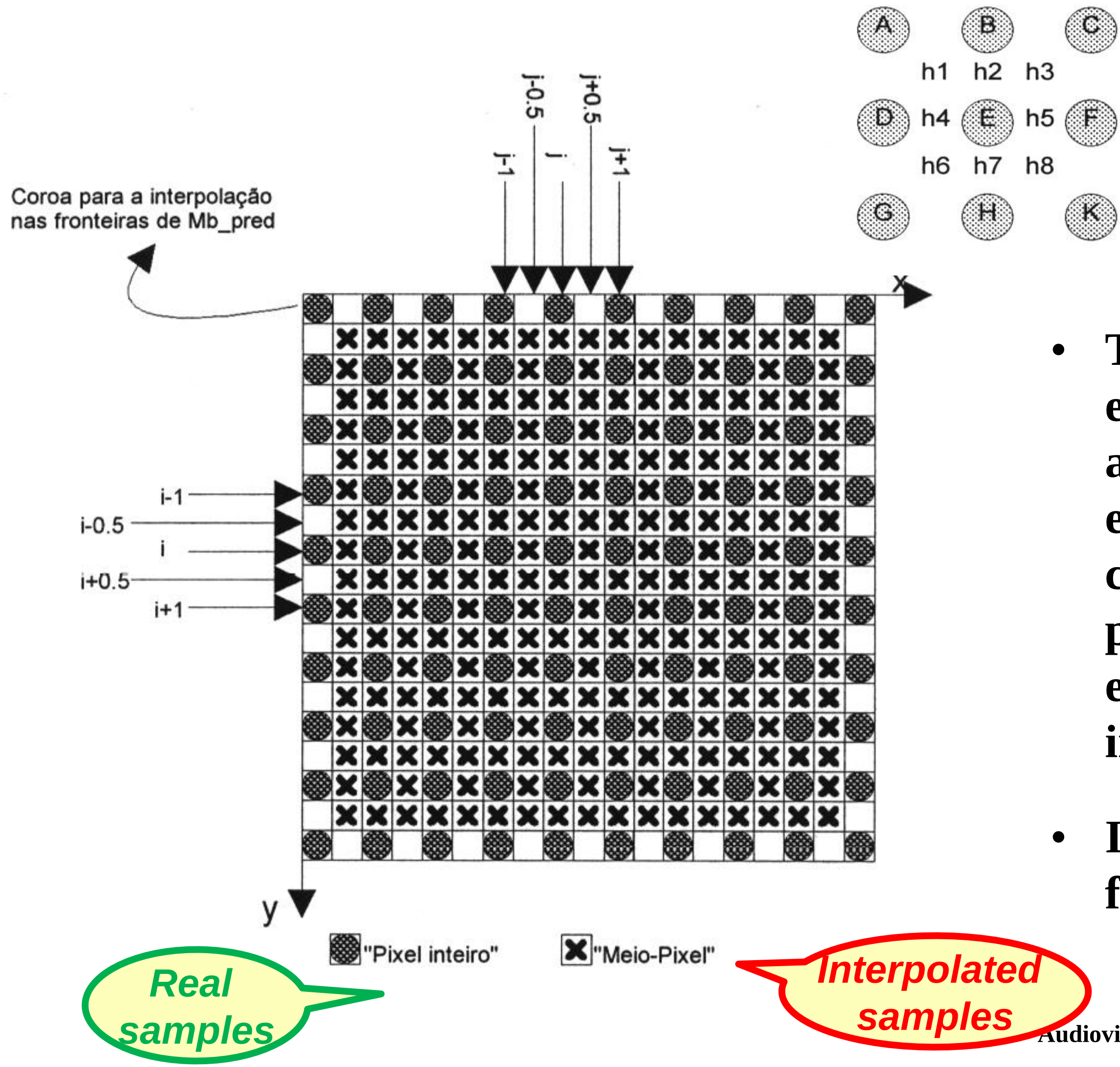


Inventing Samples: ½ and ¼ Pixel



- By interpolating additional samples, more prediction opportunities are created
- By adding complexity, compression efficiency may be increased
- The additional compression gains fade out at some stage ...

Better Predictions with More Accuracy: ½ Pel ME/MC



- **The ½ pixel motion estimation accuracy allows a more precise motion estimation with the consequent reduction of the prediction error (and encoder complexity increase).**

- **Its usage is signalled at the frame level.**

MPEG-1 Video: Motion Estimation

- Spatial support - Motion estimation and compensation is performed at the macroblock level (16×16 luminance samples).
- Optional - Motion estimation and compensation are always optional, meaning that the encoder may decide to use it or nor (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 many 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.
- Macroblock matching - Since the bitstream syntax allows using up to two motion vectors per MB, macroblock matching motion estimation is the most used solution.
- Prediction 'breaks' - Bidirectional prediction cannot be applied to all frames of a sequence due to the delay constraints; thus, there is a need to define relatively close prediction anchors (which do not predict from the future).

Exploiting the Spatial Redundancy and the Irrelevancy

After Time, the Space …

Original

Actual image

+

+

-

To be coded residual

DCT Transform

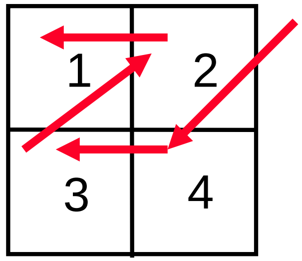
Prediction error

Decoded

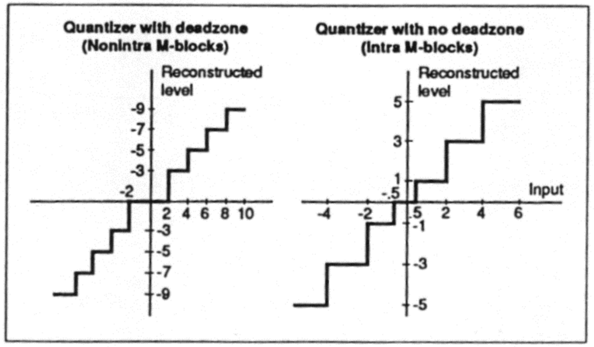
Prediction image, motion compensated at MB level

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

- **Support -** The DCT is applied to 8×8 blocks of samples (N=8).
- **Mismatch control -** 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.
- **DCT coefficients selection -** 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.
- **Zig-zag scanning -** The quantized DCT coefficients in each block are zig-zag scanned to assure that they are transmitted according to their (decreasing) subjective relevance.
- **Differential coding -** 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 reconstruction levels 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 default quantization matrix is different for Intra and Inter coded MBs. These matrices may be changed to more adequate matrices for the cases at hand, naturally paying the necessary bitrate cost.
- For Intra coded MBs (full energy), the DC coefficient is always quantized with step 8.

Default Quantization Matrices

QUANTIZATION STEP = Quantization Factor ×

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

INTRA

16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

INTER

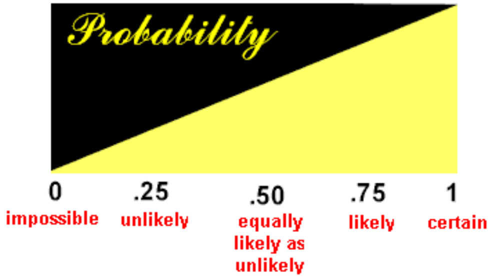
For Inter coding, the high frequency coefficients are not necessarily associated to high frequency image content since they may result from block artifacts in the reference image, poor motion compensation, or camera noise.

Thus, for Inter coding, it is not appropriate to apply psychovisual criteria in defining the quantization matrices (for the prediction error).

Exploiting the Statistical Redundancy

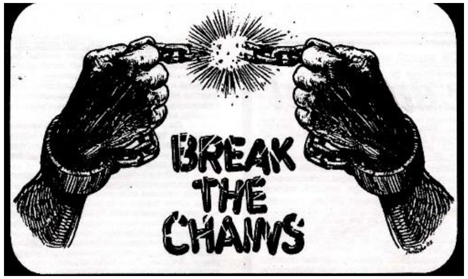
MPEG-1 Video: Huffman Coding

- **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 a symbol, the less bits it deserves, 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**
  - **...**



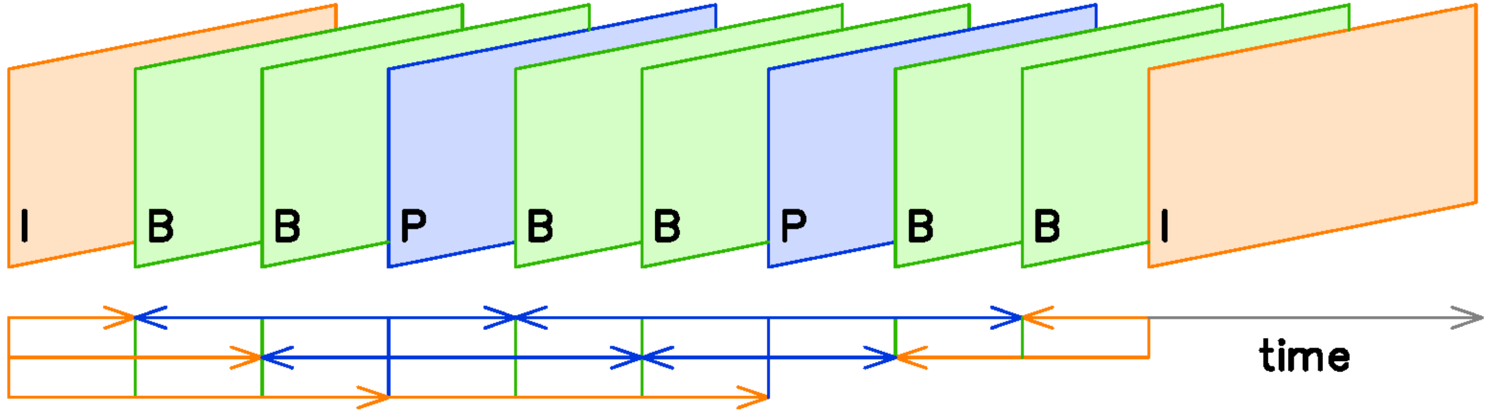
Combining the Tools …

Random Access: a Key Requirement



- Random access regards the capability to decode and visualize a specific target frame in a rather limited (access) time, e.g. 0.5 s
- This is not compatible with continuous Inter prediction coding which creates long prediction chains and thus very long access times
- Shortening the prediction chains implies regularly using Intra coding which does not exploit temporal redundancy
- Contrary to Intra refreshment for error resilience which may be made at MB level, random acess requires Intra coding at full frame level
- In summary, the random access functionality has to paid with a compression efficiency penalty (while complexity is also reduced)

Temporal Prediction Structure



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

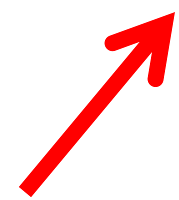
- **Random access:** Intra frames (I) – Don't use temporal prediction tools
- **Compression efficiency:**
  - Predicted frames (P) – May only use *forward* prediction from previous I/P frame (no algorithmic delay)
  - Bidirectionally predicted frames (B) – May use both forward and backward prediction from first previous and first future I/P frame (introduce algorithmic delay)

The Frame Level Syntactical Restrictions

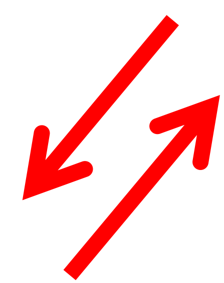
Motion

- **I (INTRA CODED) FRAMES –** All MBs in I frames are Intra coded 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.

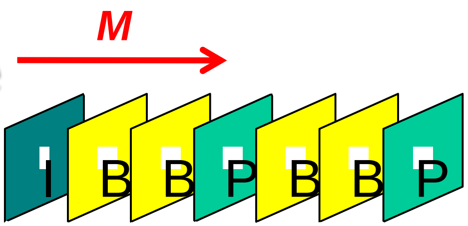
- **P FRAMES –** MBs in P frames MAY use forward prediction, this means a prediction from a past frame, with or without a motion vector; for MBs in P frames, only the previous P/I frame may be used as forward prediction (not from B frames).



- **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.



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**

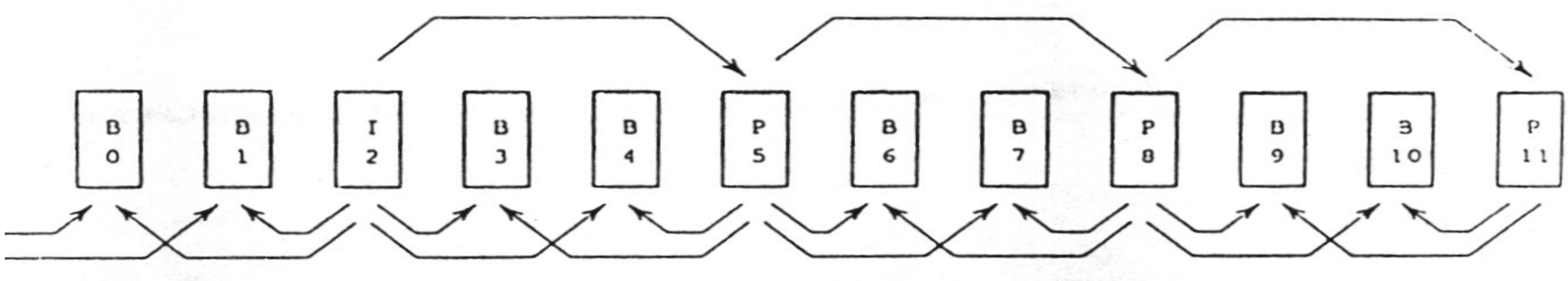
… I B B P B B P B B P B B I B B P B …

- **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 explicit 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 algorithmic delay …



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

Constant Quality: How ?

- Users prefer content with rather <u>constant quality</u> this means without noticeable quality variations along time and space.
- The simultaneous need for random access and high compression led to the definition of <u>3 frame types</u> depending on the used coding tools which have very 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 <u>need to non-uniformly allocate the bitrate resources</u> depending on the compression power of the coding tools used for each frame.
- Experience has shown that for good results are achieved for M=2-3 attributing similar quality to the I and P frames and a slightly lower quality to the B frames (as they are prediction useless).

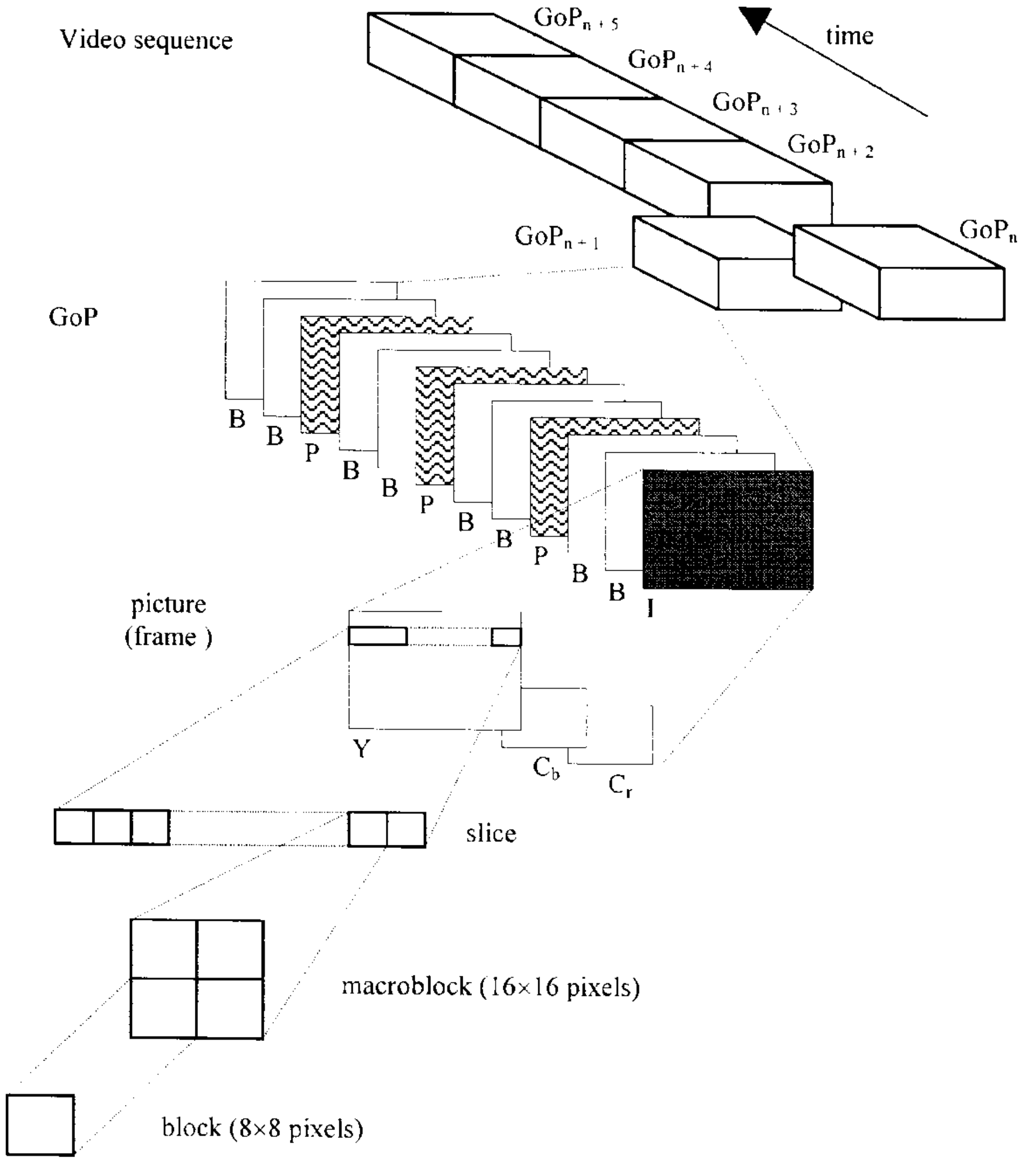
Who Does Take the Best Part ?

