# 11. VIDEO CODING[1] H.261-to-MPEG4

**General idea:** For the transmission of live video over a band-limited (fixed or variable rate) channel, compression rates need to be significantly higher, which require motion detection and compensation in addition to varying level of compression based on DCT or wavelet-based image processing architectures. There are several standards in this area, starting with H.261 codec through MPEG-4 as of now and others are in progress.

## H.261 Codec

ITU (CCITT)  Recommendation H.261 is a video compression standard based on inter-picture prediction, transform coding, and motion compensation video compression standard developed to facilitate *"videoconferencing and videophone services"* over  ISDN or equivalent (Video over IP) at *px64 kbps; where p=1,...,30.*

- Low-quality videophone with 48 kbps for video and 16 kbps for audio.
- Videoconferencing services require *p=6* or bit rate of 384 kbps or more up to 2.0 MHz.
- H.261 has been the basis for MPEG-1 and MPEG-2 and H.26X series standards has followed.

- To accommodate different TV line standards H.261 has two input picture formats:
    1. Common Intermediate Format (CIF) and for lower-bit rate cases
    2. Quarter CIF (QCIF).

| Picture Formats Supported | | | | | | | |
|---|---|---|---|---|---|---|---|
| Picture format | Luminance pixels | Luminance lines | H.261 support | Uncompressed bitrate (Mbit/s) | | | |
| | | | | 10 frames/s | | 30 frames/s | |
| | | | | Grey | Colour | Grey | Colour |
| QCIF | 176 | 144 | Yes | 2.0 | 3.0 | 6.1 | 9.1 |
| CIF | 352 | 288 | Optional | 8.1 | 12.2 | 24.3 | 36.5 |

- Inter-picture prediction removes temporal redundancy.
- Transform coding removes the spatial redundancy.
- Motion vectors are used to help the codec compensate for motion.
- To remove any further redundancy in the transmitted bit stream, variable length coding is used.
- In performance evaluations the peak signal to noise ratio (PSNR) is used as the measure.

**Some Examples of H.261 coded Video sequences:**

Here are some h261 video sequences. They were H261 coded then recoded with MPEG at high quality to make them viewable. You can use the default MPEG player in your system or an MPEG player is available from Berkeley university however this also only compiles for UNIX, but DOS/Windows versions are available. To download and display the video, just click on the movie icon in the tables below.

---

[1] The material in this chapter has been provided by Professor A. Murat Tekalp at Koc University, Istanbul, Turkey and VCDemo has been provided by Professor Inald Lagendijk of Delft University of Technology, The Netherlands.

**Miss America (QCIF):**



| Description | MPEG Sequence | Average PSNR(dB) | Compression Ratio |
|---|---|---|---|
| Original: Miss America | | n/a | 1:1 |
| High Quality | | 38 | 25:1 |
| Low Quality | | 34.7 | 104:1 |
| Low Quality with motion vectors | | 35 | 113:1 |

- Raw data rate for CIF and QCIF at 30 frames/s are 37.3 Mbps and 9.35 MBps, respectively. Aggressive compression is needed even at medium range of 384 Kbps ISDN channel.
- With QCIF imagery at 10 frames/s, a compression ratio of 48:1 is needed for videophone services over a 64 kbps channel.

**Basics of H. 261 Coding Precedure:**
- Decoded Sequence
- Two frame types: Intraframes (*I-frames*) and Interframes (*P-frames*)
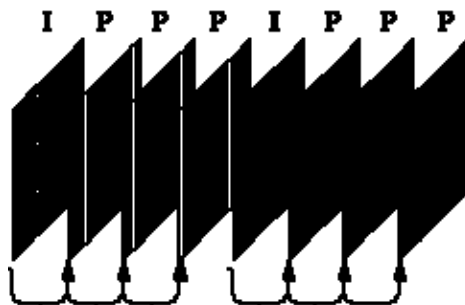


- I-frames use basically JPEG
- P-frames use *pseudo-differences* from previous frame (predicted), so frames depend on each other.
- I-frame provides user an accessing point.

**Intra Frame Coding:** Refers to various lossless & lossy compression techniques are performed relative to information that is contained only within the current frame only.

