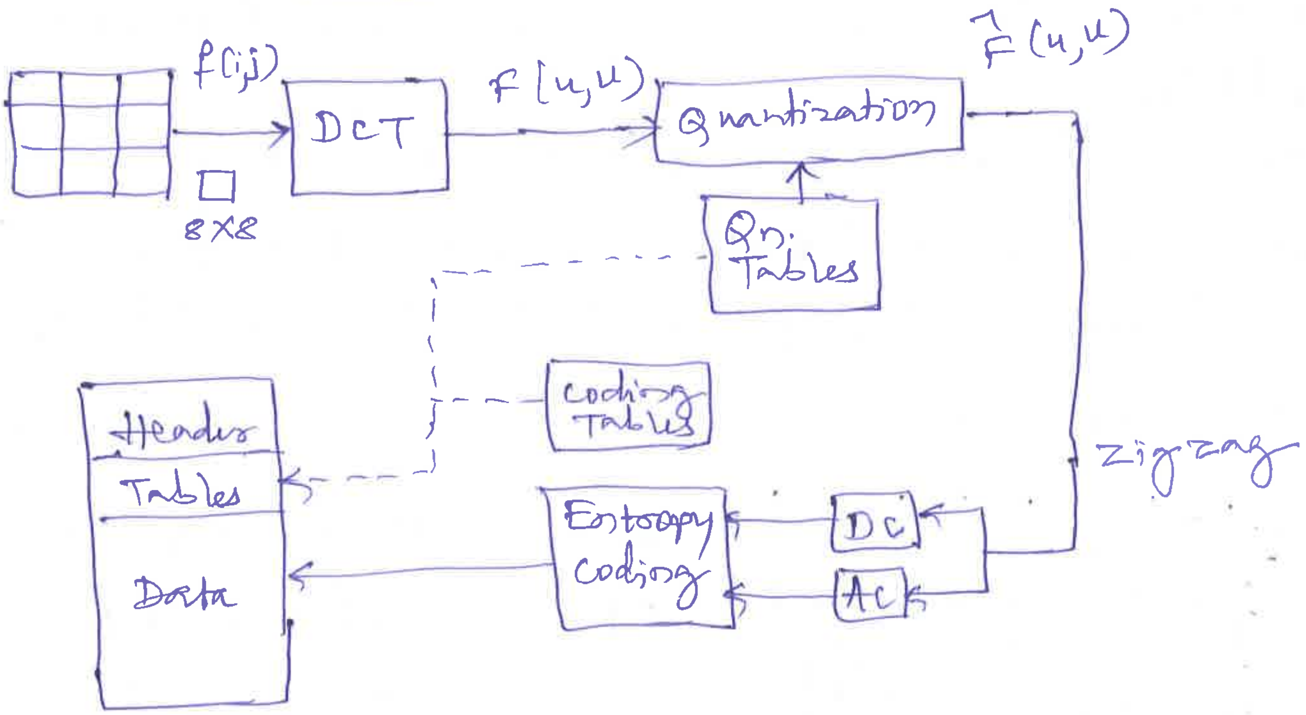
Unit 3

JPEG Image Compression

Steps for JPEG

- Transform RGB to YIQ or YUV
- Perform DCT on image blocks (8×8)
- Apply quantization
- Perform zigzag scanning & run length coding (RLC)
- Entropy coding

Block diagram of JPEG



DCT on image Block (8×8 pixel Block)

2D DCT is applied on each block image represented by $f(i,j)$ with DCT output $F(u,v)$.

Block size 8×8 is selected because it helps faster DCT & IDCT functions.

Quantization – It consists of dividing each entry of a block by an integer & rounding.

Quantized DCT coefficients → $\tilde{F}(u,v) = \text{round}\left(\frac{F(u,v)}{Q(u,v)}\right)$

$Q(u,v)$ = Quantization matrix entry

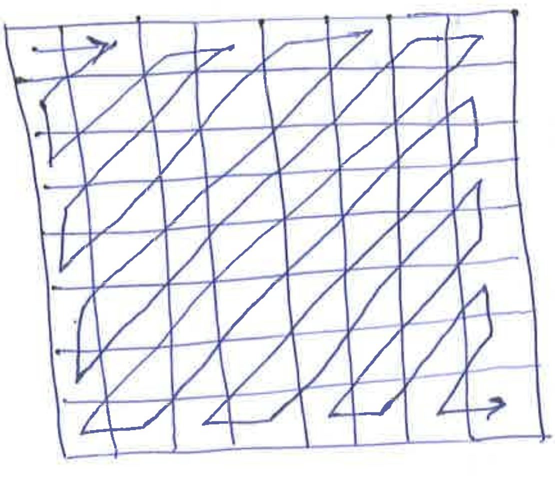
Quantisation helps in reducing total number of bits needed to represent each entry; hence produces better compression.

$\hat{F}(u,v)$ values have lesser magnitude compared to $F(u,v)$.

Elements of $Q(u,v)$ are selected in such a way that results in psychophysical studies. with the goal of maximizing the compression ratio. while minimizing perceptual losses in JPEG.

By quantization, most of the values are zeroed, along a zigzag path. It is useful for compression using Run length coding (RLC).

Zig zag Scan



Run length coding (RLC)

RLC is performed over the zigzag blocks.

Most of the zeroes along zigzag are compressed using RLC.

RLC replaces the values by a pair (RUNLENGTH, VALUE) for each run of zeroes in the AC coefficients. of $\hat{F}(u,v)$

RUNLENGTH: No. of zeros in the run

VALUE: Non-zero coefficients.

DC: component - top left corner entry in a Block (8×8) is the DC component; rest all 63 elements belong to AC components.

DPCM is performed on DC components of inter-blocks.
This helps in giving a link from one block to another block & reducing the no. of bits to represent these DC values.

eg If 5 DC coeffs of image blocks are
150, 155, 149, 152, 144

DPCM O/P → 150, 5, −6, 3, −8.

predictor for ith block $d_i = DC_{i+1} - DC_i$
$d_0 = DC_0$.

<u>Entropy coding</u>

Both DC and AC coefficients undergo Huffman entropy coding seperately.

<u>JPEG Modes</u> : 4

1. sequential mode
2. progressive mode
3. Hierarchical mode
4. Lossless mode

① sequential mode – each image component is encoded in a single left-to-right & top to bottom scan.

Used in video codecs.

② progressive mode – first few scans carry few bits and deliver a rough picture of what is to follow.

Clarity increases after each scan as more data is scanned.

Used in web browsers.

Scan 1: Encode DC & first few AC $AC_1, AC_2, \ldots$

Scan 2: AC_3 AC_4 AC_5

⋮

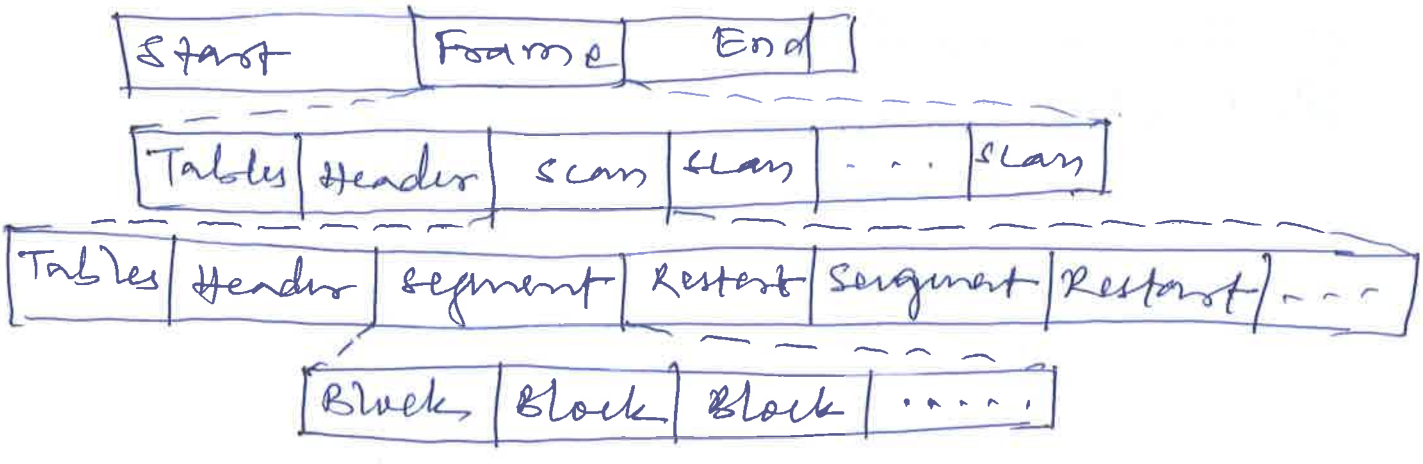
Scan k: AC_{61}, AC_{62}, AC_{63}

③ Hierarchical mode

Encodes the image in a hierarchy of several resolutions.

