Week 14

Video Compression

Ref: Fundamentals of Multimedia

Last lecture review

- Prediction from the previous frame is called forward prediction
- Prediction from the next frame is called forward prediction
- P-frames: Predicted from previous p-frame or I-frame
- B-frame: bi-directionally predicted from the I or P frames
- B-frames are never used for prediction.

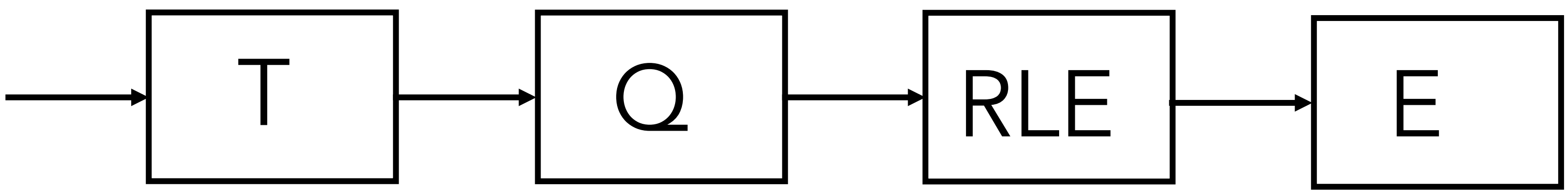
What is video?

A sequence of images displayed at a high frequency to give illusion of motion!

Observation 1: Everything we observed for images!

- Color redundancy
- Spatial redundancy
- Spectral redundancy

Transform Coding

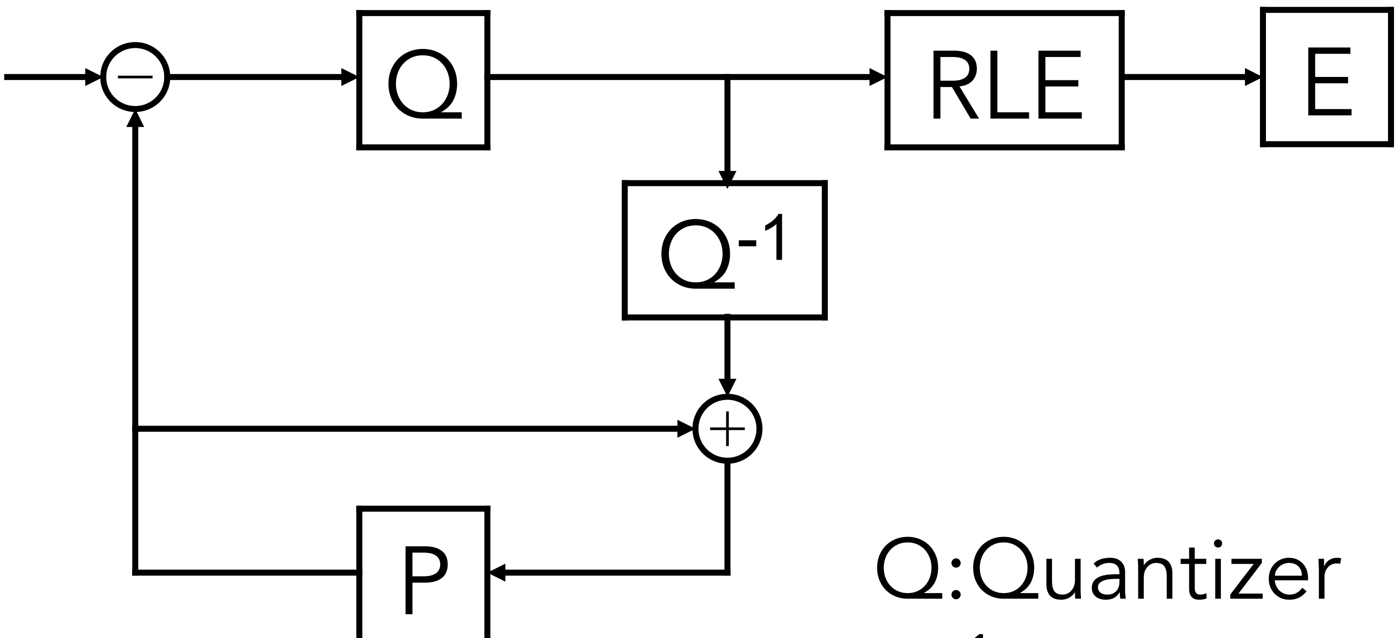


T: Transform
Q: Quantizer
RLE: Run Length Encoding
E: Entropy coder

Observation 2: Content of consecutive frames is mostly similar!

Temporal redundancy!