- **The ideal allocation of resources among the various frame types depends on the specific video content; however, the following distribution model 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.**

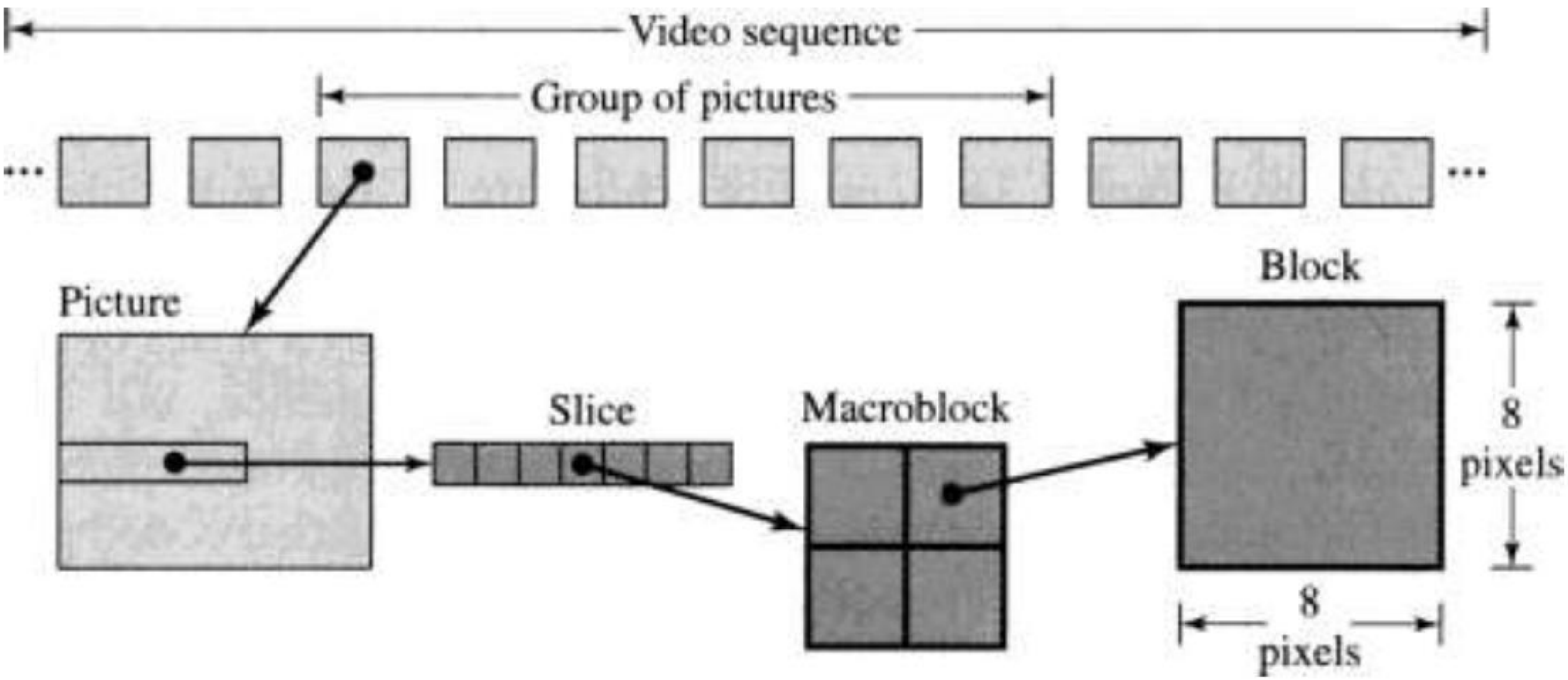
Video Structure

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

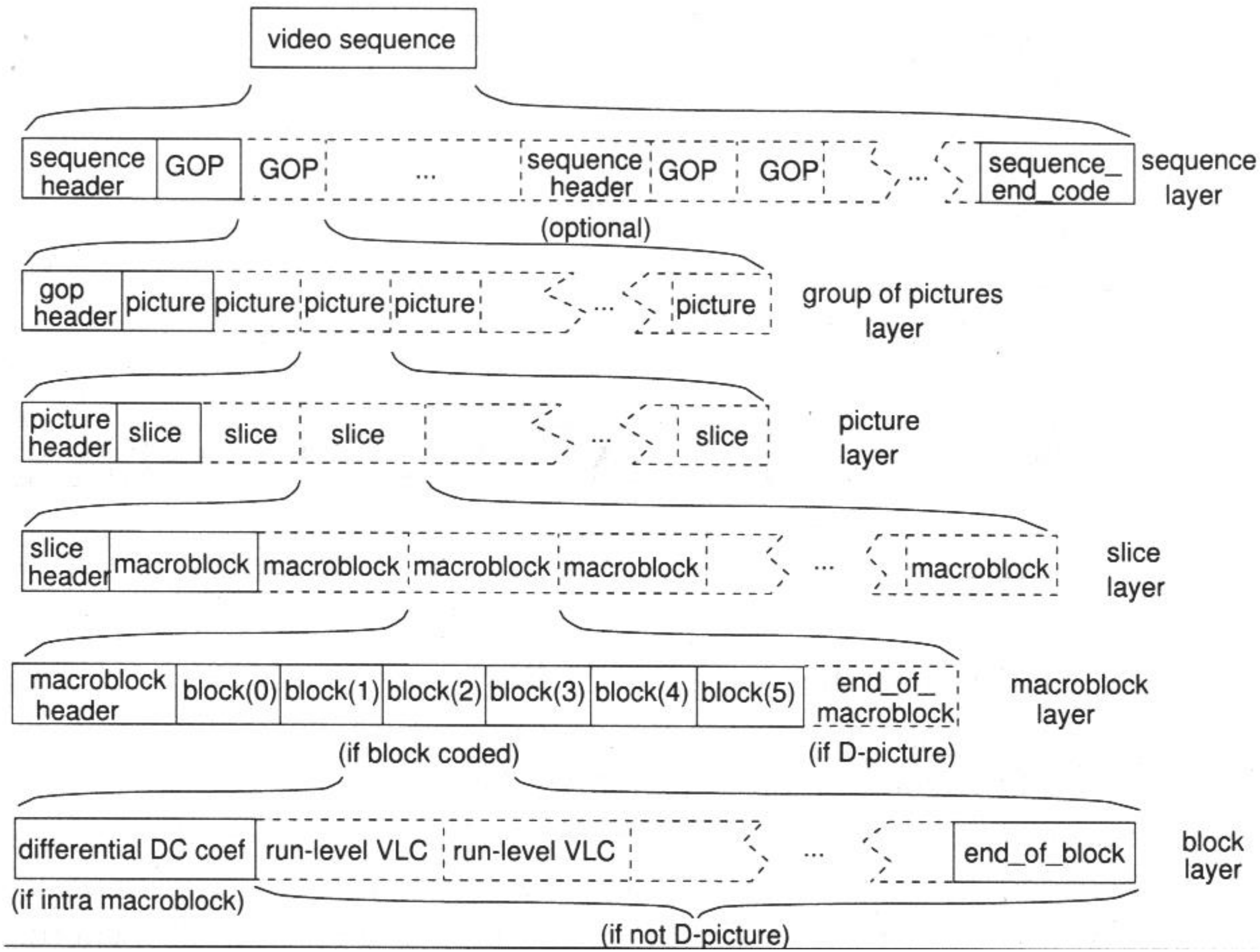
- Sequence
- Group of Pictures (GOP)
- Picture
- Slice (more flexible than H.261 GOBs)
- Macroblock (MB)
- Block



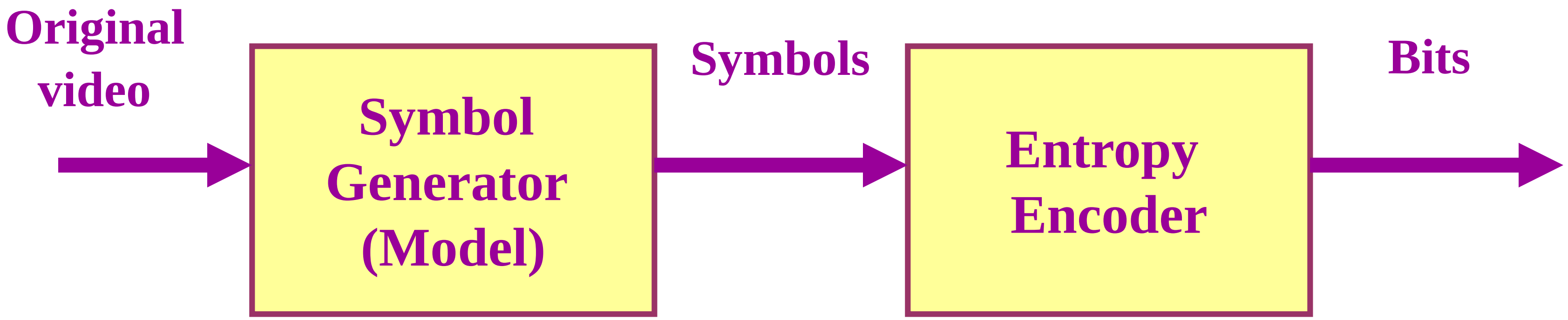
Video Structure Hierarchical Syntax



MPEG-1 Video Syntax



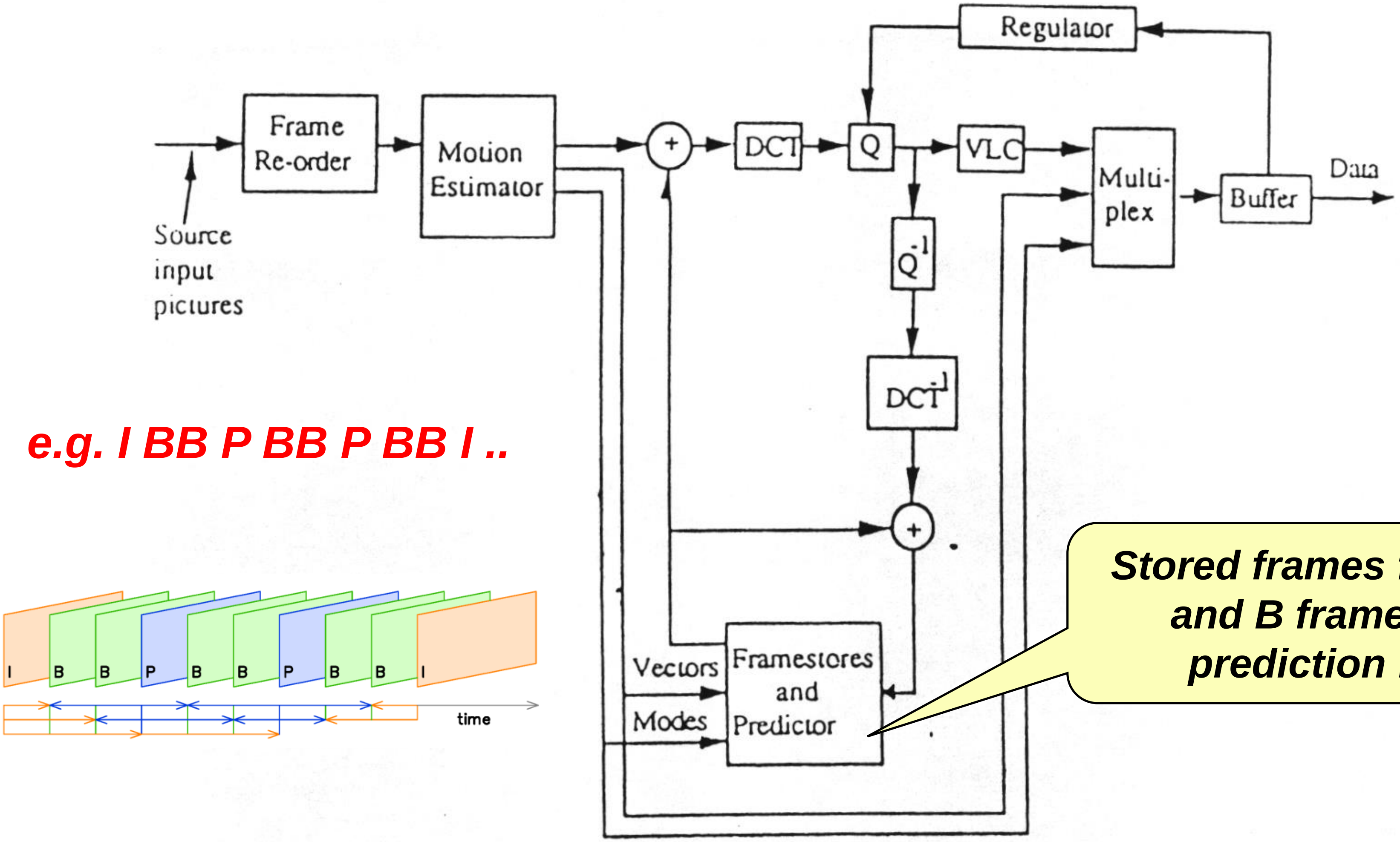
The MPEG-1 Video Symbolic Model



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 quantized coefficients, following the constraints set by the frame coding type (I, P or B).

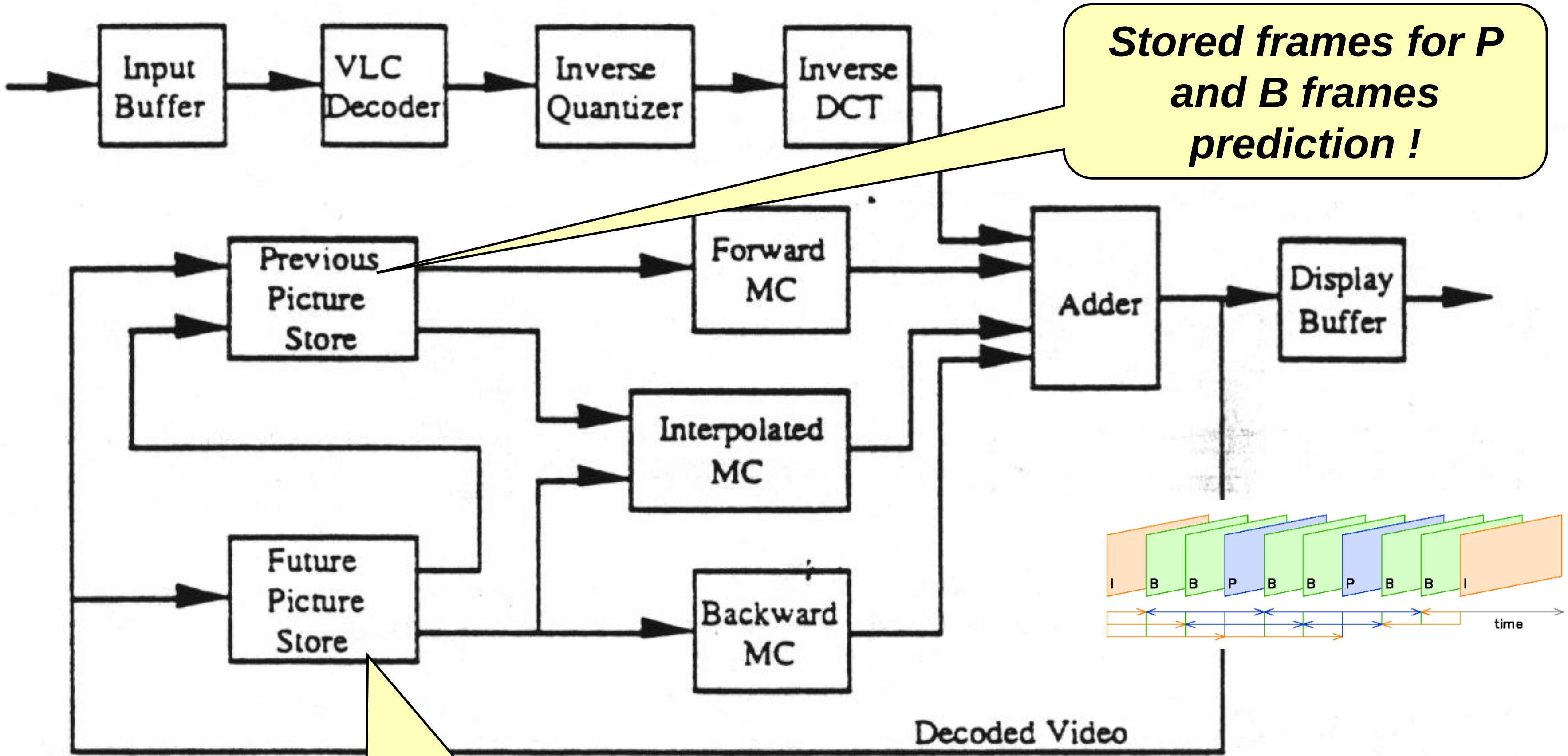
MPEG-1 Video: Encoder

e.g. I BB P BB P BB I ..



Stored frames for P and B frames prediction !

MPEG-1 Video: Decoder



Stored frames for P and B frames prediction !

Stored frame for B frames prediction !

MPEG-1 Video and H.261: What 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 random access 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.

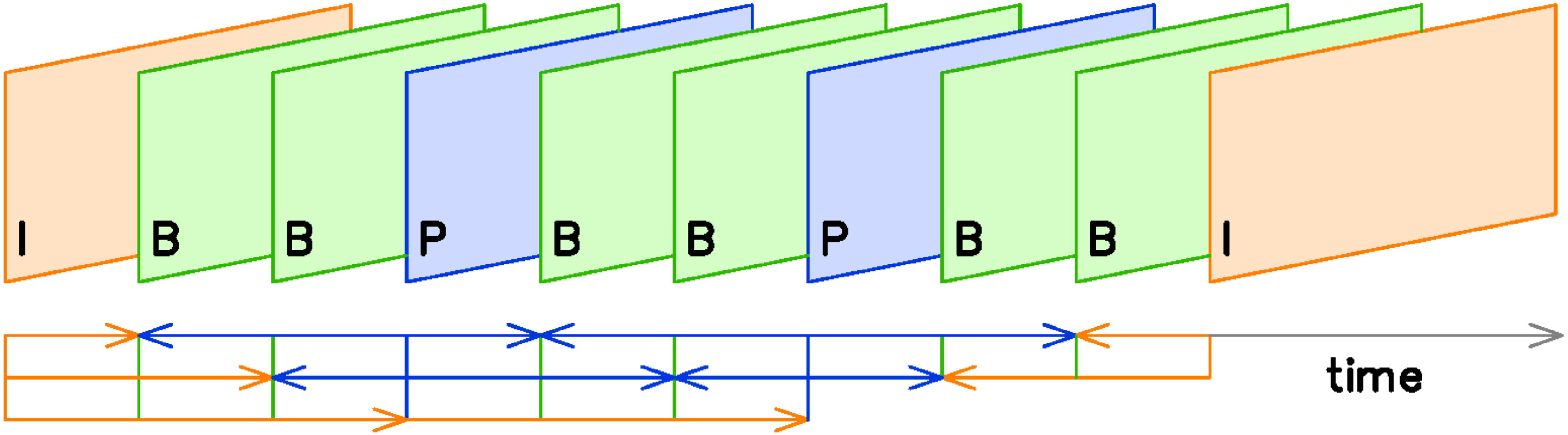
Rate Control … and Offline Encoding ...

- **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, <u>encoding may be performed *off-line*</u> (taking whatever time, iterative encoding, multiple passes, etc.), thus achieving much higher quality than real-time encoding for similar bitrate resources.

Especial Access Modes

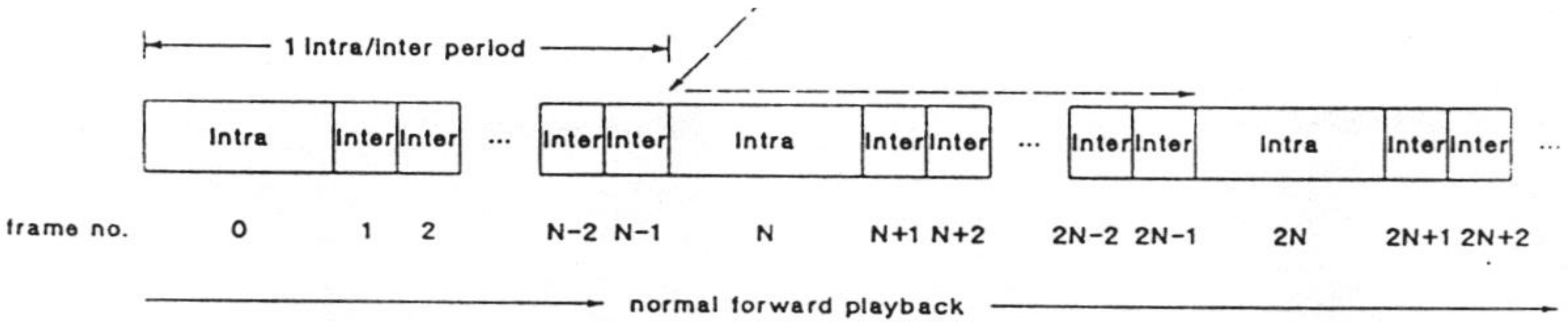
The I, P, B Frames Prediction Cocktail



Trade-off between

- **Compression efficiency**
- **Random access**
- **Complexity**
- **Delay**

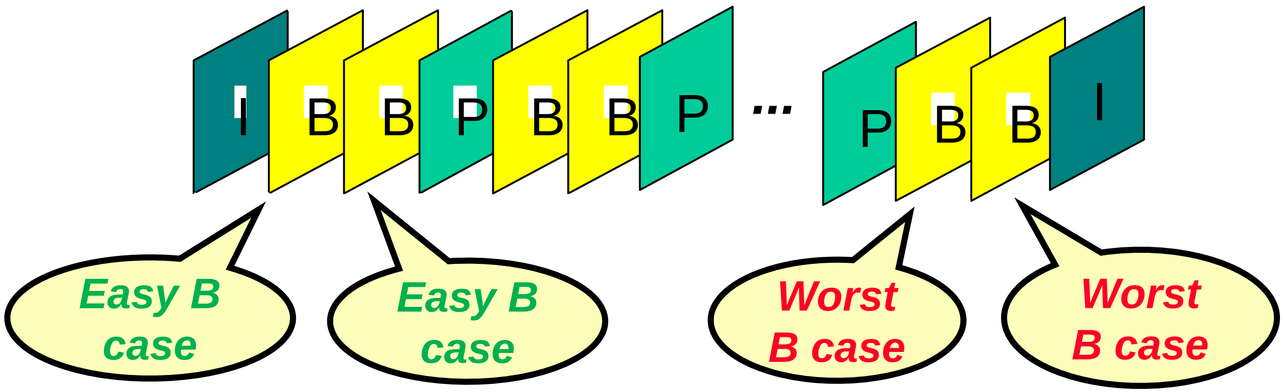
Random Access …



- **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 following the prediction chain.**

Maximum Random Access Time



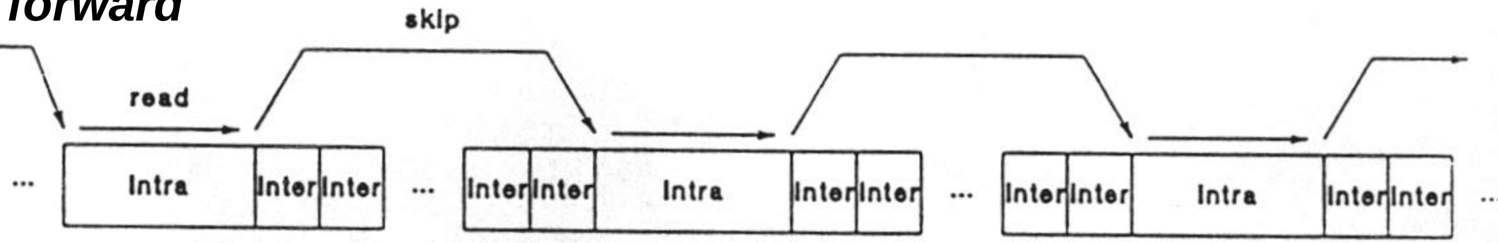
For the CD-ROM, the Maximum Random Access Time (MRAT) depends on the 'worst case' allocation of bits among the various types of frames, the frame rate and the time between I frames:

$$MRAT = T_{DSM} + [2 \times MNBF_I + (N/M-1) \times MNBF_P + 1 \times MNBF_B] / r_{leitura} \text{ (s)}$$

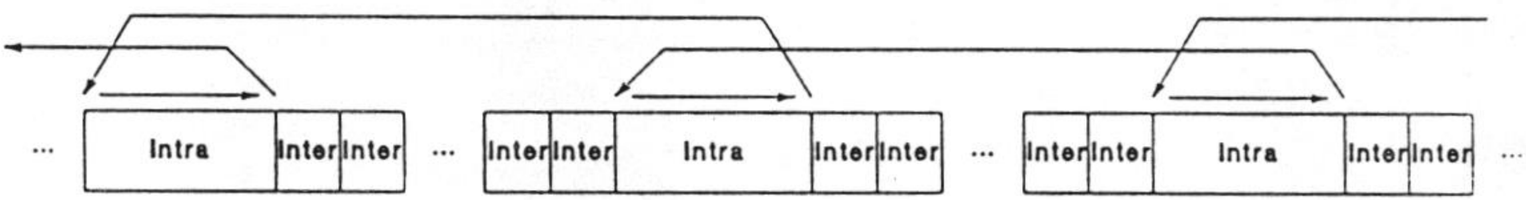
- $MNBF_X$ is the maximum number of bits for a specific frame type X, where X may be I, P or B.
- T_{DSM} is the sum of the various access times due to the need to jump in the CD-ROM to read only the strictly necessary bits (in the prediction chain).