④ Lossless mode – It involves only differential coding, no transform coding.

JPEG Bitstream



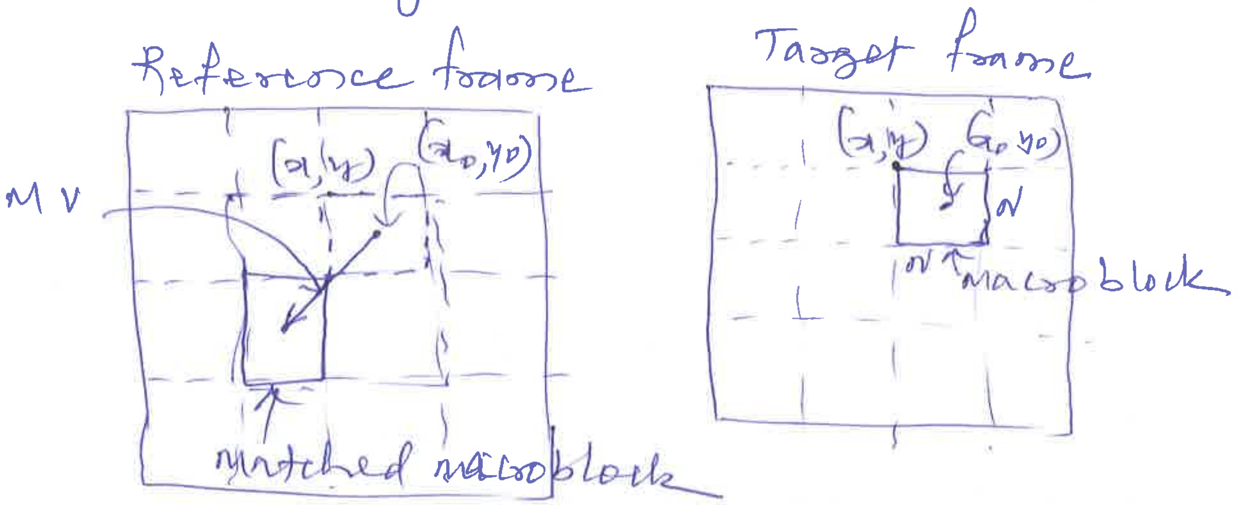
Video compression based on motion compensation

- JPEG image compression exploits spatial redundancy
- A video is a sequence of images stacked in temporal dimensions.
- Video frame rate is relatively high (≥ 15 frames / second).
- The contents in such frames do not change rapidly unless certain objects in the scene move extremely fast.
- Thus video has temporal redundancy.
- Every new image in a video need not be coded independently
- fast moving objects are "motion generators" that are to be compensated by detecting the displacement of corresponding regions in these frames and computing the differences

Motion compensation Algorithms steps

1. Motion estimation (Motion vector search)
2. Motion compensation based prediction
3. Derivation of prediction error - difference

Each image is divided into macroblocks of size (N X N), N=16 for luminance, N=8 for chrominance if 4:2:0 chroma subsampling is used.



Target frame current image frame
Reference frame - Most similar macroblock match found in previous / future from target frame.

MV - motion vector: displacement vector

search for motion vectors

sequential search - motion vector (MV)
p: +ve integer

```
Begin
    min_MAD = LARGE_NUMBER;  /* Initialization */
    for i = -p to p
        for j = -p to p
        {
            cur_MAD = MAD(i,j);
            if cur_MAD < min_MAD
            {
                min_MAD = cur_MAD;
                u = i;      /* get coordinates of MV */
                v = j;
            }
        }
End
```

In sequential search, MV is searched sequentially over the whole window in the reference frame. (full search)

Mean Absolute Difference (MAD) is computed.

Vector (i, j) offers the least MAD is designated the MV (u, v).

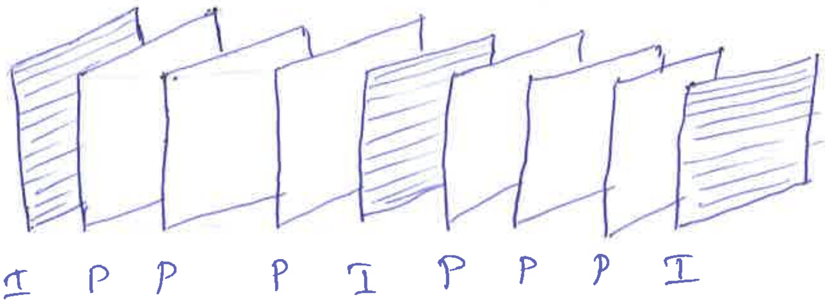
H.261 – Earlier digital video compression standard

- Designed for videophone, video conferencing over audio visual ISDN telephone lines.
- Bit rates of p x 64 kbps, p = 1 to 30.
  video encoders delay ≤ 150ms.

Other variants

H.221 audio visual channel 64 to 1920 kbps
H.230 Frame control signals for audiovisual system
H.242 Audio visual communication protocols
H.261 video encoder / decoder for audio visual services at p x 64 kbps
H.320 Narrow band audio visual terminal equipment for p x 64 kbps transmission.

H.261 Frame sequence



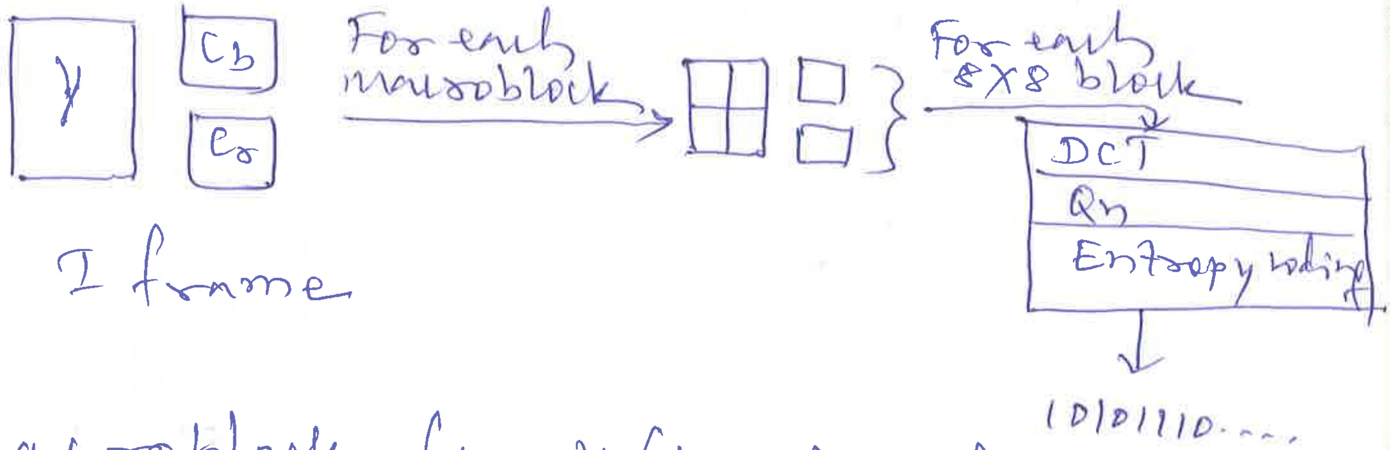
I : Intra-frames P : Inter-frames

I: independent frames (images)
JPEG compression is used to compress I frames.