Predictive Coding

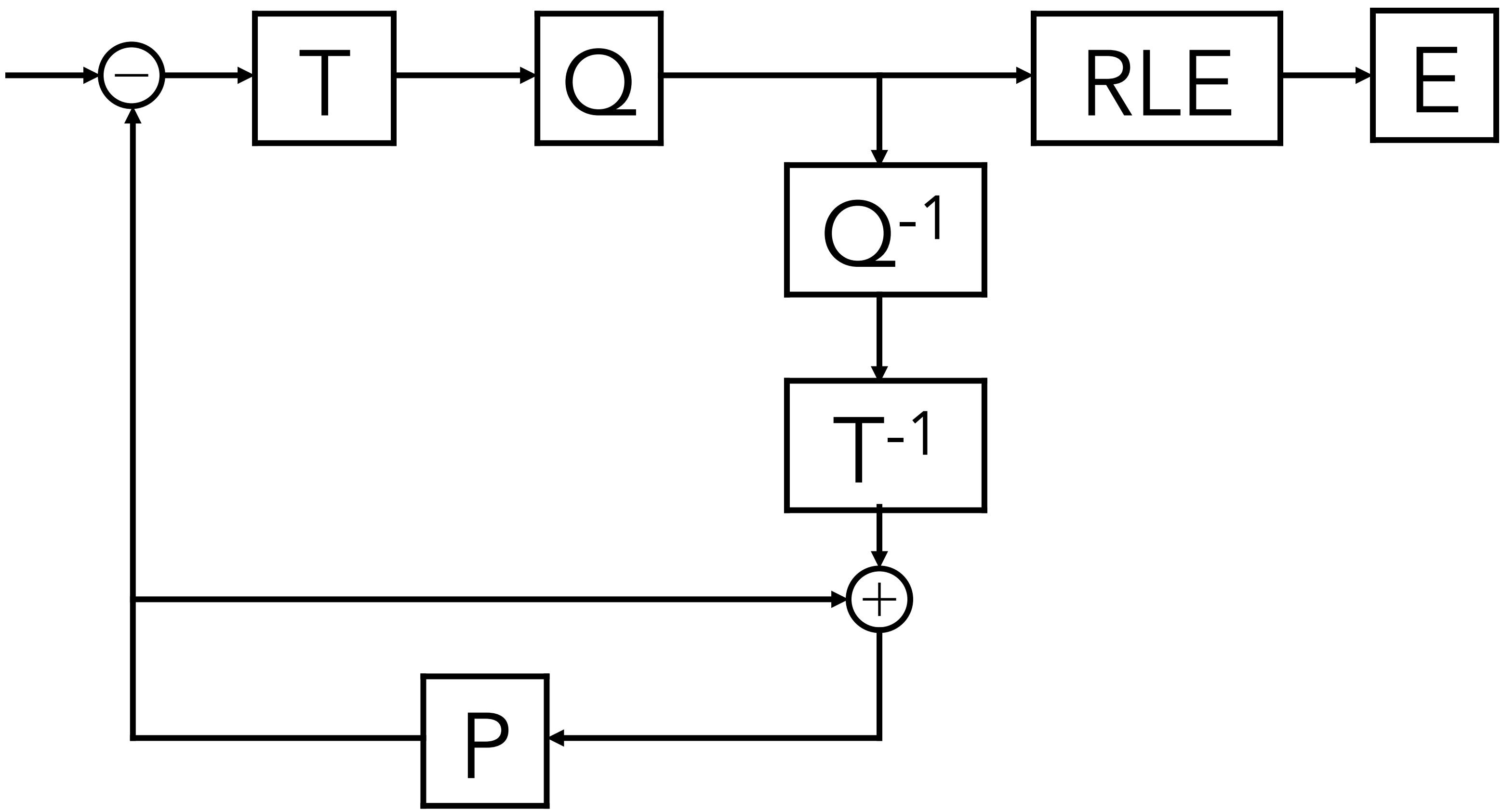


Q:Quantizer

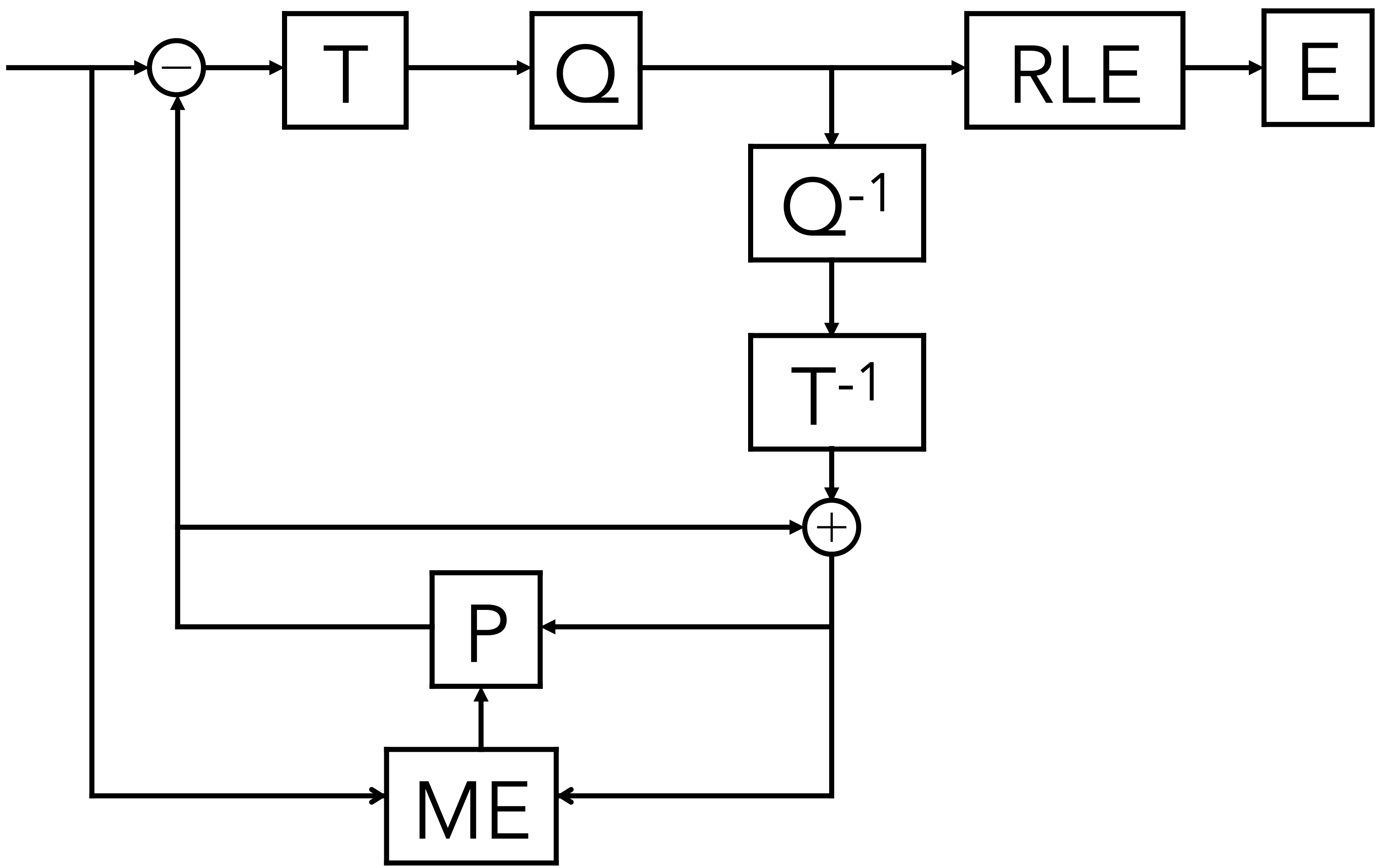
Q^{-1}: Inverse quantizer

P: Predictor

Hybrid Coding



Hybrid Coding



Video Coding Standards

- H.261
- H.263
- MPEG 1/2/4
- H.264

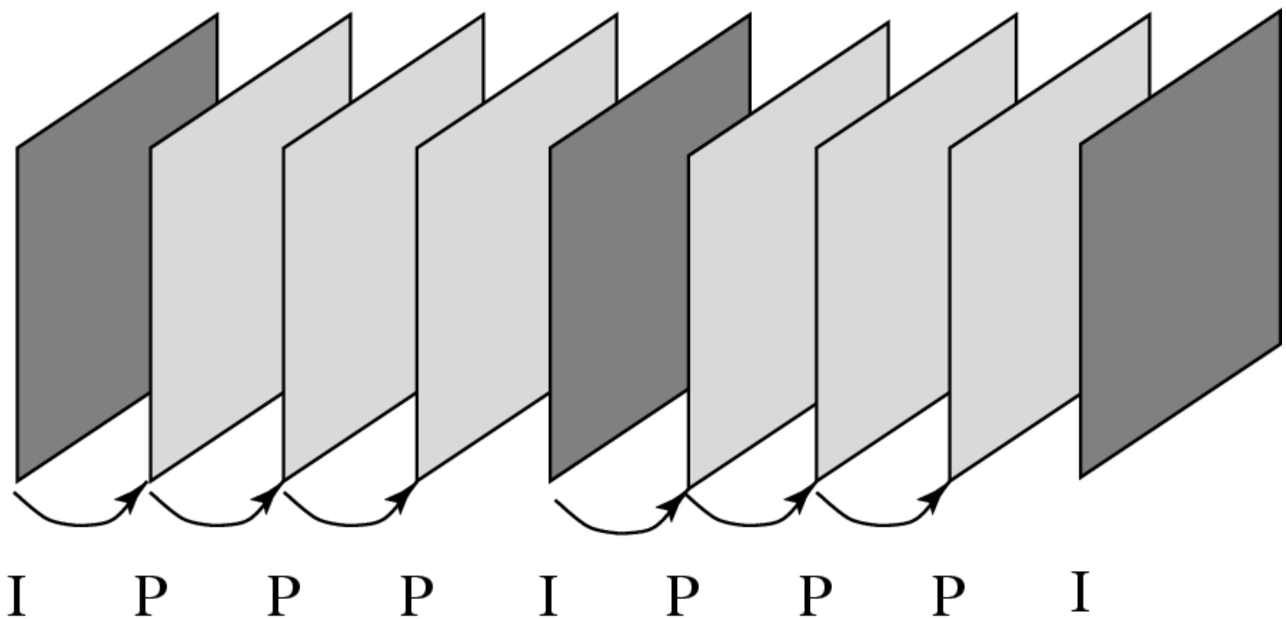
H.261

- The standard was designed for videophone, video conferencing.
- The video codec supports bit-rates of p x 64 kbps, where p ranges from 1 to 30 (Hence also known as p * 64).
- Require that the delay of the video encoder be less than 150 msec so that the video can be used for real-time bidirectional video conferencing.

