VIDEO COMPRESSION STANDARDS

The Need for Compression

Image sequences must be significantly compressed for efficient storage and transmission as well as for efficient data transfer among various components of a video system.

Examples

▷ Motion Picture:
One frame of a Super 35 format motion picture may be digitized (via Telecine equipment) to a 3112 lines by 4096 pels/color, 10 bits/color image.
As a result, 1 sec. of the movie takes $\approx$ 1 Gbytes !

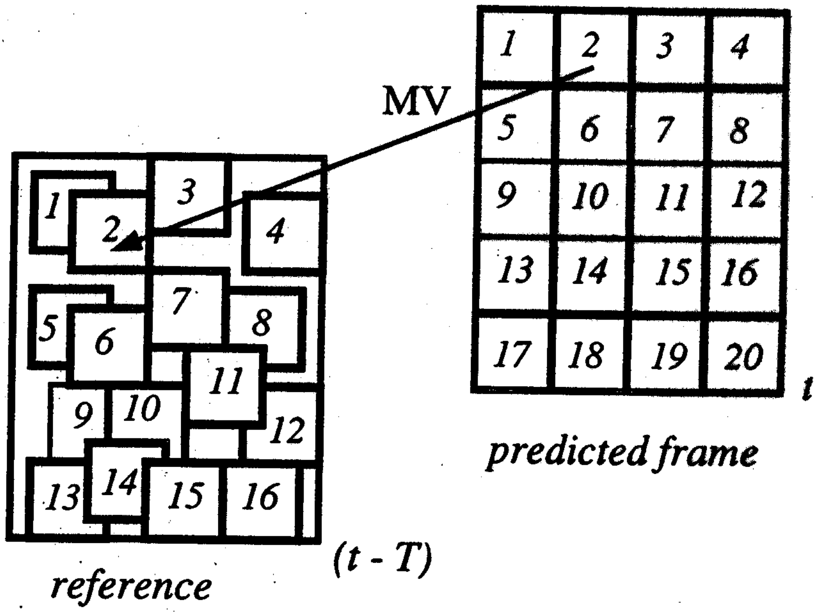
▷ HDTV:
A typical progressive scan (non-interlaced) HDTV sequence may have 720 lines and 1280 pixels with 8 bits per luminance and chroma channels.
The data rate corresponding to a frame rate of 60 frames/sec is $720 \times 1280 \times 3 \times 60 = 165$ Mbytes/sec !

20 Mbits/sec. nur erlaubt !!

Approaches to Video Compression

1. Intraframe compression treats each frame of an image sequence as a still image. Intraframe compression, when applied to image sequences, reduces only the *spatial* redundancies present in an image sequence.

2. Interframe compression employs temporal predictions and thus aims to reduce *temporal as well as spatial* redundancies, increasing the efficiency of data compression. Example: Temporal motion-compensated predictive compression.



The reference frame is available at the decoder;

MV's and Prediction Error Blocks (PEB) are encoded and transmitted;

From MV's and PEB's the predicted frame can be reconstructed at the decoder.

- Some application specific requirements, such as random access to all frames, may require intraframe compression, at the expense of decreased efficiency. z.B. Editing

Standards Relevant to Video Compression

Standard Activity	Description
JPEG	Joint (ITU-ISO) Photographic Expert Group; Primarily designed for still imagery.
H.261	International Telecommunication Union (ITU) Recommendation; Videophone Designed for ISDN applications at $p \times 64$ Kbps ($p = 1, 2, \cdots, 30$);
H.263	International Telecommunication Union (ITU) Recommendation; public switched telephone network Video Coding for $<$ 64 Kbs communication Viideophone over PSTN (e.g., via 28.8 Kbps modem).
H.263+	Extensions to H. 263 (to be finalized April 1997).
MPEG	ISO Moving Picture Expert Group MPEG 1: Efficient storage and retrieval of video + audio at about 1.5 Mbps; (rate of CD-rom) MPEG 2: Efficient storage and retrieval of video + audio at higher bit rates. (Broadcast) MPEG 4: Efficient compression + additional functionalities (to be standardized November 1998).

ISO: International Standards Organization

ITU: International Telecommunications Union.

MPEG 7: MM - Data base search standard

MPEG 3: HDTV → MPEG 2

Applications Video Compression Standards

Broadcast

- Advanced Television
  - US-Standard Definition Digital TV (SDTV) $---$ *MPEG2*
  - US-Grand Alliance HDTV $---$ *MPEG2*

Multimedia and Entertainment

- Video on CDROM (Video CD, CD-Interactive) $---$ *MPEG1*
- Video on Digital Versatile Disk (DVD) $---$ *MPEG2*
  (DVD is high density disk $\sim 4.7GB$ versus CDROM $\sim 680MB$)

Personal Communication

- Videophone and videoconferencing over ISDN$---$ *H.261*
- Videophone and videoconferencing over PSTN – *H.263, H.263+*

MPEG1: General Remarks

□ The MPEG1 standard has 5 main parts:

ISO 11172-1 Systems: Defines a multiplexed structure of the combination of video and audio data, and means of representing the timing info needed to synchronize the video and audio replay.

ISO 11172-2 Video: Specifies the coded representation of the video data and the decoding process.

ISO 11172-3 Audio: Specifies the coded representation of the audio data and the decoding process.

ISO 11172-4 Compliance: Specifies testing procedures to test compliance of decoders to Parts 1, 2, and 3 of the standard. (Description of actual tests are defined for audio only.)

ISO 11172-5 Technical Report: Software simulations of Parts 1, 2, and 3.

□ MPEG1 standardization activities are based on the premise that video and its associated audio, at satisfactory quality, can be stored and retrieved at about 1.5 Mbits/sec. Among the factors motivating this bit rate are:

- CD-ROM is an inexpensive storage medium capable of delivering data at about 1.2 Mbits/sec.

ASIDE: Summary of MPEG1 Constrained Parameters

MPEG2 Thema - dis [illegible]

Horizontal picture size	$\leq$ 768
Vertical picture size	$\leq$ 576
Picture area	$\leq$ 396 Macroblocks
Pixel rate	$\leq 396 \times 25$ Macroblocks per sec.
Picture rate	$\leq$ 30 Hz
Bit rate	$\leq$ 1.856 Mbits/sec (constant)

- A flag in the bitstream indicates whether or not it is a *Constrained Parameter Bitstream.*

MPEG1: General Remarks (cont'd.)

□ The MPEG1 standard simultaneously supports both **interframe** and intraframe compression modes.

□ The MPEG1 standard considers:

- Progressive-format video only;
- Luminance and two chroma channels representation where **chroma** channels are subsampled by a factor of 2 in both directions; (4:2:0)
- 8 bit/pixel video.

□ The standard *syntax* supports the operations of

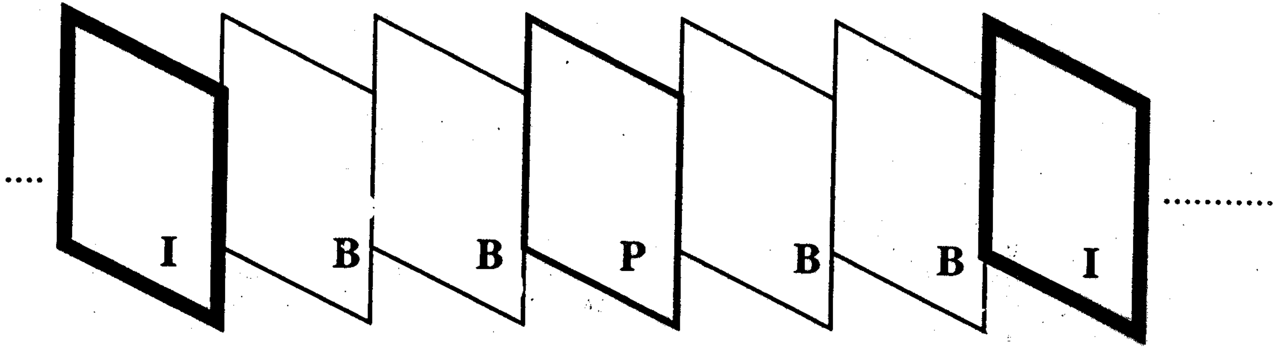
- discrete cosine transformation (DCT),
- motion-compensated prediction,
- quantization, and
- variable length coding.

- Substantial flexibility, is allowed in designing the encoder.

For example, MPEG1 does not standardize the motion estimation algorithm.

wd: Encoder ist auch ¬ vorgeschrieben,
nur DECODER!!

MPEG1: Basic Compression Modes

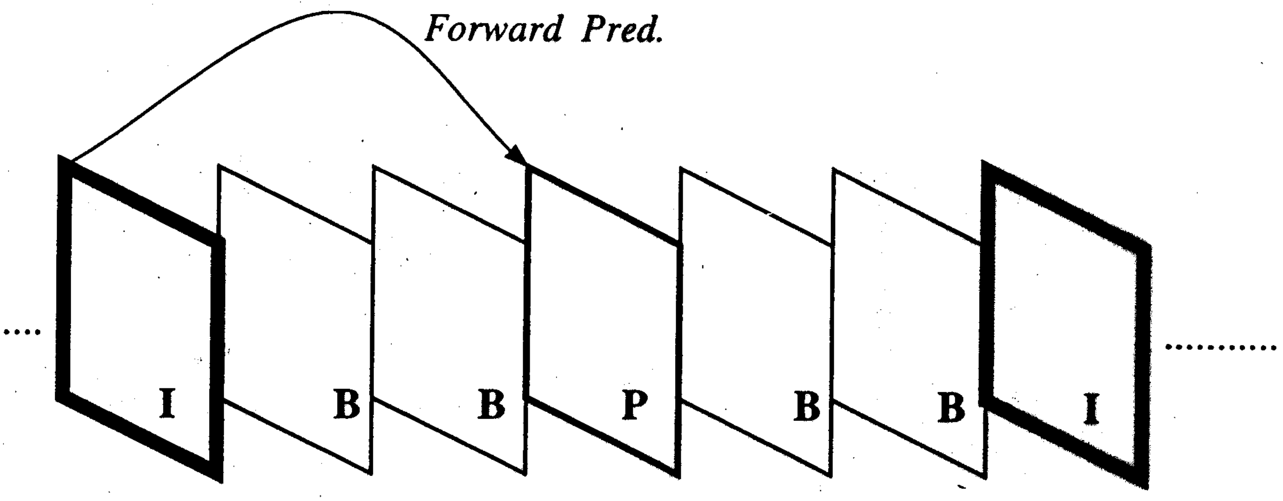


--A typical assignment of compression modes for the frames of an image sequence.

Intraframe Compression

- Frames marked by **(I)** denote the frames that are strictly <u>intraframe</u> compressed. The purpose of these frames, called the "**I** pictures", is to serve as random access points to the sequence.

MPEG1: Basic Compression Modes (cont'd.)

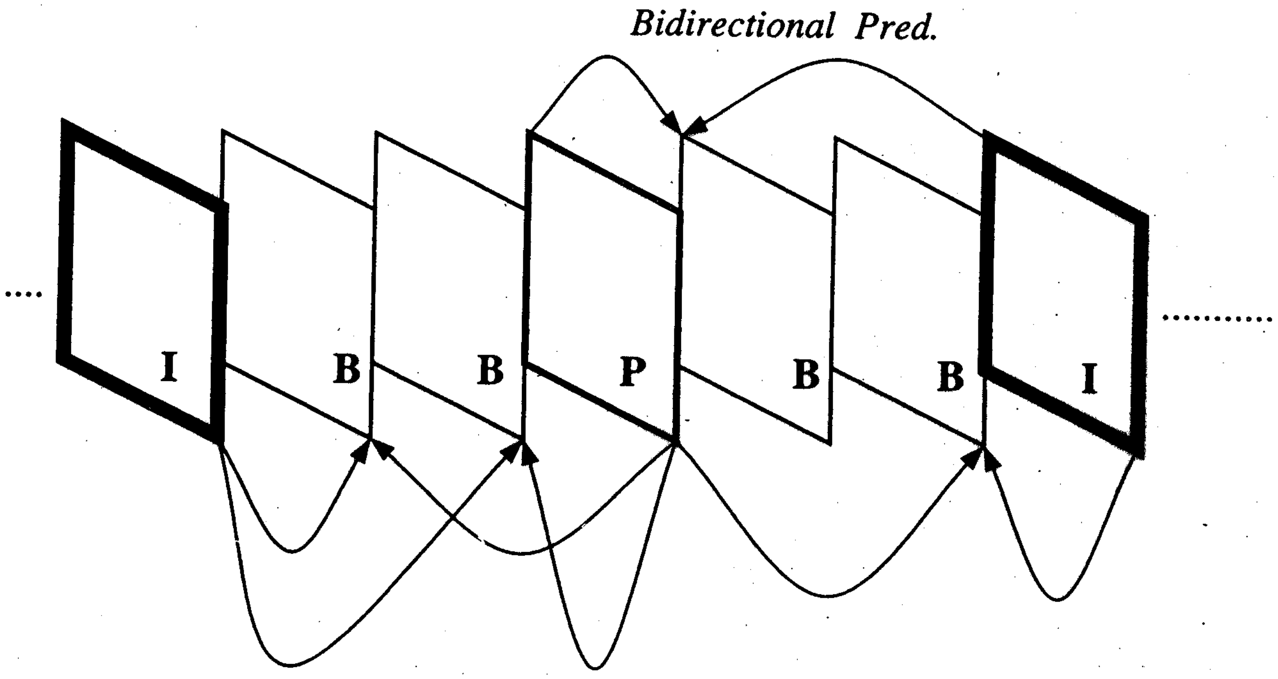


--A typical assignment of compression modes for the frames of an image sequence.

Non-Intra (Motion-Compensated Predictive) Compression

- In frames marked by (**P**), *motion compensated forward predictive* compression is performed on a block basis. Predicting blocks from closest (most recently decoded) **I** and **P** pictures are utilized. Motion vectors and prediction errors are coded. In **P** pictures, blocks are allowed to be intra compressed if the prediction is deemed to be poor.

MPEG1: Basic Compression Modes (cont'd.)



--A typical assignment of compression modes for the frames of an image sequence.

Non-intra (Motion-Compensated Predictive) Compression (cont'd.)

- In frames marked by **(B)**, *motion-compensated bidirectional predictive* compression is performed on a block basis. Predicting blocks from closest (most recently decoded) **I** and **P** pictures are utilized. Motion vectors and prediction errors are coded. In **B** pictures, blocks are allowed to be <u>intra</u> compressed if the prediction is deemed to be poor.

MPEG1: Basic Compression Modes (cont'd.)

- Relative number of **I**,**P**, and **B** pictures depends on
  - access and decoding delay requirements
  - amount of compression
  - nature of the content
- Some posibilities:

  ··· **I B B P B B I B B P B B I** ···

  ··· **I I I I I I I I I I I I I** ···

  ··· **I P I P I P I P I P I P I** ···

  ··· **I P I P I I I P I P I I I** ···

- Group Of Pictures (GOP):

A GOP contains at least one **I** picture, and an arbitrary number of **B** and **P** pictures.

I B B P B B P B B P B B P B B I

Parameters: $M = 15$ (distance between two **I** pictures) and $N = 3$ (distance between two **P** pictures)

- Maximum value of M is constrained to be 132.

MPEG1: Basic Compression Modes (cont'd.)

- Display and Encoding/Decoding order:

Encoder In./Decoder Out./Display:	1	2	3	4	5	6	7
	$\boxed{I}$	$\boxed{B}$	$\boxed{B}$	$\boxed{P}$	$\boxed{B}$	$\boxed{B}$	$\boxed{I}$
Encoder Out./Bitstream/Decoder In.:	1	3	4	2	6	7	5

Intra Compression

Quantization

- The DCT coefficients are uniformly quantized.

The DC Coefficient

The DC coefficient is divided by 8, and the result is truncated to the nearest integer in $[-256, 255]$

$$\boxed{QF(0,0) = NINT[F(0,0)/8],}$$

AC Coefficients

Each AC coefficient, $F(u,v)$ is first multiplied by 16, and the result is divided by a weight, $w(u,v)$, times the *quantizer_scale.*

$$\boxed{QF(u,v) = NINT[16 * F(u,v)/(w(u,v) * quantizer\_scale)].}$$

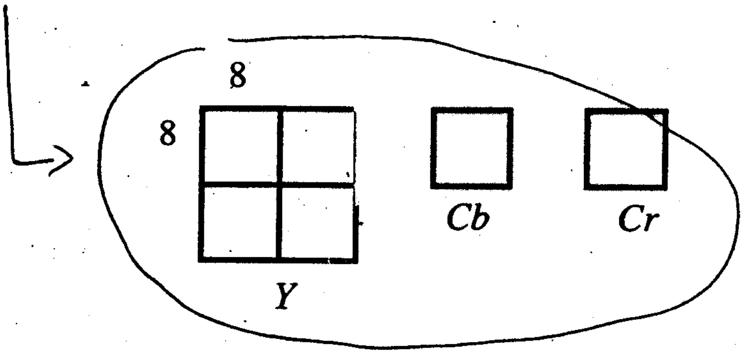
The result is then truncated to $[-256, 255]$. The 8×8 array of weights, $w(u,v)$, is caled the *quantization matrix.* The parameter *quantizer_scale* facilitates adaptive quantization.

Intra Compression

Spatially-Adaptive Quantization

- Spatially-adaptive quantization is made possible by the scale factor *quantizer_scale.*

  quantizer_scale (called MQUANT) is allowed to vary from one "macroblock" to next.



The *quantizer_scale* is chosen from a specified set of values on the basis of

- spatial activity of the block
- buffer status in constant bitrate applications.

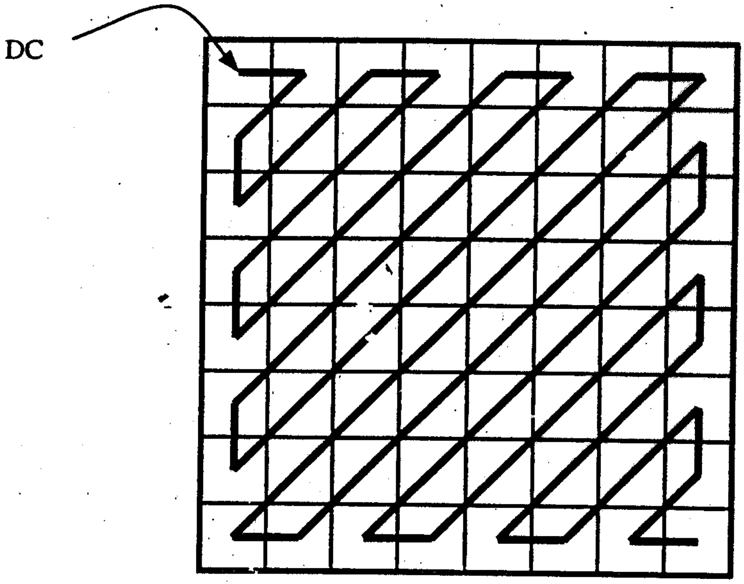
NOTE that the baseline JPEG standard does not allow for spatially-adaptive quantization (Adaptive quantization is among the extensions of JPEG.)

MPEG1: Intra Compression

Coding: AC Coefficients

- Main idea: represent the quantized coefficients by location and value of the nonzero coefficients.
- The quantized AC coefficients are scanned in a zigzag fashion and ordered into *symbol = [run, level]* pairs and then coded using variable length (Huffman) codes (VLC) (longer codes for less frequent pairs and vice versa). (The VLC tables are standardized.)

level: is the value of a nonzero coefficient;
run: is the number of zero coefficients preceding it.