In other words, no temporal processing is performed outside of the current picture or frame. This mode will be described first because it is simpler, and because non-intra coding techniques are extensions to these basics.
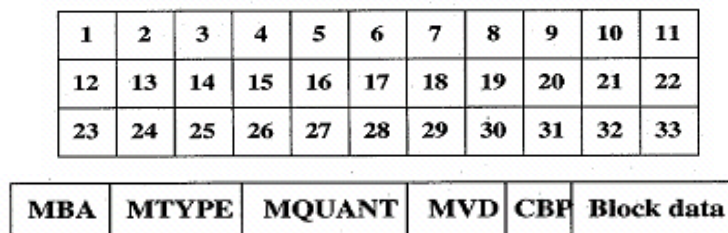
The following block diagram highlights a basic video encoder for intra frames only. It turns out that this block diagram is very similar to that of a JPEG still image video encoder, with only slight implementation detail differences.
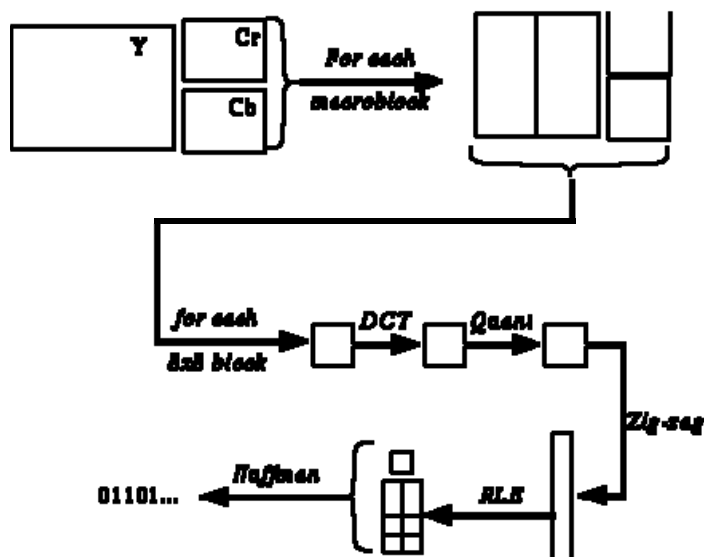


The basic processing blocks shown are the video filter, discrete cosine transform, DCT coefficient quantizer, and run-length amplitude/variable length coder. These blocks are described individually in the sections below or have already been described in JPEG Compression.

- **Videodata** is divided in a hierarchical structure consisting of a picture-level, which is divided into several Group-of-Blocks (GOB) layers. Each GOB layer is composed of macroblocks (MB); which are made up of blocks of pixels.
- **Macroblock:** Smallest unit of data for selecting a compression mode. It has four 8x8 pixels of Y (luminance) and the spatially corresponding 8x8 U and V (chrominance) blocks. There are one U & one V blocks for every four Y blocks.
- **GOB-layer** is always composed of 33 macroblocks; each MB has a header with a MB address and the compression mode, followed by the data for the blocks.

Arrangement of microblocks (MB) in a GOB and Structure of an MB layer

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|----|----|
| 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 |

| MBA | MTYPE | MQUANT | MVD | CBP | Block data |
|-----|-------|--------|-----|-----|------------|

1. MBA: Header with MB Address.
2. MTYPE: Compression modes including intraframe, interframe with zero motion vector, motion-compensated interframe, and motion-compensated interframe with loop-filtering.
3. MQUANT: Quantizer step size.

4. MVD: Motion vector data to be transmitted as side-information.
5. CBP: Coded block pattern; a pattern number signifying blocks in the MB for which at least one transform coefficient is transmitted and TCOEFF for the transform coefficients that are encoded. Finally, a VLC (variable-length code) that identifies the compression mode in the MB header.

- Variable thresholding is used to increase the number of zero coefficients before quantizing.
- Scalar uniform quantizers are used for all coefficients except DC, which is linearly quantized with a step-size 8.
- Zigzag scanning followed by an entropy coding is used for coding efficiency.
- Quantizer steps are adjusted according to buffer fullness since a max. coding delay is 150 ms.
- There are other features to combat overflow of errors, ICDT accuracy, etc.

# MPEG FAMILY VIDEO COMPRESSION STANDARDS

There are two groups of MPEG standards:
- ISO Standards: MPEG-1, MPEG-2, MPEG-4
- ITU-T Standards
  - Recommendation H.263
  - Recommendation H.263+
  - Recommendation H.263++

## MPEG-1 (ISO/IEC-11172):

It is a standard developed for storage of CIF format video and its audio at about 1.5 Mbps on CD-ROM, DAT, HD disks, optical drives and interactive multimedia systems. Some development history and requirements:
- MPEG, part of ISO IEC/JTC1/SC29/WG11, started 1988
- VHS quality video at 1-1.5 Mbit/s for storage on CD-ROM
- Oct. 1989: Competitive tests for video coding and collaborative phase soon after, for standards development
- Sept. 1990: video part becomes Committee Draft; IS in May 1993 and bitstream syntax & decoder defined, which consists of: ISO/IEC 11172-1: MPEG-1 Systems; ISO/IEC 11172-2: MPEG-1 Video