Fast Forward and Fast Reverse: Examples

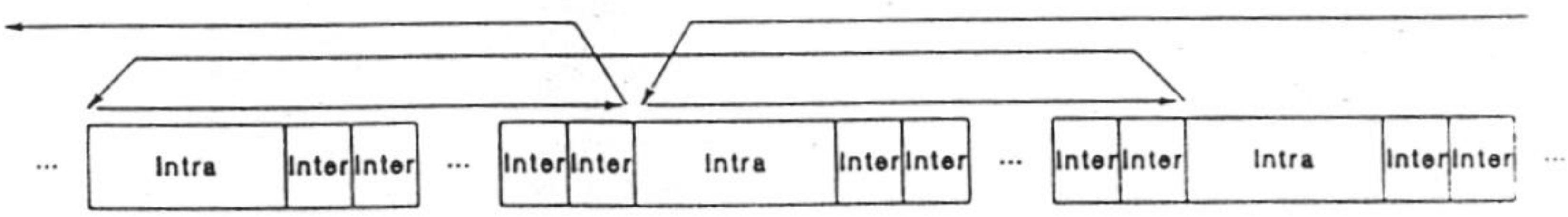
Fast forward



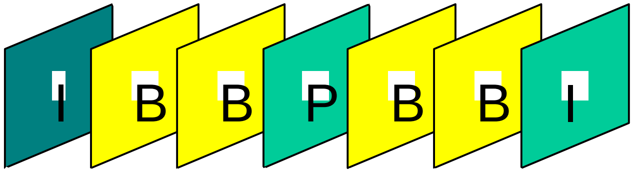
Fast reverse



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.

SUF = <real time> / <visualization time>

- **If only I frames are read ($K = N/M-1$):** $SUF = N \times 1/f \;/\; V \times 1/f = N/V$
- **If only I and P frames are read:** $SUF = [(K+1)\,M] \times 1/f \;/\; V \times 1/f$
  - ***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 visualization frame periods for each decoded frame**

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 <u>unnecessarily complex to guarantee interoperability</u>, 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 !

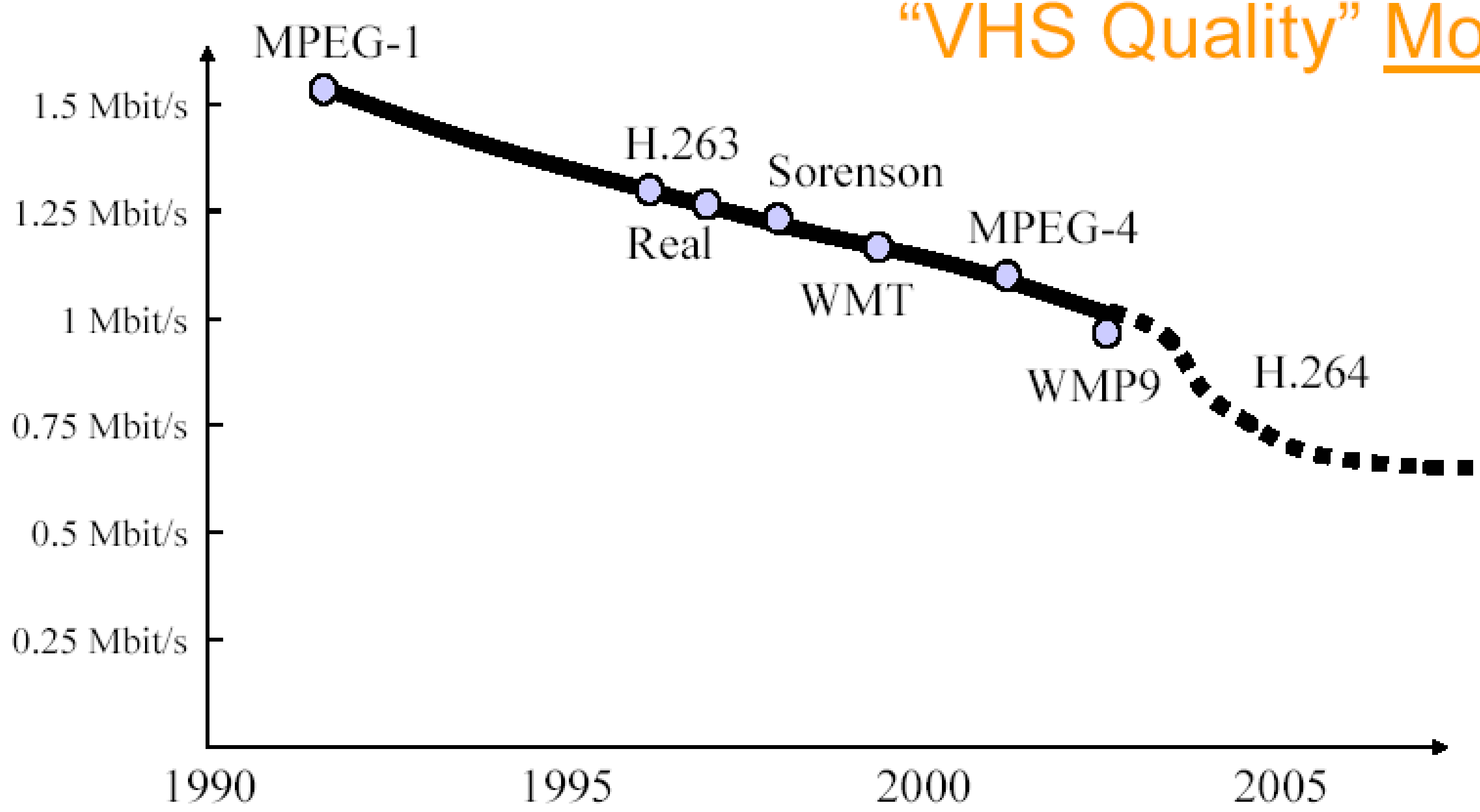
Error Concealment

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 after channel decoding (with error correction up to available power)**
- **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 ☺

Cutting the Rate for Low Resolution Video ...

"VHS Quality" Movies



TÉCNICO LISBOA

MPEG-1 Standard

Part 3: Audio

TÉCNICO LISBOA

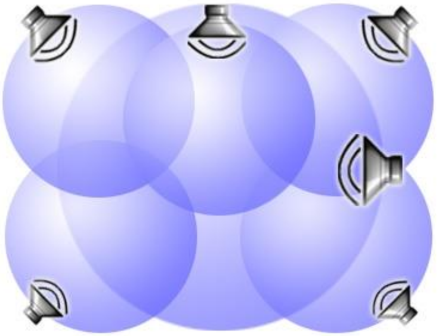
MPEG-1 Audio: Objective

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

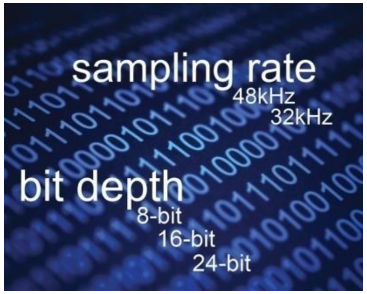
Audio PCM Parameters …

- **<u>Number of audio channels</u>** – An higher number of channels enhances the perception of sound spatialization by exploiting sound localization and thus the listener's ability to identify the location or origin of a detected sound in direction and distance.



- **<u>Sampling frequency/rate per channel</u> -** In digital audio the most common sampling rates are 44.1 kHz, 48 kHz, 96 kHz and 192 kHz; the two major common sampling rates are 44.1 kHz and 48 kHz.

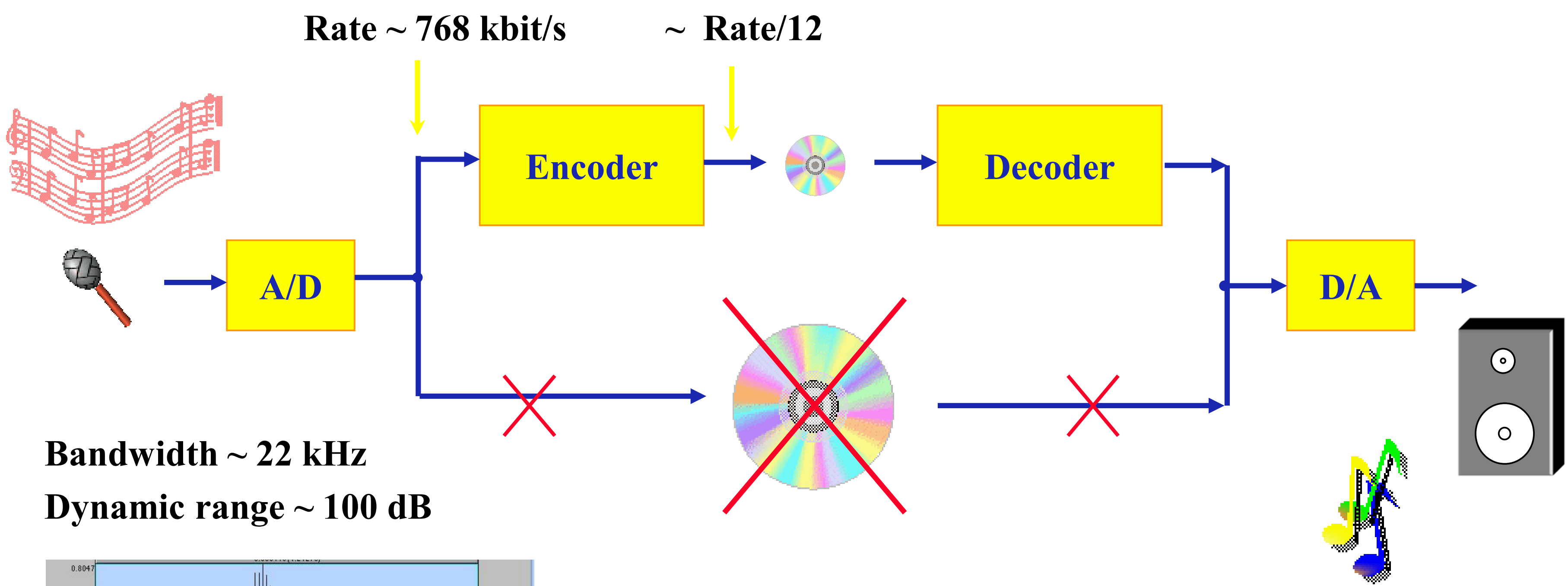
- **<u>Number of bits per sample or bit depth</u>** - Bit depth corresponds to the resolution of each sample in digital audio data; common examples of bit depth include CD quality audio, which is recorded at 16 bits, and DVD-Audio, which can support up to 24-bit audio.



PCM for Various Sound Signals …

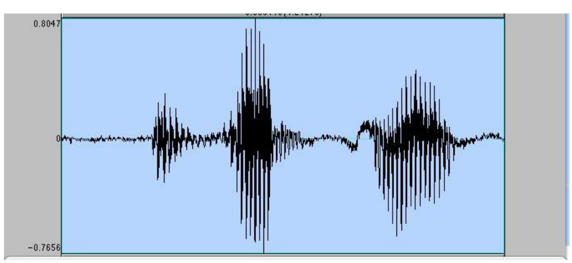
	Frequency (Hz)	Sampling rate (kHz)	bit/sample (PCM)	PCM bitrate (kbit/s)
Speech (telephone)	300-3400	8	8	64
Speech (wideband)	50-7000	16	8	128
Audio (medium band)	10-11000	24	16	384
Audio (wideband)	10-22000	48	16	768

The Audio Compression Chain …



Bandwidth ~ 22 kHz

Dynamic range ~ 100 dB



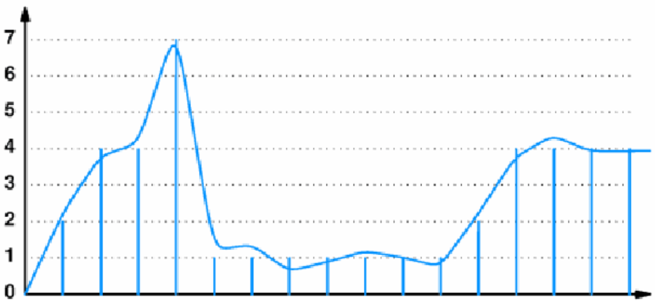
Redundancy and Irrelevance: Audio vs Video

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
- ….

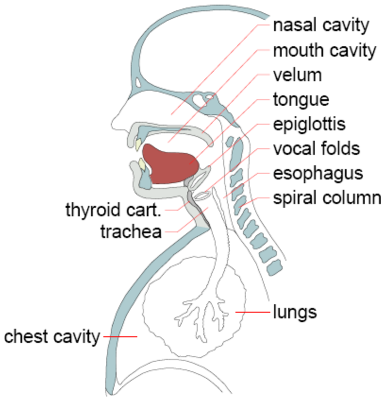
MPEG-1 Audio: Requirements

- High (decoded) signal quality independently of the spectral and amplitude characteristics of the coded signal (for target rate)
- 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



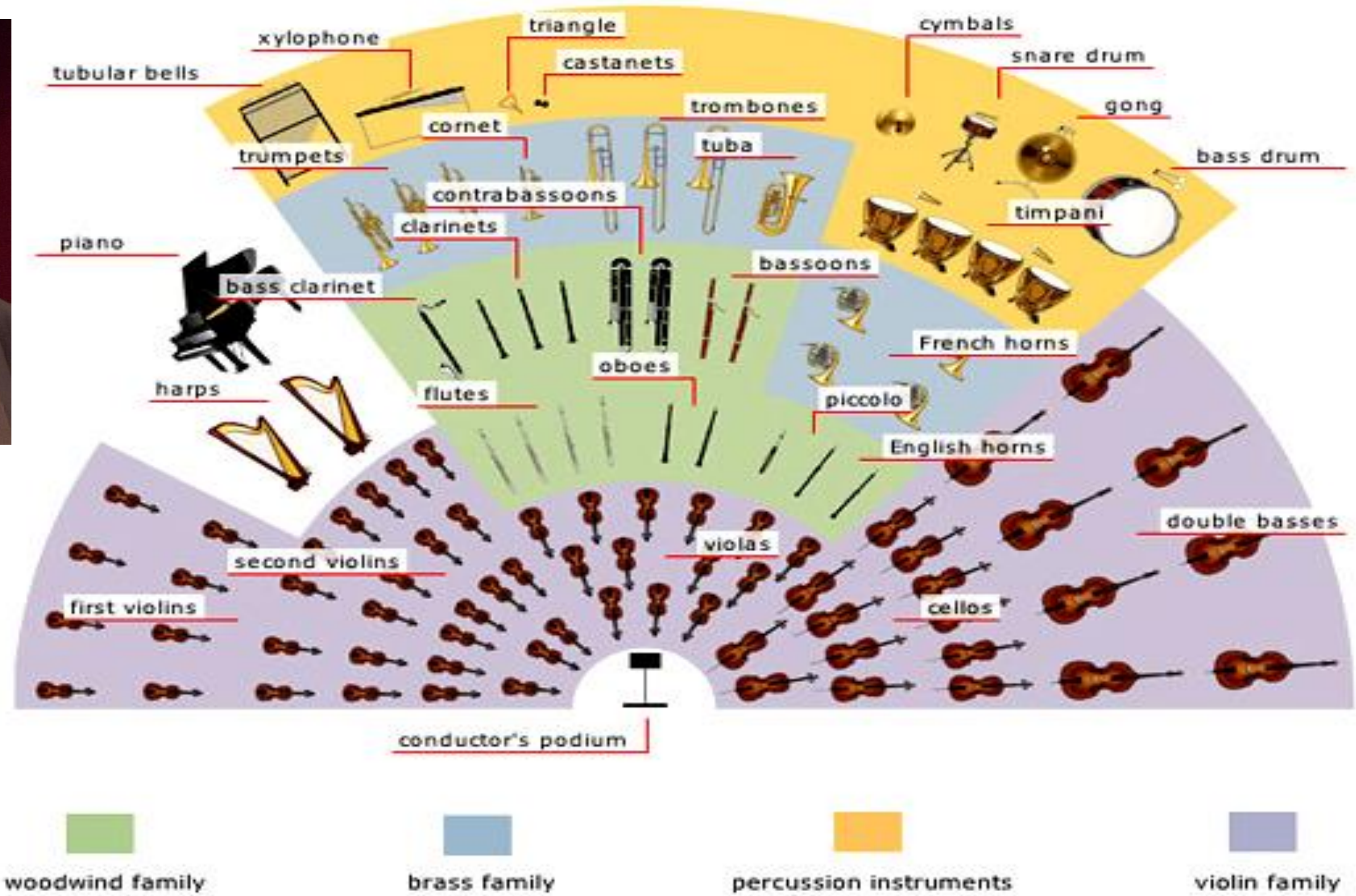
Audio Coding Peculiarities …

- Very <u>high dynamic range</u> (ratio between the maximum and minimum signal amplitudes) ~ 100 dB
- <u>Larger bandwidth</u> (in comparison with speech)
- <u>Absence of a universal source production</u> mode (speech production models allows reaching higher compression factors)
- Certain <u>simplifying assumptions</u> usually adopted for speech coding <u>are not valid</u> anymore such as:
  - Gaussianity
  - Stationarity
  - Spectral smoothness



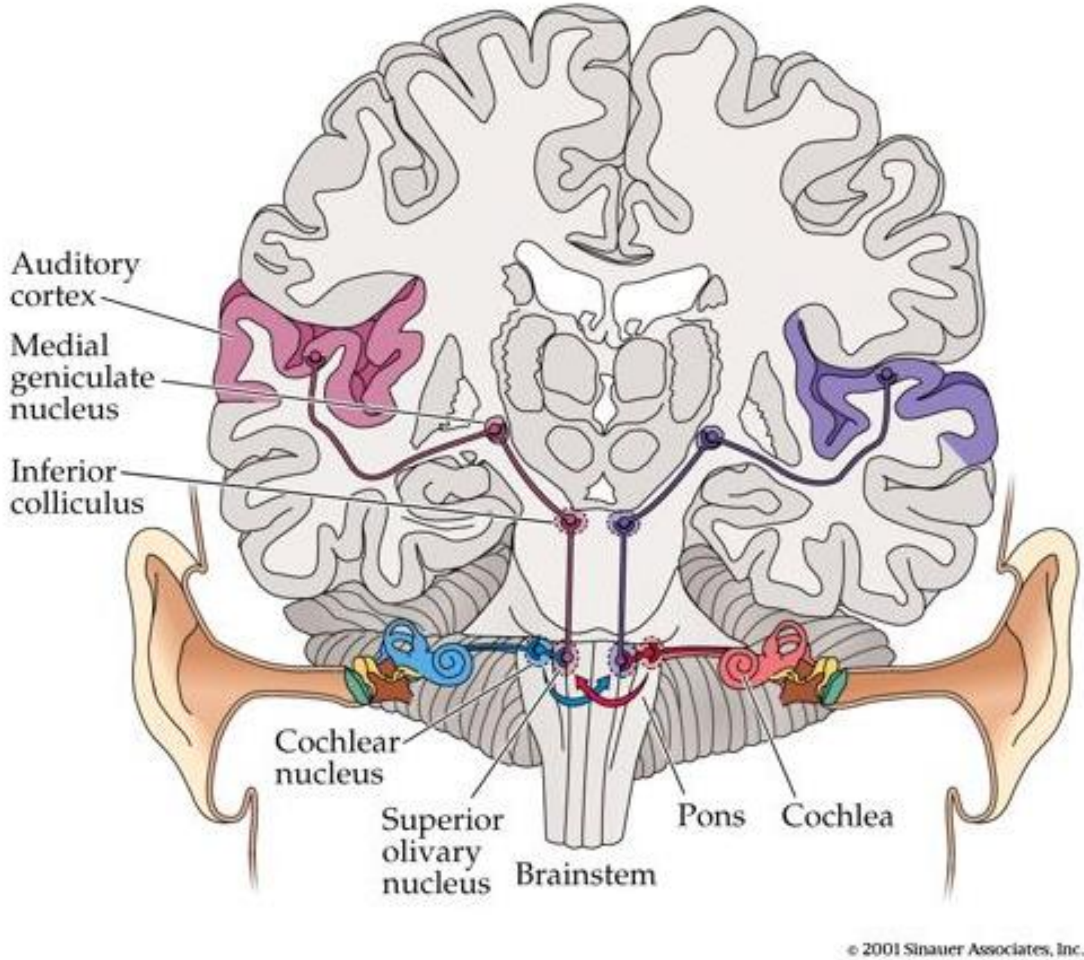
Compression gains for audio coding mainly result from irrelevancy reduction (as redundancy is short …).