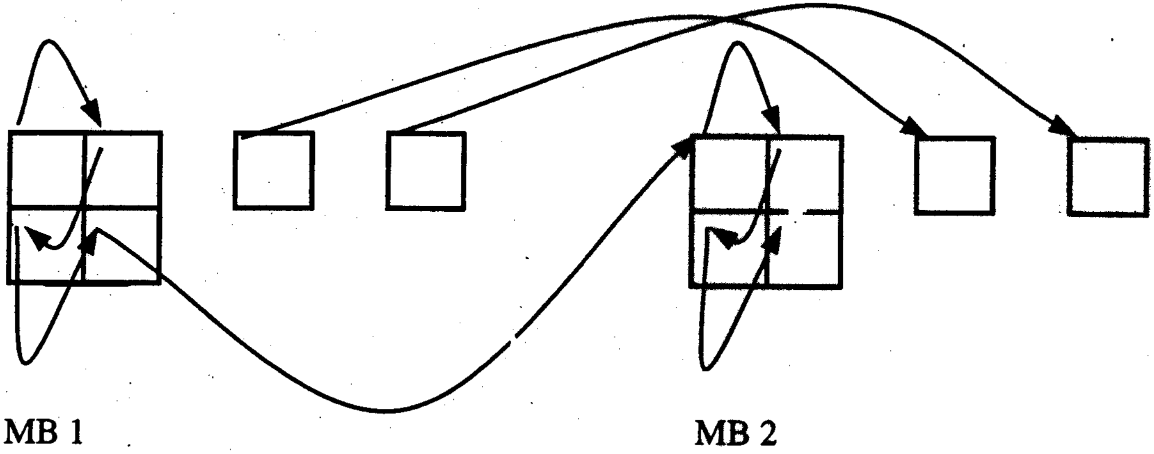


“zigzag” scan

MPEG1: Intra Compression

Coding: DC Coefficients

- Redundancy among quantized DC coefficients of 8 × 8 blocks is reduced via differential pulse coded modulation (DPCM). (**Standard VLC tables are specified. In fact, these tables are the only standard tables in MPEG1 that make a distinction between luminance and chrominance components of the data.)**



—— Prediction paths for the DC coefficients

MPEG1: Motion-Compensated, Predictive (Non-Intra) Compression

Motion Estimation

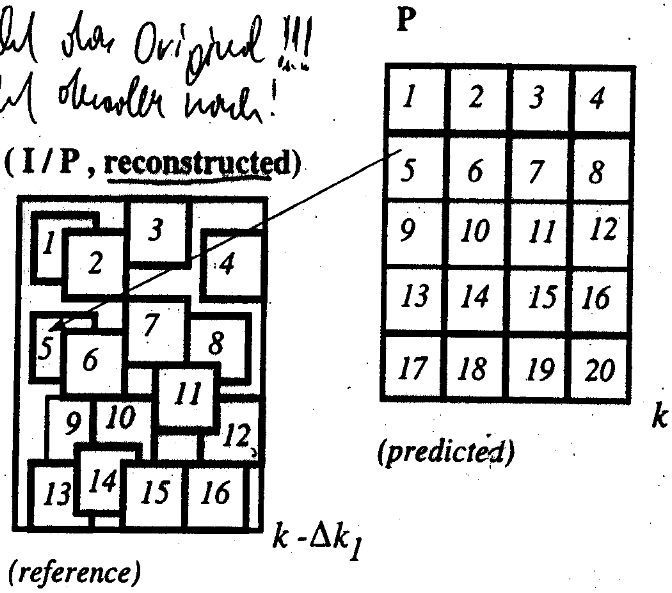
- Motion-compensated prediction is performed on the basis of macroblocks.
- Displacement vectors are assumed to be constant over a macroblock. Thus, a common displacement vector is estimated for and assigned to 16 × 16 luma and the two associated 8 × 8 chroma blocks. In case of bidirectional prediction, (i.e., in the case of **B** pictures) two vectors – one pointing to the past and the other pointing to the future – are estimated for each macroblock.
- Half (1/2) pixel accuracy is allowed for motion vectors.
- Redundancy among the displacement vectors of neighboring macroblocks is reduced by considering their consecutive differential values (i.e., applying DPCM to vector components).

Remark: Although, block matching seems to be a natural. choice for what is described above, MPEG1 does not specify (or standardize) the motion estimation algorithm. Any algorithm that is appropriate for the particular application can be used.

Motion-Compensated, Predictive Compression

Forward Prediction and P *Pictures*

nicht das Original !!!
nach decoder noch!



□ : **macroblock**

- For each current macroblock, at predicted picture k, a displacement vectors is estimated
- The displacement vector and the prediction error macroblock represent the information that is needed for reconstructing the current macroblock at the k th picture.
- DCT is applied to 8×8 blocks of the prediction-error macroblock, prior to quantization and coding.

MPEG1: Motion-Compensated, Predictive Compression

Definition of **P** *Pictures*

P Pictures are composed of macroblocks that are either

(i) forward predictive (non-intra) coded , or
(ii) intra coded (using the same quantization and VLC as macroblocks of the I pictures).

- The encoder is allowed to make a Intra/Non-Intra decision depending on the accuracy of the prediction.

A possible simple decision mechanism compares the variance of the of the original macroblock with that of the prediction error macroblock.

Motion-Compensated, Predictive Compression

Quantization and Coding of DCT of Prediction Error MB's

Steps that are different than those for the Intra Macroblocks:

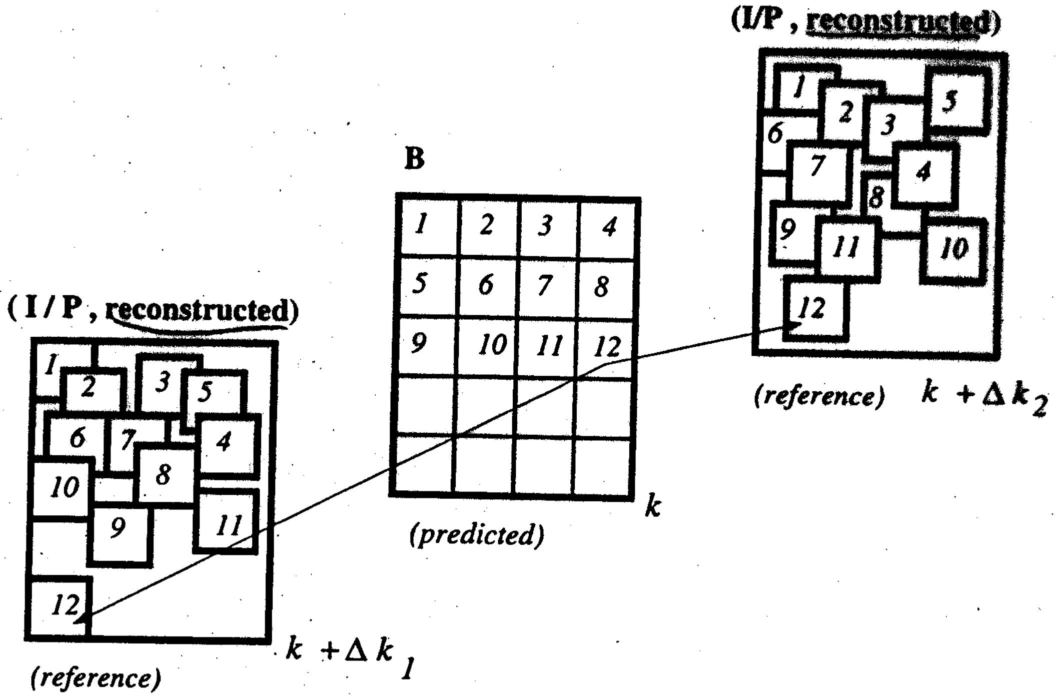
- The DC coefficient is quantized as the AC coefficients. The overall quantization is expected to have a dead-zone around zero. The default *quantization weight matrix* is the following:

$$\begin{bmatrix} 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \end{bmatrix}$$

- All quantized DCT coefficients, including the DC coefficient, are scanned to form *[run, level]* pairs which are then coded using a standard VLC table. (Single choice for the VLC table, unlike intra blocks..)

Motion-Compensated, Predictive Compression

Bidirectional Prediction and B *Pictures*



- For each current macroblock, at predicted picture k, two displacement vectors are estimated
  (one to a reference picture in the past and one to a reference picture in the future).

MPEG1: Motion-Compensated, Predictive Compression

Bidirectional Prediction (cont'd.) → Background occlusion ! Gut!!

- $pred = NINT\,[(\alpha_1)\,pred\_forward\; + \;(\alpha_2)\,pred\_backward]$

$$\begin{pmatrix} \alpha_1 = 0.5 \text{ and } \alpha_2 = 0.5 & \Longrightarrow & \text{Bidirectional Prediction} \\ \alpha_1 = 1 \text{ and } \alpha_2 = 0 & \Longrightarrow & \text{Forward Prediction} \\ \alpha_1 = 0 \text{ and } \alpha_2 = 1 & \Longrightarrow & \text{Backward Prediction} \end{pmatrix}$$

- The two displacement vectors and the prediction error macroblock represent the information that are needed for reconstructing the current macroblock.

- DCT is applied to 8 × 8 blocks of the prediction-error macroblock, prior to quantization and coding.

Definition of **B** *Pictures*

B Pictures are composed of macroblocks that are

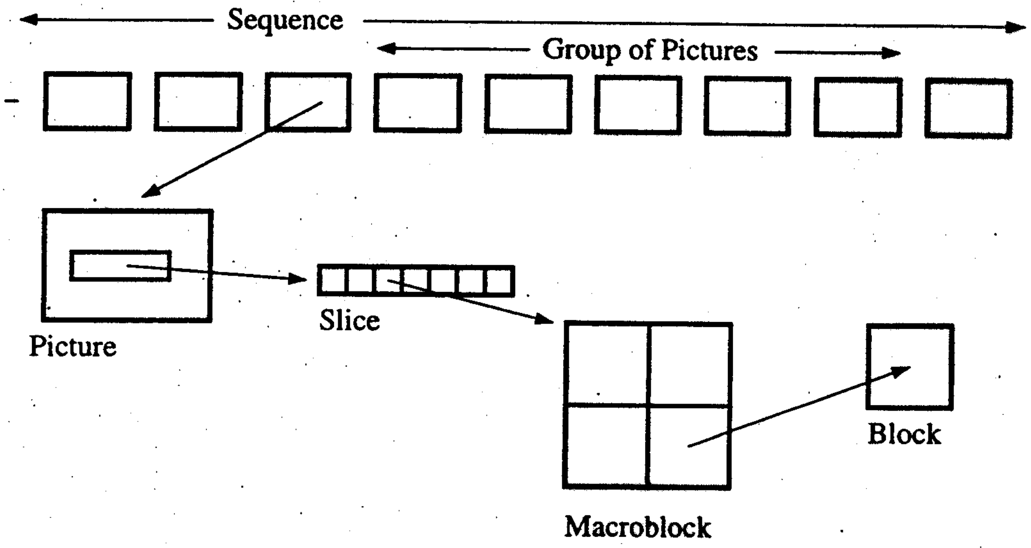
(i) bidirectional predictive coded, or
(ii) backward predictive coded, or
(iii) forward predictive coded, or
(iv) intra coded

- A possible decision mechanism is picking the mode that results in the least macroblock (luminance component) variance.

- The macroblocks in the B pictures are not used as references.

MPEG1 - Bit Stream Hierarchy

Main Hierarchy

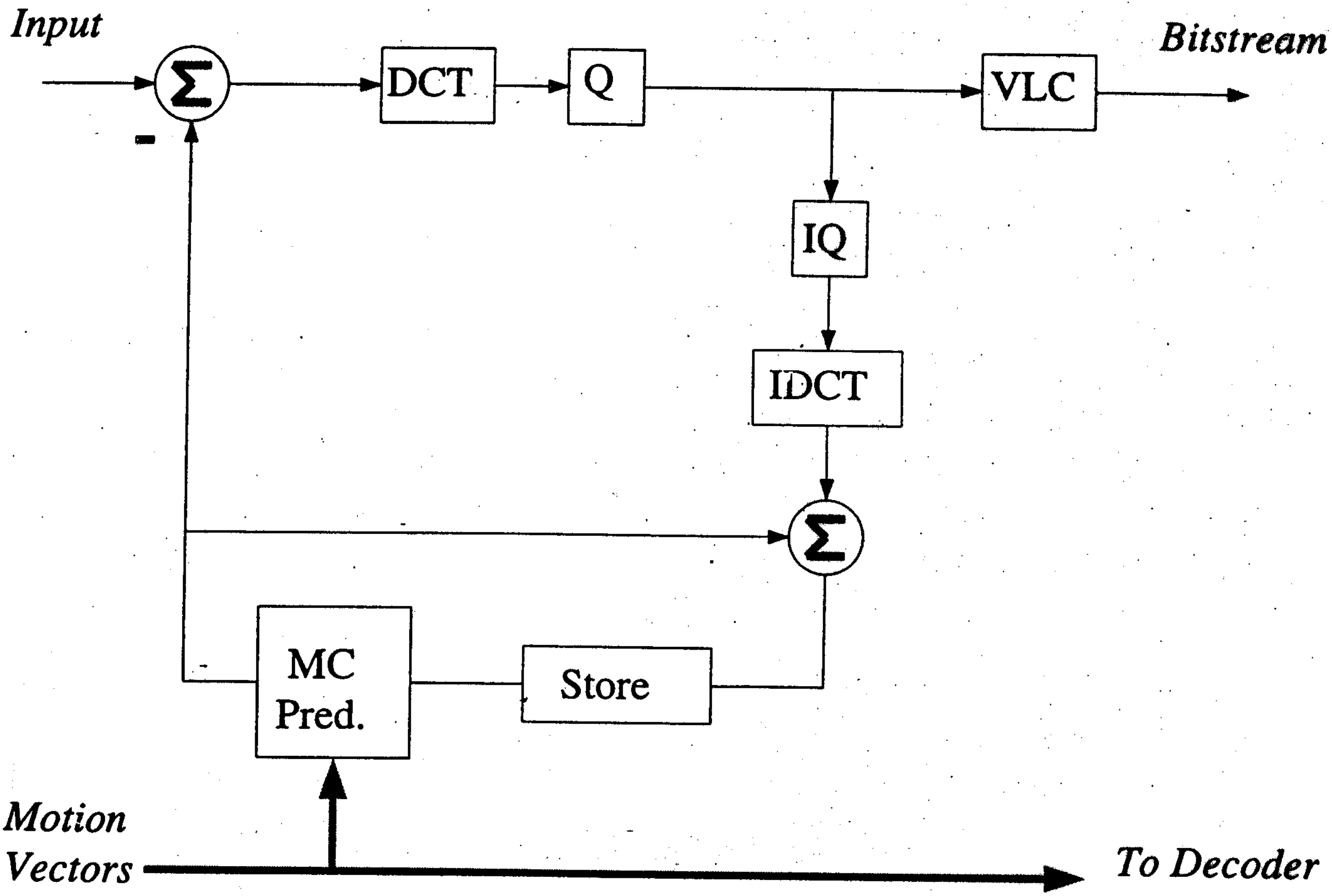


The Bit Stream

Sequence 1 Level	GOP Level	Picture Level	Slice Level	MB Level	Block Level	Sequence 2 Level	. . .

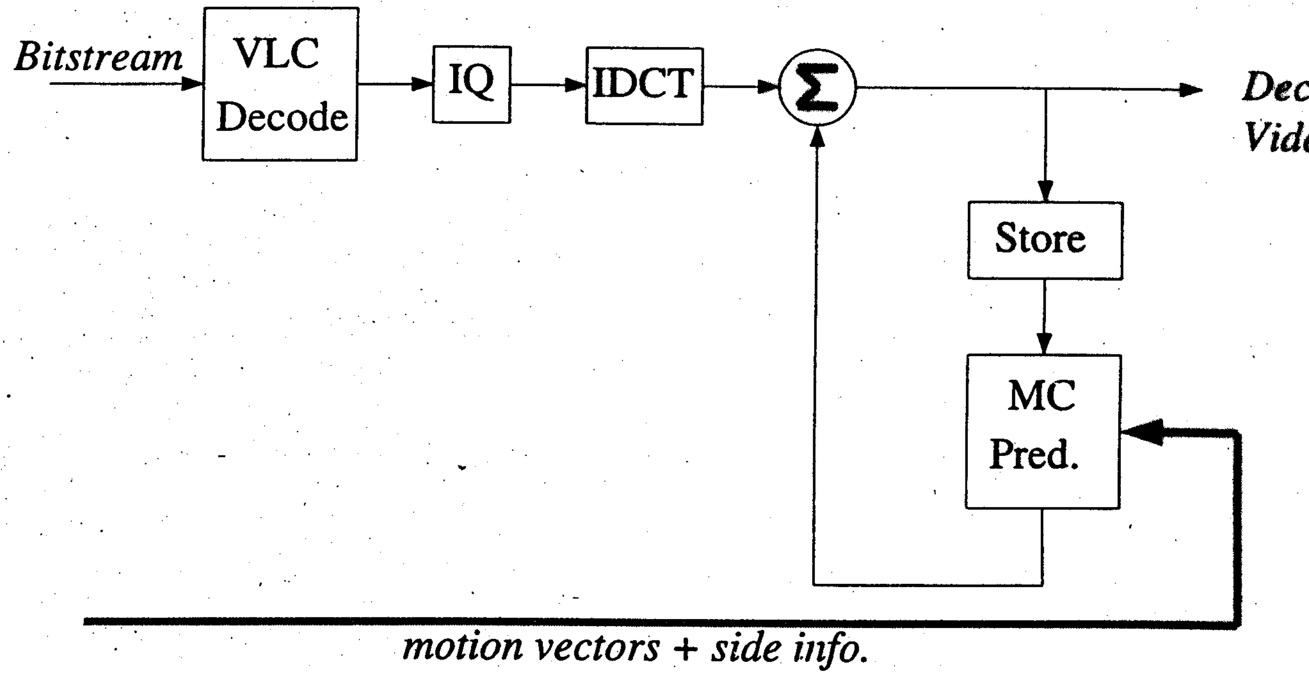
A Typical MPEG1 Encoder

- A typical MPEG encoder includes modules for motion estimation, motion-compensated prediction (predictors and framestores), quantization and dequantization, DCT and IDCT, variable length coding, a multiplexer, a buffer and a buffer regulator. A simplified diagram is given below.



MPEG1 Decoder

- The decoder basically reverses the operations of the encoder. A block diagram of a simplified generic decoder is shown below.
- The incoming bit stream (with a standard syntax) is demultiplexed into DCT coefficients and side information such as displacement vectors, quantization parameter, etc. In the case of B pictures, two reference frames are used to decode the frame.



The MPEG2 Standard

- MPEG2 is intended for higher bit rates than MPEG1.

- MPEG2 has a *Profile* and *Level* structure.