Video Formats Supported by H.261

Video format	Luminance image resolution	Chrominance image resolution	Bit-rate (Mbps) (if 30 fps and uncompressed)	H.261 support
QCIF	176×144	88×72	9.1	required
CIF	352×288	176×144	36.5	optional

H.261 Frame Sequence

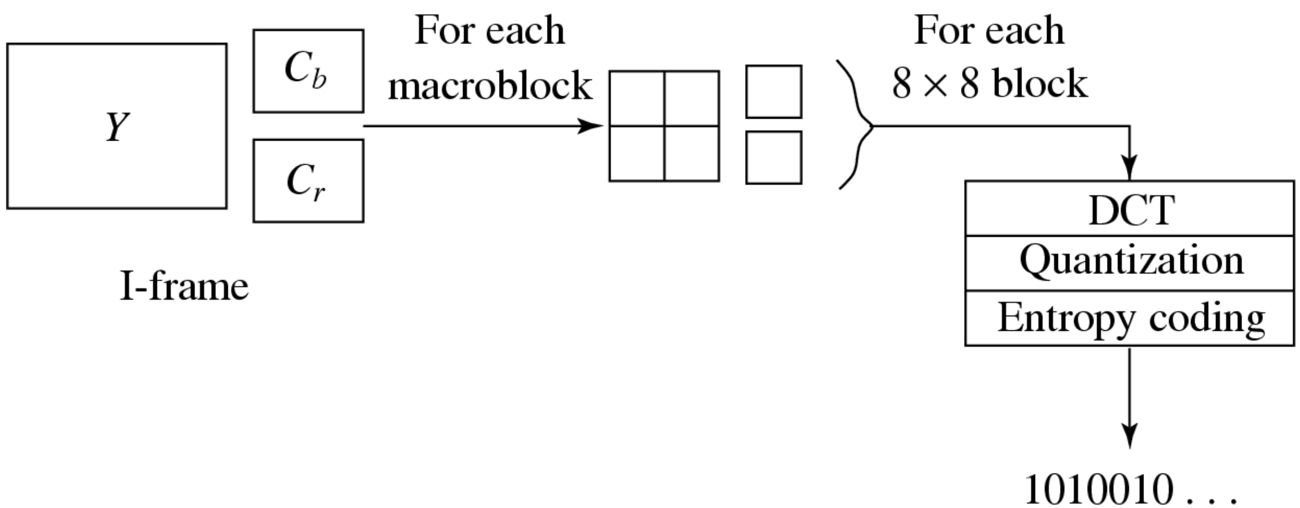


H.261 Frame Sequence

- Two types of image frames are defined: Intra-frames (I-frames) and Inter-frames (P-frames):
  - I-frames are treated as independent images. Transform coding method similar to JPEG is applied within each I-frame, hence "Intra".
  - P-frames are not independent: coded by a forward predictive coding method (prediction from a previous P-frame is allowed — not just from a previous I-frame).
  - Temporal redundancy removal is included in P-frame coding, whereas I-frame coding performs only spatial redundancy removal.
  - To avoid propagation of coding errors, an I-frame is usually sent a couple of times in each second of the video.

Motion vectors in H.261 are always measured in units of full pixel and they have a limited range of ± 15 pixels, i.e., $p = 15$!