Music is Much More than Speech …



Human Auditory System: Hearing is more than Ears ….

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



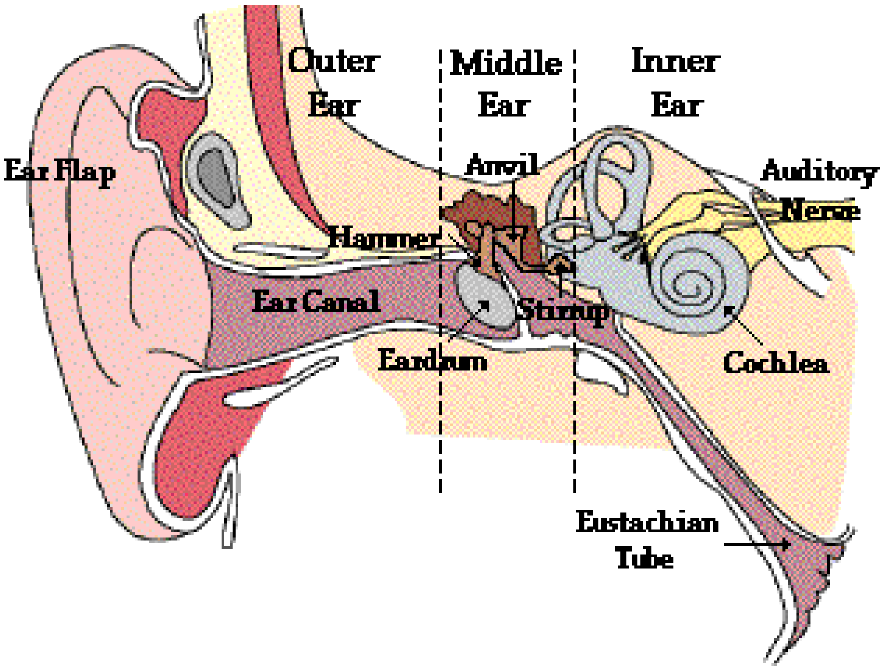
Human Auditory System: the Ear

The ear has three main sections:

1) Outer ear – Directs the sound to the eardrum.

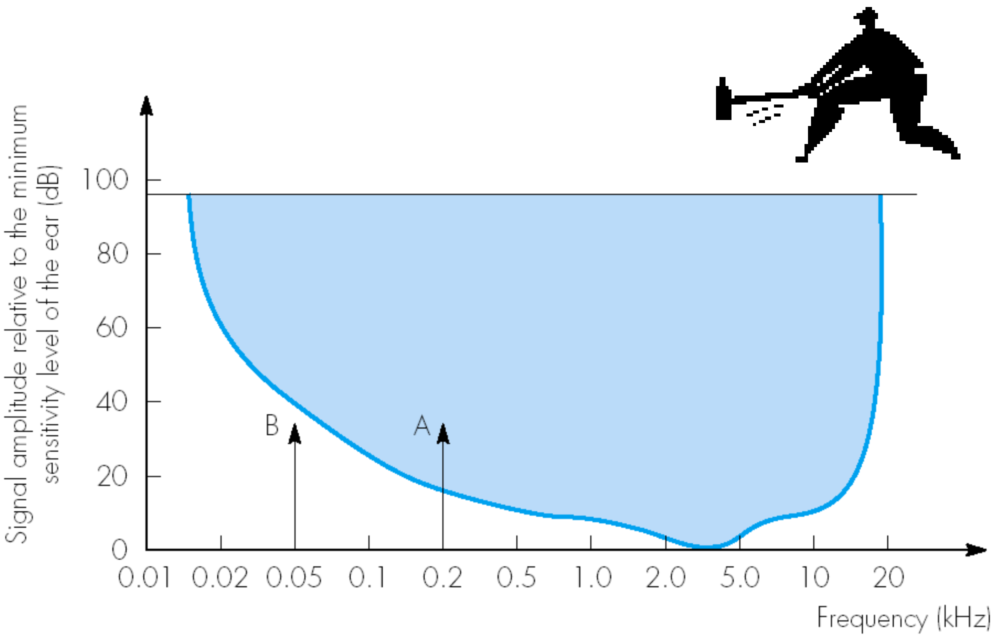
2) Middle ear – Transforms the sound pressure into mechanical vibration.

3) Inner ear – Converts these mechanical vibrations into excitations of the auditory nerves which send electrical signals to the brain.

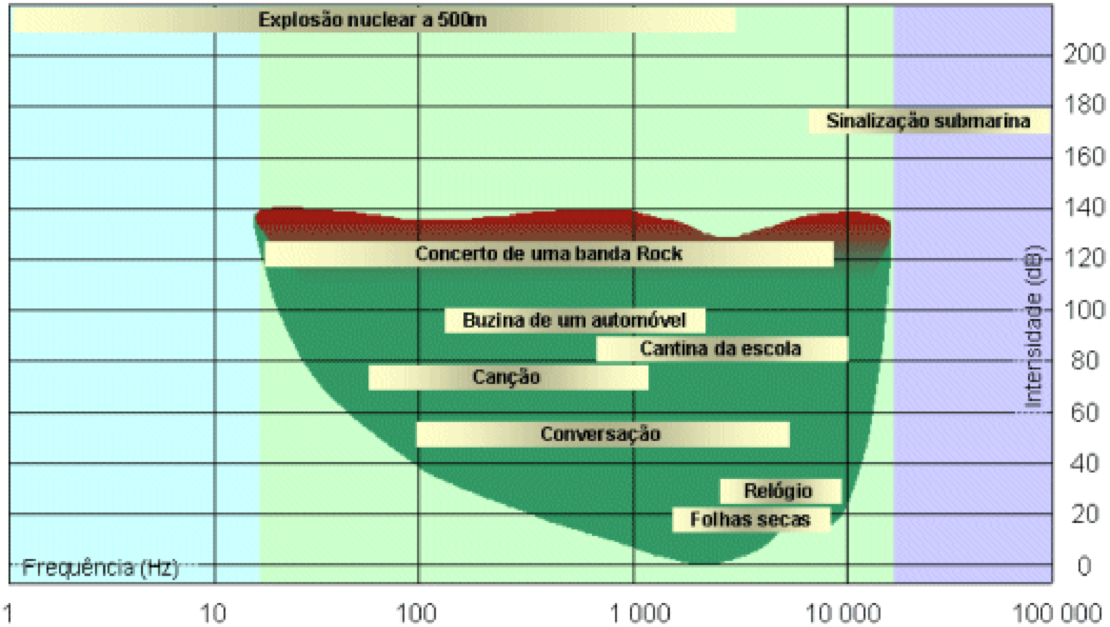


Physiological Effects: the Thresholds

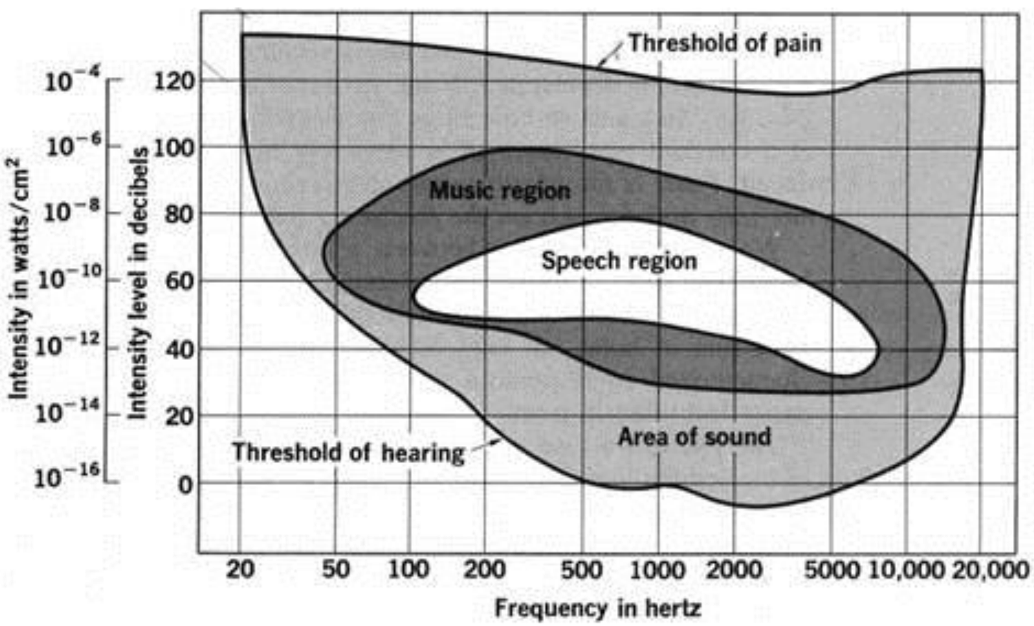
- **Threshold of Hearing** – Defines the minimum sound intensity which may be perceived; this threshold varies along the audio band.
- **Threshold of Feeling or Pain** – Defines the sound intensity above which the sounds may cause pain and provoke hearing damages.



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^2 at 1 kHz.

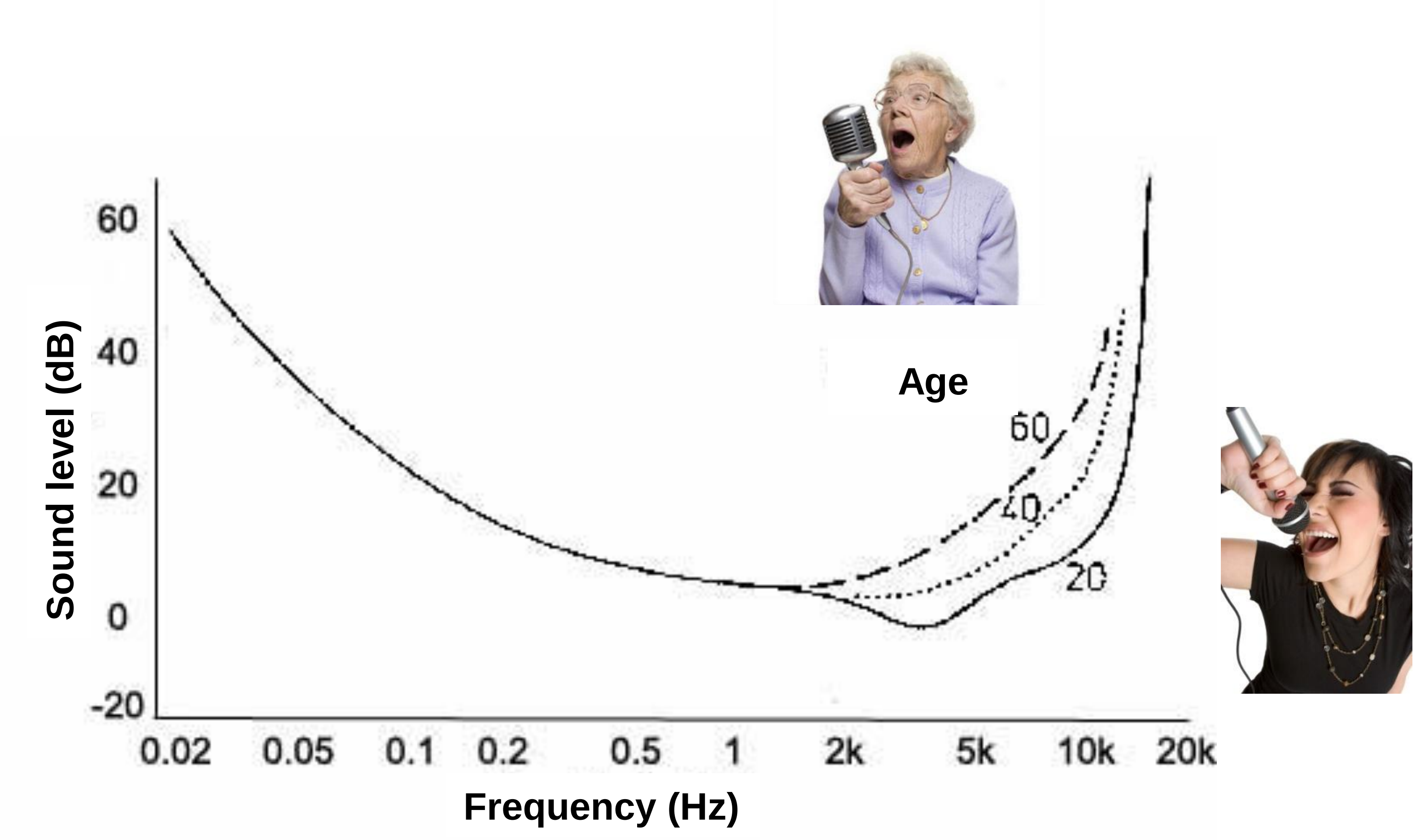


Sound Sensibility …



The human hearing dynamic range is about 100 dB.

Hearing Threshold Variation with Age …



TÉCNICO LISBOA

MPEG-1 Audio: Coding Tools

Lossless :-)

- **Redundancy**

Frequency coding

Window switching

- **Statistical Redundancy**

Huffman entropy coding

Lossy :-(

- **Irrelevancy**

Perceptive coding, masking and quantization

Dynamic allocation of bits

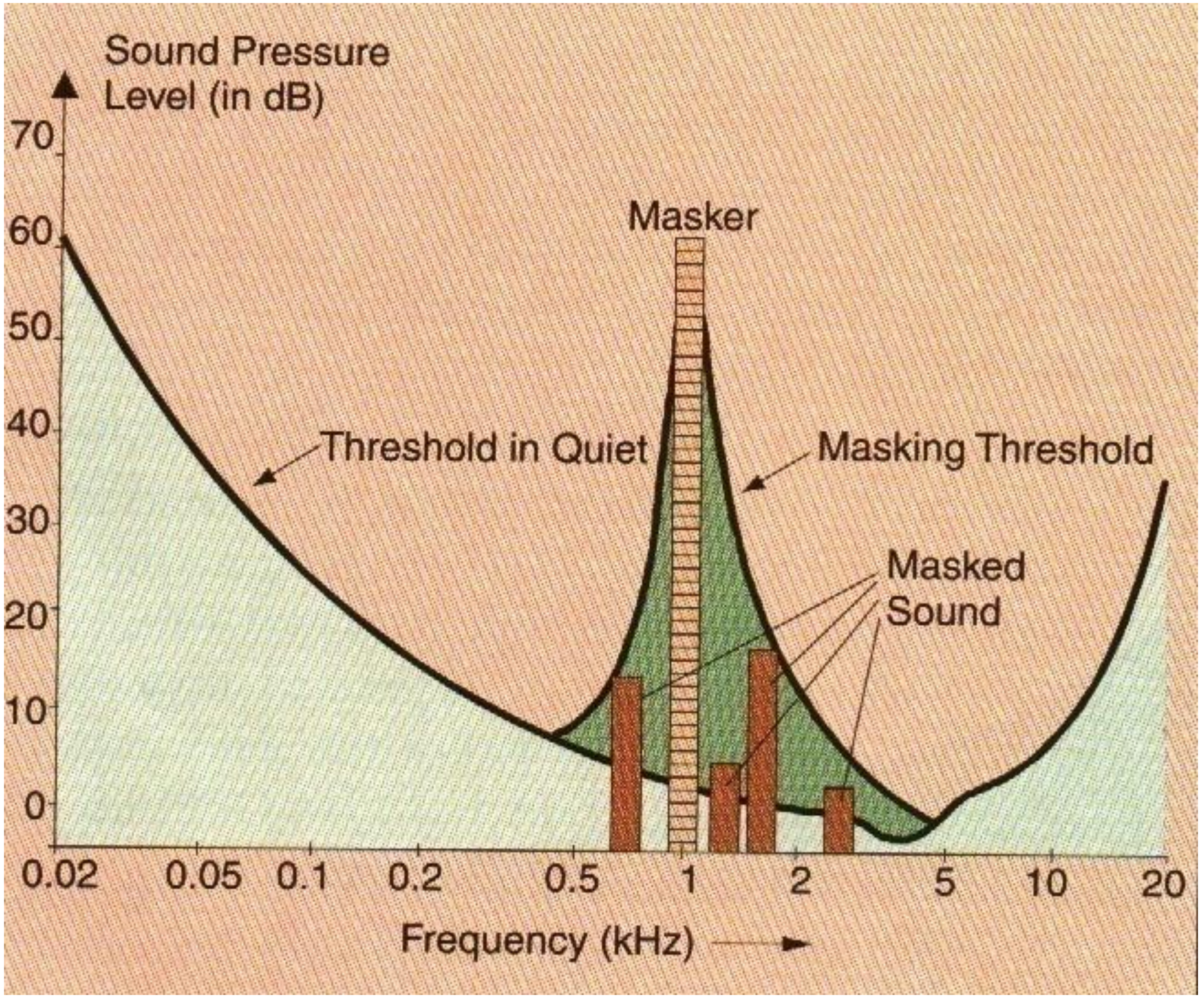
Defining Audio Masking

Auditory masking is the hearing behavior when the perception of one sound is affected (*masked*) 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.

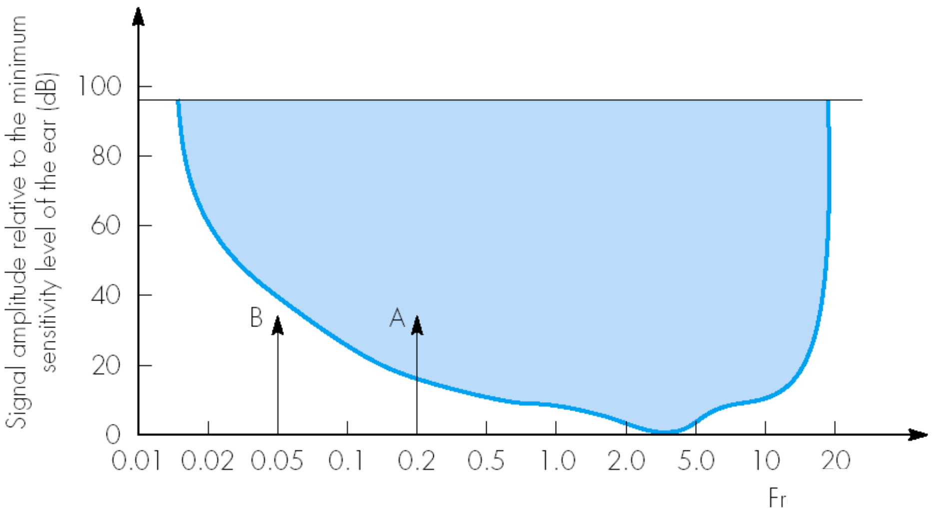
- One sound may simply totally mask another sound or increase its hearing threshold.
- The masking sound depends on the circumstances: for example, although it may be possible to speak 'normally' with someone at a party, any distraction may result in the background noise masking the voice of the other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

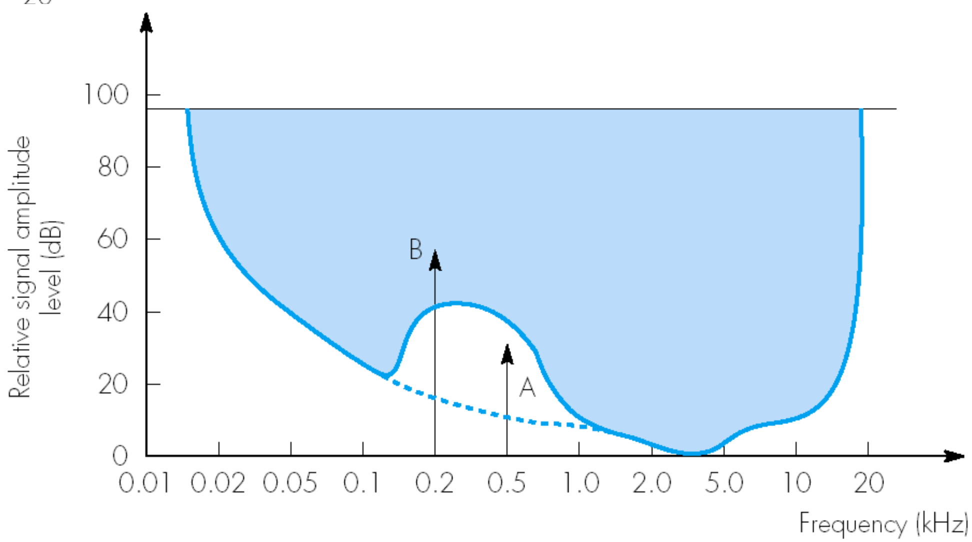
The masking effect is highly non-linear and its effects are very diverse.