- The MPEG2 standard is published in 7 parts:

13818-1 Systems

13818-2 Video

13818-3 Audio

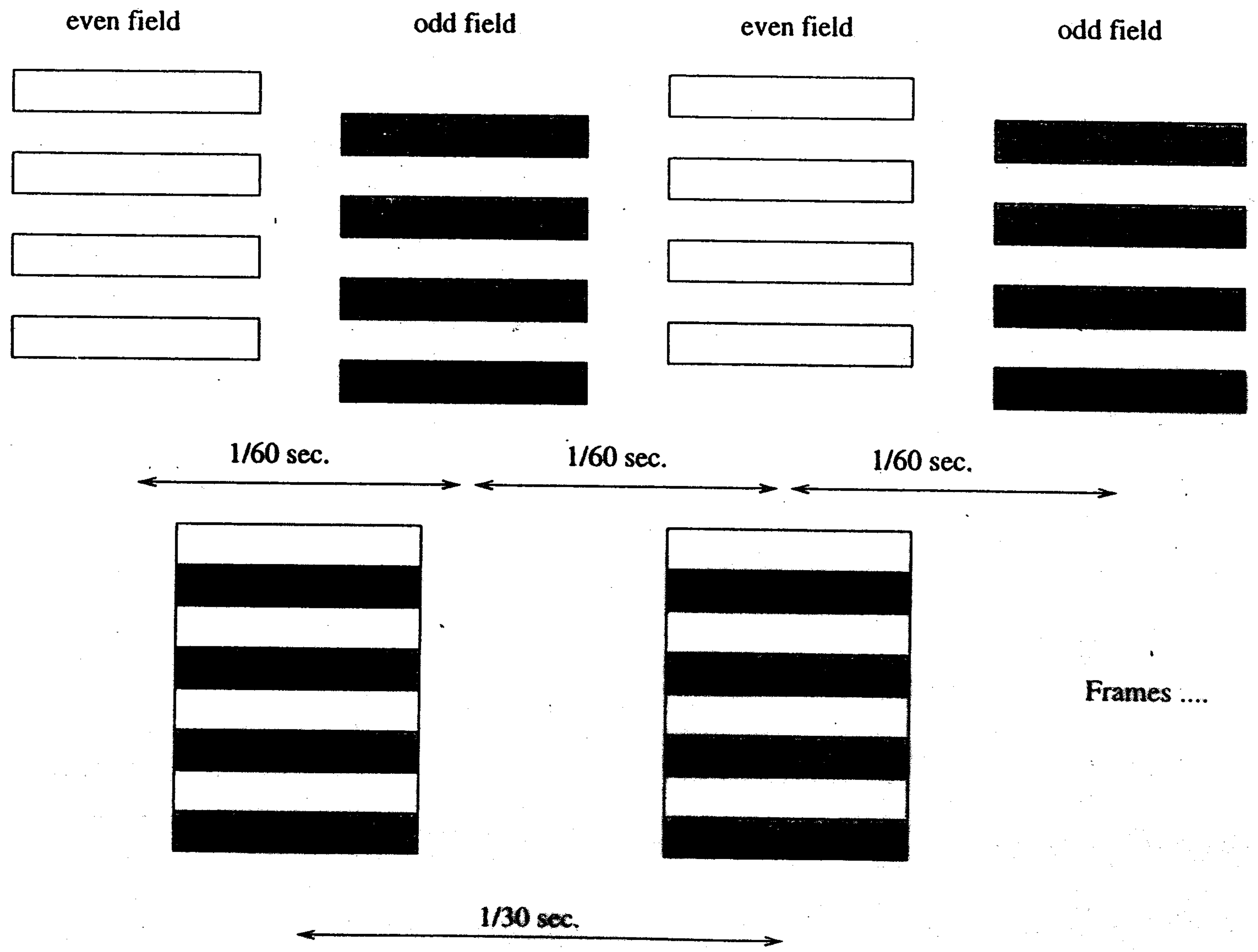
13818-4 Conformance

13818-5 Technical Report

13818-6 Digital Storage Media Command Control (DSM-CC) (Defines a set of protocols and interfaces to manipulate/playback MPEG bitstreams, e.g., in video-on-demand applications)

3818-7 Non-Backward Cor patible (w/ MPEG1) Audio compression. (Parts 6-7 are currently un^der development).

- MPEG2 allows for higher quality source material by supporting 4:2:2 (chroma channels subsampled in the horizontal dimension only) and 4:4:4 chroma format. MPEG2 allows both interlaced and progressive (non-interlaced) video format.

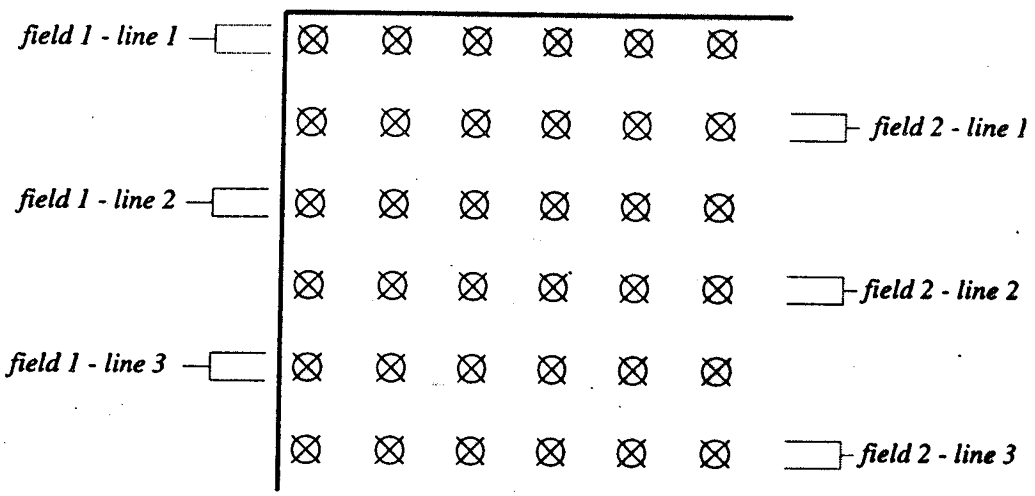


Field/Frame structure in an Interlaced Signal (60 fields/sec)

Chroma Sampling Formats

Video frames are represented as one luminance (Y) and two chrominance (C_r and C_b) channels.

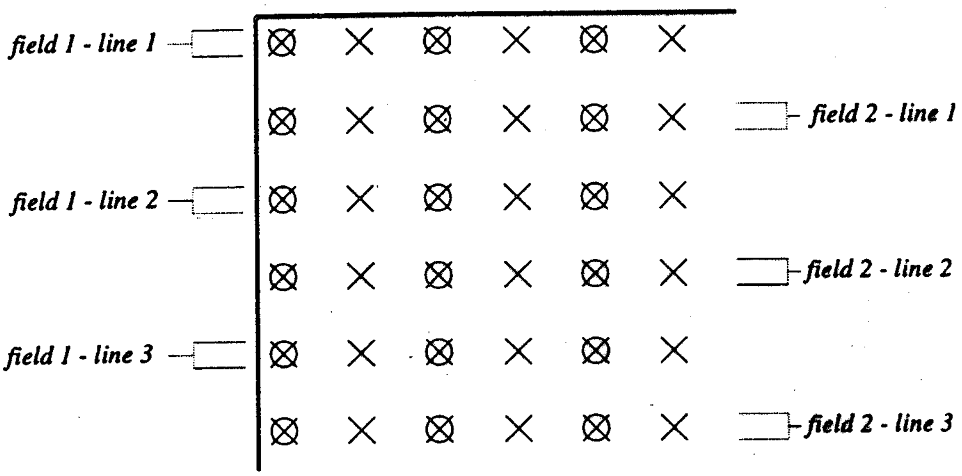
- 4:4:4



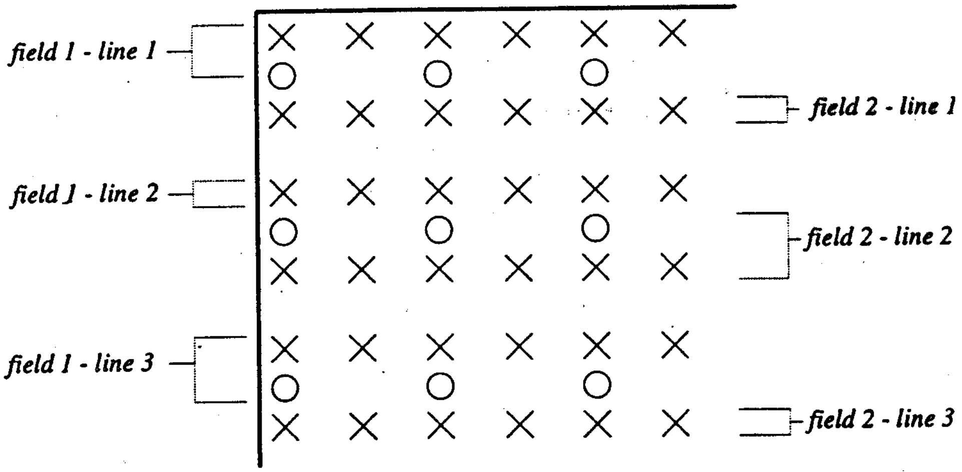
╳ : luma samples *(Y)* ○ : chrominance samples *(Cr and Cb)*

Chroma Sampling Formats (cont'd.)

- 4:2:2



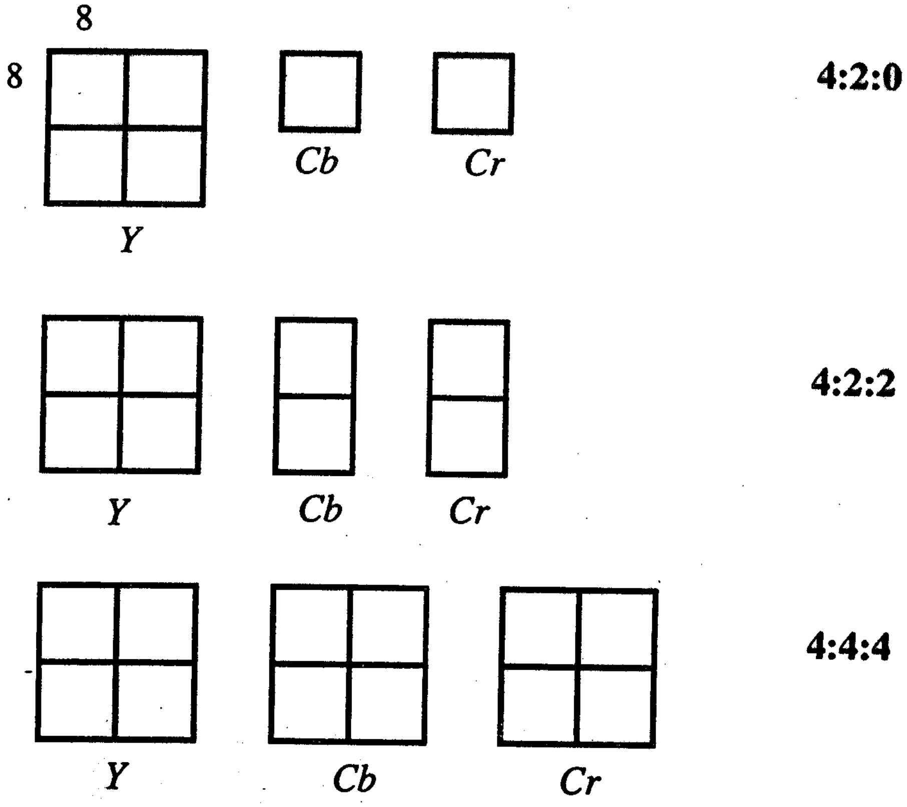
- 4:2:0



× : luma samples ○ : chroma samples.

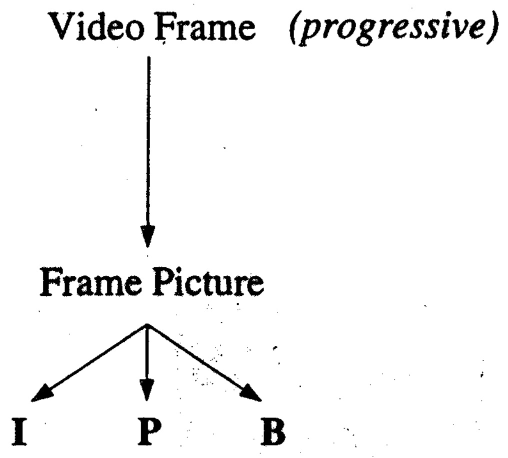
The Macroblock Structure

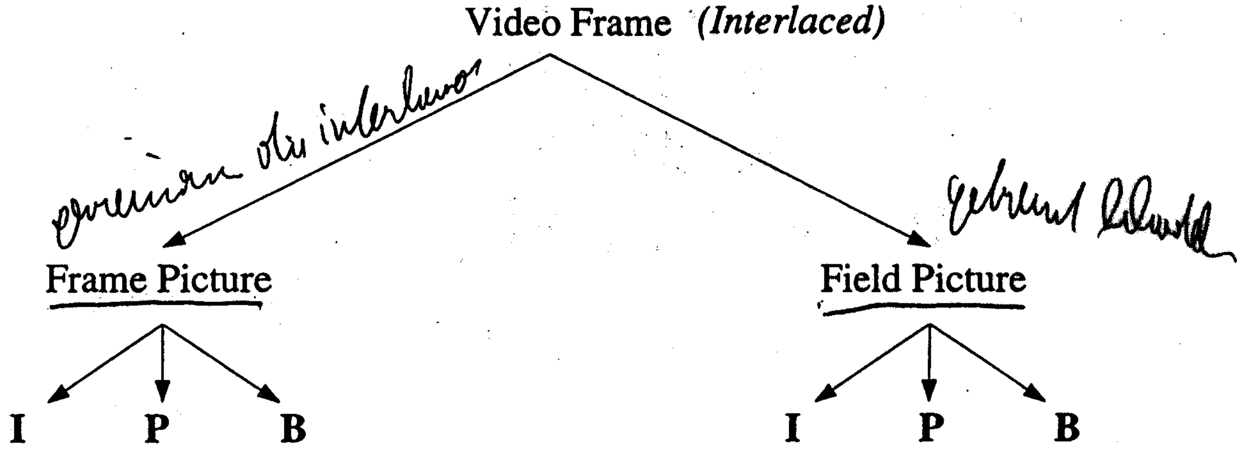
A macroblock contains a 16 × 16 block of luminance pixels and the spatially corresponding chroma samples.



New Picture Types

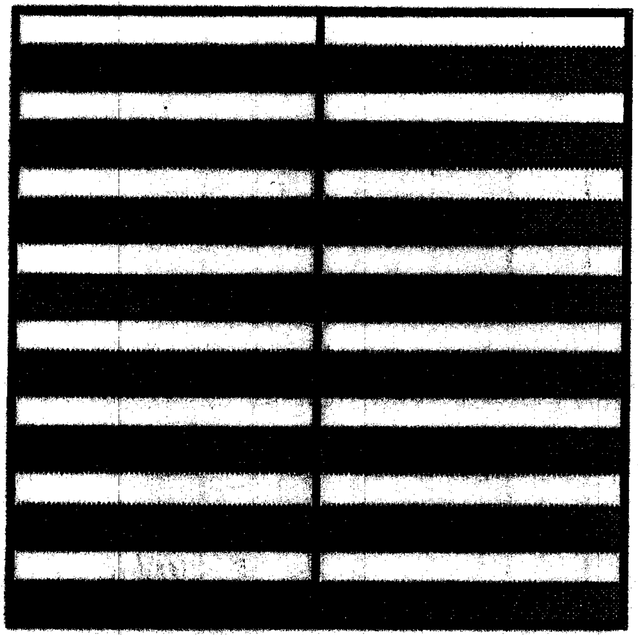
Two Types: Frame Pictures and Field Pictures





Frame Pictures

- In Frame Pictures even and odd fields are interleaved and the resulting frame is compressed. The luminance component of the macroblock is shown below.



Luma Component of the 16x16 MB

□: Even (top) field

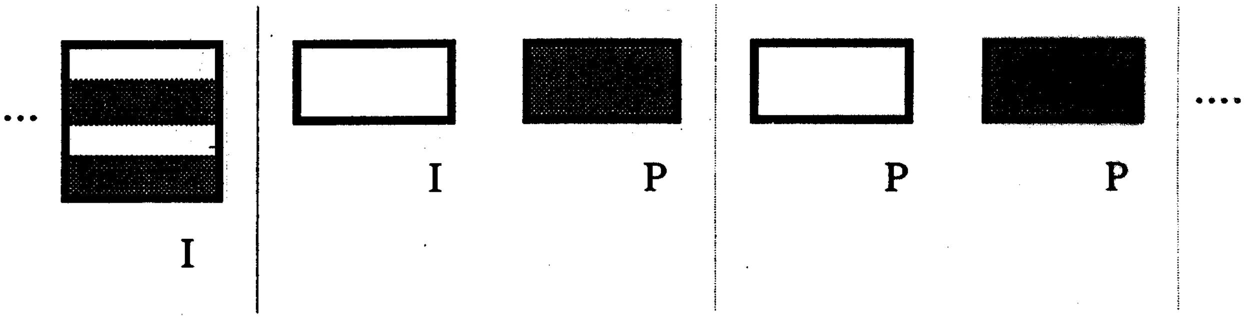
■: Odd (bottom) field

Field Pictures

- A Field Picture is either the even or the odd field belonging to a certain frame, i.e., the fields are treated as individual pictures. The macroblock in a field picture contains a 16 × 16 block of luminance data and the associated chroma data.

Field pictures occur in pairs and should be encoded such that they are transmitted in the order in which they are to be displayed.

  - If the first field of a certain frame is a P (B) picture, then the second field should also be a P (B) picture.
  - If the first field is an I picture, the second field should be either an I or a P picture.



An example of a sequence containing Frame and Field Pictures.

A sequence can contain an arbitrary mix of Field and Frame pictures.

The MPEG2 Standard

- MPEG2 allows for finer quantization of DCT coefficients. Also, it is allowed to specify (download) **separate** quantization matrices for luma and chroma data.

- MPEG2 allows for finer adjustment of quantization scale factor MQUANT, used in performing spatially adaptive quantization.

Finer quantization of DCT coefficients

Intra Macroblocks

- The quantization weight for the DC coefficient can be 8, 4, 2, and 1. That is 8, 9, 10, and 11 (in High Profile) bits resolution is allowed for the DC coefficient. (Recall that the weight is fixed to 8 in MPEG1.)

- The quantized AC coefficients can have a range of [-2048,2047]. (Recall that this range was limited to [-256,255] in MPEG1.)

- It is also allowed to download separate quantization matrices for luma and chroma data, when the color format is either 4:2:2 or 4:4:4.

Non-Intra Macroblocks

- The quantized coefficients can have a range of [-2048,2047]. (Recall that this range was limited to [-256,255] in MPEG1.)

The MPEG2 Standard

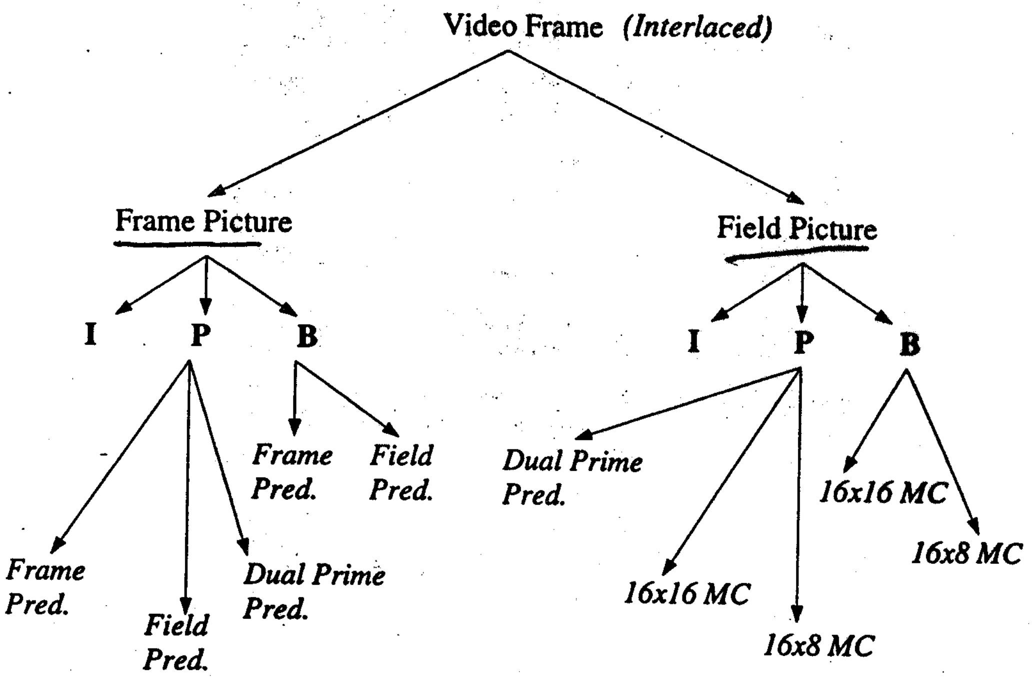
- Non-Intra (Motion-Compensated Predictive) Compression

  - Motion vector components are required to be estimated with half-pixel accuracy

  - Prediction Modes:

    – Progressive video: *Frame Prediction* is performed.