## Rates, Picture Format and Data Structure:

1. Maximum pixels/line                = 720
2. Maximum lines/picture              = 576
3. Maximum pictures/second            = 30
4. Maximum macroblocks/picture        = 396
5. Maximum macroblocks/second         = 9,900
6. Maximum Bit Rate                   = 1.8 Mbps
7. Maximum decoder buffer size        = 376,832 bits.
   - Y, Cb, Cr as noninterlaced 4:2:0; size as big as 4Kx4K
   - SIF 352x240; rates of 23.97, 24, 25, 29.97, 50, 59.94, 60 Hz
   - Layers: Group of Pictures; Picture; Slice, Macroblock, and Blocks

## MPEG 1 - Data Structure:

1. Sequences, made up of group of pictures
2. Group of pictures (GOP), made up of pictures

3.  Pictures, made up of slices
    I-pictures: intra coded
    P-pictures: relative to a preceding I or P picture
    B-pictures: forward, backward or bi-directional
    D-pictures: contain only the DC component of each block
4.  Slices, made up of macroblocks
    Introduced mainly for error recovery.
5.  Macroblocks, made up of blocks
    A MB consists of 4 Y, 1 Cr and 1 Cb blocks (same as H.261 MB)
6.  Blocks, $8 \times 8$ sample array

## Group of Pictures:

*   A GOP may contain I, P and B pictures
*   N pictures in a GOP; Number of B-pictures between consecutive anchor pictures is M-1, where M is the prediction distance
*   Processing order is different than the display order (in case of B pics)
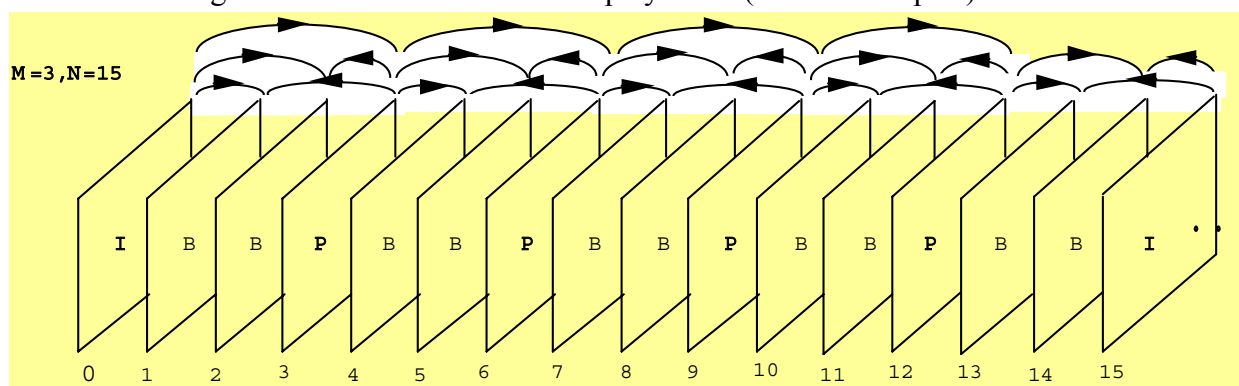


## B Frames:

MPEG encoder has the option of using forward/backward interpolated prediction. These frames are referred to as bi-directional interpolated prediction frames (B).

*   There is no defined limit to the number of consecutive B frames that may be used in a GOB, and the optimal number is application dependent.
*   Most broadcast quality applications however, have tended to use 2 consecutive B frames (I,B,B,P,B,B,P,Š) as the ideal trade-off between compression efficiency and video quality.



These notes are © Hüseyin Abut, August 2006
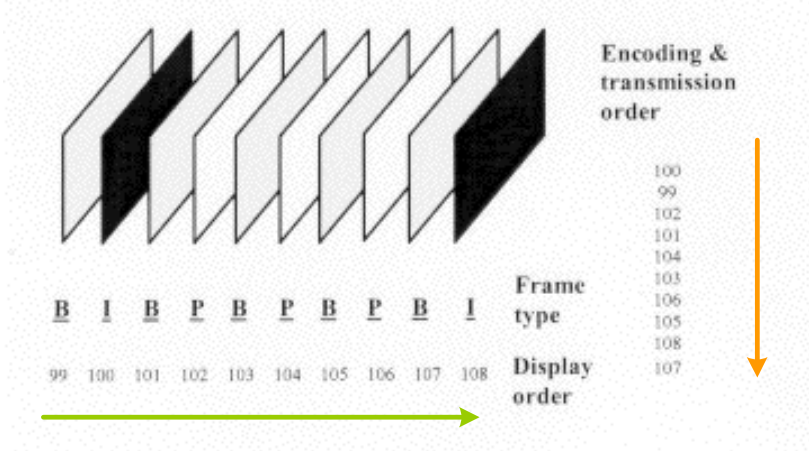
**Motion Estimation:** The temporal prediction technique used in MPEG video is based on motion estimation. In most cases, consecutive video frames will be similar except for changes induced by objects moving within the frames. In the trivial case of zero motion between frames (and no other differences caused by noise, etc.), it is easy for the encoder to efficiently predict the current frame as a duplicate of the prediction frame When there is motion in the images, the situation is not a simple case to deal with.