P: Dependent frames: obtained from forward predictive coding.
Current macroblocks are identified from previous ~~I/P~~ frame macroblocks.

I frame coding to remove spatial redundancy
P frame coding is to remove temporal redundancy between previous I frames & P frames.

I frame coding



I frame

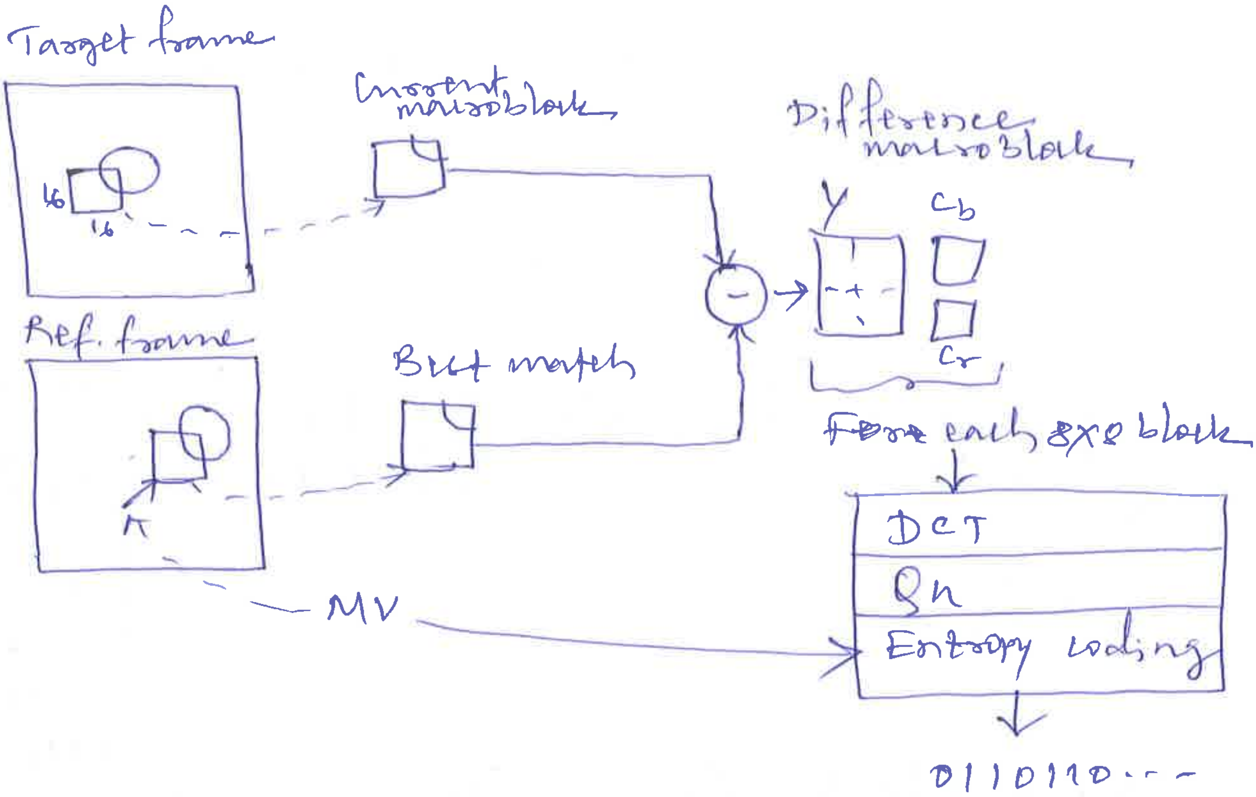
Macroblock for Y (16X16) & for Cb & Cr (8X8), ~~MAC~~ chroma subsampling 4:2:0
For each ~~macro~~ 8X8 block, JPG compression steps are used.

P- frame coding (predictive coding)

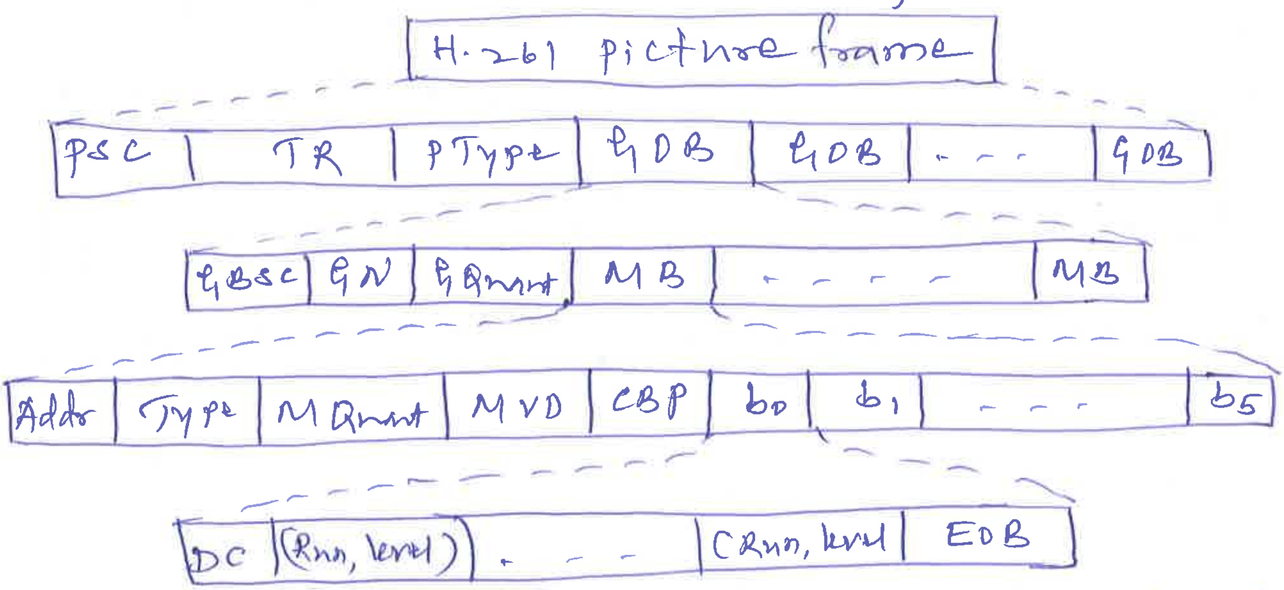
- For each macroblock in the target frame, a motion vector is allocated by using search method.
- Difference macroblock is derived to measure the prediction error.
  It contains 4 Y blocks, one Cb, one Cr
- Each 8X8 block goes to DCT, Qn, entropy coding.

$$MVD = MV_{preceding} - MV_{current}$$

MVD is sent for entropy coding.



H.261 Video Bitstream Syntax



PSC: Pict. start code
PType: Pict. type
GBSC: GOB start code
GQuant: GOB Quantizer
MQuant: MB Quantizer
CBP: coded block pattern

TR: Temporal ref.
GOB: Group of Blocks
GN: Group no.
MB: Macroblock
MVD: Motion vector data
EOB: End of block

1. Picture layer: PSC differentiates boundaries between pictures.

2. GOB Layer: H.261 pictures are divided into regions of 11×3 macroblocks – GOB.

3. Macroblock layer: Each MB has its own address in GOB, quantizer (MQuant) & six 8×8 image blocks (4Y, 1Cb, 1Cr)

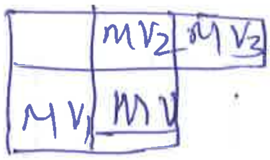
4. Block layer: Each 8×8 block bitstream has DC component & run length parameters for AC components.