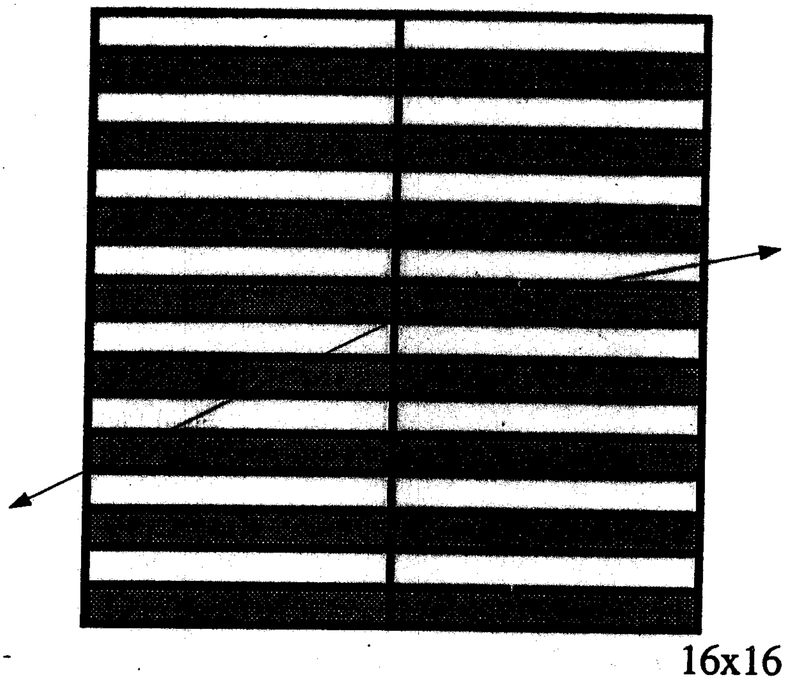
    – Interlaced video *New Modes*:



Motion-Compensated, Predictive Compression

Frame-Based prediction in Frame Pictures

One motion vector is estimated in each direction (2 directions in the case of bidirectional prediction) per macroblock of the predicted picture, corresponding to 16 × 16 pixels luma area and its associated chroma.



Motion Vectors for bidirectional *Frame Prediction* for interlaced data. (16x16 Luma component of the MB is shown.)

At each direction (backward or forward), prediction is based on the most recently reconstructed frame –– the reference frame. The fields of the reference frame may have been compressed individually as *two field pictures*, or as *a single frame picture*.

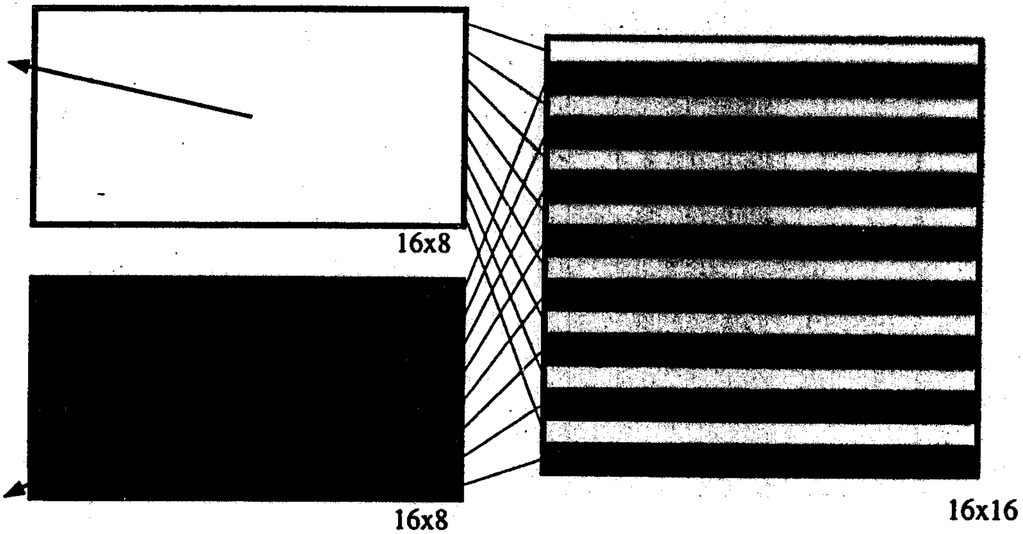
Motion-Compensated, Predictive Compression

Field-Based prediction in Frame Pictures

Two motion vectors per macroblock of the predicted picture are estimated in each direction, one for *each* one of the fields; each vector corresponds to 16 × 8 pixels luminance and its associated chroma.

At each direction (backward or forward), prediction is based on the most recently reconstructed frame —— the reference frame. The fields of the reference frame may have been compressed individually as *two field pictures*, or as *a single frame picture*.

A bit is reserved to specify which field (top or bottom) of the reference frame is used for each 16 × 8 field-block.

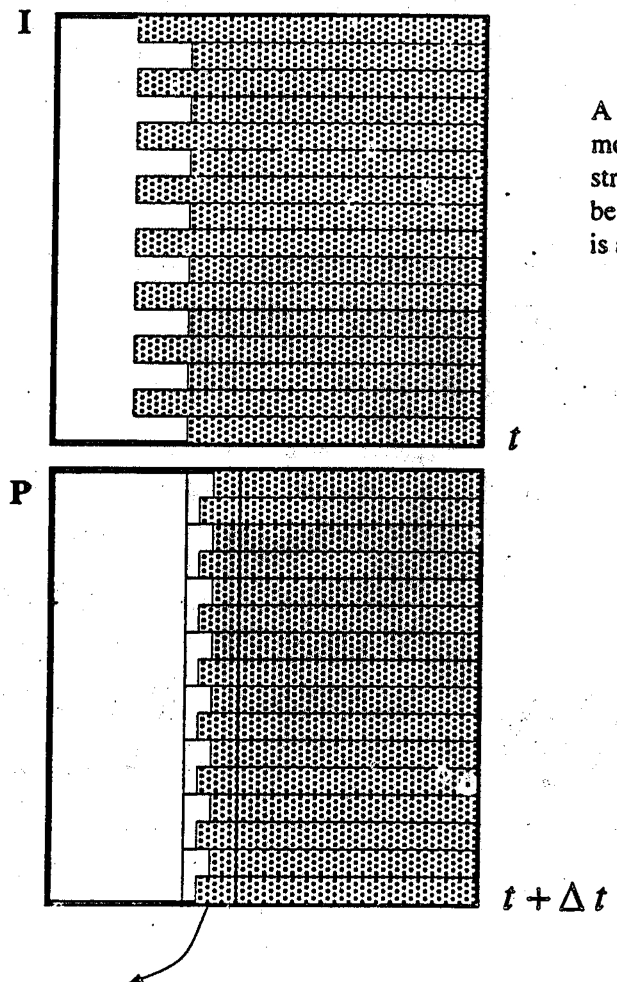


Motion vectors for forward *Field Prediction* .
(Luma components are shown.)

Motion-compensated prediction modes

Switching prediction modes in Frame Pictures

- It is possible to switch among prediction modes, on a macroblock basis, within the same frame picture. For instance, the encoder may prefer the field prediction mode over frame prediction in areas with sudden and irregular motion.



A fast rectangular structure is moving right horizontally. The structure changes its direction before the odd field at $t + \Delta t$ is acquired.

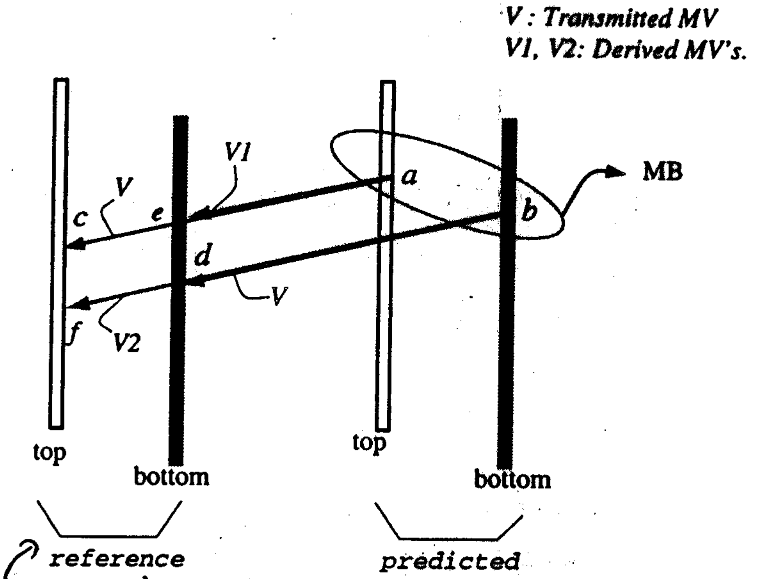
This column of MB's can be better predicted when Field Prediction is utilized.

Motion-Compensated, Predictive Compression

Dual-Prime (forward) prediction in **P** *Frame Pictures*

- Dual prime is a special kind of *forward* field prediction in P frame pictures. A single forward motion vector is estimated for each macroblock of the predicted frame picture. This motion vector (denoted as *V* in the illustration below) points at the most recently reconstructed frame–the reference frame. Using this vector, each field in the macroblock is associated with a field of same parity (top or bottom) in the reference frame.

The (simplified) main idea of dual prime prediction is illustrated below:



? ≠ 0!!

Motion vectors *V1* and *V2*, pointing at fields of opposite parity, are derived from *V* on the basis of linear motion trajectory (no acceleration) assumption, via appropriate scaling determined by the temporal distance between the reference and the predicted frames. Predictions of *a* and *b*, that constitute the prediction macroblock, are formed as

$$\hat{a} = NINT[1/2c + 1/2e] \qquad \hat{b} = NINT[1/2d + 1/2f].$$

Note that this mode is used in **P** pictures where there are no **B** pictures between the predicted and the reference pictures.

warum: ähnlich wie B-Frames nur mit weniger Delay

Motion-Compensated, Predictive Compression

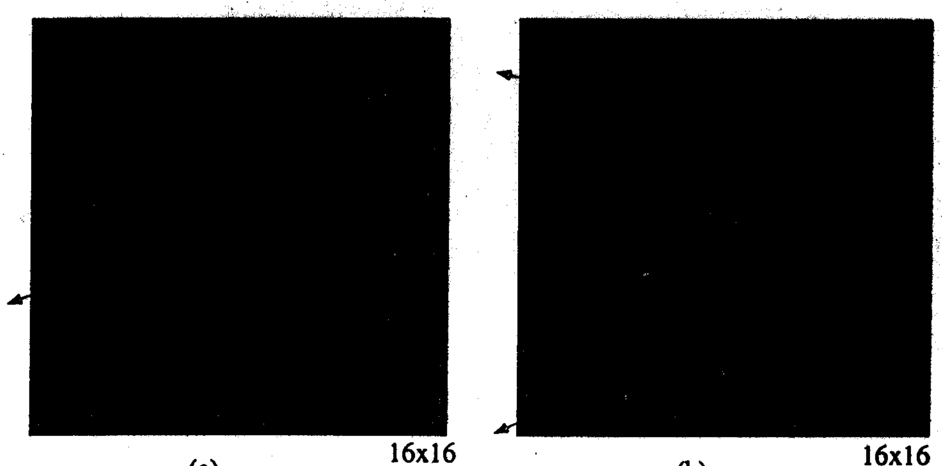
Field-Based prediction modes in Field Pictures

(a) One motion vector is estimated in each direction (2 directions for bidirectional prediction) per macroblock of the predicted picture, corresponding to 16 × 16 pixels luminance area and its associated chroma; OR

(b) Two motion vectors per macroblock are estimated in each direction, each vector corresponding to 16 × 8 pixels luminance area and its associated chroma.

Most recently reconstructed field pictures, or fields of most recently reconstructed frame pictures are used as references. A bit is reserved to specify which field (top or bottom) of the reference frame is used in prediction.

Forward Field-Based Prediction



(a) 1 MV per 16x16 Luma block.

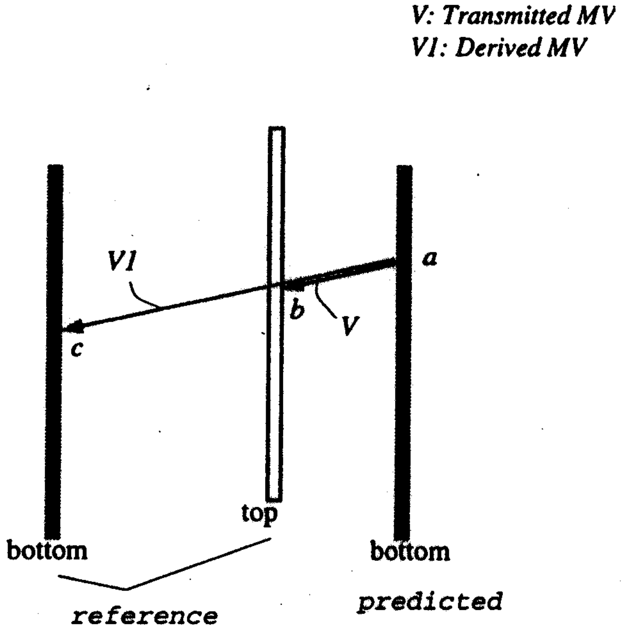
(b) 1 MV per 16x8 Luma block.

Motion-Compensated, Predictive Compression

Dual-Prime prediction in **P** *Field Pictures*

Data from two reference fields of opposite polarity is averaged to predict a field picture. A single forward motion vector is estimated for each macroblock within the predicted field picture. This motion vector (denoted as V below) points at the most recently reconstructed field –– the reference field. The other motion vector vector pointing at the other reference field of opposite polarity is derived from V.

The (simplified) main idea of dual prime prediction mode is illustrated below:



The motion vector *V1* (pointing at a field that is of opposite polarity) is derived from V on the basis of linear motion trajectory (no acceleration) assumption, via appropriate scaling determined by the temporal distance between the reference and the predicted fields. Prediction of macroblock a is formed as

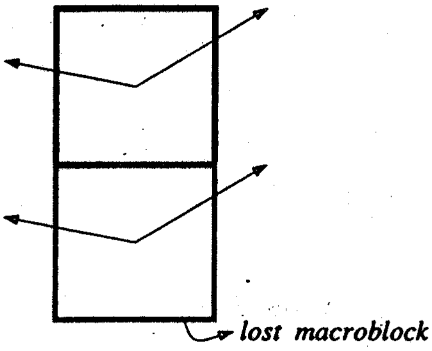
$$\hat{a} = NINT[1/2b + 1/2c].$$

Note that this mode is used in **P** pictures where there are no **B** pictures between the predicted and the reference fields.

Error Concealment

Concealment of lost Macroblocks

Within P and B pictures, the lost macroblock is substituted with the motion compensated macroblock of the previously decoded picture. In motion compensation, the motion vectors belonging to the macroblock above the lost one are used.



- If the top macroblock used in concealment is intra coded, then there are two possibilities:

1. If a (forward) motion vector for the intra block pointing at the previously decoded picture is included in the bitstream then the concealment is performed as above. *(The standard allows for motion vectors for intra blocks for concealment purpose.)*
2. Otherwise, the concealment is performed using the macroblock in the previously decoded picture that is colocated with the lost macroblock, i.e., zero-displacement is assumed. (This strategy applies to the first line of blocks as well.)

Error concealment within I pictures is performed in this same manner.

Scalability via Hierarchical Compression

Main Idea

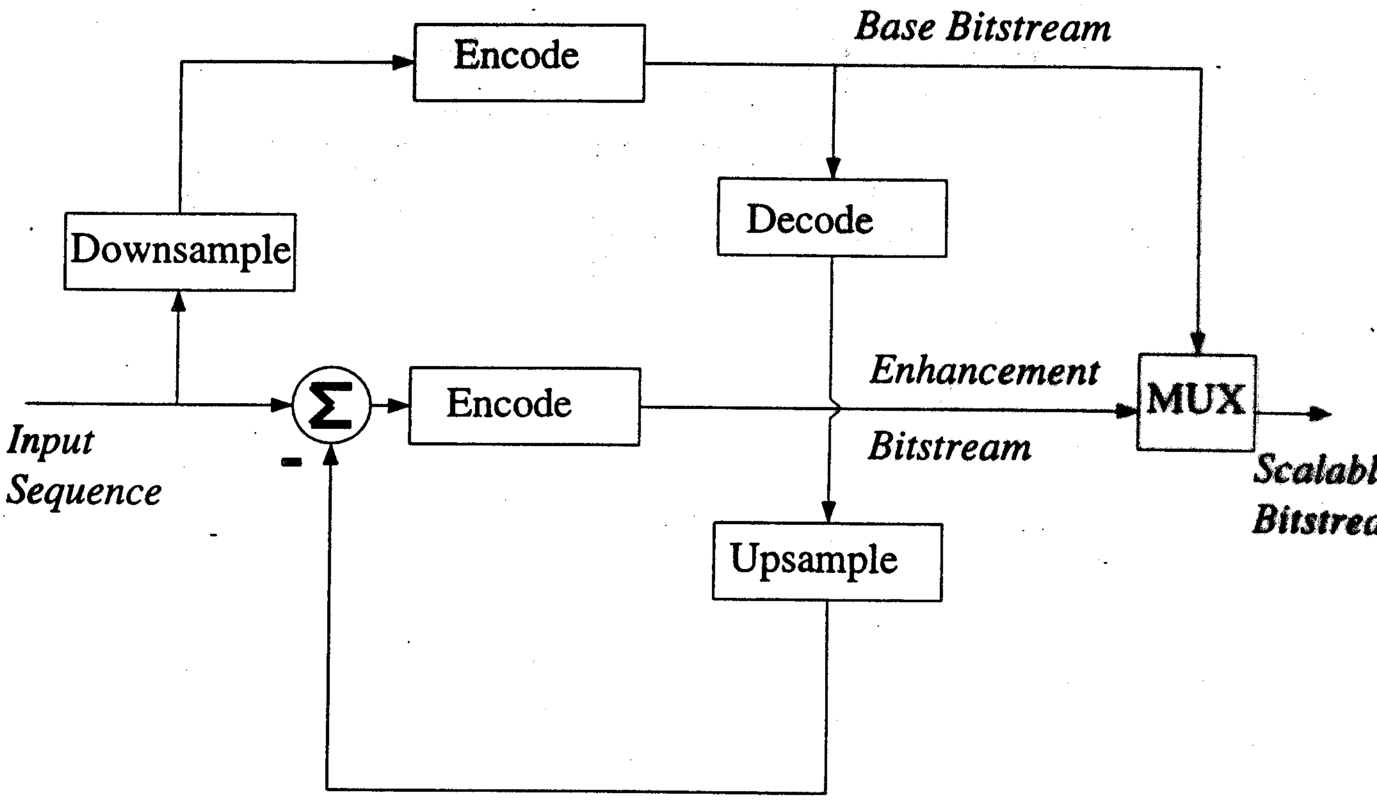
A scalable decoder is capable of decoding only a part of a scalable bitstream and generate a "useful" video. For instance, a lower resolution (amplitude/spatial/temporal) video can be decoded from the partial bitstream, and the remaining bits are used in refining it to a higher resolution, when desirable. In other words, a lower resolution video can be obtained without decoding an entire bitstream first and then lowering the resolution by postprocessing, or without simulcasting two distinct bitstreams corresponding to different resolutions.