Frequency Masking

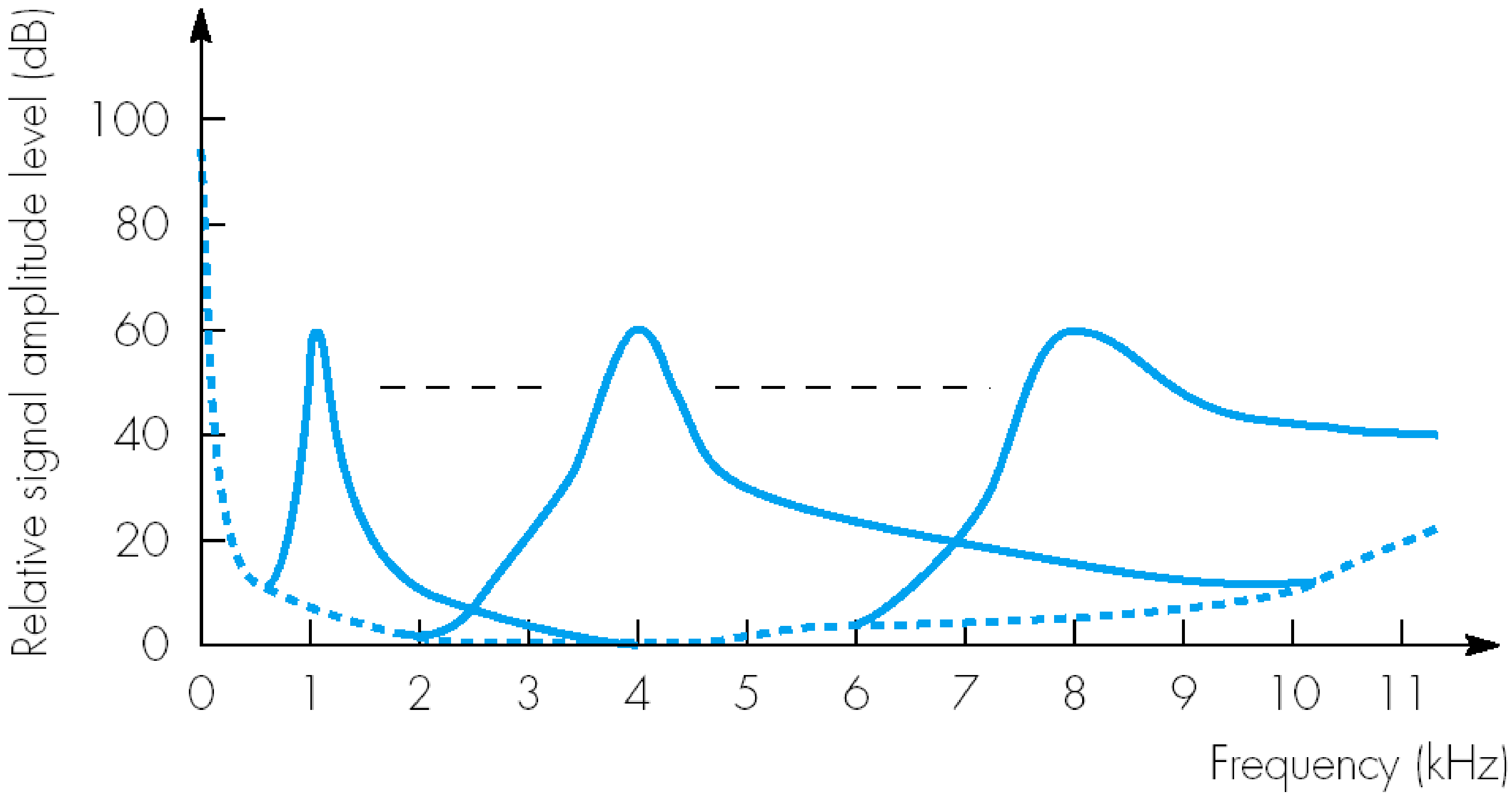


Frequency Masking at Work ...

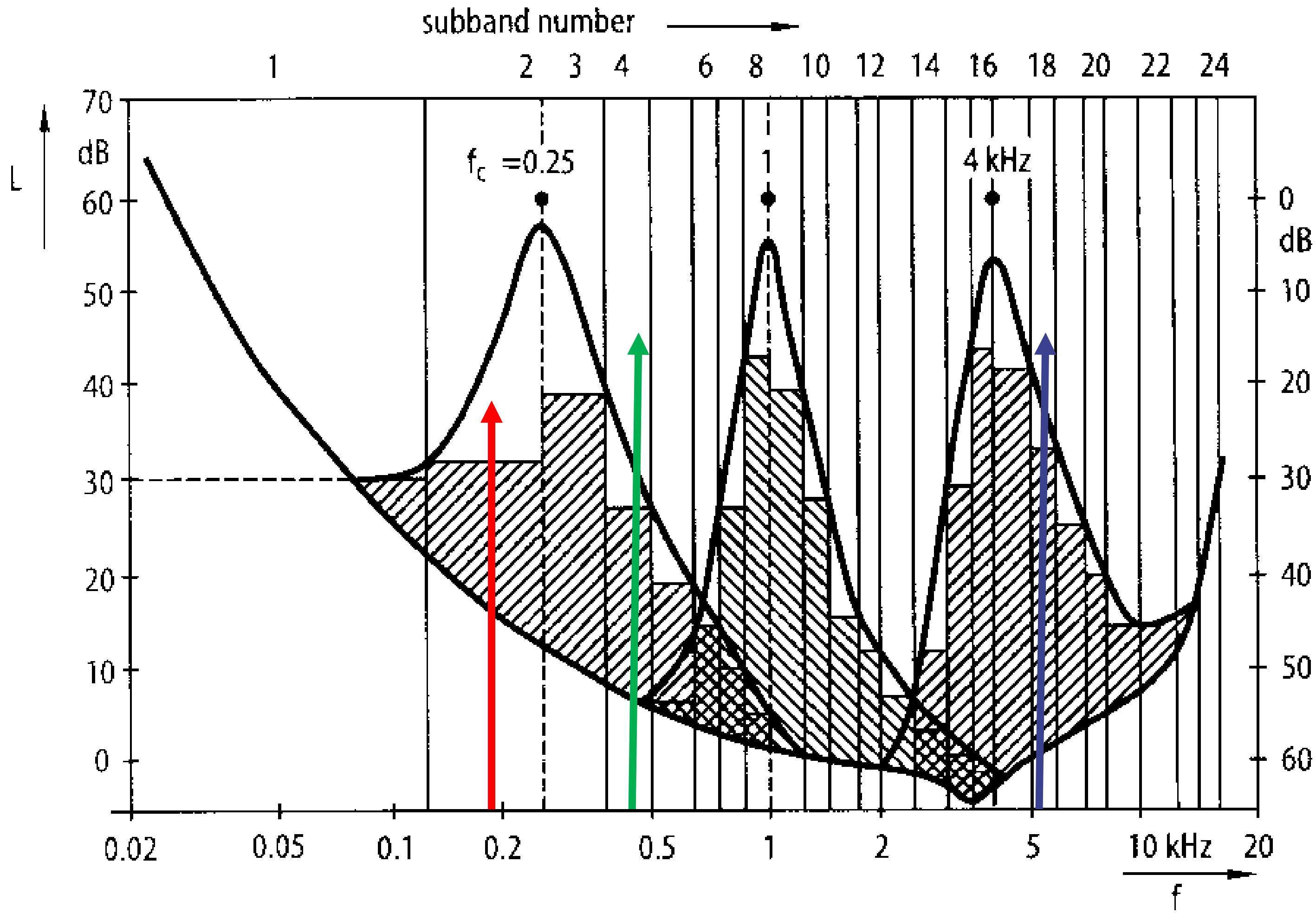




Masking Width Variation with Frequency



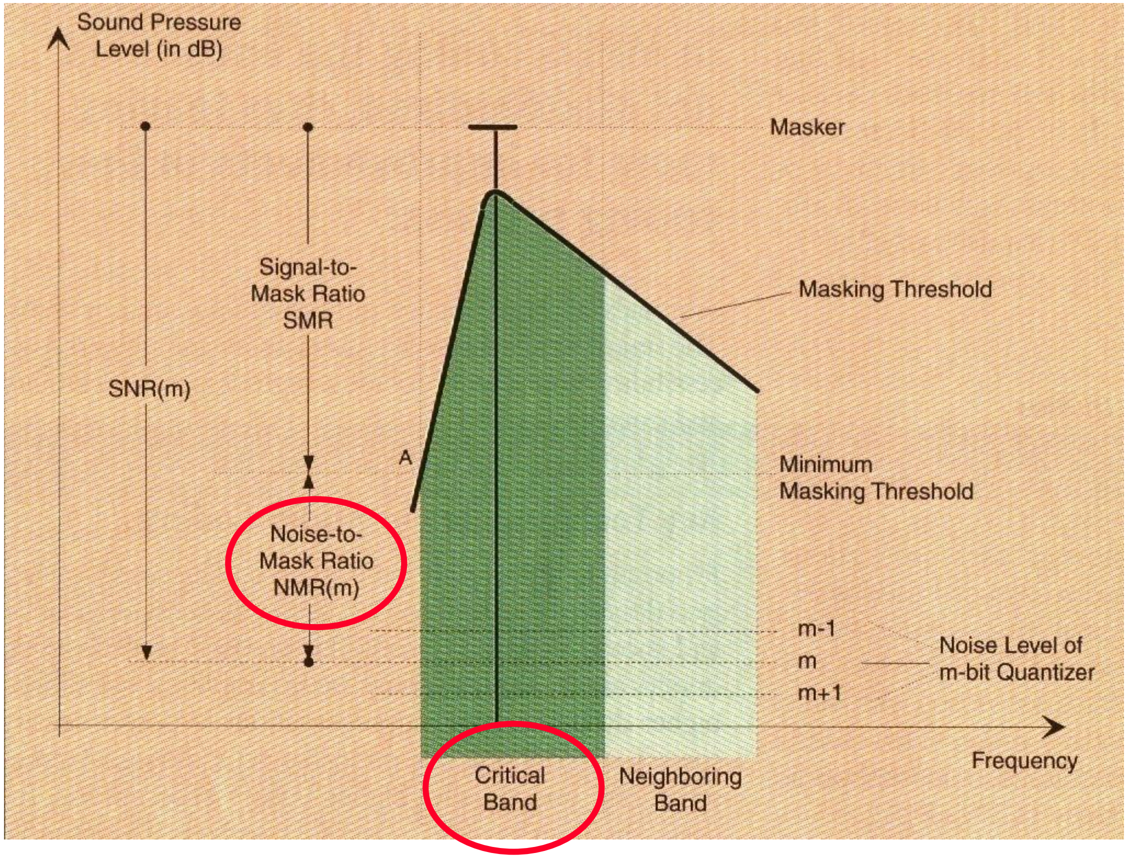
‘Killing and Hurting’ …



Frequency Masking

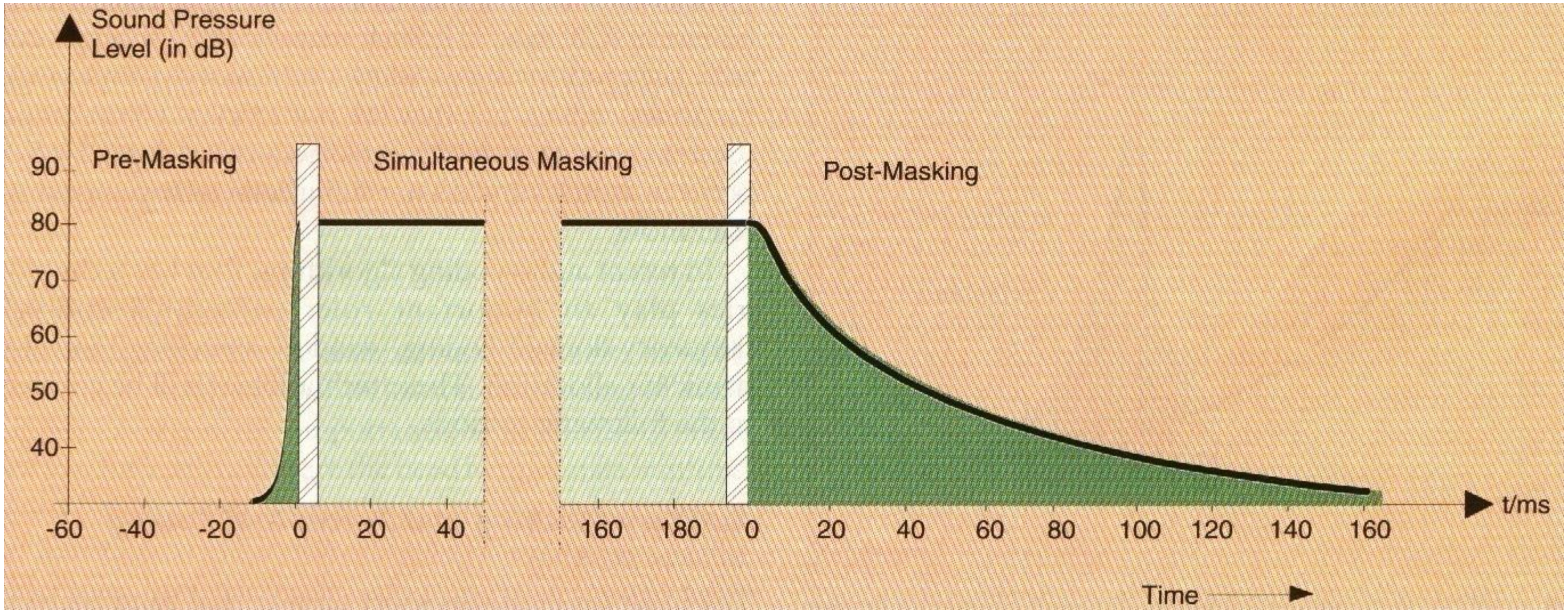
- NMR measures the difference between the quantization noise level and the level at which the distortion becomes audible for a certain band.

- 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

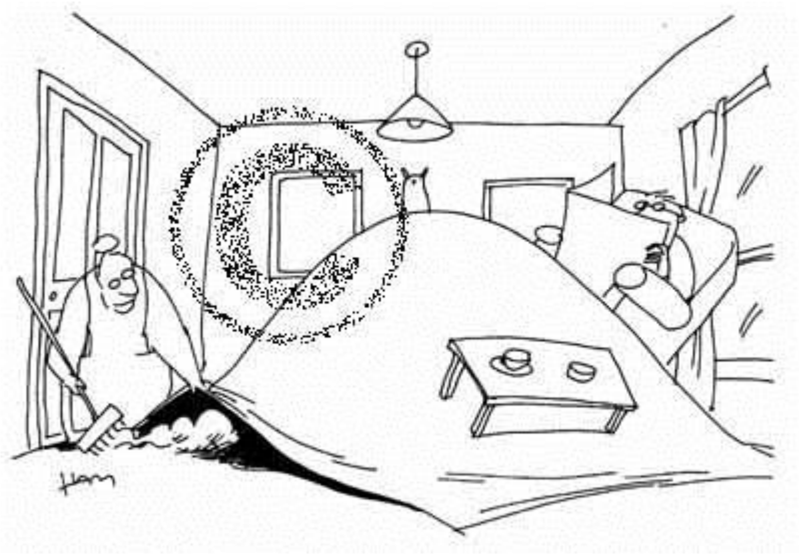


- **Temporal masking occurs when a sudden stimulus sound makes inaudible other sounds which are present immediately preceding or following the stimulus.**

- **Masking obscuring a sound immediately preceding the masker is called <u>backwards masking</u> or pre-masking (< 5 ms) and masking obscuring a sound immediately following the masker is called <u>forwards masking or post-masking</u> (≈20 ms).**

Perceptive Audio Coding

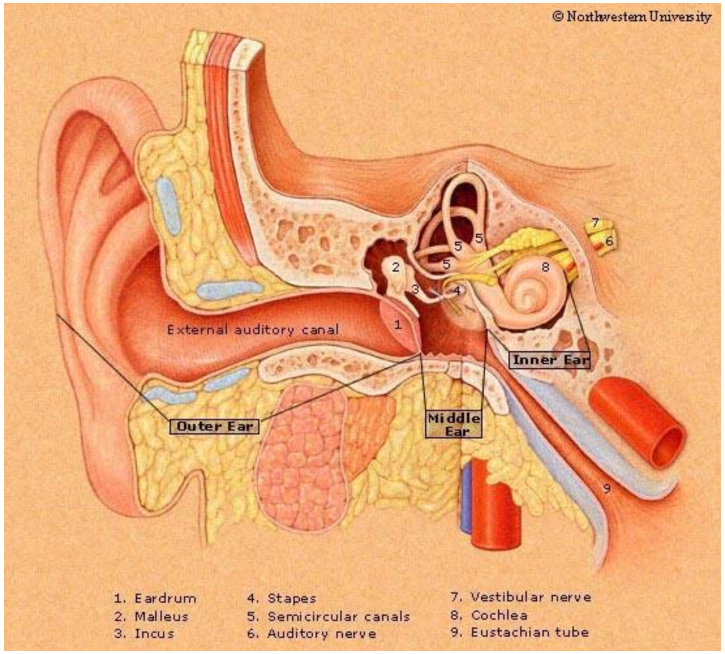
Irrelevancy manifests itself as amplitude or frequency information (resolution, detail) which cannot be perceived by humans. All masked signal components do NOT need to be coded/transmitted.



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.

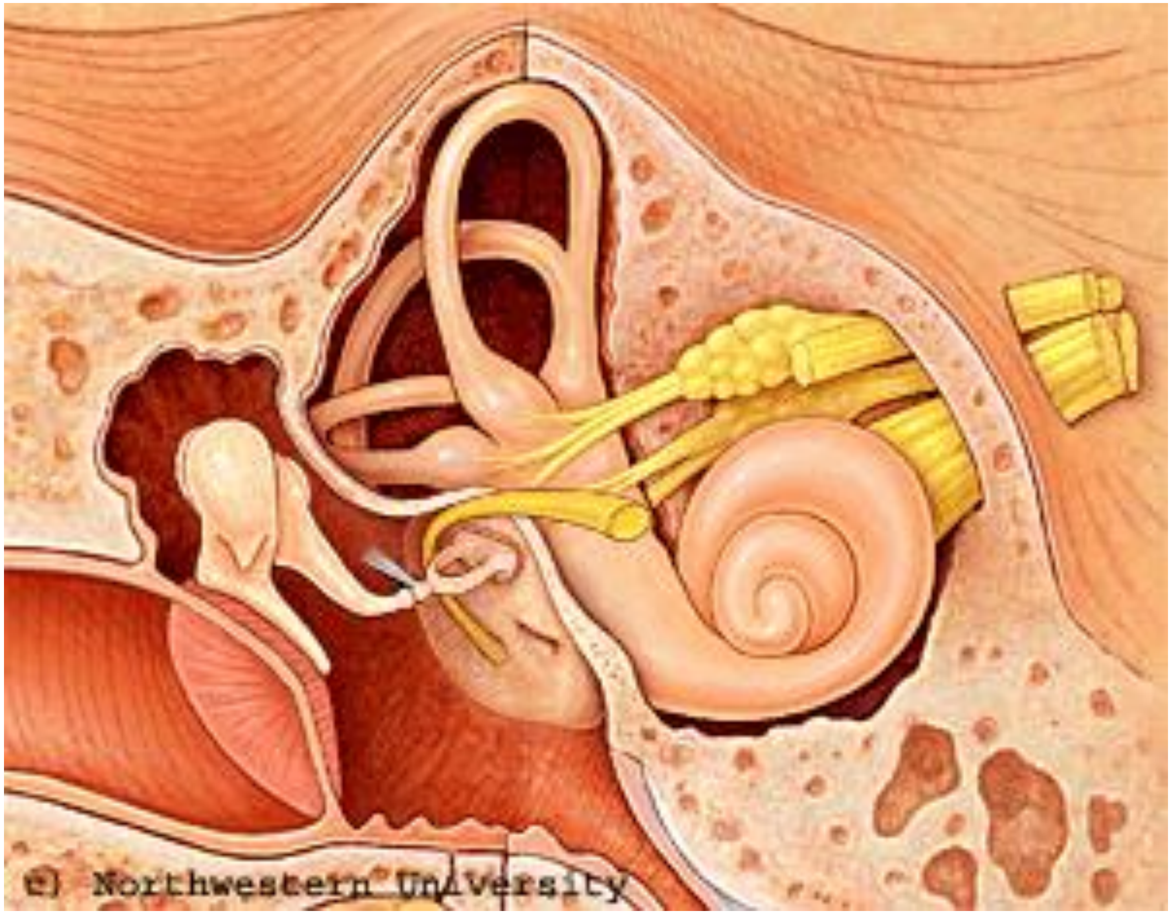
Psychoacoustic Model: the Secret !