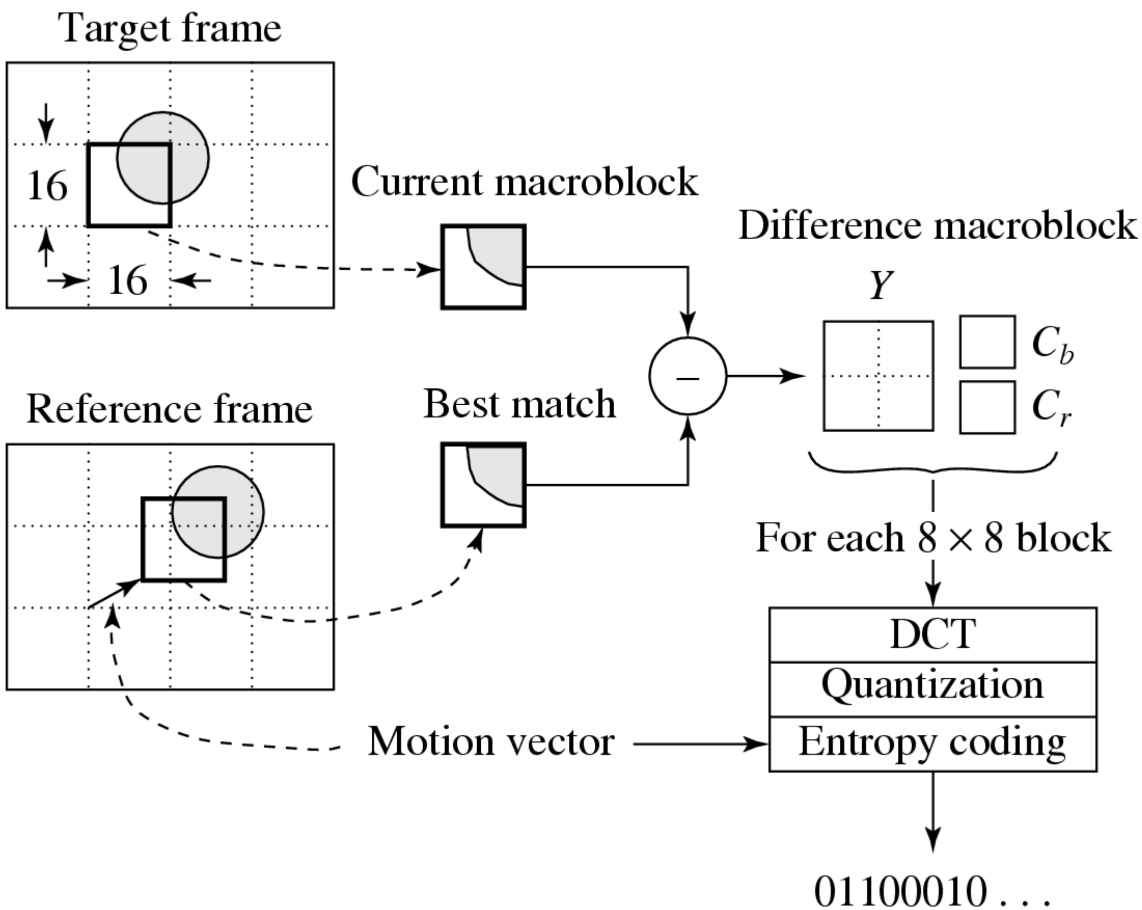
Intra-frame (I-frame) Coding



I-frame Coding.

- Macroblocks are of size 16 x 16 pixels for the Y frame, and 8 x 8 for Cb and Cr frames, since 4:2:0 chroma subsampling is employed. A macroblock consists of four Y, one Cb, and one Cr 8 x 8 blocks.
- For each 8 x 8 block a DCT transform is applied, the DCT coefficients then go through quantization zigzag scan and entropy coding.

P-frame Coding Based on Motion Compensation



Inter-frame (P-frame) Predictive Coding

- For each macroblock in the Target frame, a motion vector is allocated by one of the search methods discussed earlier.

- After the prediction, a *difference macroblock* is derived to measure the *prediction error*.

- Each of these 8 x 8 blocks go through DCT, quantization, zigzag scan and entropy coding procedures.

What if no good match is found?

- Sometimes, a good match cannot be found, i.e., the prediction error exceeds a certain acceptable level.
- The MB itself is then encoded (treated as an Intra MB) and in this case it is termed a *non-motion compensated MB*.

For a motion vector, the difference **MVD** is sent for entropy coding:

$$\mathbf{MVD} = \mathbf{MV}_{\mathbf{Preceding}} - \mathbf{MV}_{\mathbf{Current}}$$

Quantization in H.261

- The quantization in H.261 uses a constant *step_size*, for all DCT coefficients within a macroblock.
- If we use *DCT* and *QDCT* to denote the DCT coefficients before and after the quantization, then for DC coefficients in Intra mode:

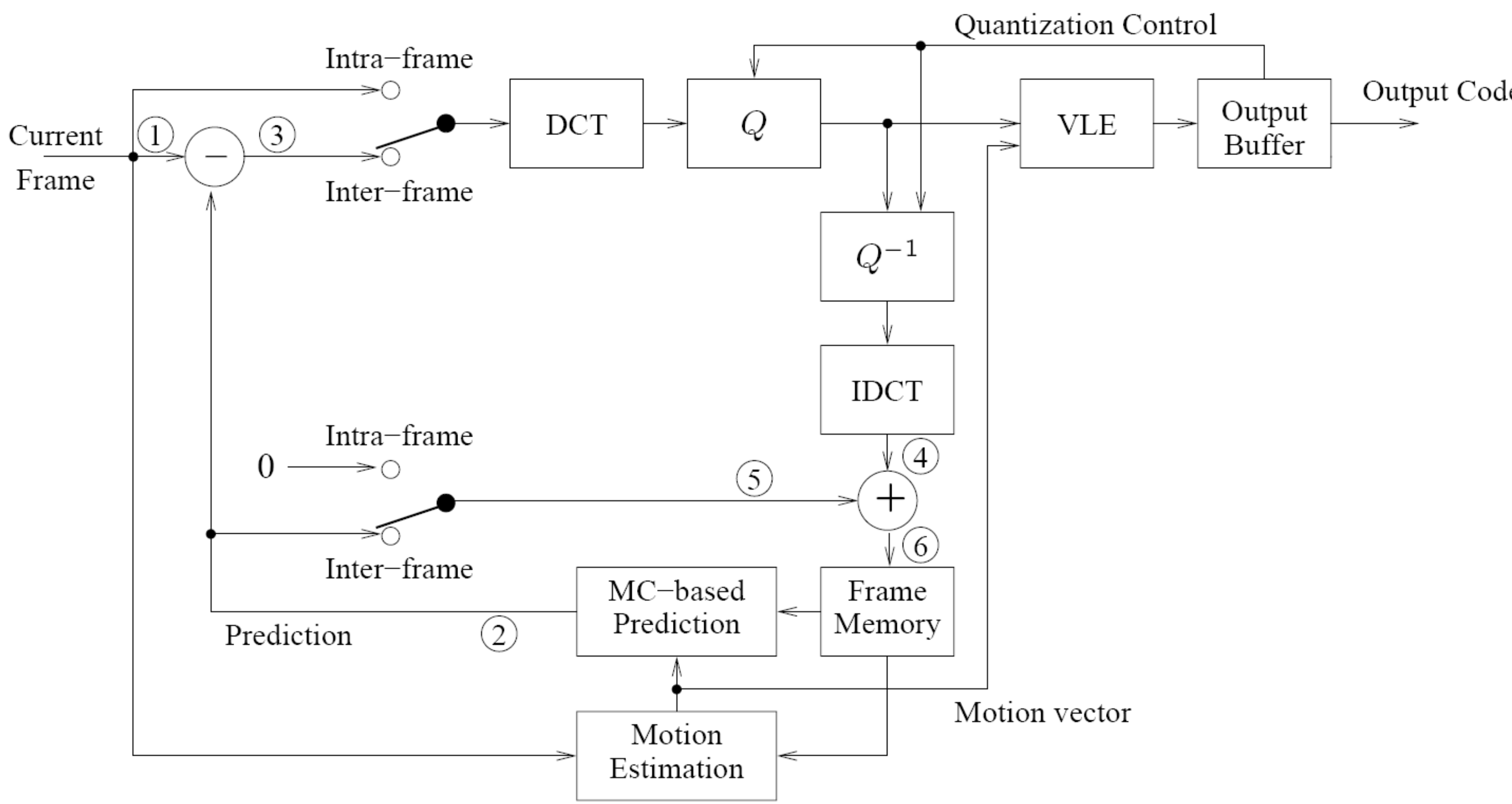
$$QDCT = round\left(\frac{DCT}{step\_size}\right) = round\left(\frac{DCT}{8}\right)$$

for all other coefficients:

$$QDCT = \left\lfloor \frac{DCT}{step\_size} \right\rfloor = \left\lfloor \frac{DCT}{2*scale} \right\rfloor$$

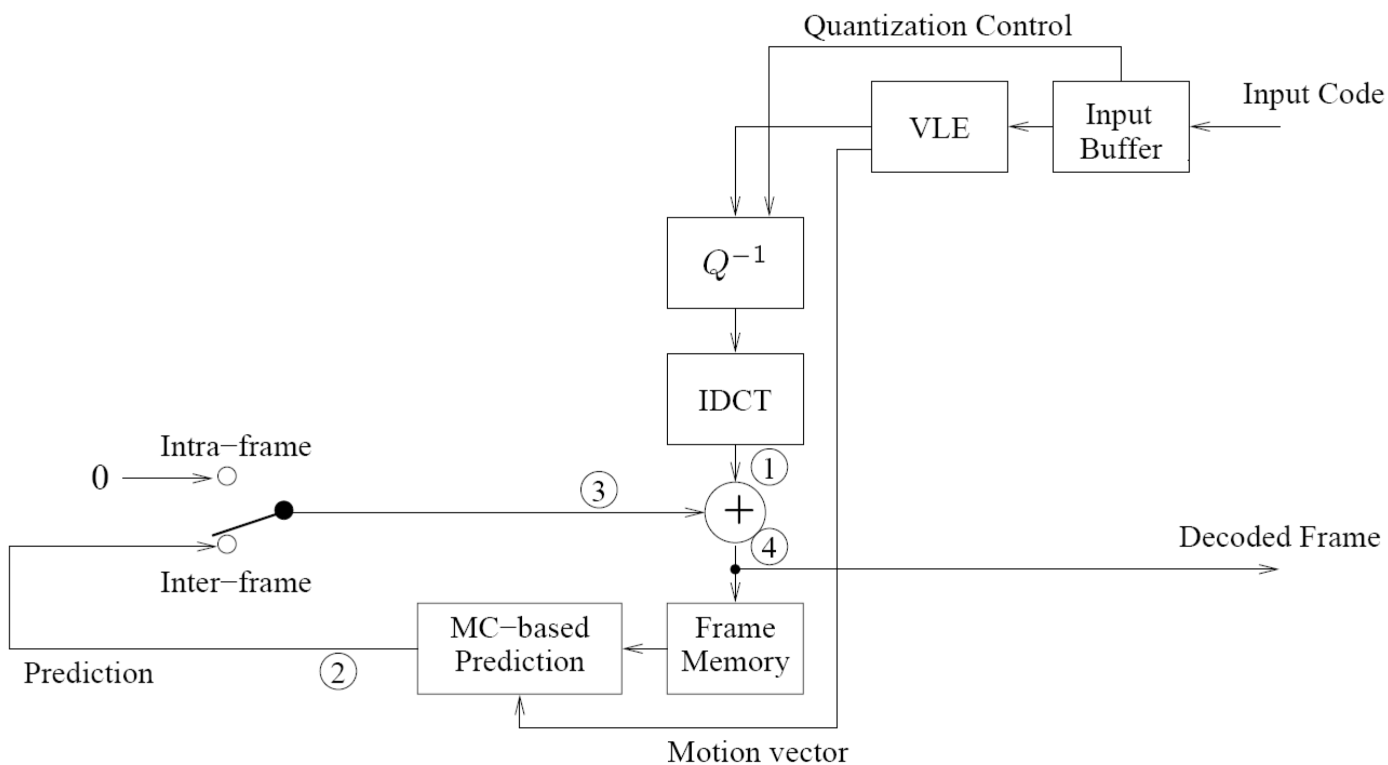
scale — an integer in the range of [1, 31].