H.263

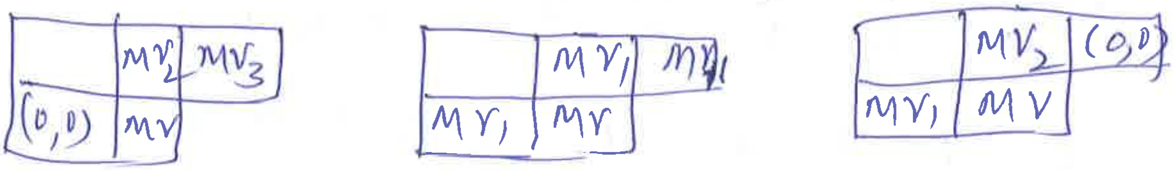
- It is a improved videoconferencing on PSTN networks.
- Low bit rate up to 64 kbps.

Motion compensation

- H(2) and V(2) MV components are predicted from the median values (MV1, MV2, MV3) refer to previous, above, above right.



MV – current motion vector
MV1 – Previous motion vector
MV2 – above motion vector
MV3 – above & right

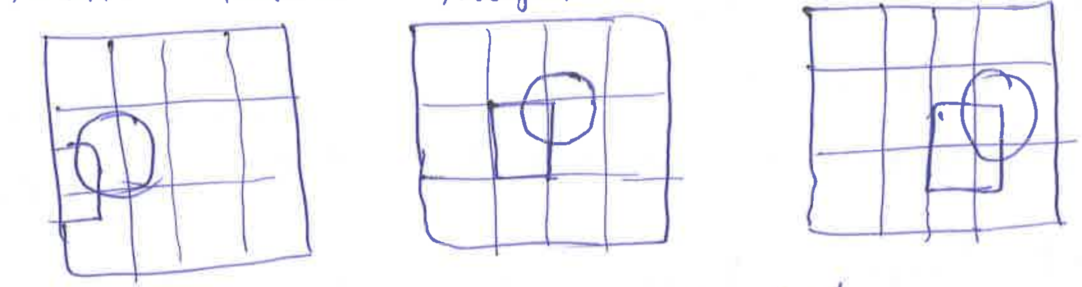


Current Macroblock is @ the border of picture or GOB.

MPEG1 – Moving pictures Expert Group. (1991)

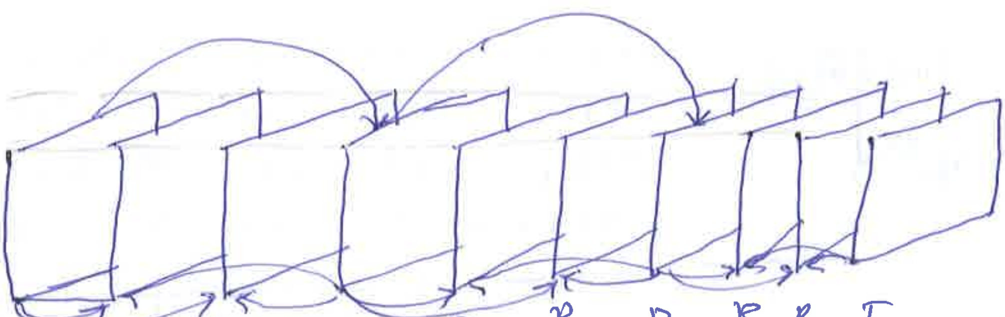
- coding for Moving pictures and associated audio for digital storage media up to 1.5 Mbps.

Motion compensation in MPEG1



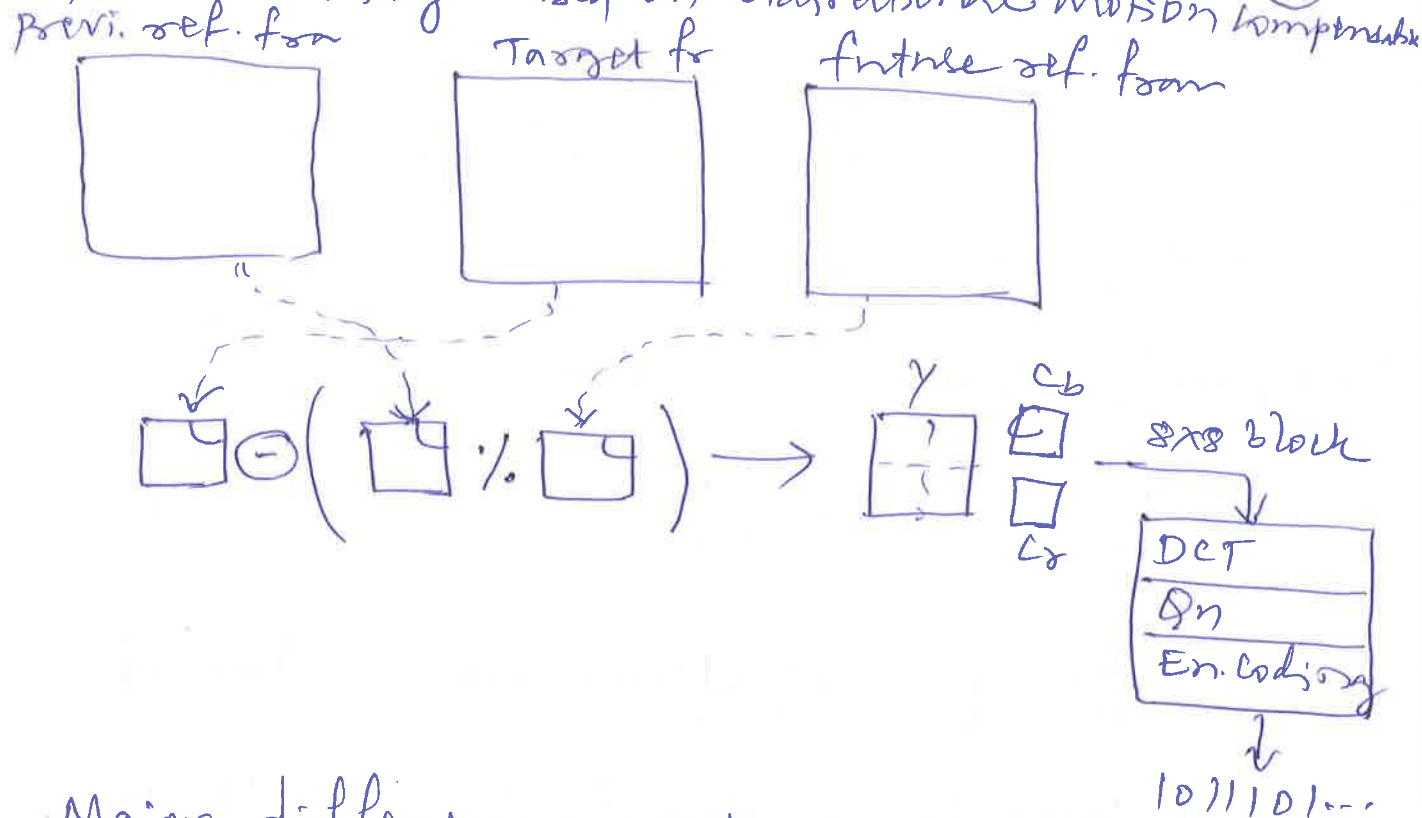
Need for bidirectional search

- If target macroblock is matched with both next frame and previous frame, it will find a good matching.
- Apart from similar I and P frames as in H.261, MPEG1 adopts 3rd frame known as B frame.
- Each macroblock from a B frame will specify up to 2 motion vectors, one from forward and one from backward prediction.