To achieve compression, the temporal redundancy between adjacent frames can be exploited. That is, a frame is selected as a reference, and subsequent frames are *predicted* from the reference. The process of video compression using motion estimation is also known as interframe coding.

When using motion estimation, an assumption is made that the objects in the scene have only translational motion. This assumption holds as long as there is no camera pan, zoom, changes in luminance, or rotational motion (quite an assumption!). However, for scene changes, interframe coding does not work well, because the temporal correlation between frames from different scenes is low. In this case, a second compression technique is used, known as intraframe coding.

**Example:** Frame 1260 of the "Jane" Sequence. Below are two sets of images. Each set shows a frame that has been reconstructed by using a reference frame that is 1, 5, or 10 frames prior to the current frame. The original image is also shown.



Original Frame



Reference frame is 1 frame away from current



Reference frame is 5 frames away from current



Reference frame is 10 frames away from current

**Frame 1280 of the "Jane" Sequence**



Original Frame



Reference frame is 1 frame away from current



Reference frame is 5 frames away from current



Reference frame is 10 frames away from current

**Note:** You can see the errors better if you look closely around Jane's eyes and mouth. This is more obvious in the images for frame 1280 of the sequence. This is because the reference frame for these reconstructed images was very different from the current frame. For example, Jane may have gone from speaking a word with her lips open to lips closed or she may have blinked her eyes.

**Example:**
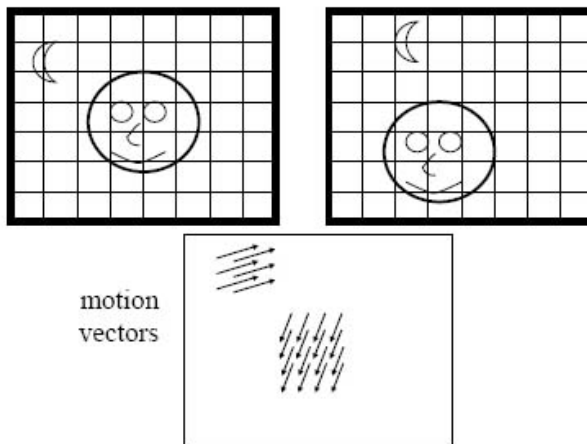
**Final Motion Estimation Prediction:** Not every macroblock search will result in an acceptable match. If the encoder decides that no acceptable match exists (again, the "acceptable" criterion is not MPEG defined, and is up to the system designer) then it has the option of coding that particular macroblock as an intra macroblock, even though it may be in a P or B frame. In this manner, high quality video is maintained at a slight cost to coding efficiency.

**What Happens During Reconstruction? Motion Compensation:** During reconstruction, a reference frame is used to predict the current one using the motion vectors and called motion compensation. During motion compensation, the macroblock in the reference frame that is referenced to by the motion vector, is copied into the reconstructed frame. The difference values are added to a reconstructed area from the previous frame. The motion vector information is used to pick the correct area (the same reference area that was used in the encoder). The result is a reconstruction of the "original frame". Of course, this is not identical to the original because of the "lossy" quantization stage, i.e. the image quality will be poorer than the original. The reconstructed frame is placed in a frame store and it is used to motion-compensate the next received frame.
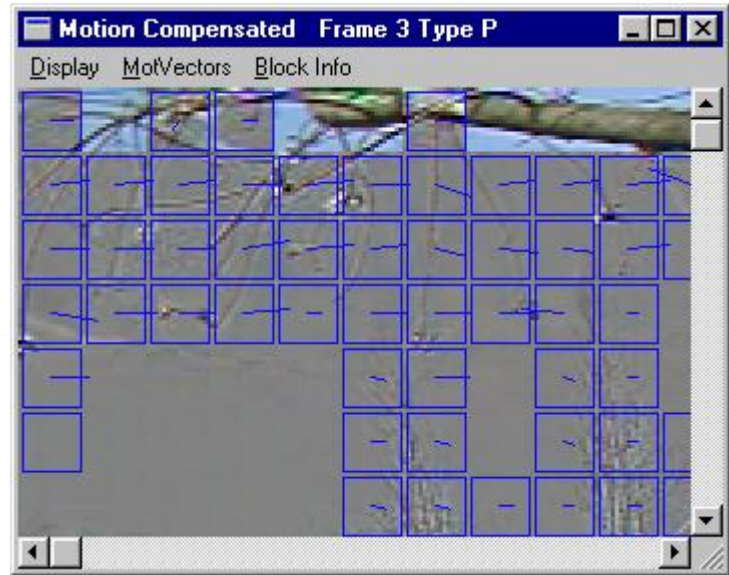