Encoder and Decoder



(a) Encoder

Decoder



(b) Decoder

Data Flow at the Observation Points in H.261 Encoder

Current Frame	Observation Point					
	1	2	3	4	5	6
I	I			$\tilde{I}$	0	$\tilde{I}$
P_1	P_1	P_1'	D_1	$\tilde{D}_1$	P_1'	$\tilde{P}_1$
P_2	P_2	P_2'	D_2	$\tilde{D}_2$	P_2'	$\tilde{P}_2$

Data Flow at the Observation Points in H.261 Decoder

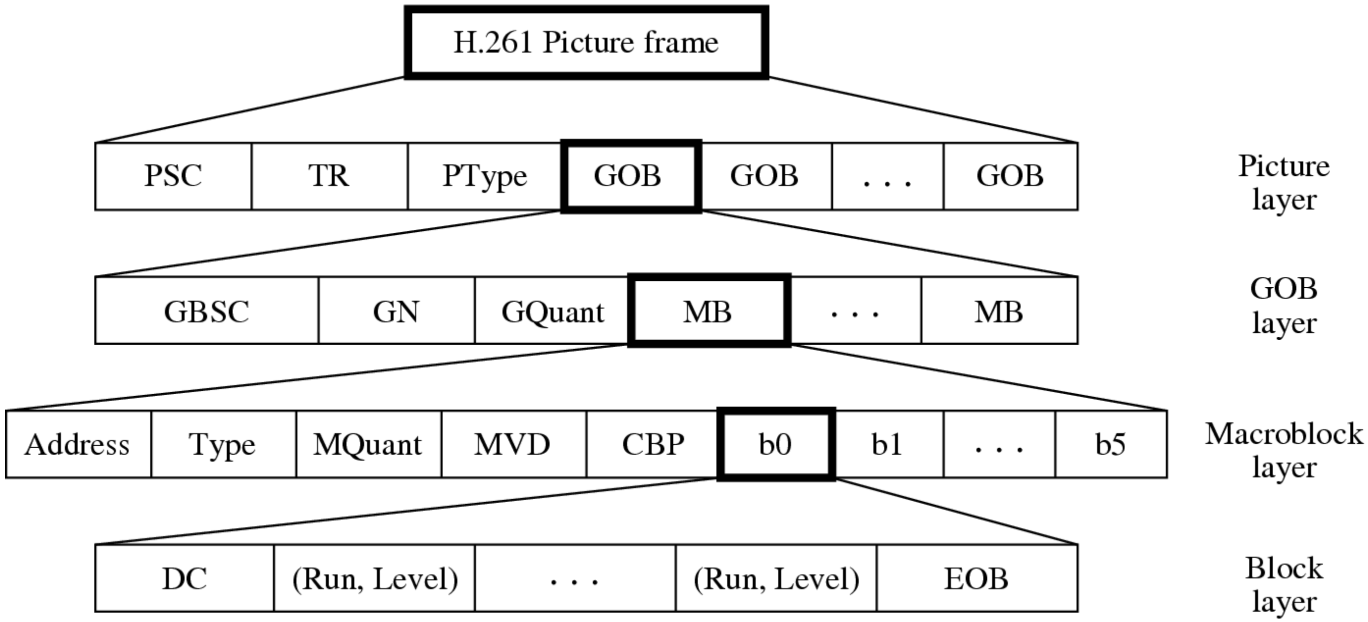
Current Frame	Observation Point			
	1	2	3	4
I	$\tilde{I}$		0	$\tilde{I}$
P_1	$\tilde{D}_1$	P_1'	P_1'	$\tilde{P}_1$
P_2	$\tilde{D}_2$	P_2'	P_2'	$\tilde{P}_2$

A Glance at Syntax of H.261 Video Bitstream

Syntax of H.261 video bitstream consists of a hierarchy of four layers:

1. Picture layer
2. Group of Blocks (GOB) layer
3. Macroblock layer
4. Block layer

Syntax of H.261 Video Bitstream.