displayed order	I	B	B	P	B	B	P	B	B	I
coding & trans. order	I	P	B	B	P	B	B	I	B	B

B frame coding based on bidirectional motion compensation



Major differences with H.261

- source formats

  H.261 → CIF & QCIF

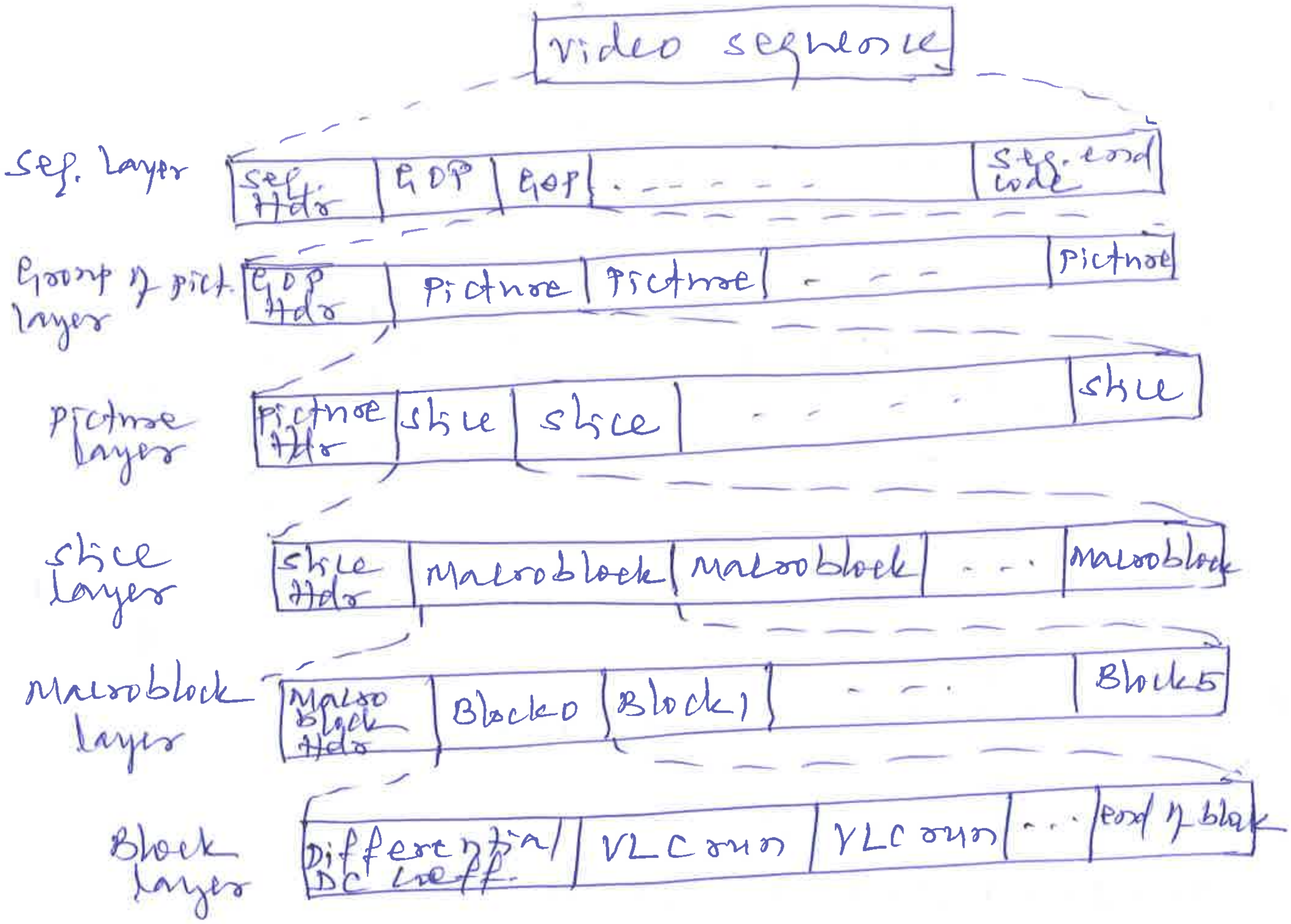
  MPEG1 → SIF (NTSC, PAL)

- slices

  H.261 → GOB

  MPEG1 → slices: contain variable number of GOB's & macroblocks. In a single picture, slices may start and end anywhere in the picture

MPEG1 Video Bitstream



1. sequence layer:
seqn Header contains - the information about picture such as hzl. size, vtl-size, pixel aspect ratio, frame rate, bit rate, buffer size, Qn. matrix, etc. as GOP's

2. GOP layer: contains one or more pictures, one of which must be I frame.
GOP header contains - time code for hr, min, sec from the start of the sequence.

3. Picture layer: I, P, B frames. D frame is for DC coeff's.

(4) Slice layer - It is introduced for bit rate control, recovery and synchronization after lost or corrupted bits.

Have variable number of macroblocks.

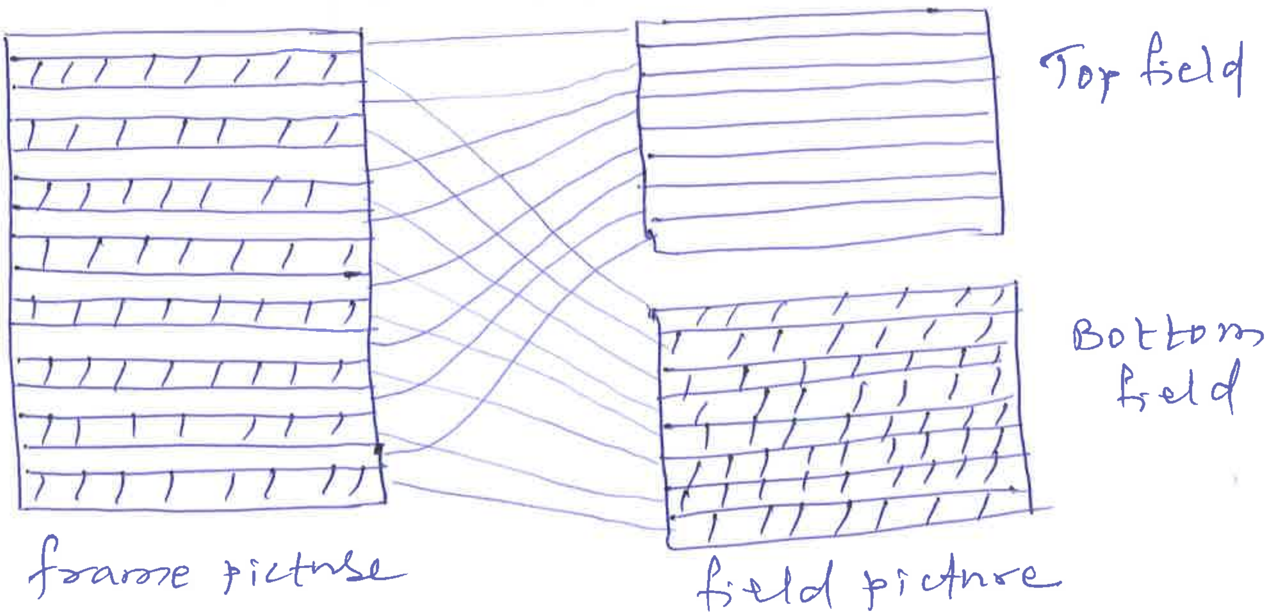
(5) Macroblock layer: 4 Y, 1 Cb, 1 Cr (8x8)

(5) Block layer - If the blocks are intra-coded, differential DC coefficients is sent first, followed by VLC AC coeff.

MPEG 2

- Higher bit rates up to 4 Mbps.
- Used for interactive TV, HDTV, DVD's, scalable video.
- 7 Profiles – simple, main, SNR scalable, spatially scalable, High, 4:2:2, Multiview
  Each profile has up to 4 levels.