- Spatial scalability (e.g.,pixel resolution and picture format (progressive/interlaced) scalability)
- SNR Scalability
  (i.e., different levels of picture quality)
- Temporal Scalability
  (e.g., different frame rates)

- Scalability/Compatibility in MPEG-2 is achieved through *layered* (***hierarchical***) compression.

Example: Spatial Scalability

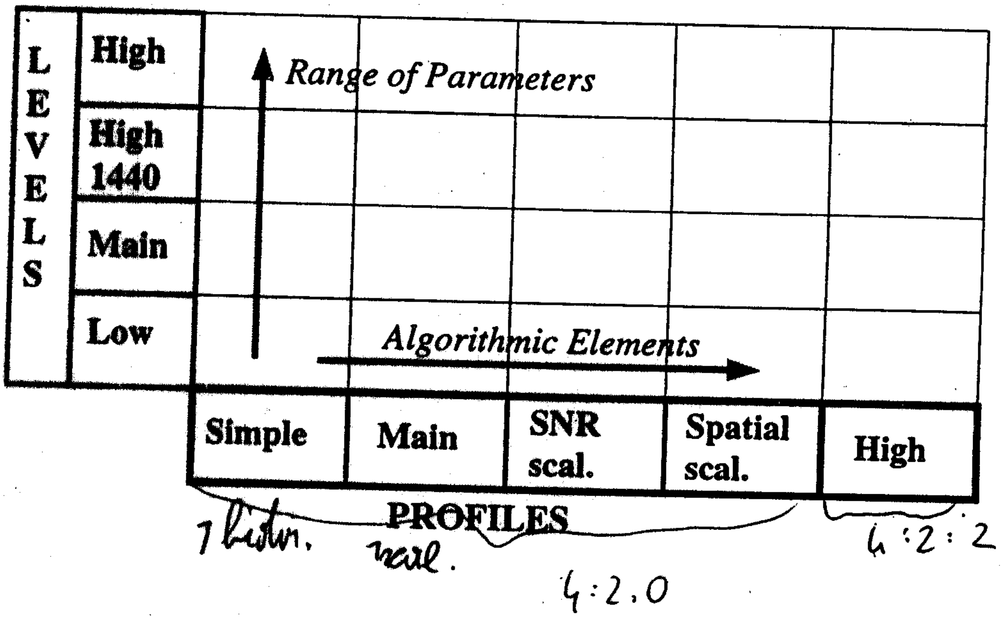
Spatial Scalability is aimed at providing pixel-resolution scalability. The base layer sequence is a lower-spatial resolution sequence obtained by downsampling the input. Scalable bitstream contains both the base layer and the residual enhancement bitstream. A simplified diagram of a spatially scalable encoder is given below.



The MPEG2 Standard

- An MPEG2 decoder complies with a predefined *Profile* and *Level*.

--- ***Profiles and Levels in MPEG2***



♠ *New Profiles:*

- **4:2:2 Profile:** Adds 4:2:2 chroma sampling to Main Profile, at bit rates up to 50 Mbs. IBM chip.
- **Multiview Profile:** Compression of stereoscopic sequences by taking inter-view correlations into account. (Based on Temporal Scalability algorithm of MPEG2).

Notes:

- The upper bounds given in the above table are required to be simultaneously satisfied.

- SDTV (standard definition digital TV) compression adheres to **Main Profile at Main Level**

Lines	Pels	Temporal Rate
480	704	60I, 60P, 30P, 24P
480	640	60I, 60P, 30P, 24P

- The video compression module of the **Grand Alliance (GA) HDTV system** adheres to **Main Profile at High Level.**

Lines	Pels	Temporal Rate
720	1280	60P, 30P, 24P
1080	1920	60I, 30P, 24P

- DVD compression adheres to **Main Profile at Main Level**

- All MPEG2 decoders are required to decode (constrained-parameter) MPEG1 bitstreams.

Implementations of MPEG Standards : Some Examples

Silicon

- C-Cubed Microsystems
  - MPEG1, MPEG2 (MP@ML) Video encoders and decoders (www.c-cube.com)
- IBM
  - MPEG2 encoder chip: I-picture only encoder supporting 4:2:2 and 11-bit DC precision at about 40 Mbps.
  - MPEG2 encoder chip set: IBP with field/frame adaptive prediction and dual prime prediction.
  - MPEG2 I frame only and IBP decoders; Single chip MPEG2 video/audio decoder.
    (www.chipsibm.com)
- Others....

Main Principles of The H.261 Standard

- Primarily intended for videophone and videoconferencing over ISDN channels at rates $p \times 64$ kbits/sec for video and audio combined, e.g., 16 kbits/sec audio and 112 kbits/sec video ($p = 2$) over the *the Basic Rate Channel.*

- The H.261 standard is very similar to the MPEG1 video compression standard. It employs I and P pictures, but **not** B pictures. An H. 261 encoder may decide to skip frames and/or reduce spatial resolution to maintain picture quality at a given bitrate. (This info is transmitted to the decoder).

- Picture Format:

Picture Format	Pixels	Lines
QCIF (mandatory)	176 (lum.) 88 (chr.)	144 (lum.) 72 (chr.)
CIF (optional)	352 (lum.) 176 (chr.)	288 (lum.) 144 (chr.)

The above picture formats are for 25 Hz temporal rate. For 30 Hz, the CIF format is 352 × 240 and the above numbers should be modified accordingly. The chroma sampling is 4:2:0.

For details of the H.26' Standard, see

A. Netravali and B. Haskell, *Digital Pictures, 2nd Edition*, Plenum Press, NY, 1995.

CCITT Recommendation H.261:"Video Codec for Audio Visual Services at $p \times 64$ kbits/sec," 1990. *Recent update on ISDN:* June 1995 Issue of IEEE Spectrum Magazine.

Improvements in The H.263 Standard

- Increased compressior efficiency to support 28.8 Kbps over PSTNs.

- Additional Picture Formats (mandatory formats are s-QCIF and QCIF):

Picture Format	Pixels	Lines
sub-QCIF	128 (lum.) 64 (chr.)	96 (lum.) 48 (chr.)
QCIF	176 (lum.) 88 (chr.)	144 (lum.) 72 (chr.)
CIF	352 (lum.) 176 (chr.)	288 (lum.) 144 (chr.)
4CIF	704 (lum.) 352 (chr.)	576 (lum.) 288 (chr.)
16CIF	1408 (lum.) 704 (chr.)	1152 (lum.) 576 (chr.)

- Motion vectors can be estimated with 1/2 pel accuracy.

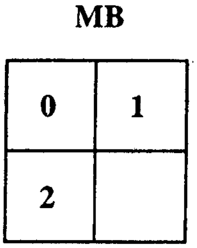
- Better predictive coding of motion vectors (a median of 3 vectors belonging to 3 neighboring macroblocks is used as a predictor:5-10savings in motion vector data).

- 3-D symbols and 3-D VLC tables: (run,level,LAST=0,1), where last=0 means "there are no more nonzero coefficients" during the zig-zag scan, i.e., no EOB code is used. A different VLC table is used for the last nonzero coefficient, resulting in about 5efficiency.

H. 263 *(Cont'd.)*

- Optional arithmetic coding (Removing the "integral number of bits per symbol" requirement of Huffman coding.) Experiments indicate apossible 10cost of higher complexity.

- Optional advanced prediction mode and overlapped motion compensation: Associating MVs with 8 × 8 blocks, and defining the prediction of an 8×8 block as a weighted average of motion-compensated predictions obtained by its own MV, and the MVs of the two neighboring 8×8 blocks. The result of OBMC is reduced blocking artifacts.



-The prediction of the luma component of B0 is a weighted sum of predictions obtained using the motion vectors for blocks B0, B1, and B2.

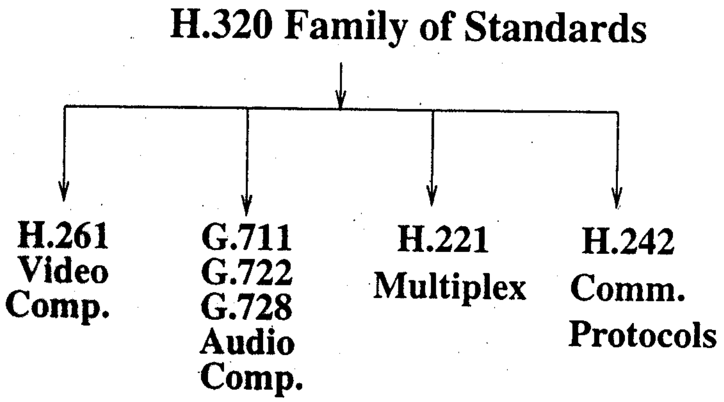
The associated chroma block is predicted using an average of vectors belonging to 4 blocks.

- A form of bidirectional prediction is allowed as an option. (See the ITU-T document for details of PB frames.)

For details of the H.263 Standard, see

ITU-T DRAFT H.263, Video Coding for Low Bitrate Communication, July 1995.
(ftp://ftp.std.com/vendors/PictureTel)

More on H.261 and H.263



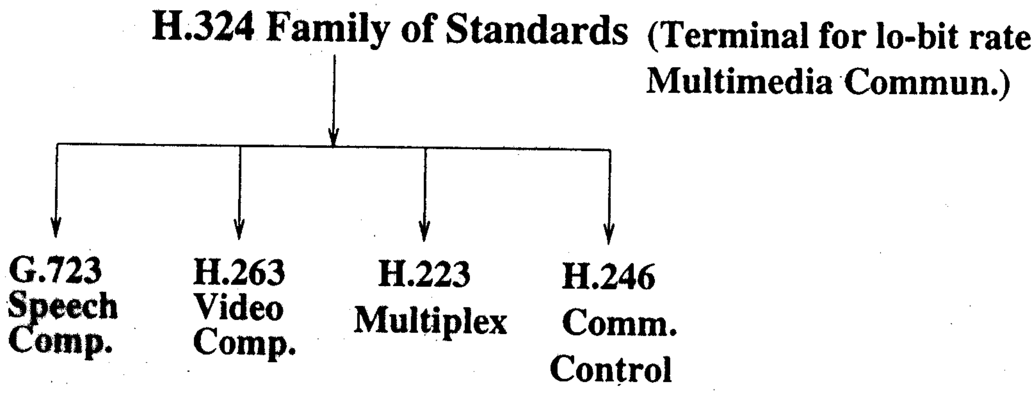
H.320 Family of Standards

H.261
Video
Comp.

G.711
G.722
G.728
Audio
Comp.

H.221
Multiplex

H.242
Comm.
Protocols



H.324 Family of Standards (Terminal for lo-bit rate Multimedia Commun.)

G.723
Speech
Comp.

H.263
Video
Comp.

H.223
Multiplex

H.246
Comm.
Control

It is mandatory for an H.324 terminal to feature both H.261 and H.263 Codecs.

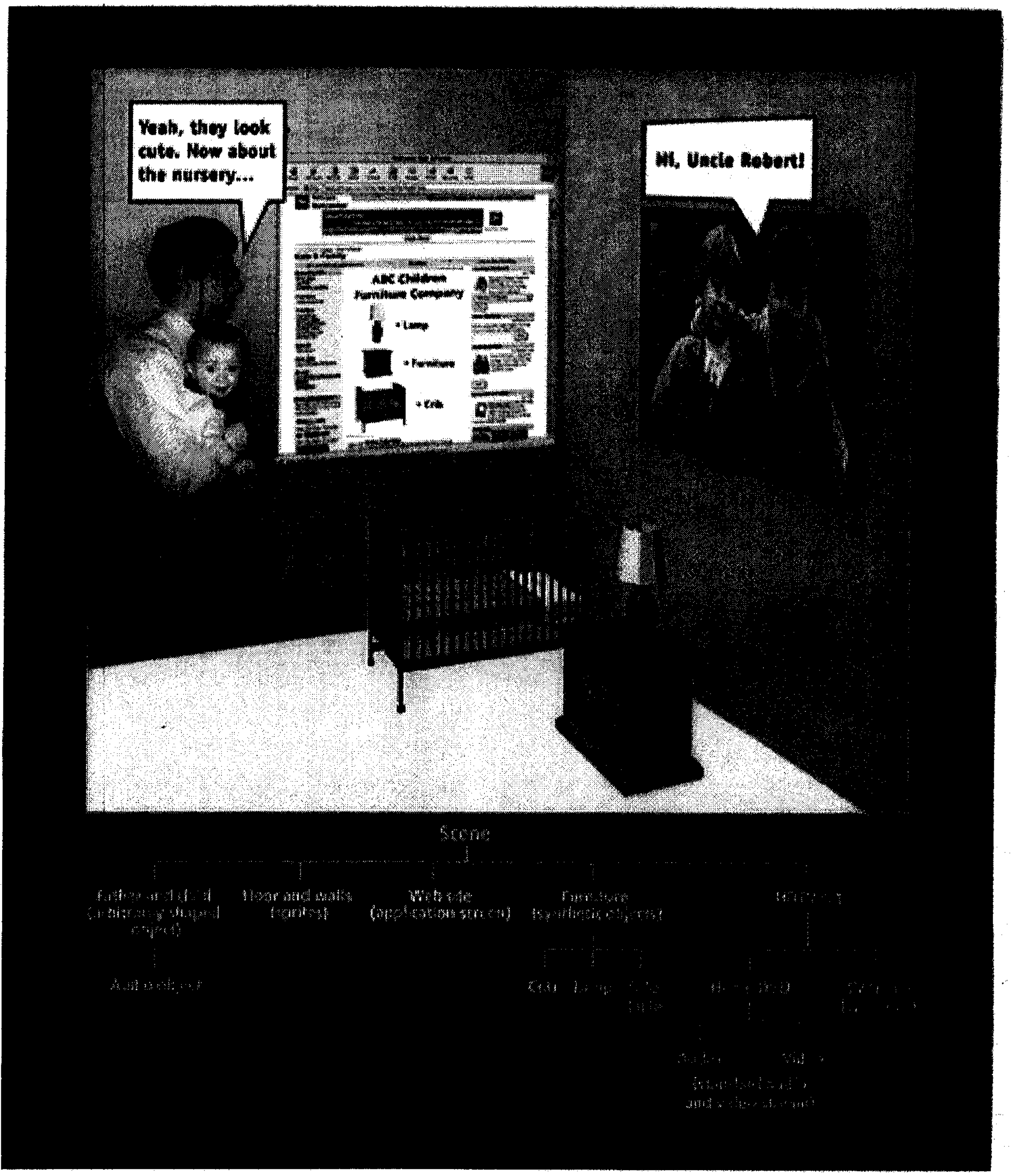
Enhancements to H.263: H.263+

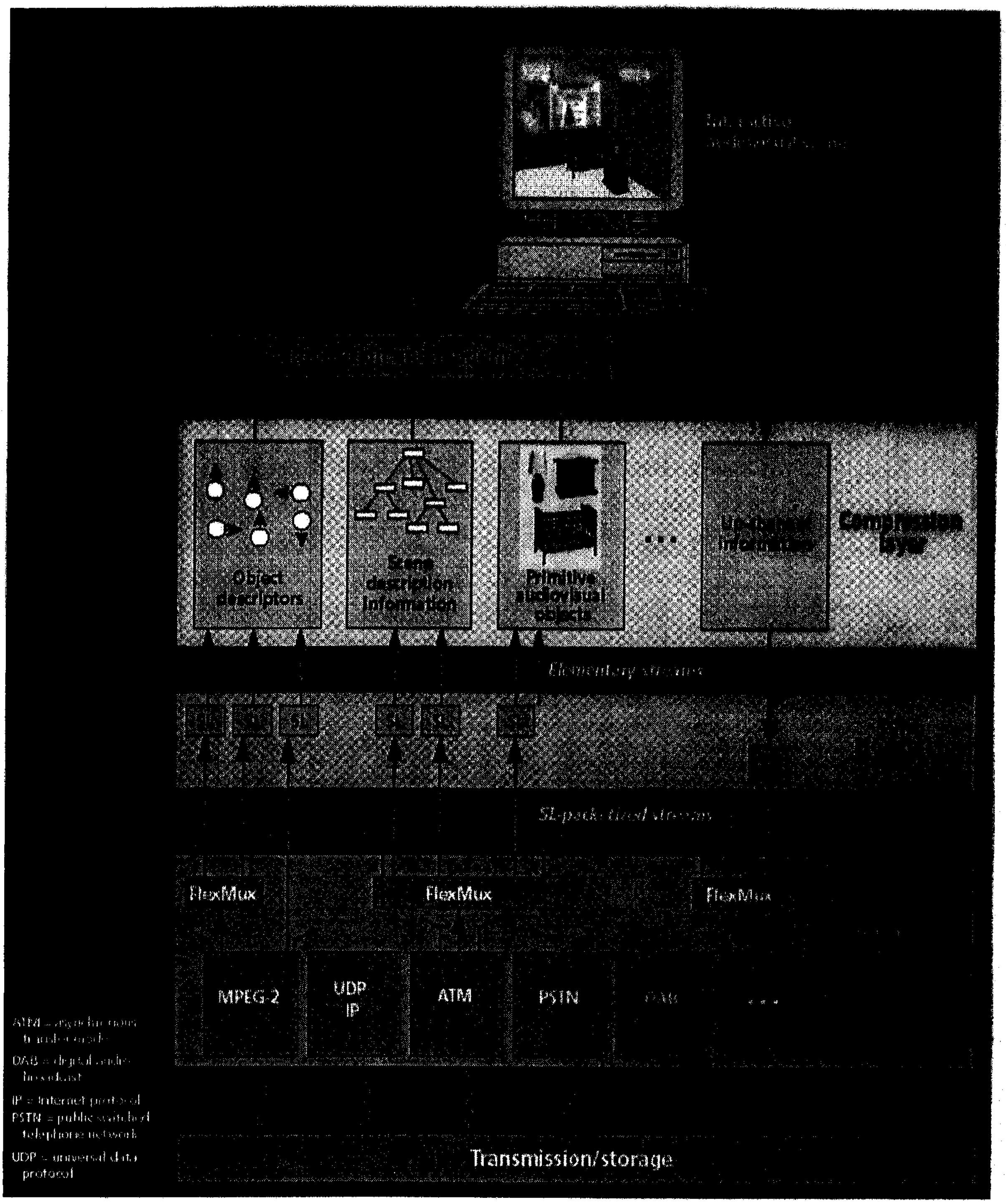
- Additional increase in compression efficiency as well as additional functionalities.
- Arbitrary frame sizes from 4 × 4 to 2048 × 1152, in 4 pixels increments.
- Square pixels
- An additional feature that has provided increased compression efficiency is advanced intra coding, where intra blocks are predicted from neighboring blocks. (There are other additional features).
- For documents and more info:
  ftp://ftp.std.com/vendors/PictureTel/h324/h263plus
- Final draft to be completed in February 199~~6~~7.