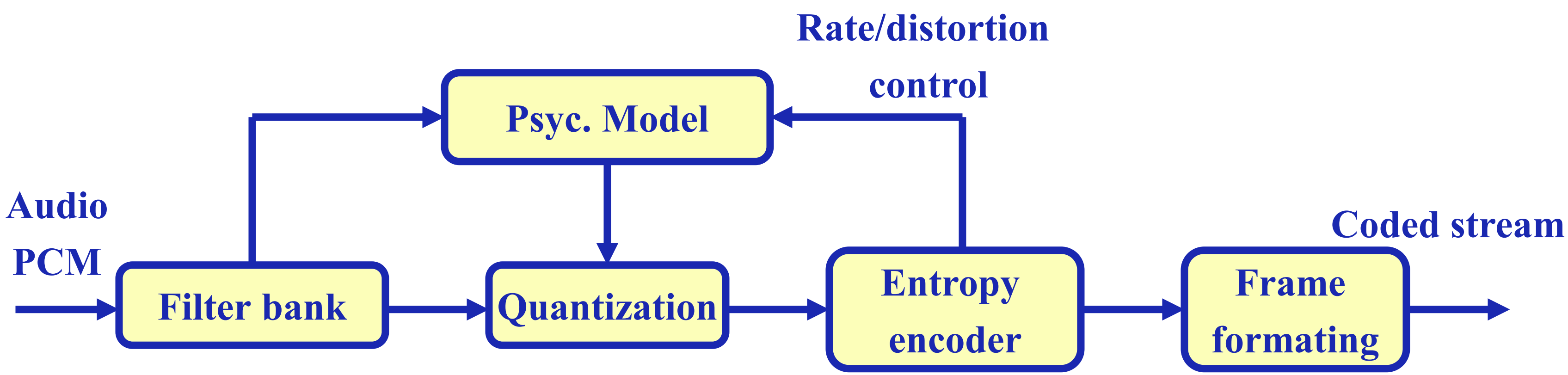
A psychoacoustic model is a mathematical model which defines, in a more or less simplified way, the main properties and tolerances of the human auditory model, notably its sound intensity perception, its spectral selectivity and, especially, the masking effect.

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.

The Audio Perception Miracle …



Perceptive Encoder Architecture



The filter bank allows to organize the signal in several bands as the Human Auditory System is differently sensitive along the frequency bands.

The psychoacoustic model controls/shapes 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.

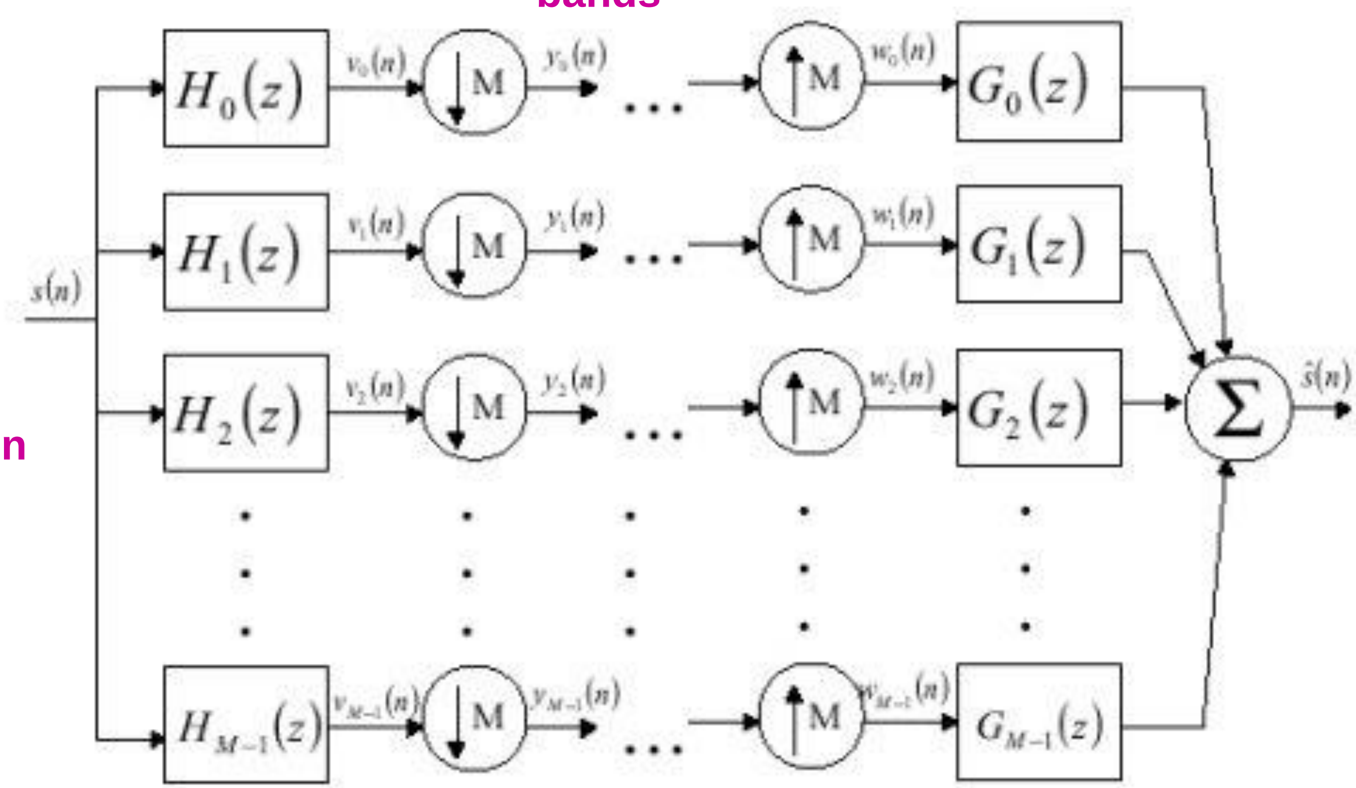
There are two main ways to perform frequency coding:

- **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.**
- **TRANSFORM CODING – A block of samples is (1D) linearly transformed using a discrete transform into a set of quasi-uncorrelated coefficients.**

Subbanded Frequency Decomposition

Audio samples in bands

Audio samples in



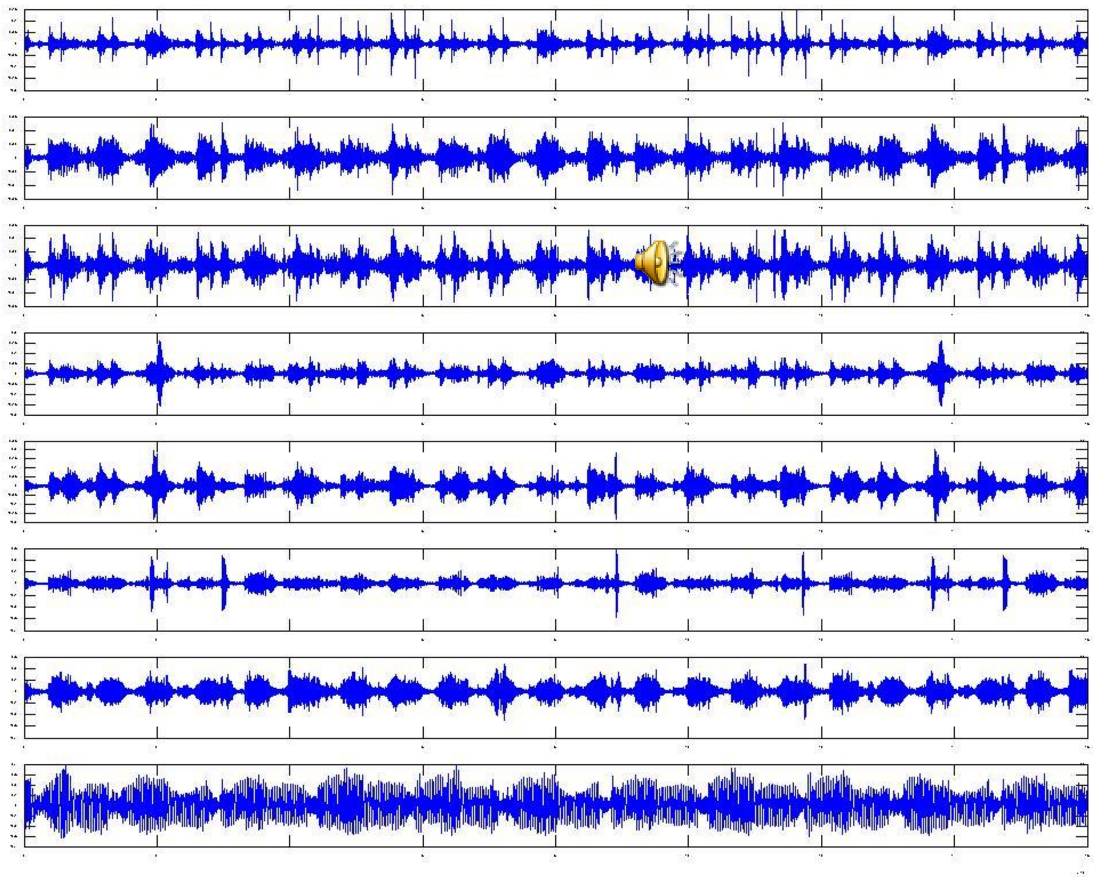
Audio samples out

Filter bank Analysis (M filters)

Filter bank Synthesis (M filters)

Signals in Different Frequency Bands …

Use of filterbanks (e.g. separating the frequency information…); musical signal in different bands with Fs=16 kHz



Bands 8-16 (3500 – 8000 Hz)

Band 7 (3000 – 3500 Hz)

Band 6 (2500 – 3000 Hz)

Band 5 (2000 – 2500 Hz)

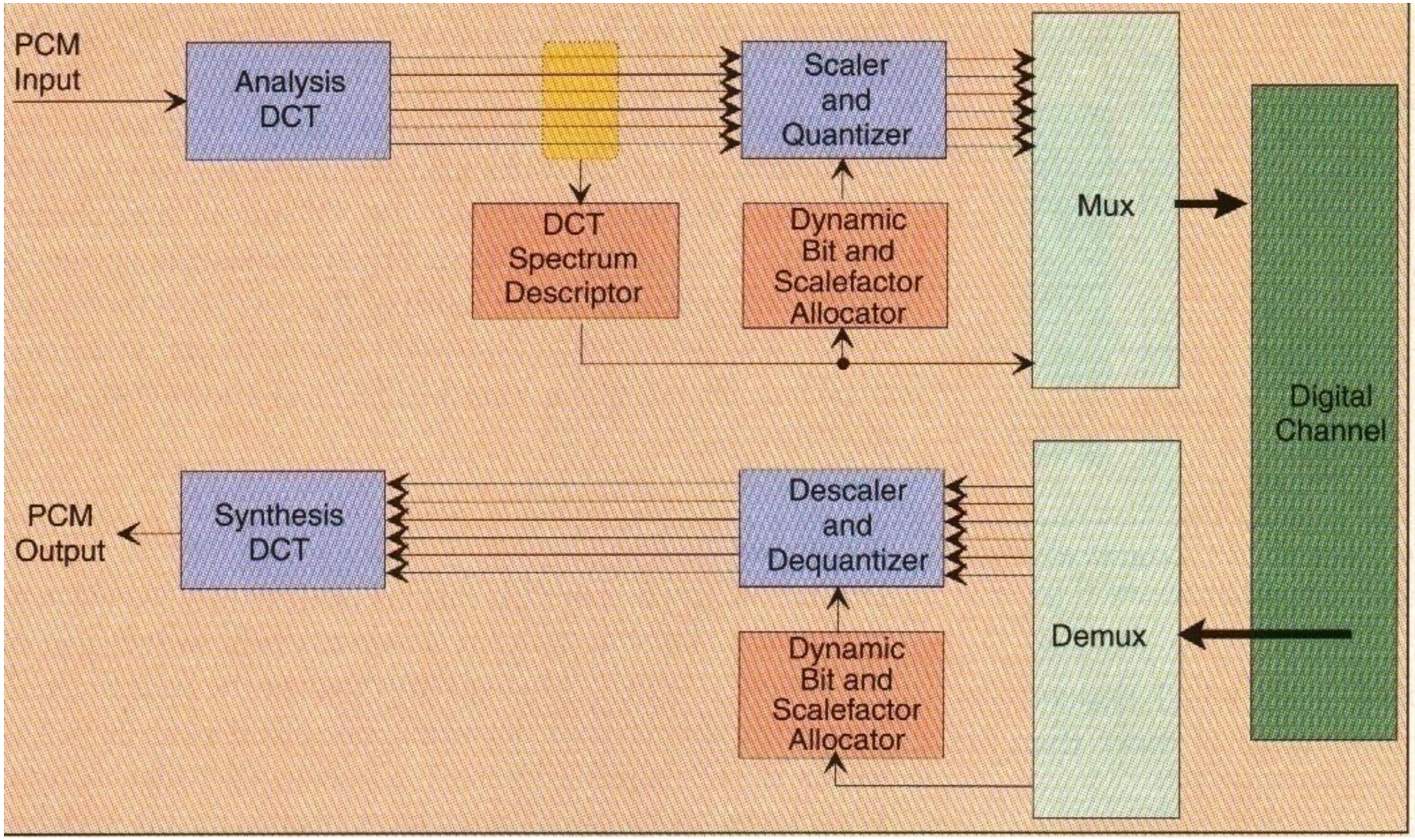
Band 4 (1500 – 2000 Hz)

Band 3 (1000-1500 Hz)

Band 2 (500 – 1000 Hz)

Band 1 (0 – 500 Hz)

Transform Coding Architecture



Depending on their frequency band, the various frequency coefficients are differently quantized. Transform coding may lead to block effects.

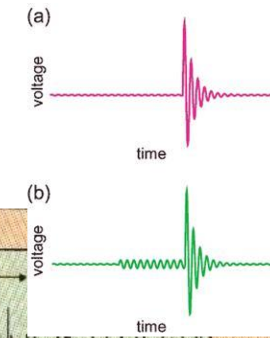
Echoes and Window Switching

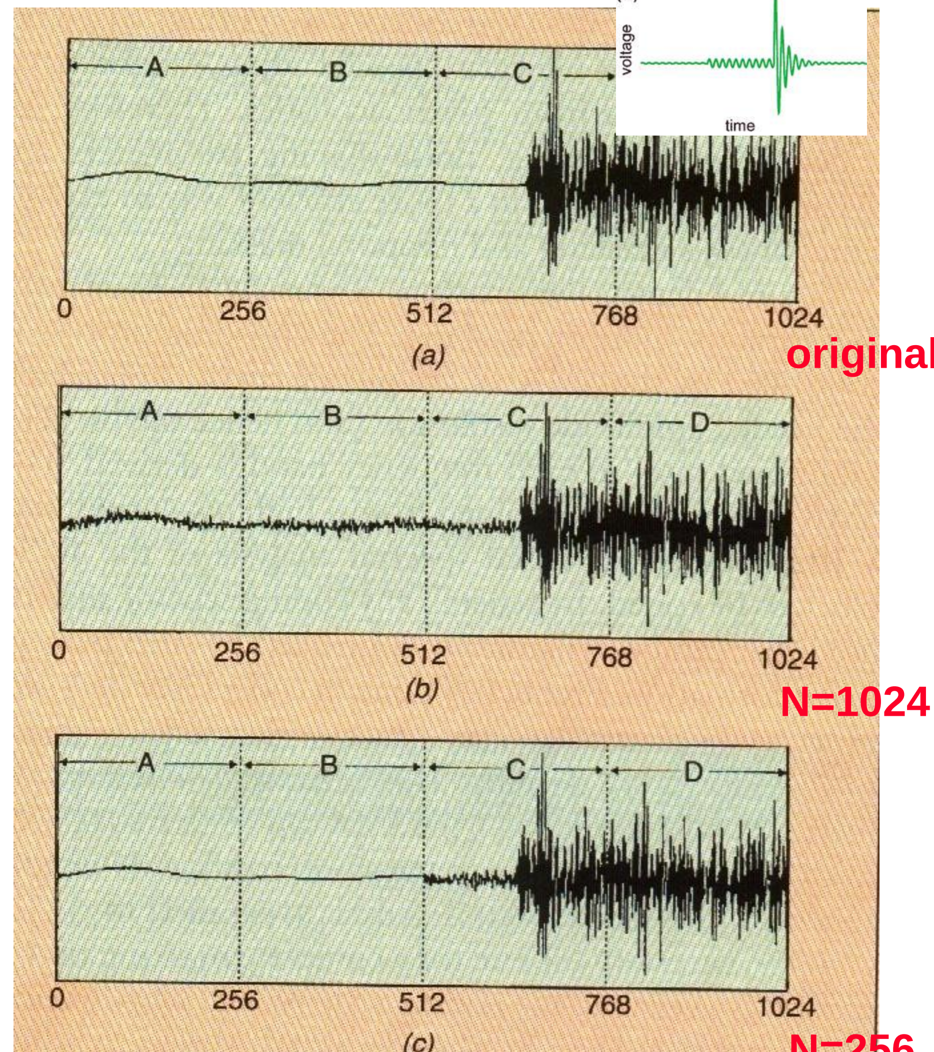
- **PROBLEM:**

  Pre-echoes/post-echoes: The usage of transform coding for blocks of samples where silence is followed by a strong signal (or vice-versa) creates the so-called <u>pre-echoes (or post-echoes)</u> since the signal synthesis may significantly change the silent part of the signal (in a more or less stronger way depending on the quantization).

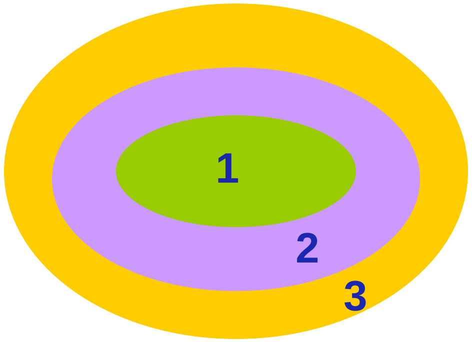
- **SOLUTION:**

  Variable size windowing - To limit this (annoying) phenomenon, variable size transform windows may be used with the encoder selecting the adequate window size depending on the signal characteristics.