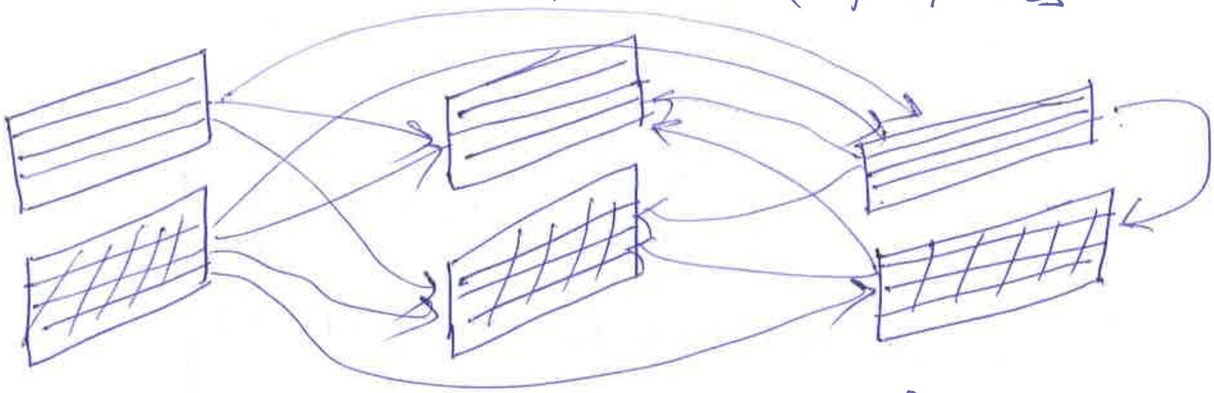
Interlaced Video



- Each ~~field~~ frame picture is divided into 2 ~~fram~~ field pictures.
- 16 col. x 16 row field picture macroblock correspond to 16x32 block area in frame picture.

Five modes of predictions

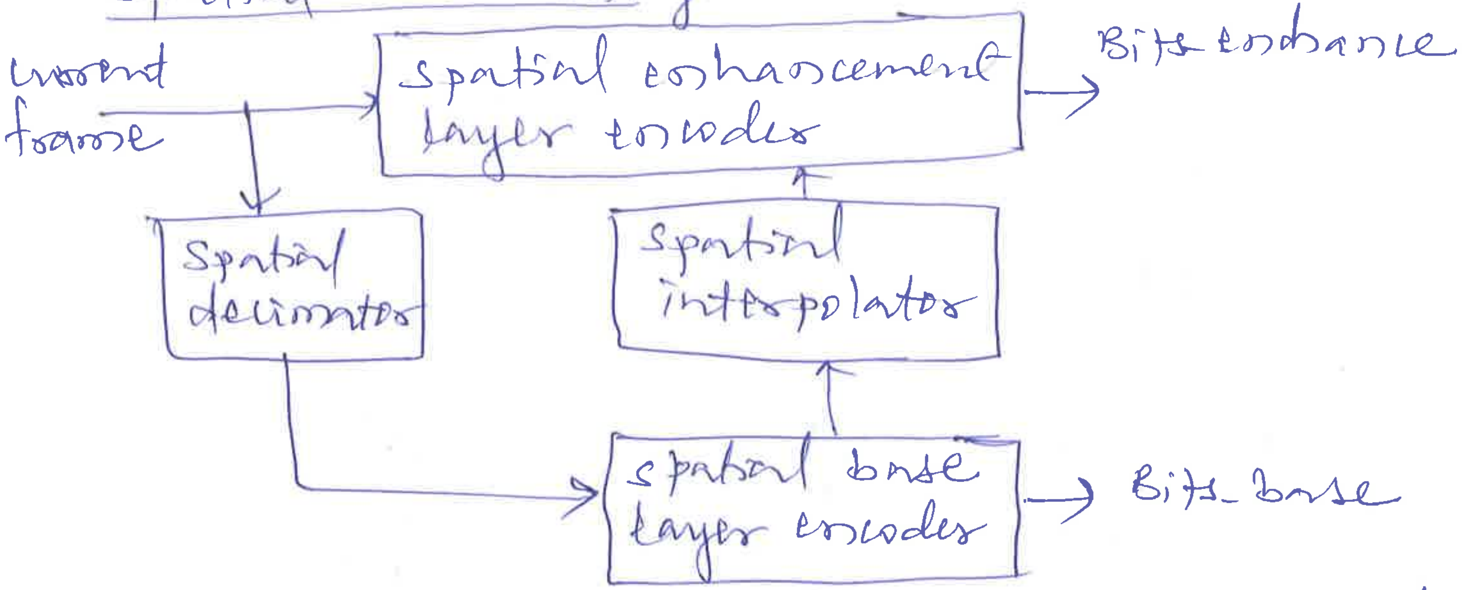
1. Frame predictions for frame picture
   same as MPEG1 to predict P & B frames
2. field prediction for field pictures



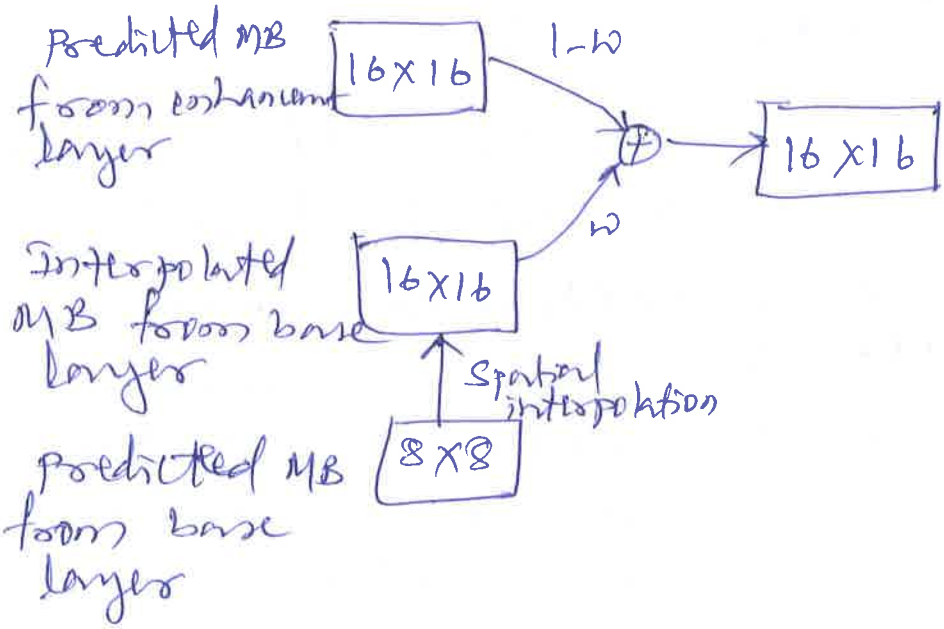
- P fields pictures ~~are~~ are predicted from the 2 most ~~frequently used~~ recently encoded fields.
- Macroblocks in the top field pictures of the are ~~forwarded~~ predicted from top field or bottom field pictures of the preceding I or P frames.
- Macroblocks in the bottom field picture predicted from top field picture of the same frame or bottom field picture of the preceding I or P frame.
- B field pictures are predicted from top and bottom field pictures of future or previous. ~~fiel~~ pictures.