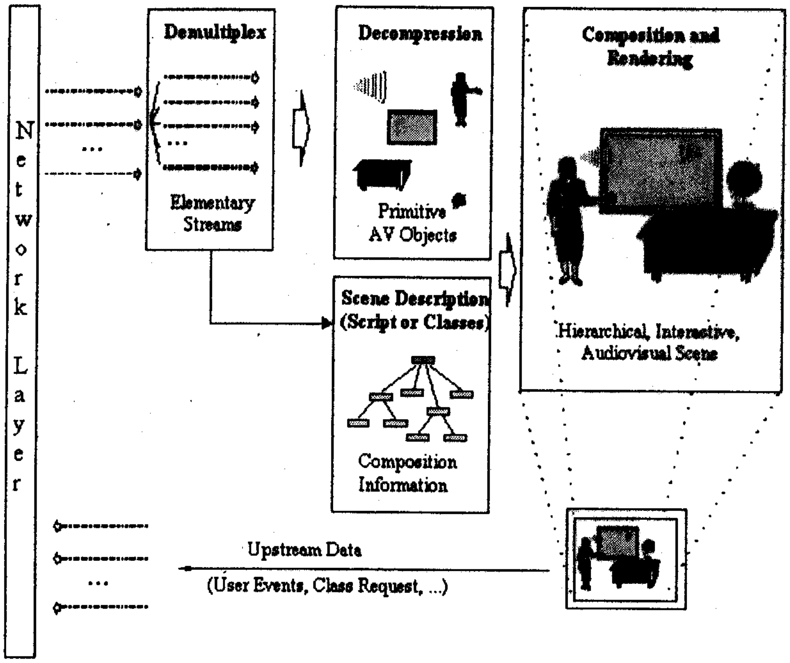
Functionality	MPEG-4 Video-Requirements
Content-Based Interactivity	
Content-Based Manipulation and Bitstream Editing	Support for content-based manipulation and bitstream editing without the need for transcoding.
Hybrid Natural and Synthetic Data Coding	Support for combining synthetic scenes or objects with natural scenes or objects. The ability for compositing synthetic data with ordinary video, allowing for interactivity.
Improved Temporal Random Access	Provisions for efficient methods to randomly access, within a limited time and with fine resolution, parts, e.g. video frames or arbitrarily shaped image content from a video sequence. This includes 'conventional' random access at very low bit rates.
Compression	
Improved Coding Efficiency	MPEG-4 Video shall provide subjectively better visual quality at comparable bit rates compared to existing or emerging standards.
Coding of Multiple Concurrent Data Streams	Provisions to code multiple views of a scene efficiently. For stereoscopic video applications, MPEG-4 shall allow the ability to exploit redundancy in multiple viewing points of the same scene, permitting joint coding solutions that allow compatibility with normal video as well as the ones without compatibility constraints.
Universal Access	
Robustness in Error-Prone Environments	Provisions for error robustness capabilities to allow access to applications over a variety of wireless and wired networks and storage media. Sufficient error robustness shall be provided for low bit rate applications under severe error conditions (e.g. long error bursts).
Content-Based Scalability	MPEG-4 shall provide the ability to achieve scalability with fine granularity in content, quality (e.g. spatial and temporal resolution), and complexity. In MPEG-4, these scalabilities are especially intended to result in content-based scaling of visual information.

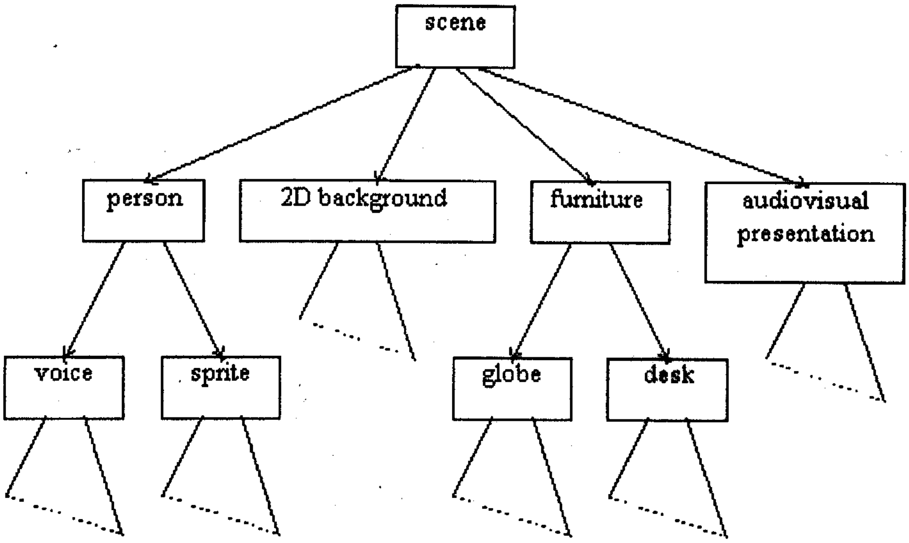
TABLE III
CORE EXPERIMENTS

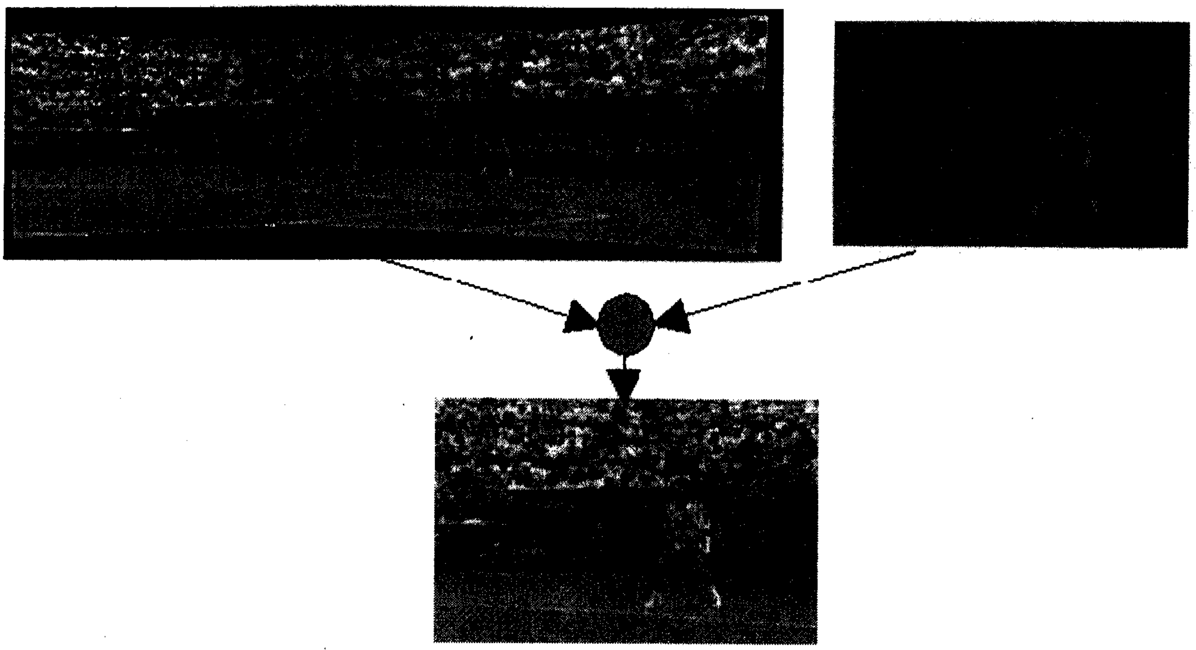
Subject	Techniques compared in Core Experiments
Motion Prediction	Global motion compensation, Block partitioning, Short-term/long-term frame memory, Variable block size motion compensation, 2D Triangular mesh prediction, Sub-pel prediction.
Frame Texture Coding	Wavelet transforms, Matching pursuits, 3D-DCT, Lapped transforms, Improved Intra coding, Variable block-size DCT.
Shape and Alpha Channel Coding	Gray scale shape coding, Geometrical transforms, Shape-adaptive region partitioning, Variable block-size segmentation.
Arbitrary-Shaped Region Texture Coding	Padding DCT, Mean-replacement DCT, Shape-adaptive DCT, Extension/interpolation DCT, Wavelet/subband coding.
Error Resilience/Robustness	Resynchronization techniques, Hierarchical structures, Back channel signaling, Error concealment.
Bandwidth and Complexity Scaling	Generalized temporal-spatial coding, content-based temporal scalability.
Misc.	Rate control, Mismatch corrected stereo/multiview coding, 2D triangular mesh for object and content manipulation, Noise removal, Automatic segmentation, Generation of sprites.

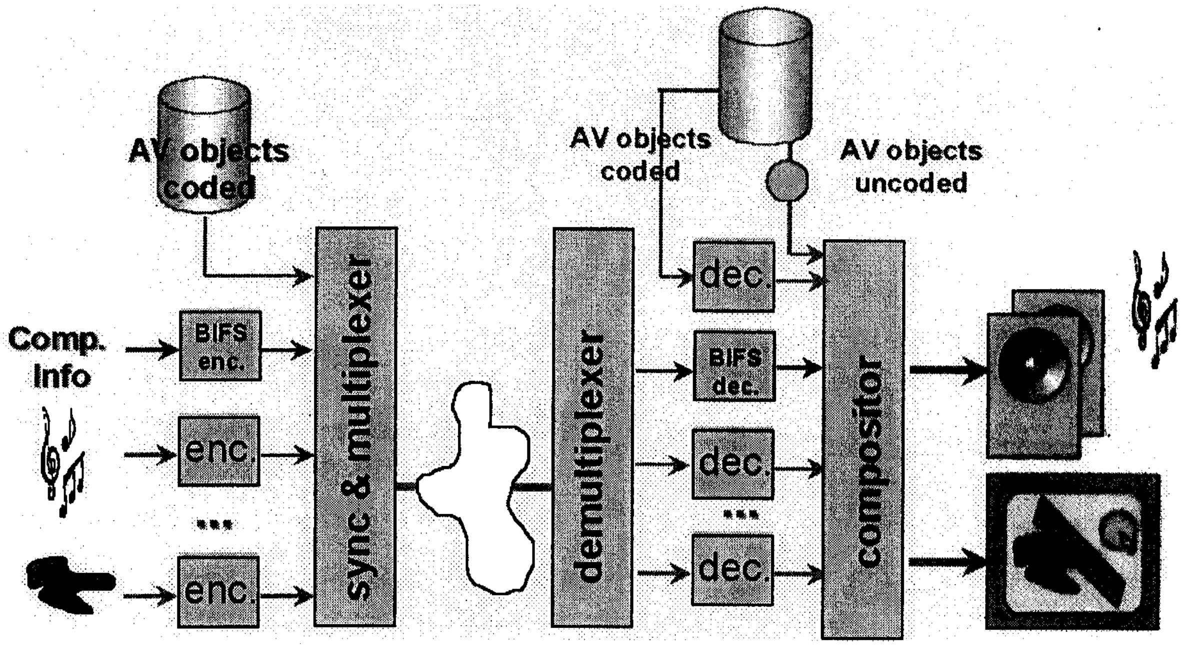

Yeah, they look cute. Now about the nursery...
ABC Children Furniture Company
• Lamp
• Furniture
• Crib
Hi, Uncle Robert!
Scene
Father and child (arbitrarily shaped object)
Floor and walls (sprites)
Web site (application screen)
Furniture (synthetic objects)
Audio object
Crib
Lamp
Side table
Audio
Video
(standard audio and video streams)


Interactive
audiovisual scene
Object
descriptors
Scene
description
information
Primitive
audiovisual
objects
. . .
information
Compression
layer
Elementary streams
SL
SL
SL
SL
SL
SL
SL-packetized streams
FlexMux
FlexMux
FlexMux
MPEG-2
UDP
IP
ATM
PSTN
DAB
. . .
Transmission/storage
ATM = asynchronous
transfer mode
DAB = digital audio
broadcast
IP = Internet protocol
PSTN = public switched
telephone network
UDP = universal data
protocol

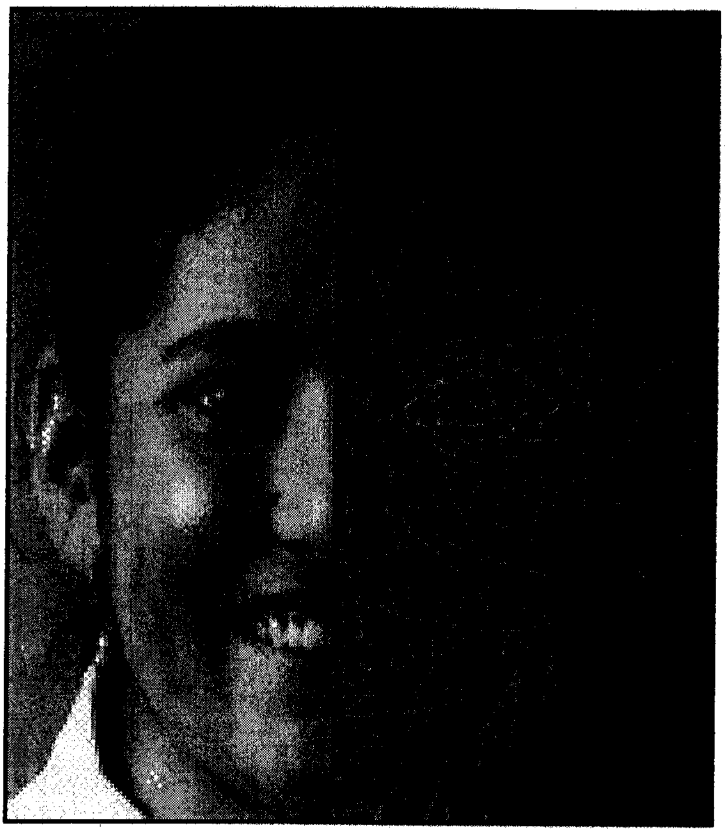

Network Layer
Demultiplex
Elementary Streams
Decompression
Primitive AV Objects
Scene Description (Script or Classes)
Composition Information
Composition and Rendering
Hierarchical, Interactive, Audiovisual Scene
Upstream Data
(User Events, Class Request, ...)

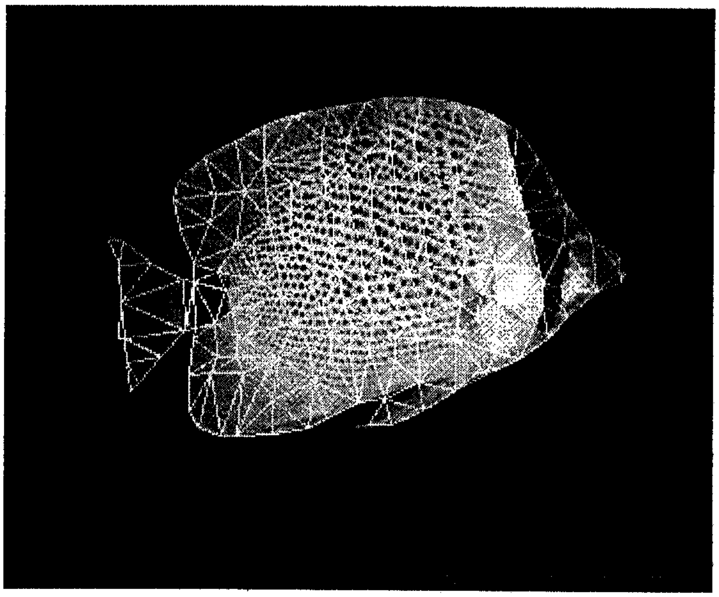

scene
person
2D background
furniture
audiovisual presentation
voice
sprite
globe
desk

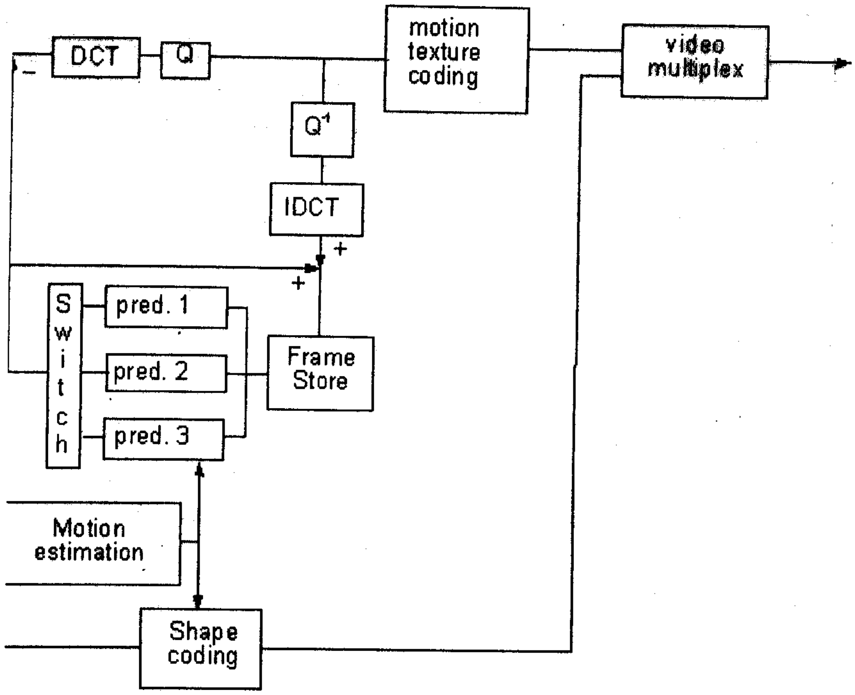




AV objects coded
Comp. Info
BIFS enc.
enc.
...
enc.
sync & multiplexer
demultiplexer
AV objects coded
AV objects uncoded
dec.
BIFS dec.
dec.
...
dec.
compositor