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:

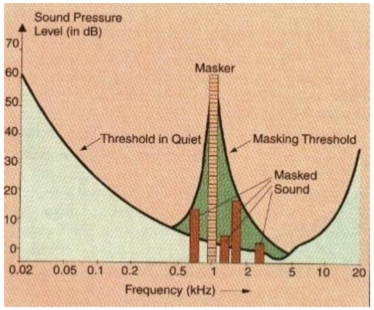
- **Each layer offers a rate/quality/complexity trade-off**
- **Higher layers have higher complexity, delay and coding efficiency**
- ***Nth* layer decoders are able to decode *(N-1)th* layers coded streams, defining a hierarchy of decoders and bitstream syntaxes**

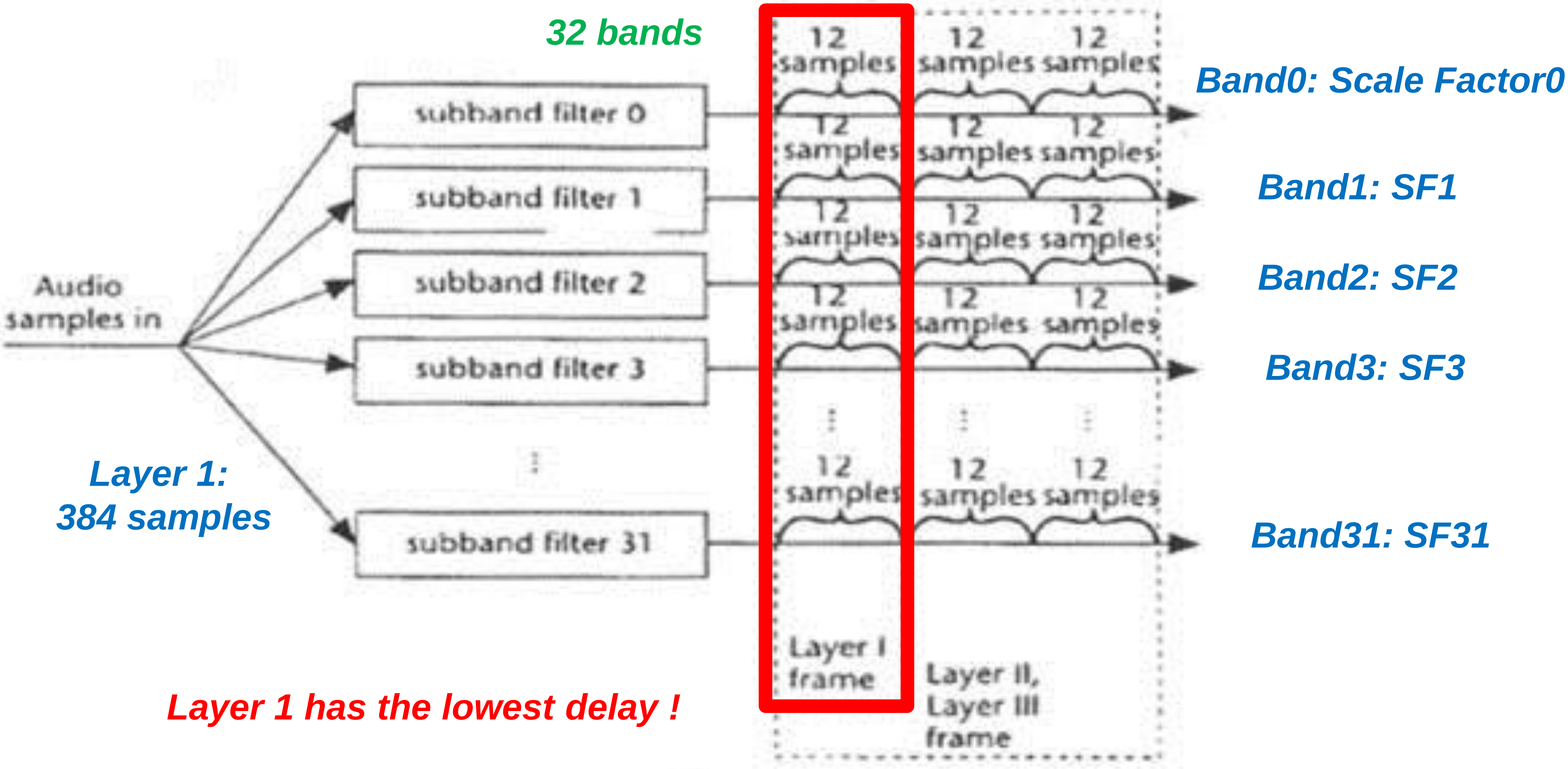
Layer	Typical rate	Minimum coding delay
1	32-448 kbit/s	$(256+256+12\times32)/48k \approx 19$ ms
2	32-384 kbit/s	$(256+256+12\times32\times3)/48k \approx 35$ ms
3	32-320 kbit/s	$(256+256+18\times2\times32\times2)/48k \approx 59$ ms

MPEG-1 Audio: Layer 1

- **<u>Blocks with 384 audio samples</u> are coded (corresponding to 8 ms at 48 kHz)**
- **Signal is decomposed into <u>32 uniform subbands</u>**
- **Fixed segmentation of <u>12 samples per subband</u> (total of 12×32=384 samples)**
- **<u>APCM type quantization</u> with adaptive block companding using a *scale factor* for each subband with 0-15 bit/sample; this value may adaptively change for each subband; each *scale factor* 'costs' 6 bits (maximum 6×32=192 bits/frame)**
- **<u>Psychoacoustic models</u> 1 and 2 suggested in the standard (there are two models 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; typical <u>compression factor of 4</u>**

Samples, Frames and Subbands …

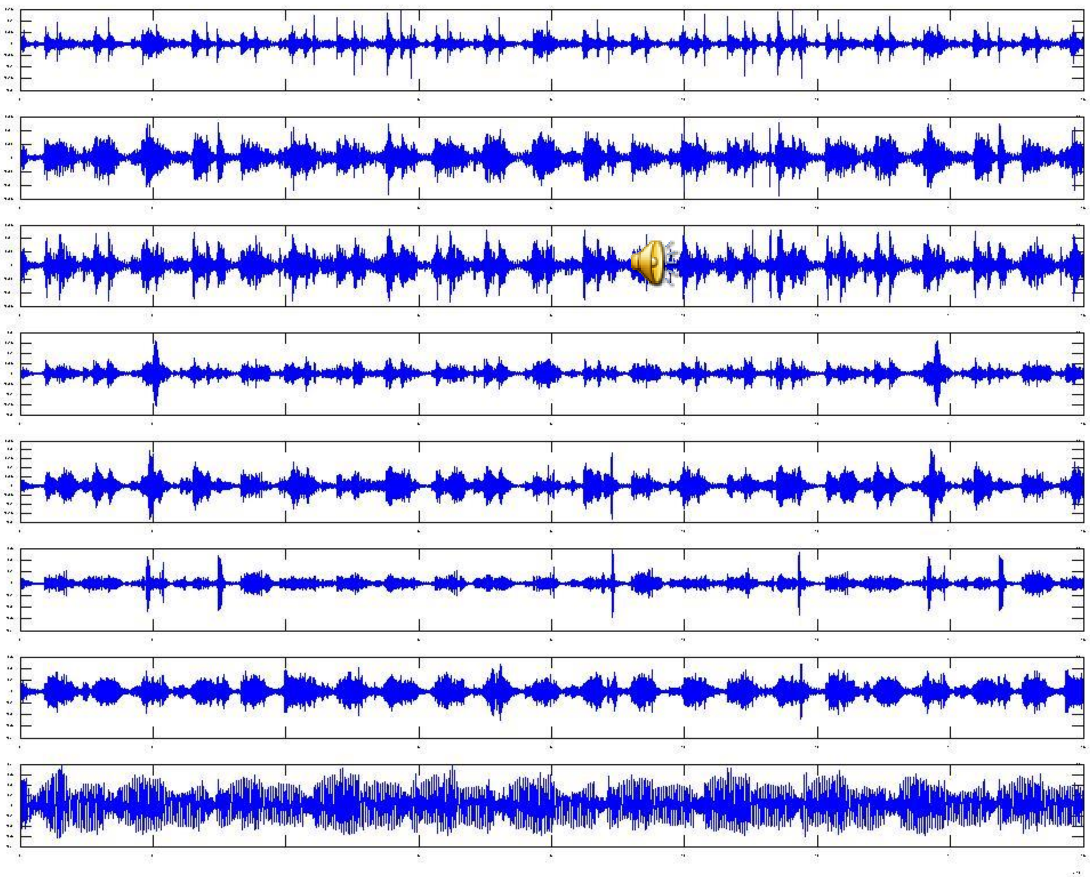




Layer 1 has the lowest delay !

Signals in Different Frequency Bands …

Use of filterbanks (e.g. separating the frequency information…); musical signal in different bands with Fs=16 kHz



Bands 8-16 (3500 – 8000 Hz)

Band 7 (3000 – 3500 Hz)

Band 6 (2500 – 3000 Hz)

Band 5 (2000 – 2500 Hz)

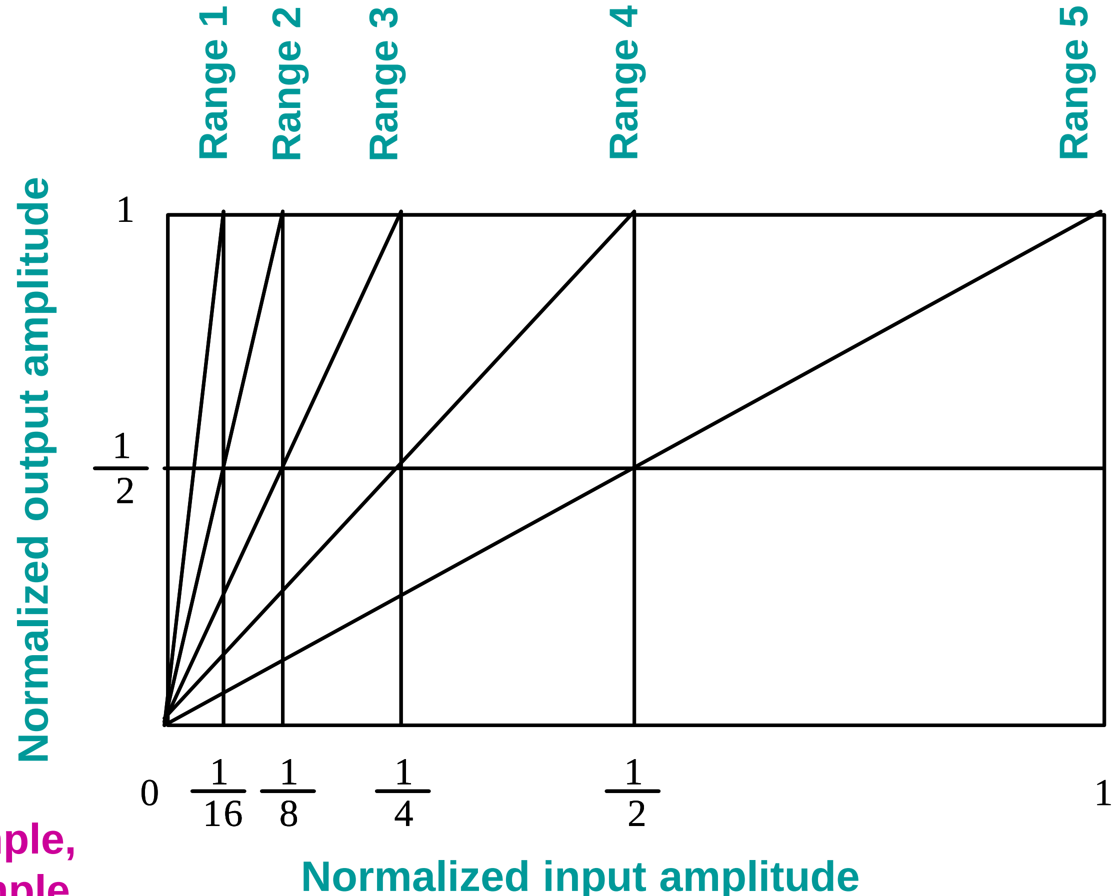
Band 4 (1500 – 2000 Hz)

Band 3 (1000-1500 Hz)

Band 2 (500 – 1000 Hz)

Band 1 (0 – 500 Hz)

Adaptive Block Companding with Scale Factors …



The basic idea is to reduce the quantization impacts for limited dynamic range signals; notably lower amplitude samples will be less penalized.

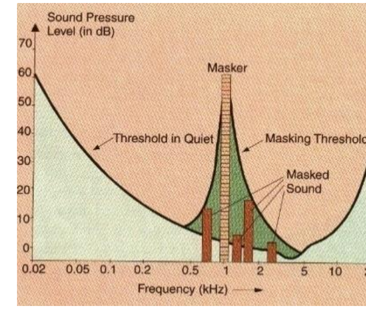
K<N bit/sample, 0-15 bit/sample

N bit/sample, typically 16 bit/sample

MPEG-1 Audio: Layer 2

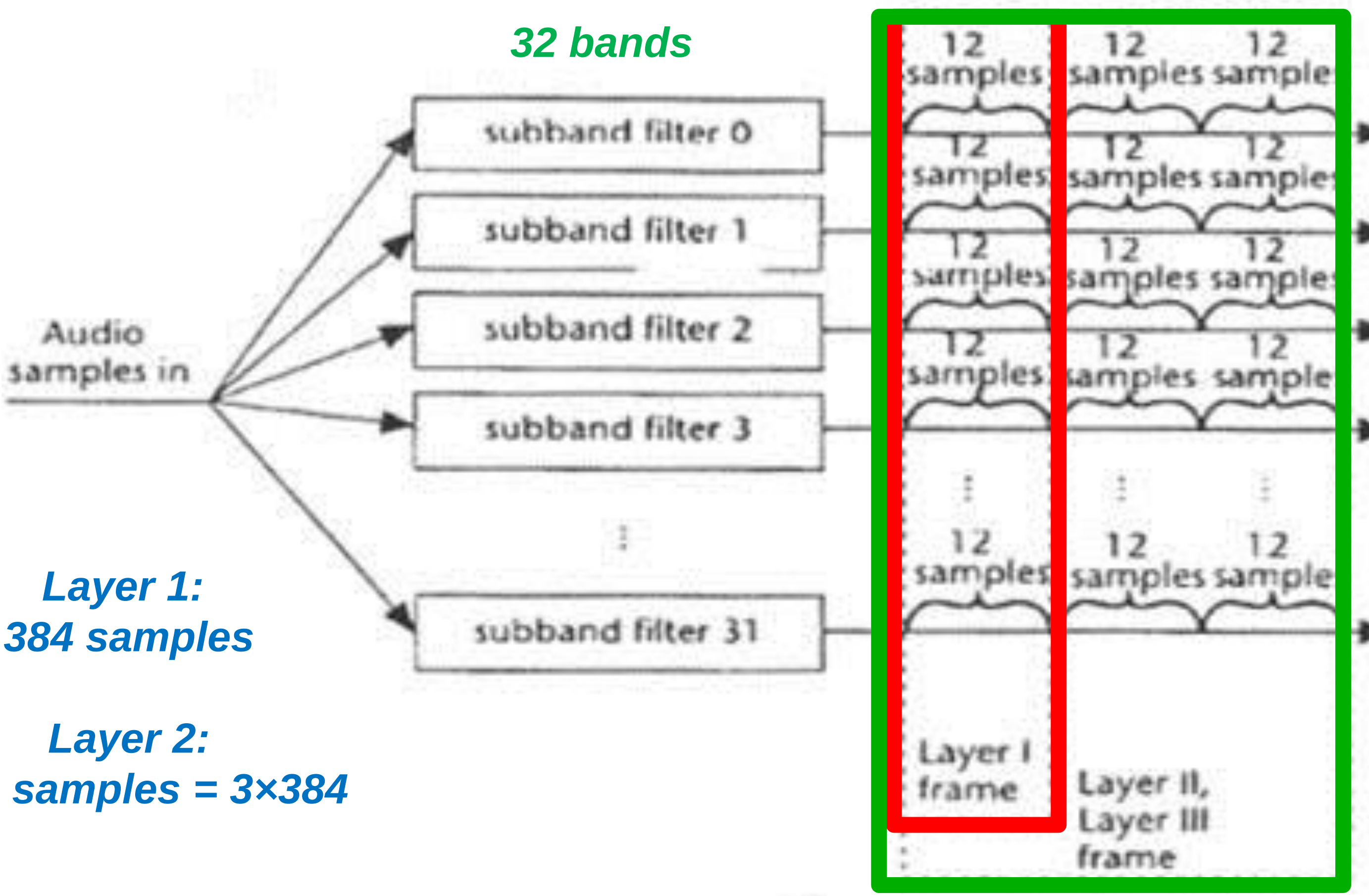
- **<u>Blocks with 3 × 384 = 1152 audio samples</u> are coded (corresponding to 24 ms at 48 kHz; 3 times more than for Layer 1)**
- **Fixed segmentation with <u>3 × 12 = 36 samples per subband</u>**
- **Coding algorithm as for Layer 1 with the exception of using more efficient methods to code the quantization *scale factors* by <u>exploiting the redundancy between the adjacent *scale factors*</u> within the 3 sub-blocks of 12 samples in each band.**
  - **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 may need to be coded.**
- **Transparent quality regarding the CD quality (PCM) at 192 kbit/s; typical <u>compression factor of 8</u>**

Samples, Frames and Subbands …



Delay at work !

32 bands



B0: SFa, SFb, SFc

B1: SFa, SFb, SFc

B2: SFa, SFb, SFc

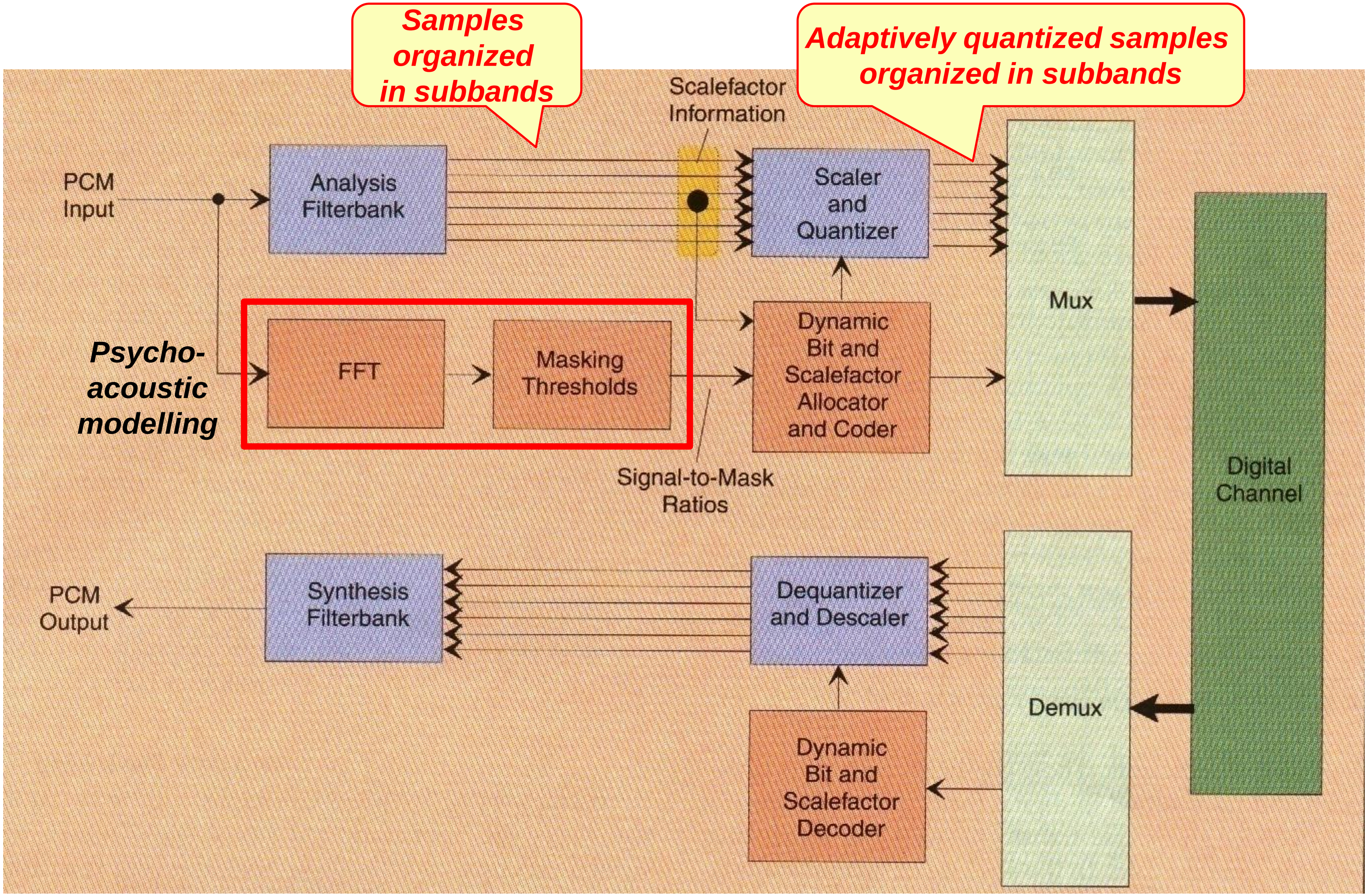
B3: SFa, SFb, SFc

B31: SFa, SFb, SFc

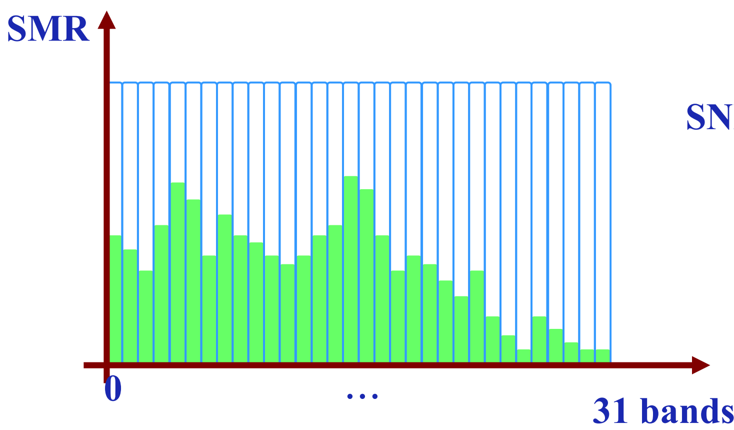
Layer 1:
384 samples

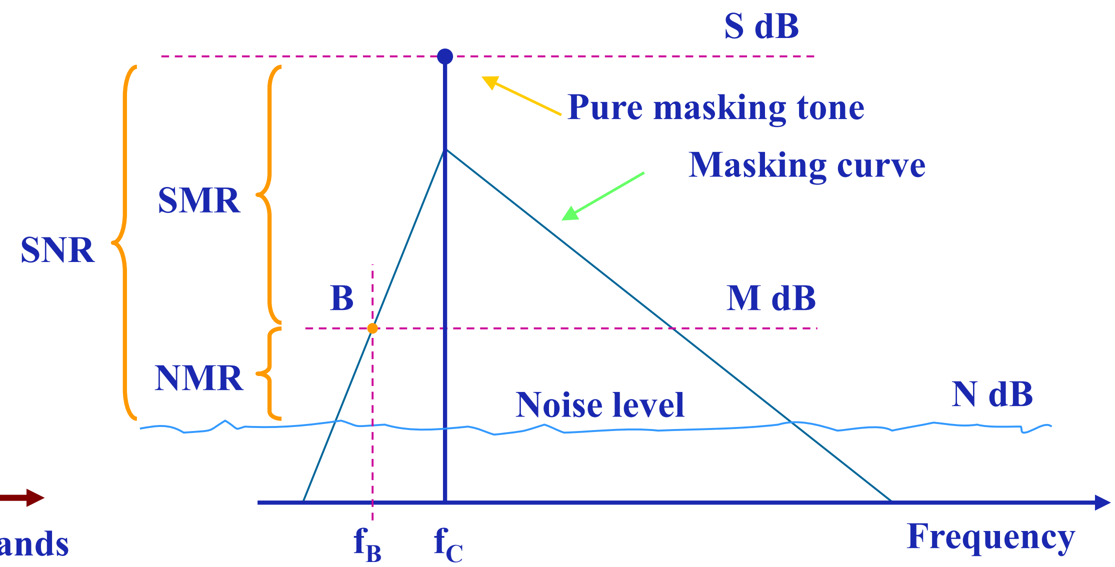
Layer 2:
1152 samples = 3×384

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.