3. Field predictions for frame pictures
4. 16x8 MC for field pictures
5. Dual ~~fo~~ Prime for P pictures

Spatial scalability

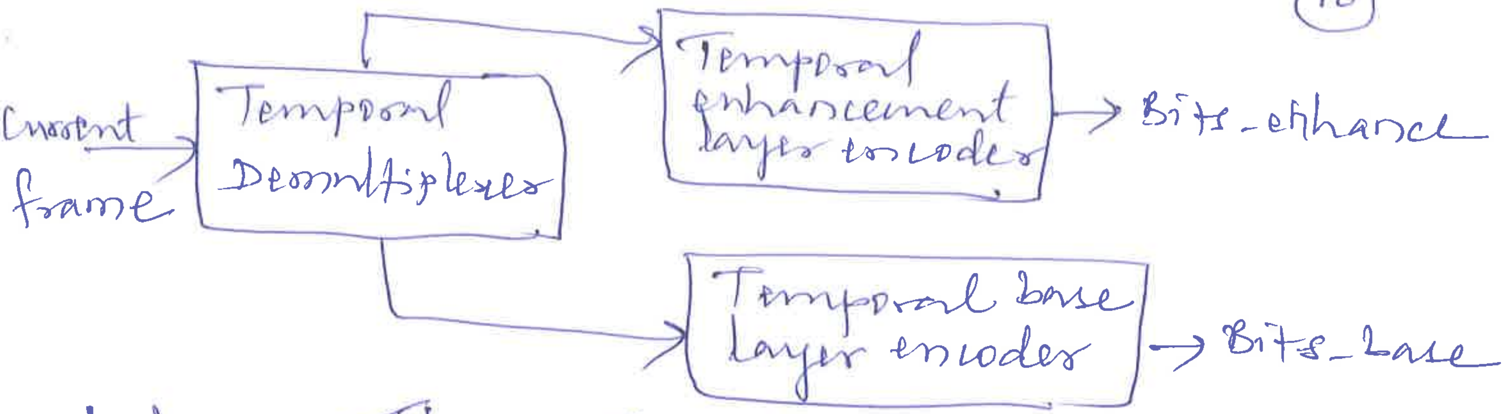


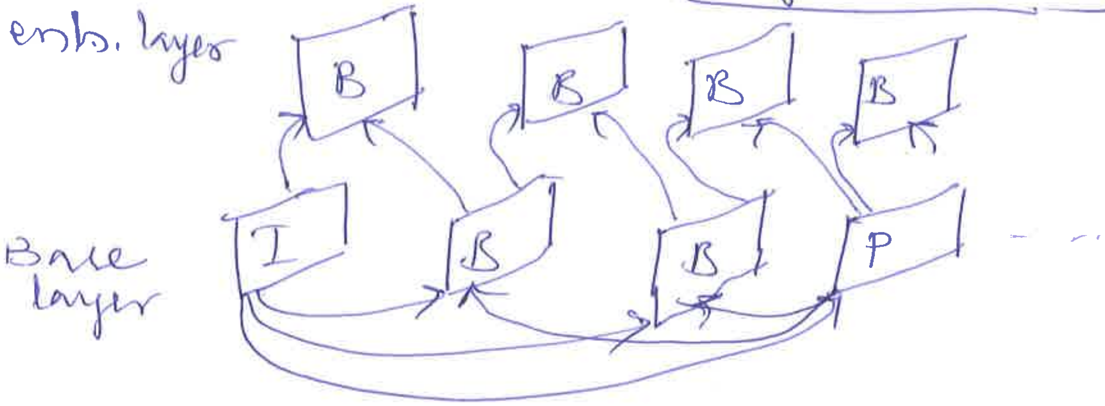
Original video data is spatially decimatted by 2 & sent to base layer encoder. After motion compensation, DCT on prediction errors, quantization & entropy coding are performed to get Bits_base.



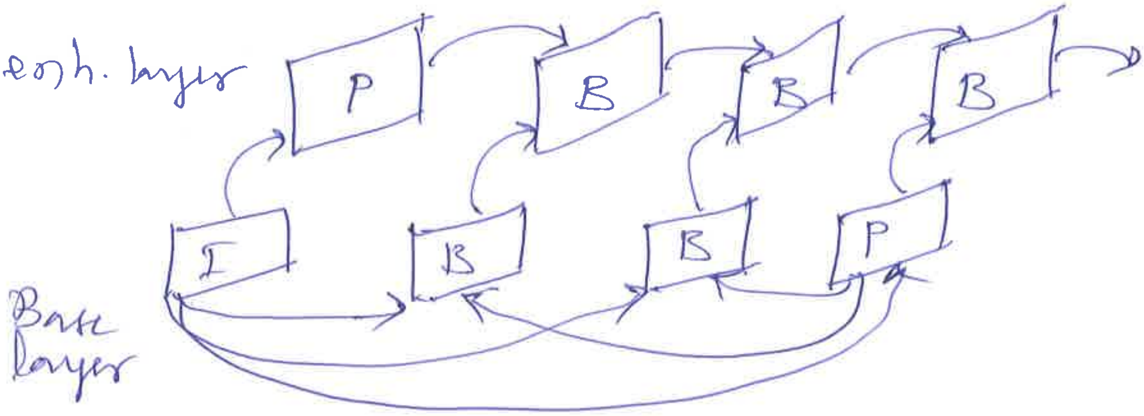
Temporal scalability

Temporally scalable coding has both the base and enhancement layers of video at reduced temporal rate (frame rate).





interlayer motion-compensated prediction



combined motion compensated prediction & interlayer motion compensated prediction

Differences with MPEG1

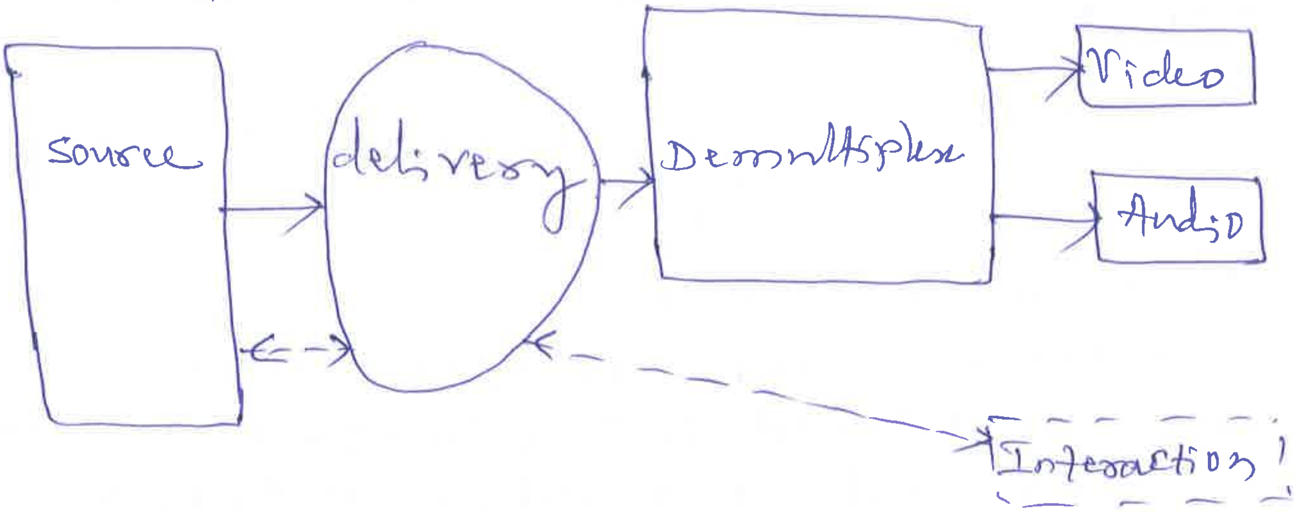
1. Better resilience of bit errors
2. support of 4:2:2 & 4:4:4 chroma subsam
3. non-linear quantization
4. More restricted slice structure
5. More flexible video formats.