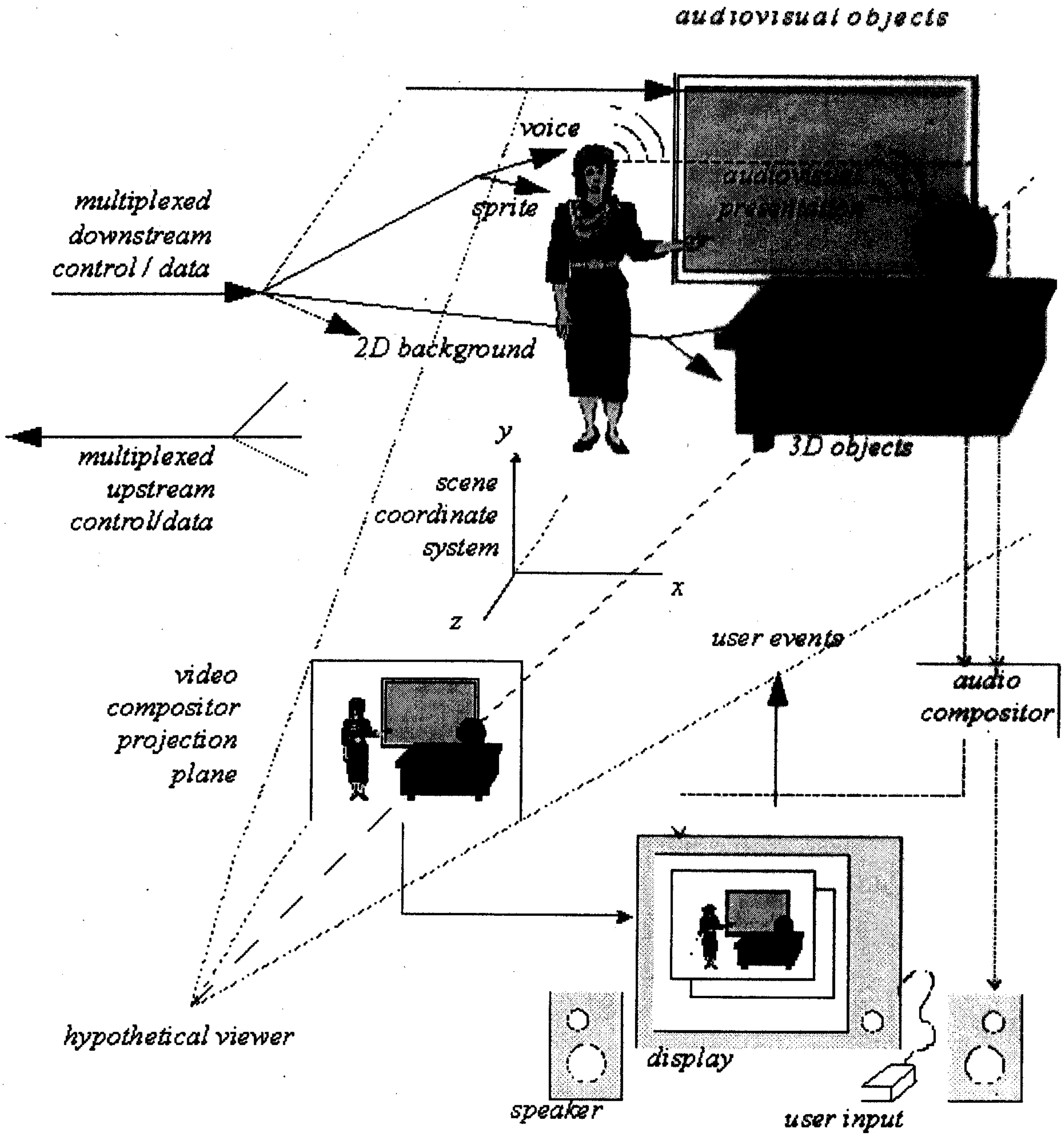


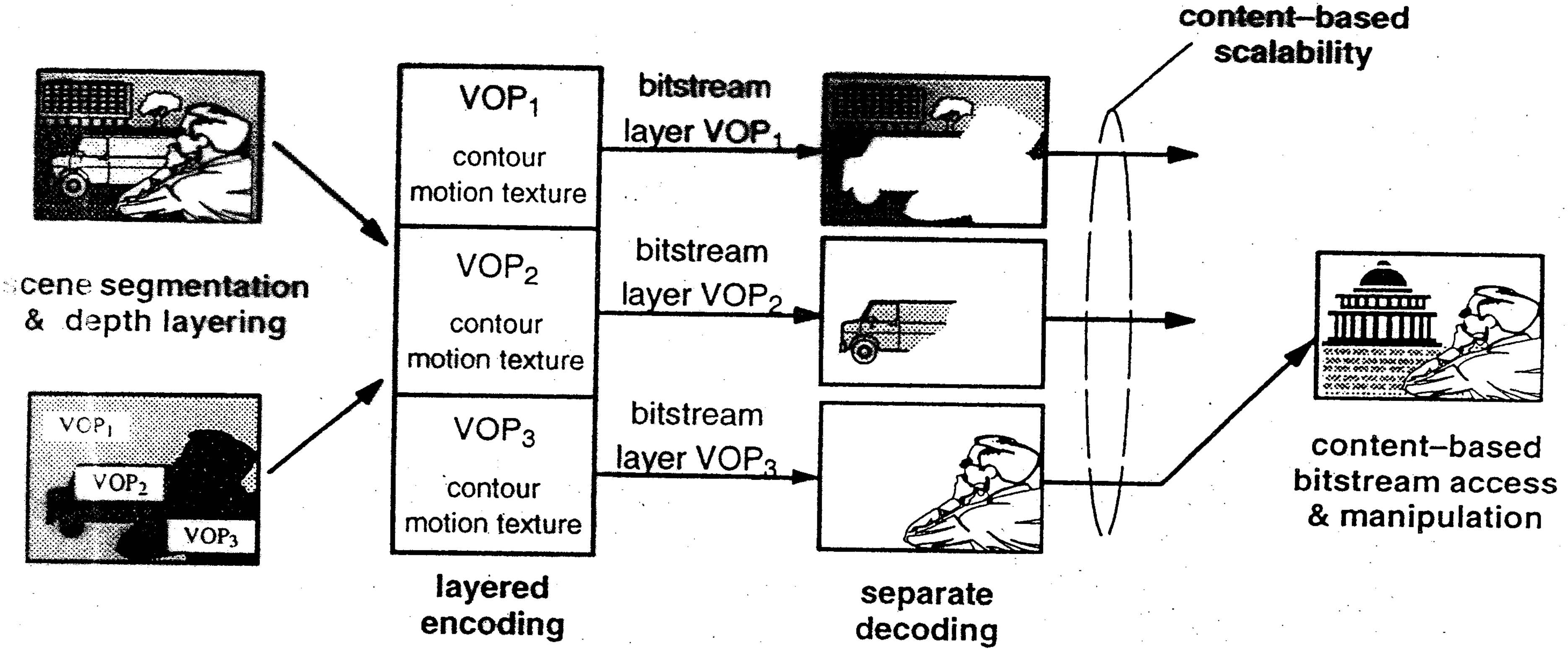
MPEG-4 VLBV Core Coder

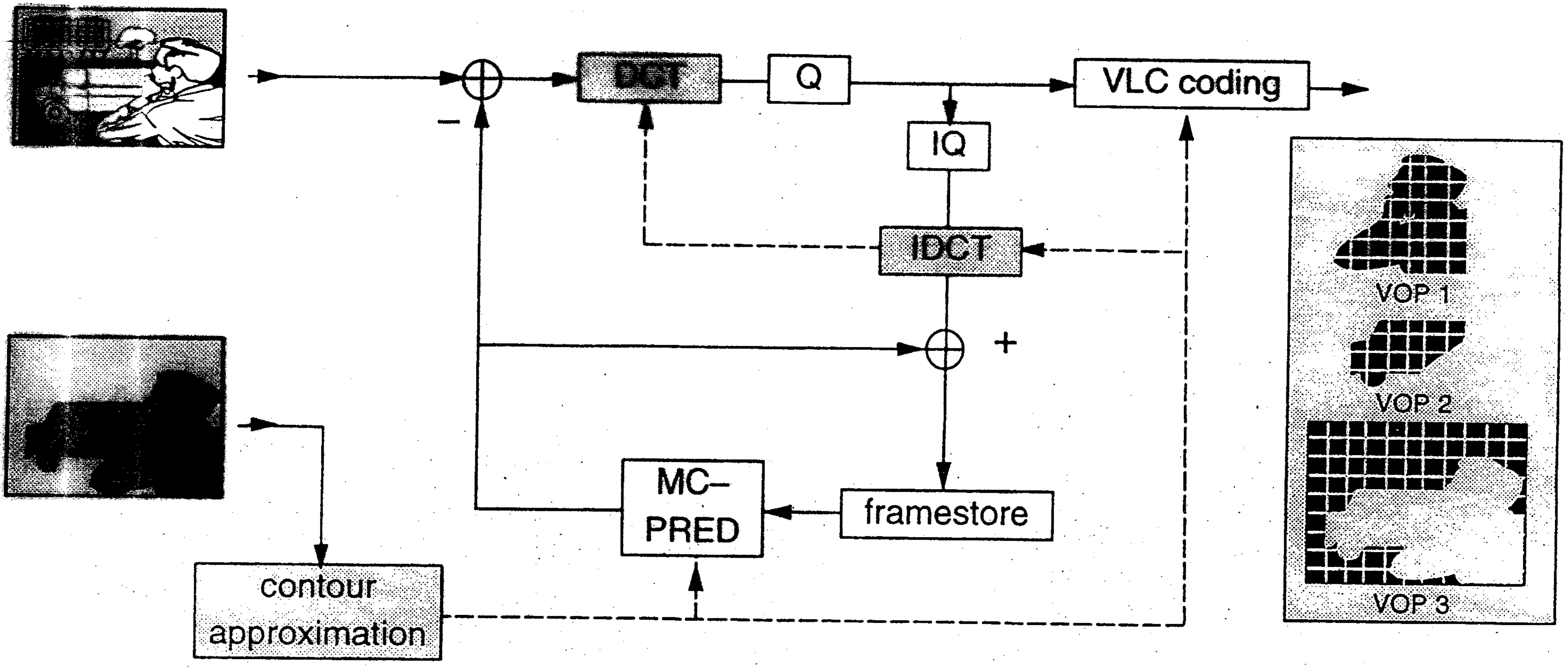
deo
ject
ane
Motion
(MV)
bitstream

(Similar to H.263/MPEG-1)

Generic MPEG-4 Coder


audiovisual objects
voice
audiovisual presentation
sprite
multiplexed downstream control / data
2D background
multiplexed upstream control/data
3D objects
y
scene coordinate system
x
z
user events
video compositor projection plane
audio compositor
hypothetical viewer
display
speaker
user input


content–based
scalability
cene segmentation
& depth layering
VOP$_1$
VOP$_2$
VOP$_3$
VOP$_1$
contour
motion texture
VOP$_2$
contour
motion texture
VOP$_3$
contour
motion texture
layered
encoding
bitstream
layer VOP$_1$
bitstream
layer VOP$_2$
bitstream
layer VOP$_3$
separate
decoding
content–based
bitstream access
& manipulation


DCT
Q
VLC coding
IQ
IDCT
–
+
MC–
PRED
framestore
contour
approximation
VOP 1
VOP 2
VOP 3

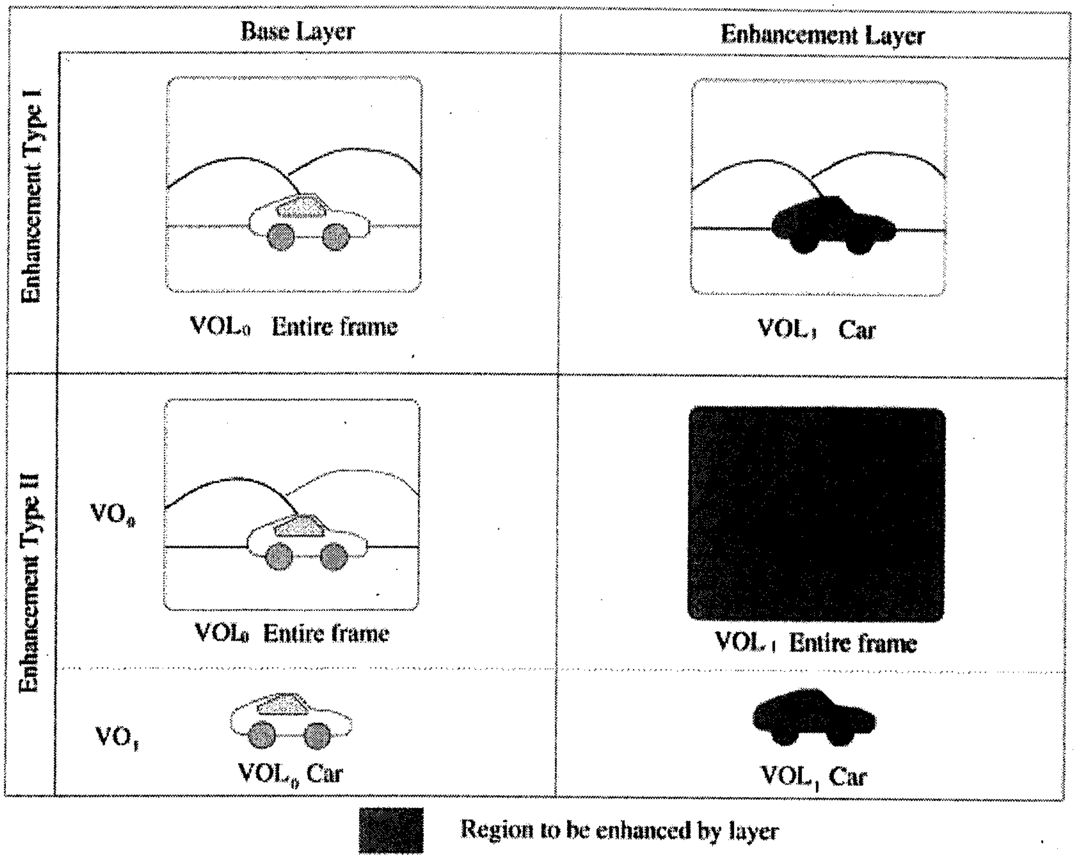


Figure 15: The two enhancement types in MPEG-4 temporal scalability. In enhancement type I, only a selected region of the VOP (i.e. just the car) is enhanced, while the rest (i.e. the landscape) is not. In enhancement type II, enhancement is applicable only at entire VOP level.

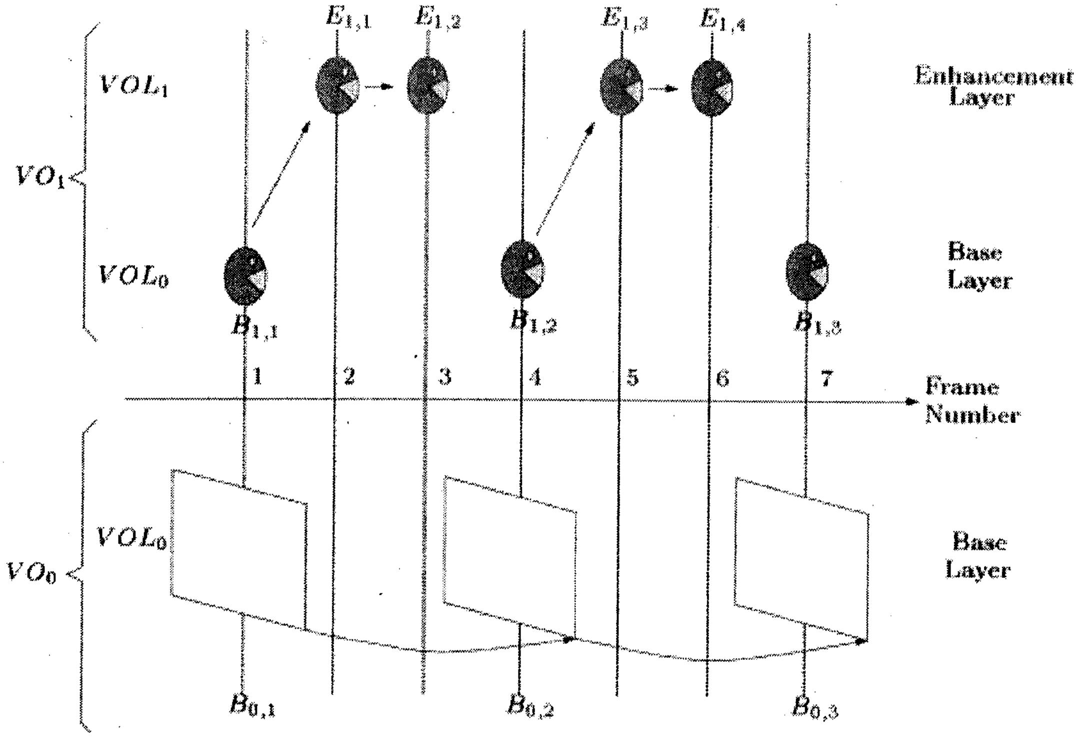


Figure 17: Base and enhancement layer behavior for temporal scalability for type II enhancement (improving the entire base-layer).}

9 Conformance points

Conformance points form the basis for interoperability, the main driving force behind standardization. Implementations of all manufacturers that conform to a particular conformance point are interoperable with each other. Conformance tests can be carried out to test and to show conformance, and thus promoting interoperability. For this purpose bitstreams should be produced for a particular conformance point, that can test decoders that are conforming to that conformance point.

MPEG-4 defines conformance points in the form of profiles and levels. Profiles and levels define subsets of the syntax and semantics of MPEG-4, which in turn define the required decoder capabilities. An MPEG-4 natural video profile is a defined subset of the MPEG-4 video syntax and semantics, specified in the form of a set of tools, or as object types. Object types group together MPEG-4 video tools that provide a certain functionality. A level within a profile defines constraints on parameters in the

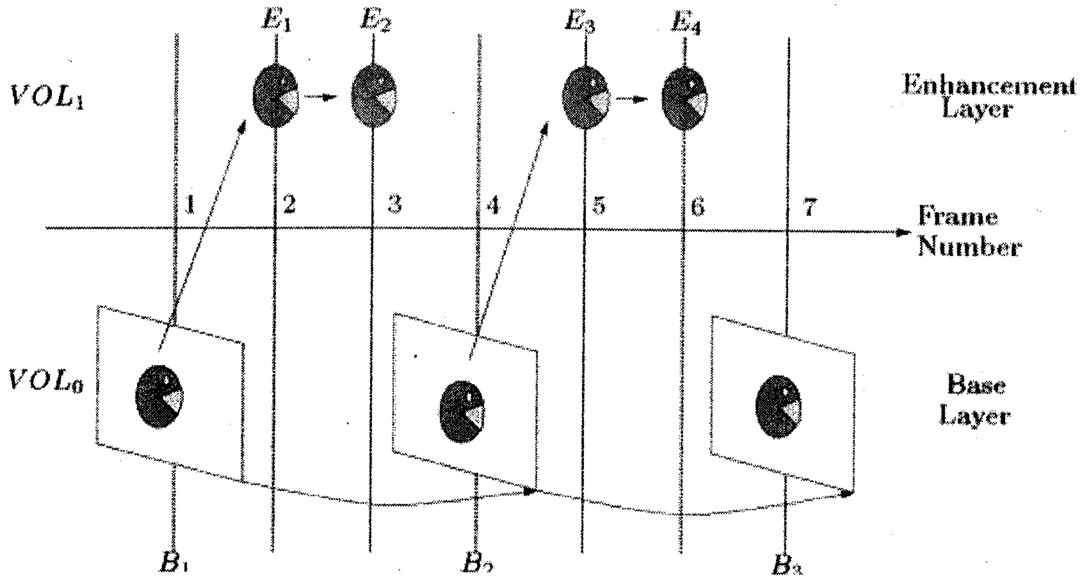


Figure 16: Base and enhancement layer behavior for temporal scalability for type I enhancement (improving only a portion of the base-layer).

An example of type I enhancement is shown in Fig. 16, whereas Fig. 17 illustrates an example of type II enhancement.

	MPEG-4 video object types					
MPEG-4 video tools	Simple	Core	Main	Simple Scalable	N-bit	Still Scalable Texture
Basic(I and P-VOP, coefficient prediction, 4-MV, unrestricted MV)	x	x	x	x	x	
Error resilience	x	x	x	x	x	
Short Header	x	x	x		x	
B-VOP		x	x	x	x	
P-VOP with OBMC (Texture)						
Method 1/Method 2 Quantization		x	x		x	
P-VOP based temporal scalability		x	x		x	
Binary Shape		x	x		x	
Grey Shape			x			
Interlace			x			
Sprite			x			
Temporal Scalability (Rectangular)				x		
Spatial Scalability (Rectangular)				x		
N-Bit					x	
Scalable Still Texture						x

Table 1: MPEG-4 video object types

MPEG-4 video object types	**MPEG-4 video profiles**					
	Simple	Core	Main	Simple Scalable	N-Bit	Scalable Texture
Simple	x	x	x	x	x	
Core		x	x		x	
Main			x			
Simple Scaleable				x		
N-Bit					x	
Scalable Still Texture			x			x

Table 2: MPEG-4 video profiles

Profile and Level		**Typical scene size**	**Bitrate** **(bit/sec)**	**Maximum number of objects**	**Total mblk memory** **(mblk units)**
Simple Profile	L1	QCIF	64 k	4	198
	L2	CIF	128 k	4	792
	L3	CIF	384 k	4	792
Core Profile	L1	QCIF	384 k	4	594
	L2	CIF	2 M	16	2376
Main Profile	L2	CIF	2 M	16	2376
	L3	ITU-R 601	15 M	32	9720
	L4	1920x1088	38.4 M	32	48960

Table 3: Subset of MPEG-4 video profile and level definitions

TABLE 40.1 Basic Parameters for Three Classes of Acoustic Signals

	Frequency range in Hz	Sampling rate in kHz	PCM bits per sample	PCM bit rate in kb/s
Telephone speech	300 - 3,400[a]	8	8	64
Wideband speech	50 - 7,000	16	8	128
Wideband audio (stereo)	10 - 20,000	48[b]	2 × 16	2 × 768

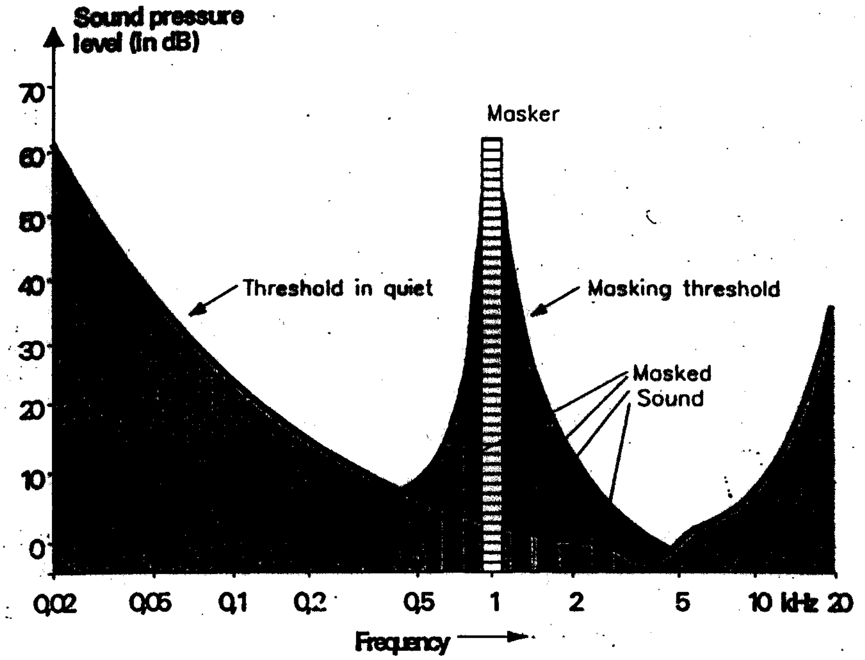
[a] Bandwidth in Europe; 200 to 3200 Hz in the U.S.