MP3

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

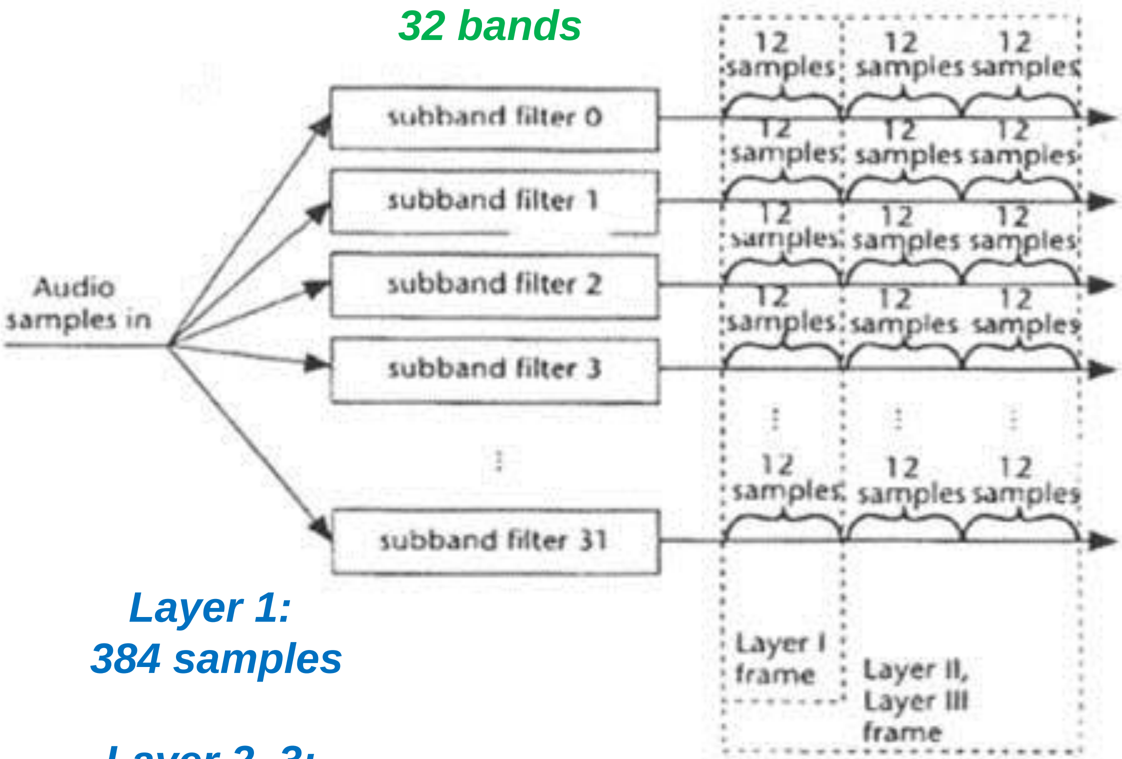
- **Blocks with 1152 audio samples are coded corresponding to 24 ms at 48 kHz.**
- **Hybrid time/frequency coding structure -** The filter bank (creating the subbands) is followed by transform coding (Modified DCT) to have a finer (frequency) granularity characterization of the signal (*using a time-frequency hierarchy*).
- **Dynamic (transform)window switching** – To increase the frequency resolution, each of the 32 subbands is characterized with more (frequency) detail by applying to each of them a transform with 6 or 18 MDCT coefficients; this results into a maximum number of frequency components of 32×18 (6) = 576 (or 192). The smallest window allows to control the temporal resolution and, thus, to reduce the pre-echo effect, eventually at the cost of some block effect.
- **Overlapping (transform) windows -** The MDCT is applied with 50% window overlapping to reduce the block artifacts meaning that the MDCT is applied to sets of 12 or 36 subband samples.

Hybrid Time/Frequency Coding Structure

time

frequency

32 bands



36 samples per band → MDCT → *MDCT coefficients*

...

...

...

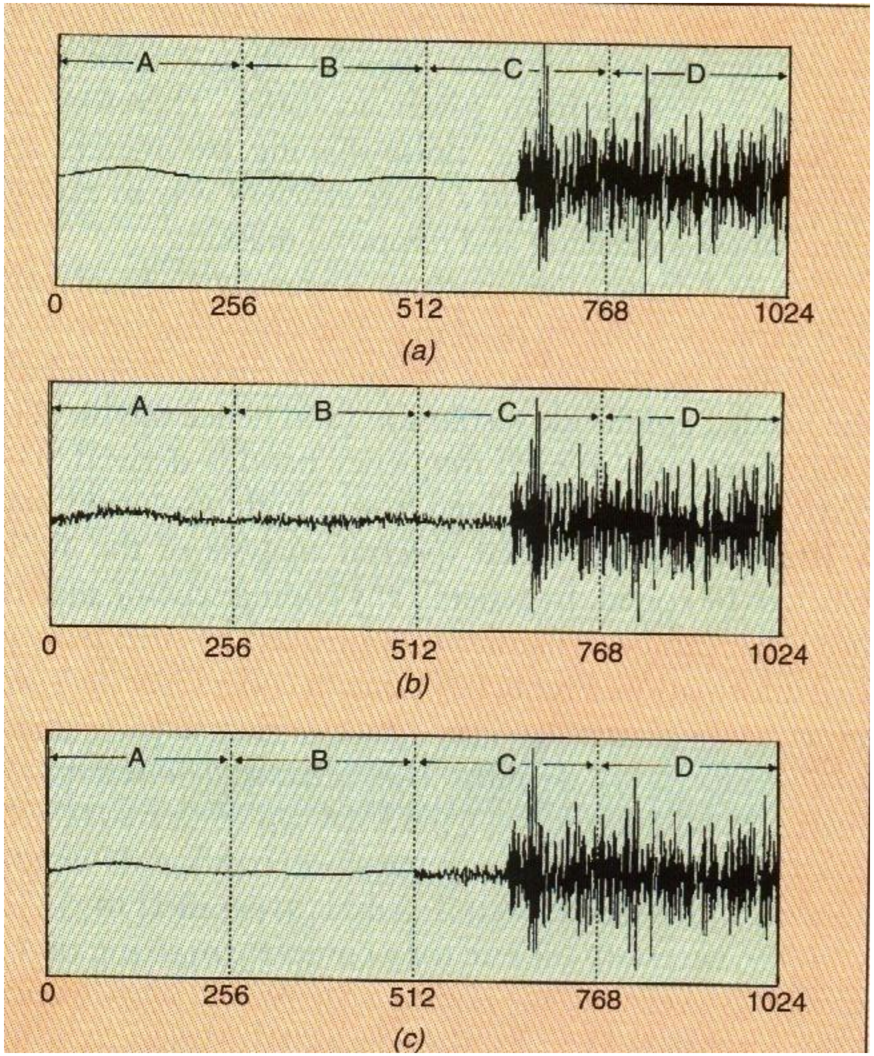
36 samples per band → MDCT → *MDCT coefficients*

Layer 1: 384 samples

Layer 2, 3: 3×384=1152 samples

Provides a chirurgical frequency characterization of the signal ...

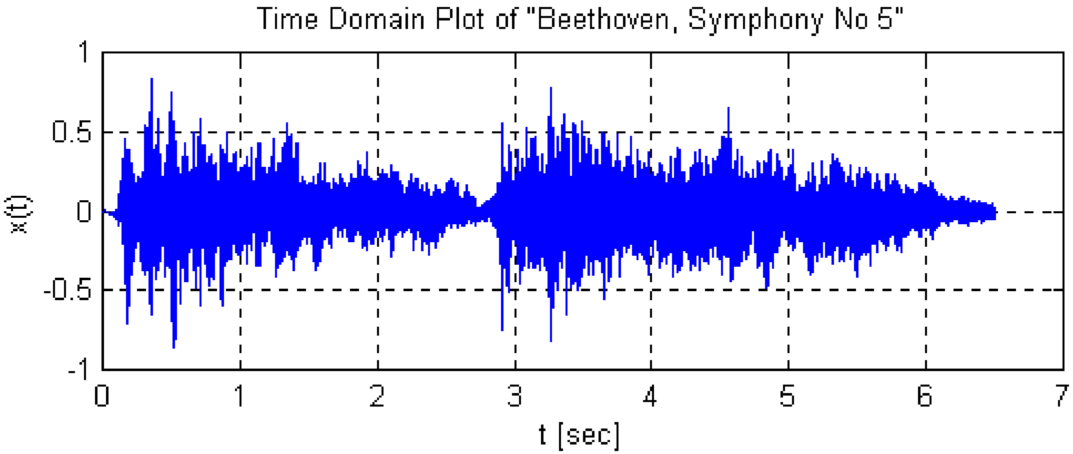
Dynamic Window Switching



The transform window size has to be dynamically adapted to avoid pre-echos !

Pre-echos are subjectively very annoying ...

Overlapping Transform Windows

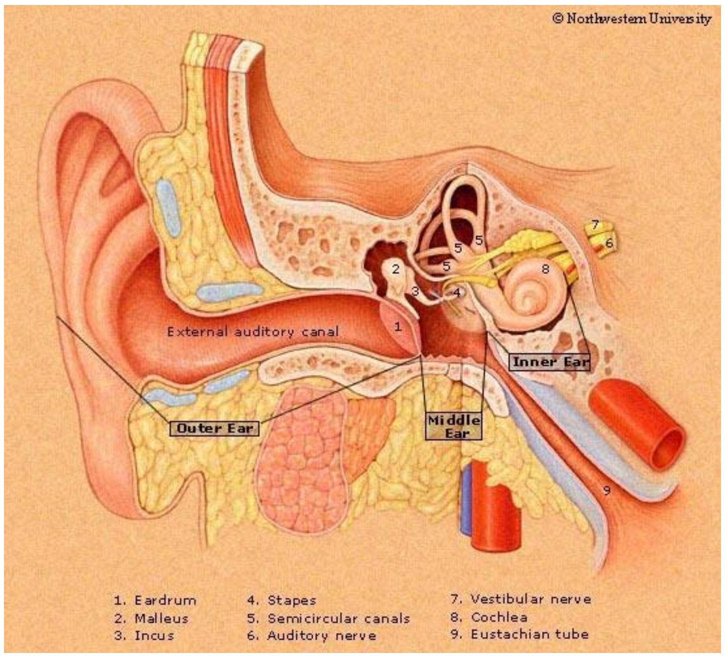


The overlapping transform windows allow to reduce the subjectively annoying block effect, especially at lower rates.

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; typical compression factor of 12

Quantization Control: the Psychoacoustic Model



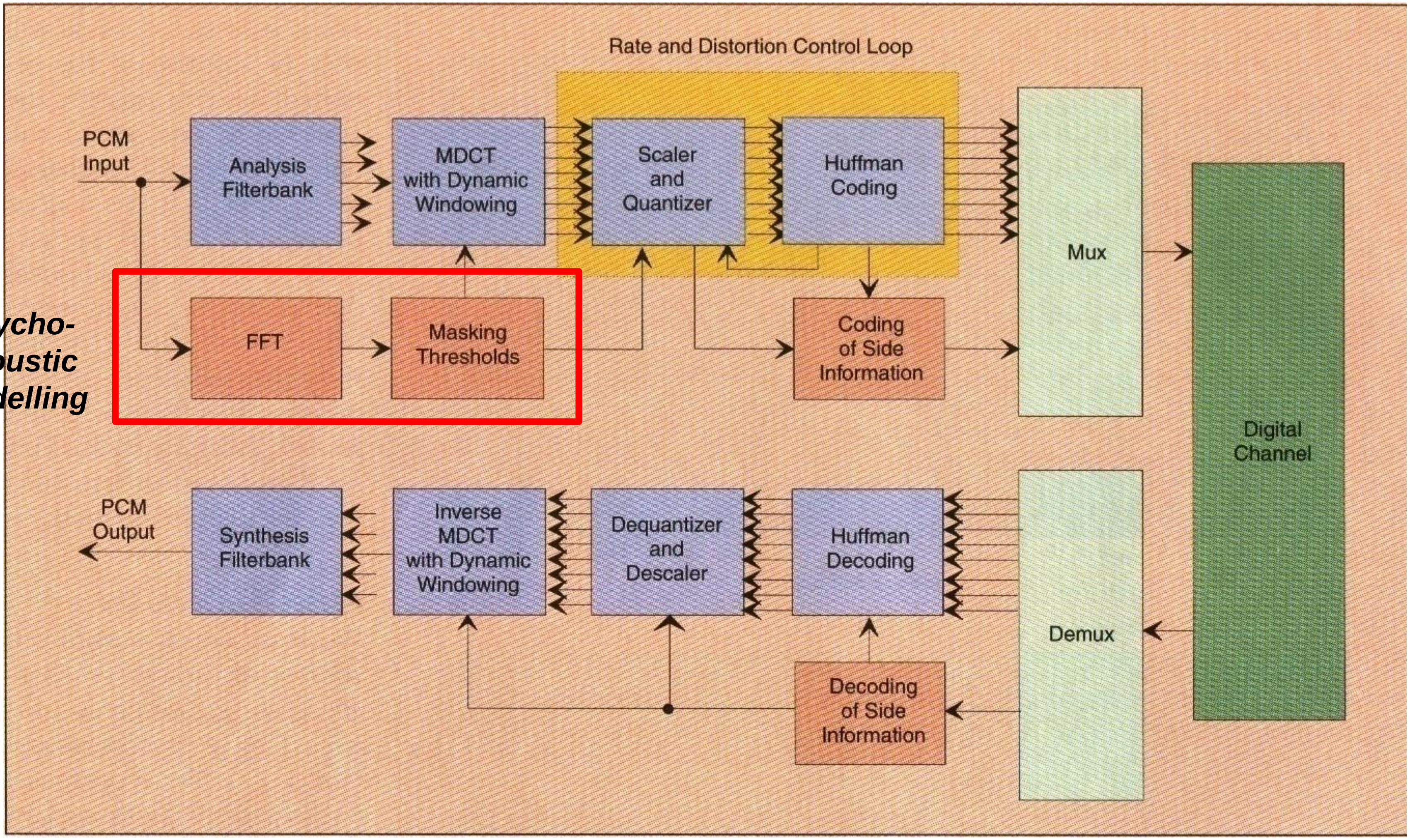
A psychoacoustic model is a mathematical model which defines, in a more or less simplified way, the main properties and tolerances of the human auditory model, notably its sound intensity perception, its spectral selectivity and, especially, the masking effect.

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.

MP3 Encoder Architecture

Check the similarities with a JPEG encoder !

Psycho-acoustic modelling



MP3 Encoding Walkthrough …

- **FREQUENCY DECOMPOSITION - Divide the audio spectrum into 32 frequency bands (known as *sub-bands*) using a filter bank.**

- **MDCT TRANSFORM – Apply a 2×6 or 2×18 DCT window to compute the frequency components for each sub-band; 6 frequency components are used 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 (and thus the allowed noise) for each spectrum partition.**

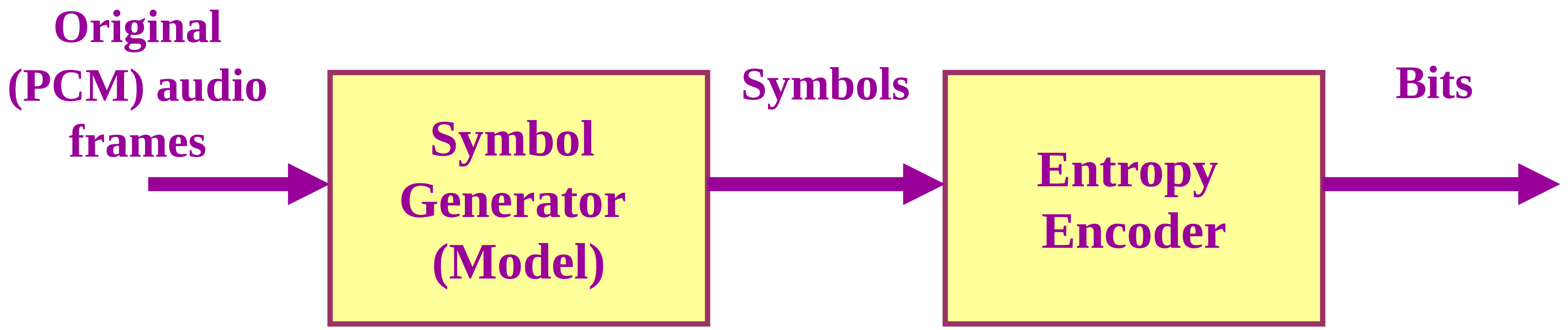
- **QUANTIZATION - Quantize the DCT components for each band using the defined quantization step and scale factor. If the quantization noise can be kept below the masking threshold, then the compression results should be indistinguishable from the original signal.**

- **ENTROPY CODING – The quantized DCT components for each band are entropy coded.**

MP3 Performance …

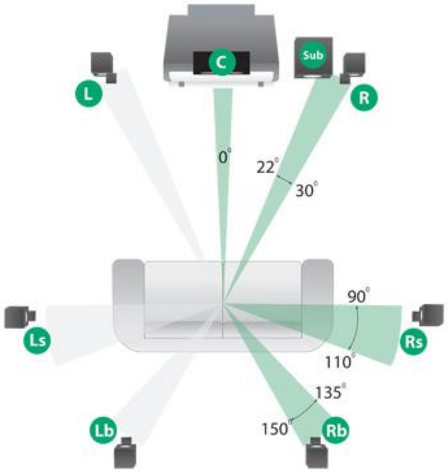
Sound quality	Bandwidth	Mode	Bitrate	Compression factor
telephone sound	2.5 kHz	mono	8 kbps *	96:1
better than shortwave	4.5 kHz	mono	16 kbps	48:1
better than AM radio	7.5 kHz	mono	32 kbps	24:1
similar to FM radio	11 kHz	stereo	56...64 kbps	26...24:1
near-CD	15 kHz	stereo	96 kbps	16:1
CD	>15 kHz	stereo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

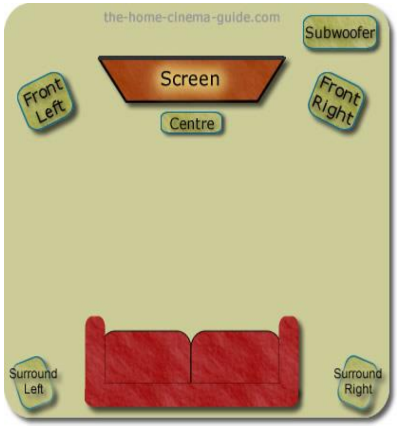
The MP3 Symbolic Model



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

Stereo … and More …





MPEG-1 Audio: Stereo Coding

iPod Not Included

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.

There are 5 MPEG-1 Audio coding modes:

- **Mono**
- **Dual Stereo – Channels are independently coded, e.g. 2 different languages.**
- **Stereo – <u>Independent coding</u> but sharing certain fields in the coded frame.**
- **Joint Stereo – <u>Channel dependency is exploited</u> through the so-called *intensity stereo technique*; above 2 kHz, the L+R signal is coded together with scale factors for the two channels (L and R) since there is lower hearing sensibility. Joint stereo is used in cases where only low bitrates are available but stereo signals are desired.**
- **Mono/Stereo (MS) (only layer 3) – <u>Channel dependency is exploited</u> with the two channels coded as L+R (middle) and (side) L-R, thus allowing to better control the spatial location of the quantization noise; this provides backward compatibility with mono decoders.**

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 us at 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

Patent terms for MPEG-1 MP3 have now expired …

The Snow Ball Effect …

- Easy exchange of music
- Piracy
- Peer-to-peer file sharing service, *Napster*
- Digital Rights Management
- New business models
- …

"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 …

TÉCNICO LISBOA

Final Remarks

- **MPEG-1 Audio Layer (MP3) is commonly used for music in the Web and much more ...**
- **MP3 players are used in a very large number of devices, applications, etc., notably portable.**
- **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 issues this means digital rights management … and Napster … and peer-to-peer … new business models …**

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