**Image 1**

This image represents an image that has been carved up into about 300 squares to perform the compressions mentioned above. This image is part of a well known test motion picture sequence that depicts passage from left to right along a row of houses fronted by a strip of colorful flowers. The gray patches represent areas that were found in the previous frame and shifted to new positions to form this frame. A uniform gray patch indicates an exactly matching area was found. Mottled gray patches indicate a near match. This operation is performed to compress temporal redundancy, the nearly complete replication of the same picture in two adjacent frames of a sequence.

Note that close matches for some squares could not be found in the previous frame. There was no close match for the new image detail that shifted into view on the right edge or for the tree that is passing by rapidly in the foreground. The process of finding the squares in another picture is called motion compensation and is very computationally intensive. Applying motion compensation to reconstruct an compressed MPEG sequence is, by contrast, a much easier task. This characteristic is called asymmetrical compression or encoding: hard to compress but easy to decompress or decode.

**Image 2**

A blowup of a section of **Image 1** with the blue lines (called motion vectors) indicating the exact location of the best matching group of pixels in the previous frame used in motion compensation. This motion vector (maybe 4 bits of data) is recorded in the MPEG stream instead of the group of pixels (2048 bits).



The original picture with a tremendous amount of detail and color has been transformed to the representation shown in **Image 1** with the first compression step. It is important to notice that resultant picture is now nearly one uniform color (the gray) with some finely mottled detail.

# MPEG Block-Based Motion Compensation



GOP - Group of Pictures

I frame - Independent encoding

P frame - Predictive encoding based on most recent past I or P

B frame - Bidirectional encoding based on past and future I or P

( Each frame is encoded as a series of 16 x 16 pixel macroblocks )
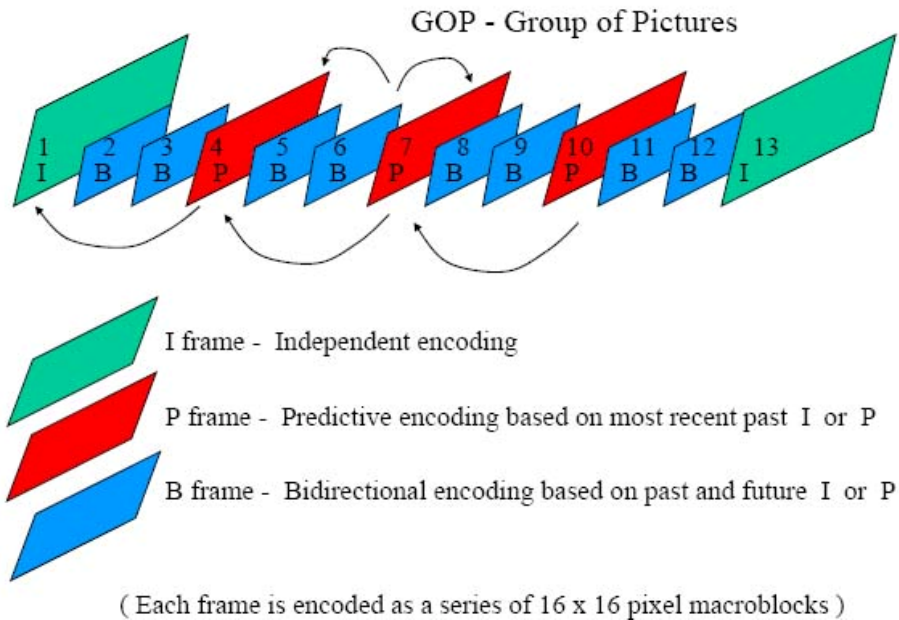
**MPEG 1 - MB Coding Modes:**

| I Pictures | P Pictures | B-Pictures |
|---|---|---|
| Intra-d | Intra-d | Intra-d |
| Intra-q | Intra-q | Intra-q |
| | Pred-c | Pred-i |
| | Pred-cq | Pred-ic |
| | Pred-m | Pred-icq |
| | Pred-mc | Pred-b |
| | Pred-mcq | Pred-bc |
| | Skipped | Pred-bcq |
| | | Pred-f |
| | | Pred-fc |
| | | Pred-fcq |
| | | Skipped |

P-Pictures:
"c" indicates that DCT of the prediction error will be coded.
"m" indicates that MC is ON.
"q" indicates that adaptive quantization will be used.