MPEG 4

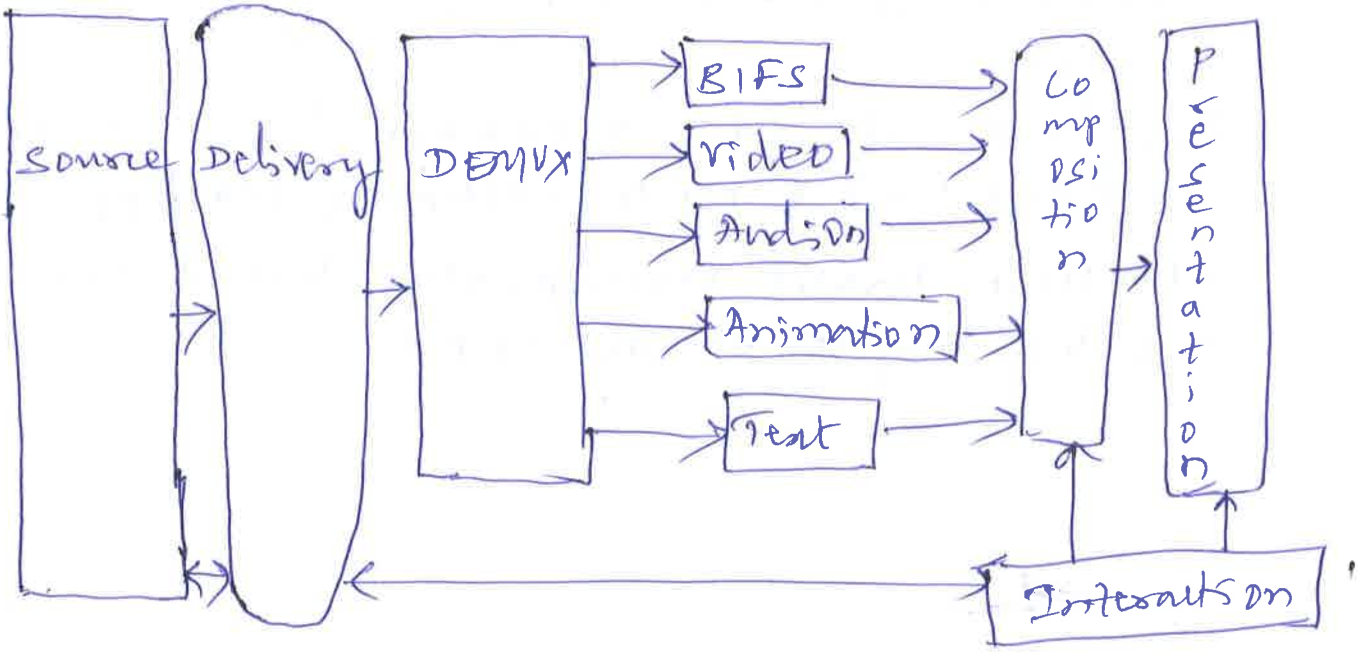
- MPEG1 & 2 are rectangular frame based coding.
  Concerns: compression ratio & satisfactory video quality.
- Benefits of MPEG4
  - Provision for user interactivities
  - Object based coding (media objects: audio & visual - natural or synthetic)
  - possible manipulations - insertion, deletion, translation, rotation, scaling on visual objects
  - Low bit rate (4.8 to 64kbps for mobile applications) & upto 2Mbps & 10Mbps.
  - Content based manipulation based on Video object planes (VOPs).

MPEG4 interactivities

MPEG1 & 2



MPEG4



In MPEG4, there is entirely new concept of creating video objects for audio-visual scenes.

- BIFS - Binary format for scenes - that facilitates the composition of media objects into a scene.
- Efficient interaction is achieved using among various media objects to achieve synchronised presentation.

Video object based hierarchical description of a scene in MPEG4 visual bitstreams

Video object sequence (VS)
Video object (VO)
Video object layer (VOL)
Group of VOPs (GOV)
Video object plane (VOP)

VS – delivers the complete MPEG4 visual scene – 2D/3D or synthetic

VO – Any arbitrary shaped object in the scene or background of scene

VOL – supports multilayer, scalable coding. VO can have multiple VOLs.

GOV – groups video object planes.

VOP – A snapshot of a VO at any moment. Reflects VO's shape, structure, texture, & motion parameters.

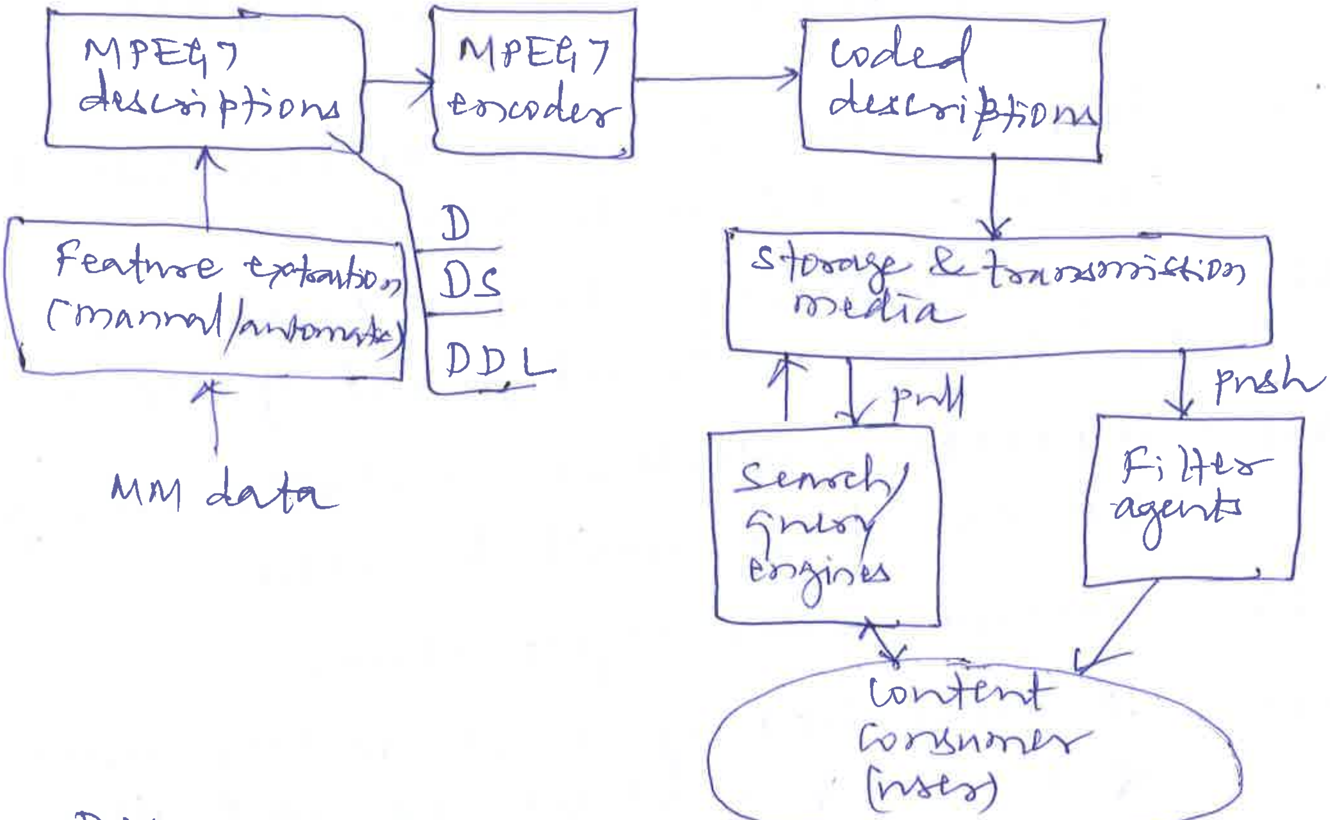
VOP Based Coding

- VOP objects are identified separately. MV's are generated for each object.

MPEG7

- Audio visual content based retrieval
- Generation & usage of multimedia data
- supports multimedia applications
  still pictures, graphics, 3D models, audio, speech, video, composition information.

Possible applications in MPEG7



Descriptor (D) - definition of feature
Low features - color, texture, shape & motion
high features - semantic objects - events & abstract concepts.

Description scheme (DS) : Basic elements, content management, content description, navigation & access, content organization, user interaction.

Basic elements
- Data types & maths structures
- Constructs
- Schema tools

Content management
- media description
- Creation & production description
- Content usage description

Content description
- Structural description
- Conceptual description

Navigation access
- Summaries
- partitions & decompositions

Content organization
- collections
- models

DDL - Description Definition Language

XML Schema structure components
- defns & declarations
- primary structural components - complex type defns, elements declarations
- secondary structural components

XML schema datatype components
- Primitive & derived
- Mechanisms for the user to derive new data types

MPEG7 Extensions
- Array & matrix data types
- multiple media types
- Enumerated data types

MPEG 21 : Multimedia framework

Key elements

- Digital item declaration - flexible schema
- Digital item identification - a framework for standardized identification & description of digital items regardless of their origin, type or granularity.
- Content management & usage
- Intellectual property management & protections - security features
- Terminals & networks - interoperable QoS n/c wide range of networks & terminals.
- Content presentation - to present the content anytime anywhere.
- Event reporting - metrics & interfaces for reporting events.