[b] Other sampling rates: 44.1 kHz, 32 kHz.

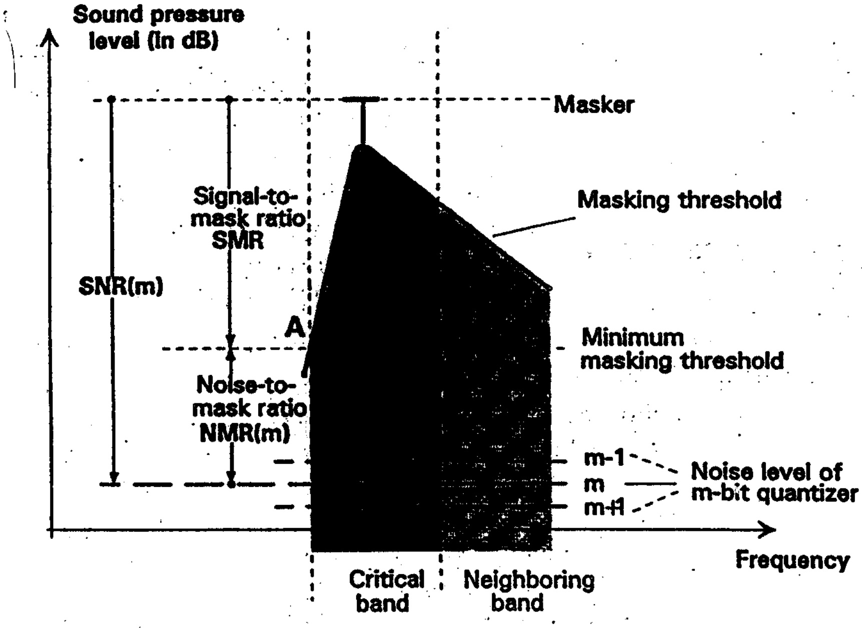
TABLE 40.2 CD and DAT Bit Rates

Storage device	Audio rate (Mb/s)	Overhead (Mb/s)	Total bit rate (Mb/s)
Compact disc (CD)	1.41	2.91	4.32
Digital audio tape (DAT)	1.41	1.05	2.46

Note: Stereophonic signals, sampled at 44.1 kHz; DAT supports also sampling rates of 32 kHz and 48 kHz.



Threshold in quiet and masking threshold. Acoustical events in the shaded areas will not be audible.


Sound pressure level (in dB)
Masker
Masking threshold
Signal-to-mask ratio SMR
SNR(m)
A
Minimum masking threshold
Noise-to-mask ratio NMR(m)
m-1
m
m+1
Noise level of m-bit quantizer
Frequency
Critical band
Neighboring band

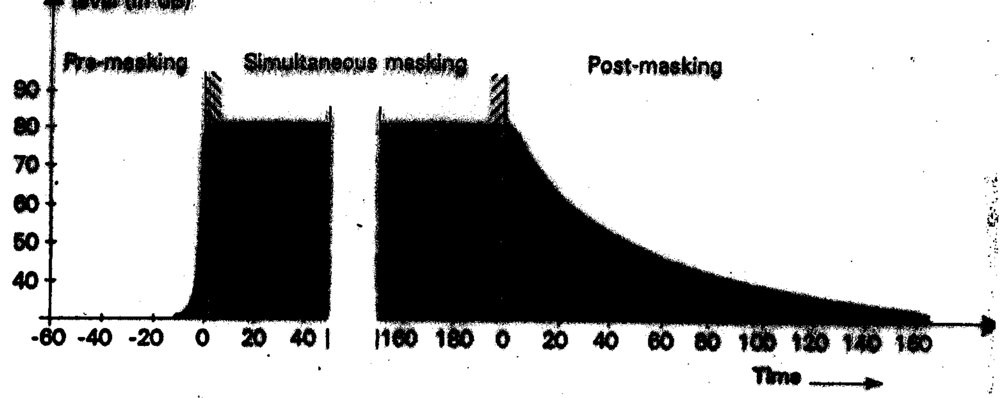
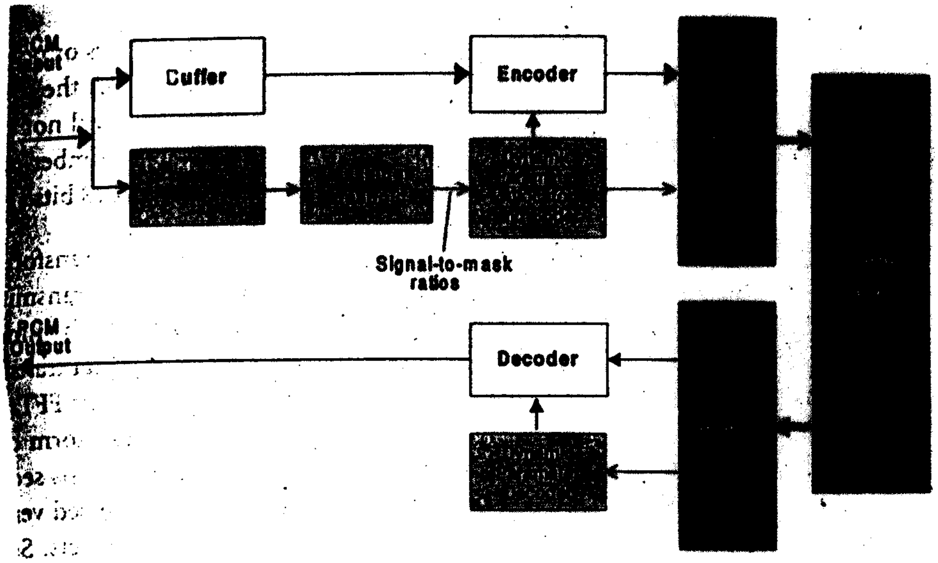
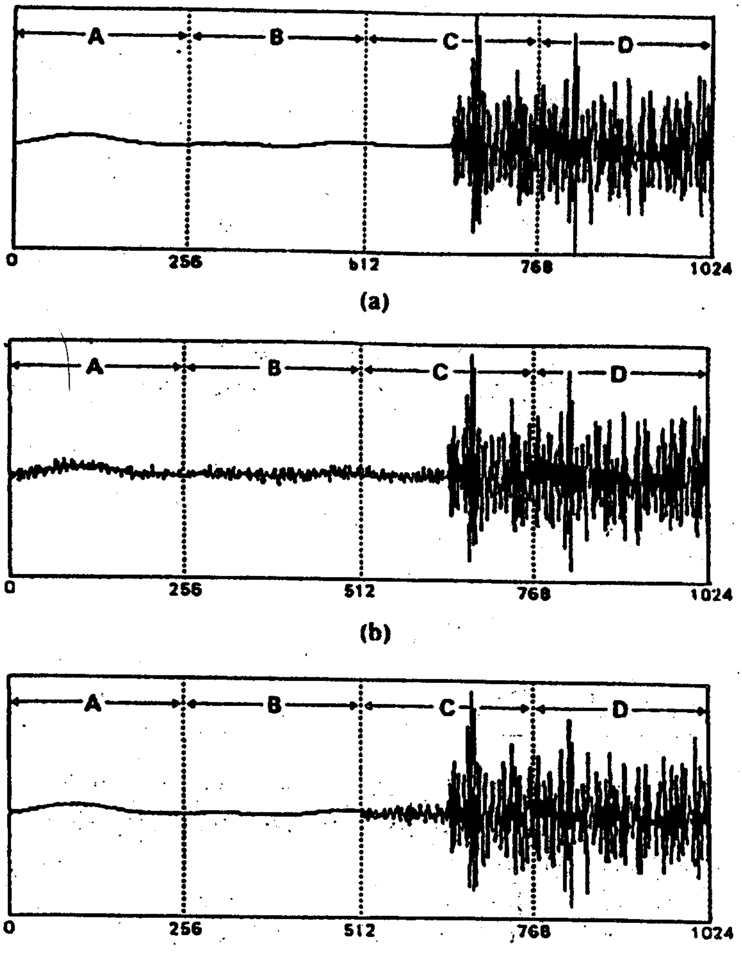


FIGURE 40.3 Temporal masking. Acoustical events in the shaded areas will not be audible.

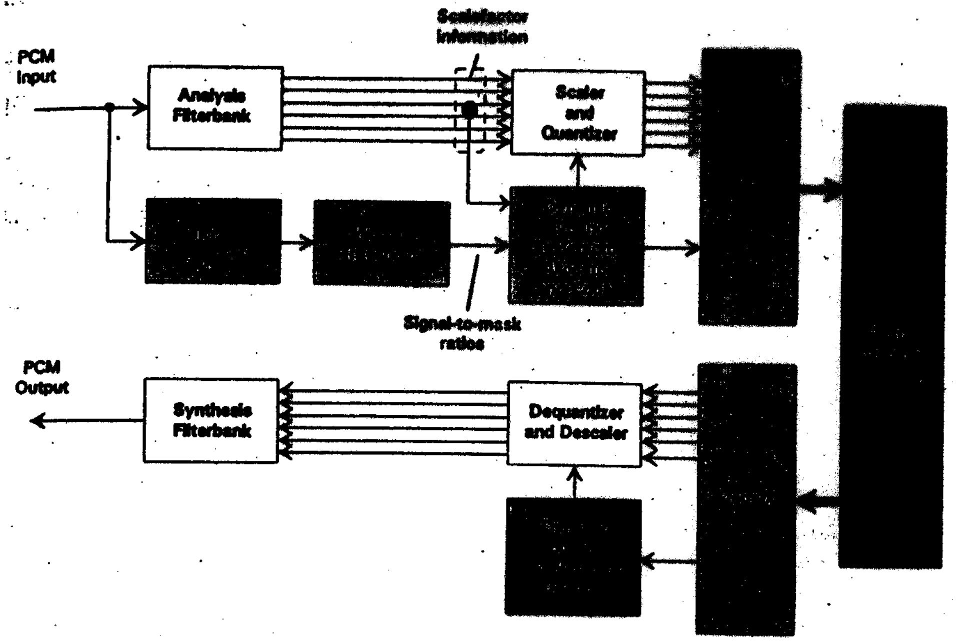


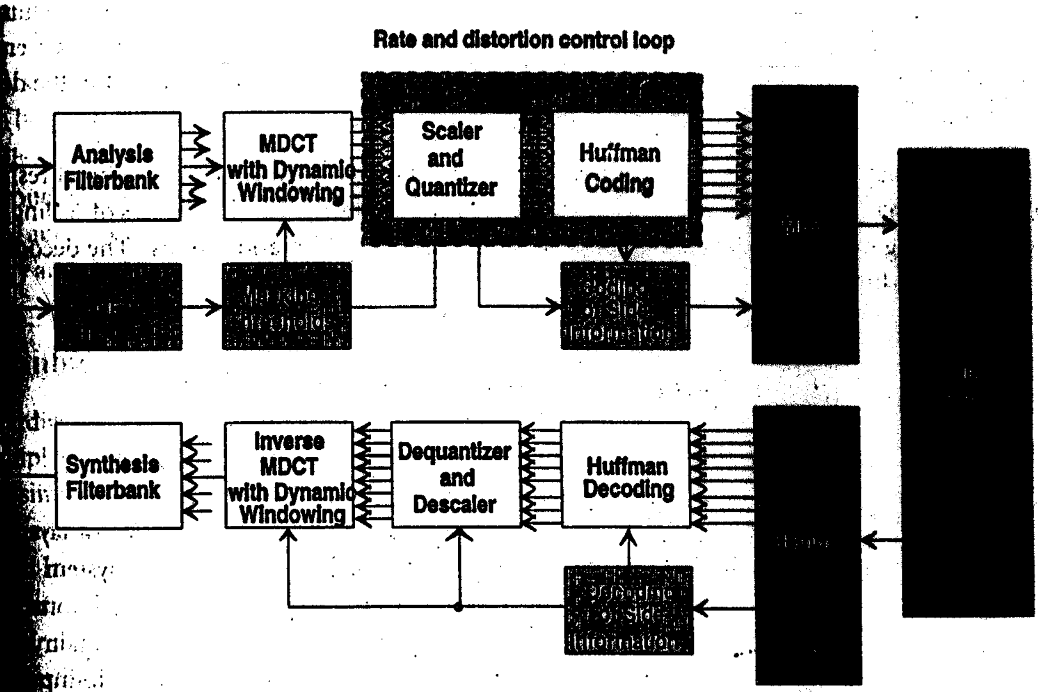
Block diagram of perception-based coders.



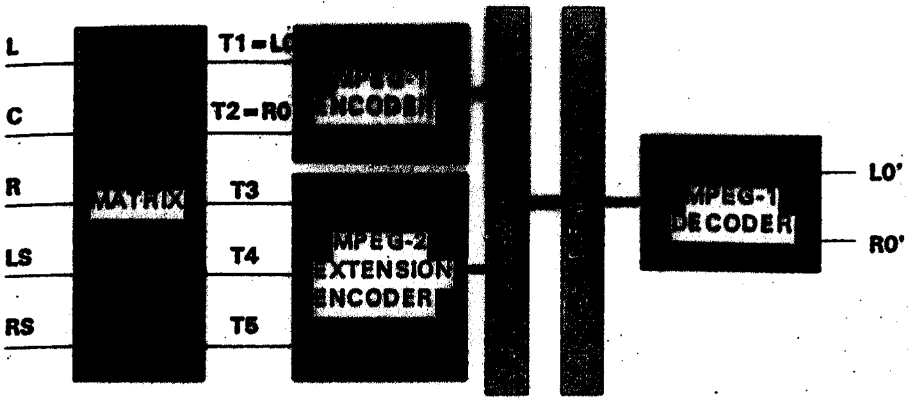
(a)

(b)

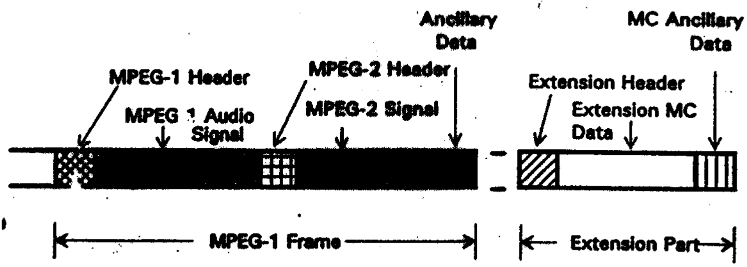



Rate and distortion control loop
Analysis Filterbank
MDCT with Dynamic Windowing
Scaler and Quantizer
Huffman Coding
Masking threshold
Coding of Side Information
Synthesis Filterbank
Inverse MDCT with Dynamic Windowing
Dequantizer and Descaler
Huffman Decoding
Decoding of Side Information

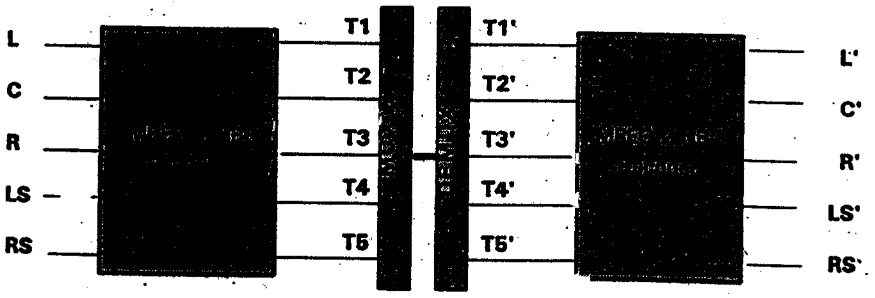
Structure of MPEG-1/Audio encoder and decoder, layer III.



MPEG-1 stereo decoding of MPEG-2 multichannel bit stream.



Data format of MPEG-2 audio bit stream with extension part.



Non-backward-compatible MPEG-2 multichannel audio coding (advanced audio coding).

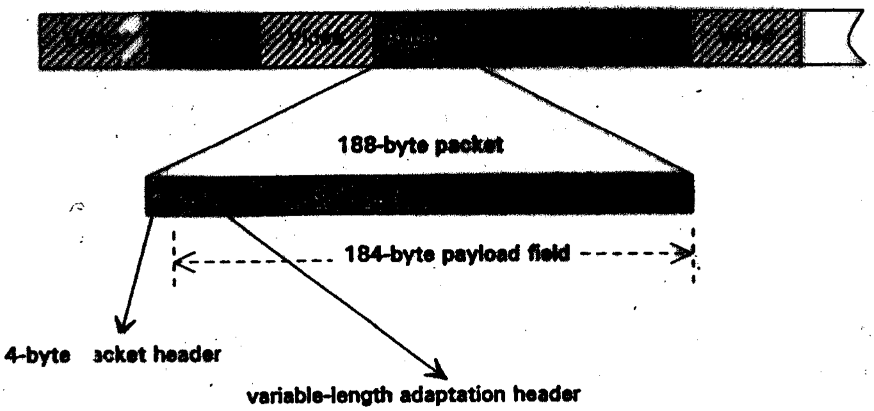


FIGURE 40.14 MPEG packet delivery.

• 1 channel	1/0-configuration:	centre (mono)
• 2 channels	2/0-configuration:	left, right (stereophonic)
• 3 channels	3/0-configuration:	left, right, centre
• 4 channels:	3/1-configuration	left, right, centre, mono-surround
• 5 channels:	3/2-configuration:	left, right, centre, surround left, surround right

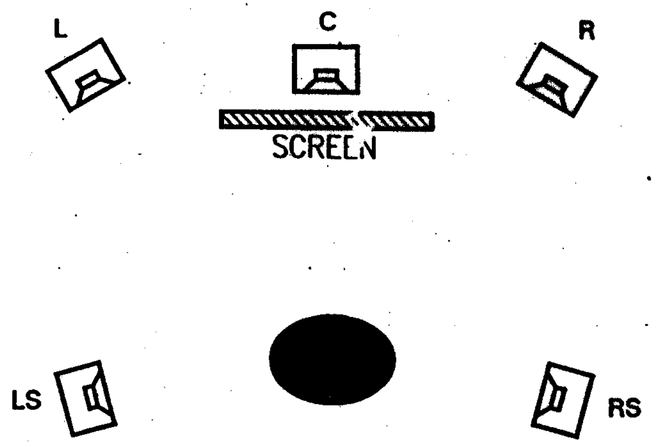


FIGURE 40.15 3/2 Multichannel loudspeaker configuration.