B-Pictures:
"f" indicates forwards prediction with motion-compensation.
"b" indicates backward prediction with motion-compensation.
"i" indicates interpolated prediction with bi-directional Motion compensation (MC).

**Quantization:**
- Visual Weighting

| Intra Macroblocks $W_I(i,j)$ | | | | | | | | NonIntra Macroblocks $W_{NI}(i,j)$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | 16 | 19 | 22 | 26 | 27 | 29 | 34 | 16 | 17 | 18 | 19 | 21 | 23 | 25 | 27 |
| 16 | 16 | 22 | 24 | 27 | 29 | 34 | 37 | 17 | 18 | 19 | 21 | 23 | 25 | 27 | 29 |
| 19 | 22 | 26 | 27 | 29 | 34 | 34 | 38 | 18 | 19 | 20 | 22 | 24 | 26 | 28 | 31 |
| 22 | 22 | 26 | 27 | 29 | 34 | 37 | 40 | 19 | 20 | 22 | 24 | 26 | 28 | 30 | 33 |
| 22 | 26 | 27 | 29 | 32 | 35 | 40 | 48 | 20 | 22 | 24 | 26 | 28 | 30 | 32 | 35 |
| 26 | 27 | 29 | 32 | 35 | 40 | 48 | 58 | 21 | 23 | 25 | 27 | 29 | 32 | 35 | 38 |
| 26 | 27 | 29 | 34 | 38 | 46 | 56 | 69 | 23 | 25 | 27 | 29 | 31 | 34 | 38 | 42 |
| 27 | 29 | 35 | 38 | 46 | 56 | 69 | 83 | 25 | 27 | 29 | 31 | 34 | 38 | 42 | 47 |

- Forward Quantization

| **Intra Macroblocks** | **Non-Intra Macroblocks** |
|---|---|
| $Q_{dc} = X_{dc}/8$ <br> $\hat{X}_{ac}(i,j) = (16 * X_{ac}(i,j))/W_I(i,j)$ <br> $Q_{ac}(i,j) = [(\hat{X}_{ac}(i,j) + Sign(X(i,j)) * mquant]/(2 * mquant)$ <br> $Q_{ac}(i,j)$ is restricted to $:[-255, \cdots 0, \cdots, 255]$ | $\hat{X}_{ac}(i,j) = (16 * X_{ac}(i,j))/W_{NI}(i,j)$ <br> $Q_{ac}(i,j) = \hat{X}_{ac}(i,j)/(2 * mquant)$ <br> $Q_{ac}(i,j)$ is restricted to $:[-255, \cdots 0, \cdots, 255]$ |

**Coding of I Pictures:**
- DC Prediction and DCT coefficients are computed with 11 bits accuracy, i.e., the DC coefficient is in the range [0,2047] & AC coefficients are in the range [-1024,1023].
- Quantized DC coefficient is represented with 8 bits since its weight in the QM is always 8. The AC coefficients are represented with less than 8 bits by using weights larger than 8.
- Spatially-Adaptive Quantization:

    – Intra-d are coded with the current quantization matrix.
    – Intra-q are coded with a scaled quantization matrix.
      mquant can be determined on the basis of spatial activity (e.g., MBs, which contain textured areas are coarsely quantized).

## Coding of P or B Pictures:

- Quantization matrix is such that the effective quantization is relatively coarser compared to those used for I frames.
- All DCT coefficients, *including the DC coefficient*, are zig-zag scanned to form [run, level] pairs which are then coded using VLC.
- VLC tables are needed for the type of MB, the differential motion vector, and the MB prediction error. There is no separate DC code table; instead, the differential DC value and AC coefficients are coded together.
- Displacement vectors are DPCM encoded.

## Rate Control:

- Global Target bit  allocation and update for each picture type
  - Initialize bits for I-, P- and B-pictures
  - Code each picture in GOP and update target for each picture type
  - Subtract excess bits used in this GOP from budget for next GOP
- Tracking local bit generation behavior within each frame
  - Update content of virtual buffer of each picture assuming linear rate
  - Based on the fullness of virtual buffer derive linear rate control based quantizer
- Mquant  based on local scene content and bits generated
  - Compute spatial activity such as variance of original macroblock
  - Compute normalized spatial activity of macroblock wrt entire picture
  - Obtain mquant values based on normal activity and linear rate control

## MPEG-2 STANDARD (ISO/IEC-13818)

- Interlaced Video at 4-15 Mbit/s; Digital TV, Cable/Satellite TV, DVD, Video on ATM, HDTV (15-30 Mbit/s)
- Nov. 1991: Competitive tests for video, Nov. 1993: Video part, stable as the Committee Draft
- Standard specifies bitstream syntax and decoding semantics
- MPEG-2 standard mainly consists of
  - ISO/IEC 13818-1: MPEG-2 Systems, ISO/IEC 13818-2: MPEG-2 Video
  - ISO/IEC 13818-3: MPEG-2 Audio; ISO/IEC 13818-7: AAC Audio
  - ISO/IEC 13818-4: MPEG-2 Conformance, ISO/IEC l3818-5: MPEG-2 Software
  - ISO/IEC 13818-6: MPEG-2 DSM-CC

## Summary of Key Features in MPEG-2

- supports frame and field picture types for interlaced video
- allows 4:2:2 and 4:4:4 chroma in addition to the 4:2:0 format.
- supports new MC prediction modes for interlaced video
- supports field/frame DCT option per MB for frame pictures
- allows for finer quantization of the DCT coefficients.
- allows for finer adjustment of the quantizer scale factor.
- allows for a separate VLC table for the DCT coefficients for the intra macroblocks.
- allows alternate scan in addition to the zigzag scan.
- supports scalability/backward compatibility/error resilience

**MPEG-2 target applications:**

```
-------------------------------------------------------------------
Level        size      Pixels/sec   bit-rate(Mb)      Application
-------------------------------------------------------------------
Low          352 x 240       3 M           4        Consumer tape equiv.
Main         720 x 480      10 M          15        Studio TV
High 1440  1440 x 1152      47 M          60        Consumer HDTV
High       1920 x 1080      63 M          80        Film production
-------------------------------------------------------------------
```

**Interlaced and Non-interlaced Formats:**
– Vertical/temporal position of samples in 4:2:0, 4:2:2, 4:4:4

**MC Prediction Modes for Interlaced Video:**
- Frame Pictures:
    – Frame-based prediction mode (P or B type, same as progressive video)
    – Field-based prediction mode
    – Dual prime prediction mode

It is possible to switch between frame/field/dual prime prediction modes within the same frame picture on a macroblock basis.

- Field Pictures:
    – Two types of field-based prediction modes
    – Dual prime prediction mode

**Scanning Option Alternatives:**
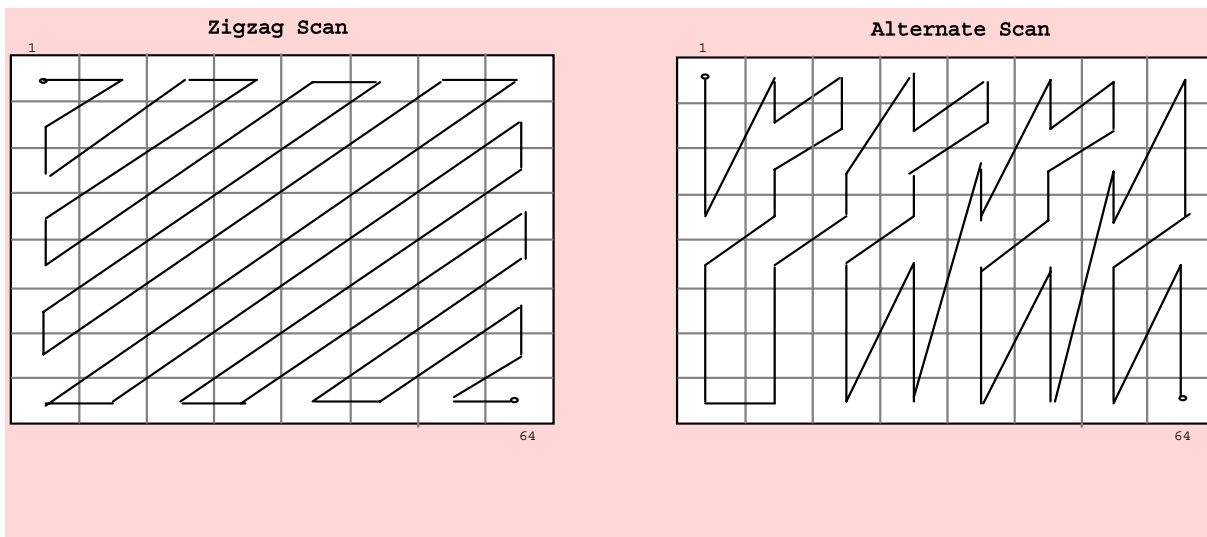


**Finer Quantization of DCT Coefficients:**
- Intra Macroblocks
    – The quantization weight for the DC coefficient can be 8, 4, 2 or 1, i.e., 11 bits (full) resolution is allowed for the DC coefficient.
    – AC coefficients are quantized into the range [-2048,2047].
- Non-Intra Macroblocks
    – All coefficients are quantized into the range [-2048,2047].
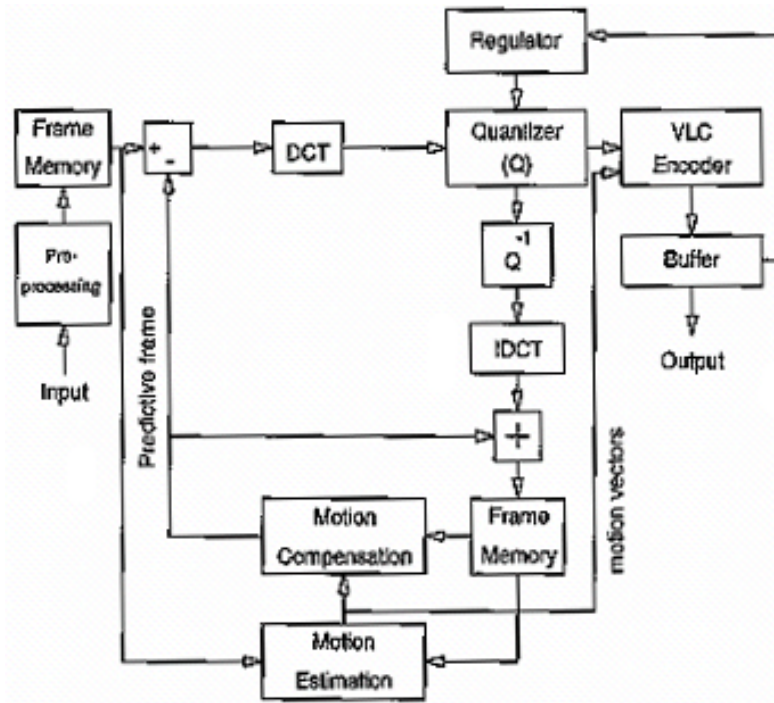- Finer Adjustment of *mquant*.

**Scalable Video:**
- Data Partitioning: Single layer coded bitstream partitioned into layers, such that more important data is assigned higher priority.
- SNR Scalability: 2 layers with lower layer coded by itself and higher layer coded with respect to lower layer; both layers with same spatial resolution but different qualities.
- Spatial Scalability: 2 layers with lower layer coded by itself and the higher layer coded with respect to lower layer; the lower layer has lower spatial resolution then higher layer.
- Temporal Scalability: 2 layers with lower layer coded by itself and the higher layer coded with respect to lower layer; lower and higher layer are both at lower temporal resolution and are temporally multiplexed for display.

**Summary MPEG-2 Profiles and Levels:** The **encoder** and **decoder** for MPEG family

| Level | | SIMPLE (nonscalable 4:2:0, no B- pictures) | MAIN (nonscalable 4:2:0) | SNR (scalable 4:2:0) | SPATIAL (scalable 4:2:0) | HIGH (nonscalable 4:2:2, scalable 4:2:0/4:2:2) |
|---|---|---|---|---|---|---|
| HIGH | | | 1920 pels/line<br>1152 lines/frame<br>60 frames/s<br>62.7 Msamples/s<br>80 Mbit/s | | | 1920 pels/line<br>1152 lines/frame<br>60 frames/s<br>62.7 Msamples/s @<br>83.5 Msamples/s *<br>100 Mbit/s for 3 layers |
| HIGH-1440 | | | 1440 pels/line<br>1152 lines/frame<br>60 frames/s<br>47.0 Msamples/s<br>60 Mbit/s | | 1440 pels/line<br>1152 lines/frame<br>60 frames/s<br>47.0 Msamples/s<br>60 Mbit/s for 3 layers | 1440 pels/line<br>1152 lines/frame<br>60 frames/s<br>47.0 Msamples/s @<br>62.7 Msamples/s *<br>80 Mbit/s for 3 layers |
| MAIN | | 720 pels/line<br>576 lines/frame<br>30 frames/s<br>10.4 Msample/s<br>15 Mbit/s | 720 pels/line<br>576 lines/frame<br>30 frames/s<br>10.4 Msample/s<br>15 Mbit/s | 720 pels/line<br>576 lines/frame<br>30 frames/s<br>10.4 Msample/s<br>15 Mbit/s for 2 layers | | 720 pels/line<br>576 lines/frame<br>30 frames/s<br>11.06 Msamples/s @<br>14.75 Msamples/s *<br>20 Mbit/s for3 layers |
| LOW | | | 352 pels/line<br>288 lines/frame<br>30 frames/s<br>3.04 Msamples/s<br>4 Mbit/s | 352 pels/line<br>288 lines/frame<br>30 frames/s<br>3.04 Msamples/s<br>4 Mbit/s for 2 layers | | |

\* refers to 4:2:0
@ refers to 4:2:2

Profi ⟶

These notes are © Hüseyin Abut, August 2006

MPEG CODEC Block Diagram


**Example: VCDemo** from Delft University of Technology, Holland.