PSC	Picture Start Code	TR	Temporal Reference
PType	Picture Type	GOB	Group of Blocks
GBSC	GOB Start Code	GN	Group Number
GQuant	GOB Quantizer	MB	Macroblock
MQuant	MB Quantizer	MVD	Motion Vector Data
CBP	Coded Block Pattern	EOB	End of Block

The Picture layer

PSC (Picture Start Code) delineates boundaries between pictures. TR (Temporal Reference) provides a time-stamp for the picture.

The GOB layer

- H.261 pictures are divided into regions of 11 x 3 macroblocks, each of which is called a Group of Blocks (GOB).

- For instance, the CIF image has 2 x 6 GOBs, corresponding to its image resolution of 352 x 288 pixels. Each GOB has its Start Code (GBSC) and Group number (GN).
- In case a network error causes a bit error or the loss of some bits, H.261 video can be recovered and resynchronized at the next identifiable GOB.
- GQuant indicates the Quantizer to be used in the GOB unless it is overridden by any subsequent MQuant (Quantizer for Macroblock). GQuant and MQuant are referred to as scale.

The Macroblock layer

Each Macroblock (MB) has its own Address indicating its position within the GOB, Quantizer (MQuant), and six 8 x 8 image blocks (4 Y, 1 Cb, 1 Cr).

The Block layer

For each 8 x 8 block, the bitstream starts with DC value, followed by pairs of length of zerorun (Run) and the subsequent non-zero value (Level) for ACs, and finally the End of Block (EOB) code. The range of Run is [0; 63]. Level reflects quantized values — its range is [–127, 127] and Level ≠ 0.

Arrangement of GOBs in H.261 Luminance Images

GOB 0	GOB 1
GOB 2	GOB 3
GOB 4	GOB 5
GOB 6	GOB 7
GOB 8	GOB 9
GOB 10	GOB 11

CIF

GOB 0
GOB 1
GOB 2

QCIF

H.263

- H.263 is an improved video coding standard for video conferencing
- Aims at low bit-rate communications at bit-rates of less than 64 kbps
- Uses predictive coding for inter-frames transform coding for the remaining

Video Formats Supported by H.263

Video format	Luminance image resolution	Chrominance image resolution	Bit-rate (Mbps) (if 30 fps and uncompressed)	Bit-rate (kbps) BPPmaxKb (compressed)
sub-QCIF	128×96	64×48	4.4	64
QCIF	176×144	88×72	9.1	64
CIF	352×288	176×144	36.5	256
4CIF	704×576	352×288	146.0	512
16CIF	$1,408 \times 1,152$	704×576	583.9	1024

Motion Vectors are predicted from the “previous”, “above” and “above and right” MBs

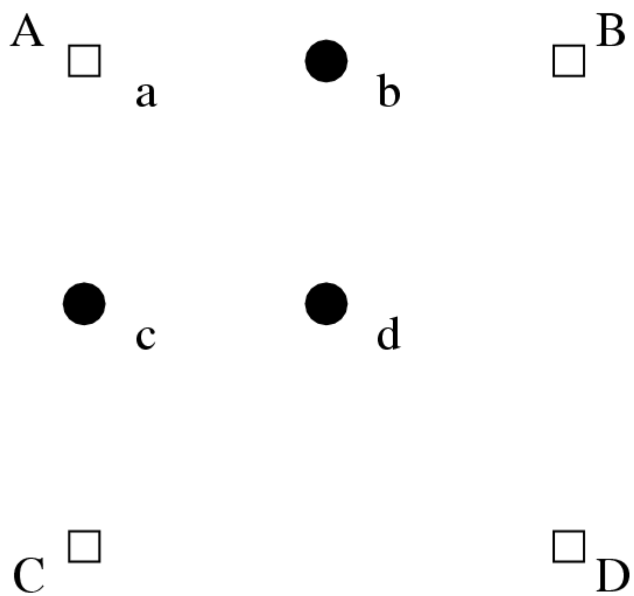
	MV2	MV3
MV1	MV	

MV Current Motion Vector
MV1 Previous Motion Vector
MV2 Above Motion Vector
MV3 Above and Right Motion Vector

For the Macroblock with MV(u, v) the error vector (δu, δv) is coded

- $\delta u = u - u_p$
- $\delta v = v - v_p$
- $u_p = median(u_1, u_2, u_3)$,
- $v_p = median(v_1, v_2, v_3)$.

Half-pixel Prediction



□ Full-pixel position

● Half-pixel position

$$a = A$$

$$b = (A + B + 1) / 2$$

$$c = (A + C + 1) / 2$$

$$d = (A + B + C + D + 2) / 4$$

Half-Pixel Precision

- In order to reduce the prediction error, half-pixel precision is supported in H.263 vs. full-pixel precision only in H.261.
- The default range for both the horizontal and vertical components u and v of $\mathbf{MV}(u, v)$ are now [−16, 15.5].
- The pixel values needed at half-pixel positions are generated by a simple bilinear interpolation method.

MPEG Overview

- MPEG: *Moving Pictures Experts Group*, established in 1988 for the development of digital video.
- It is appropriately recognized that proprietary interests need to be maintained within the family of MPEG standards:
  - Accomplished by defining only a compressed bitstream that implicitly defines the decoder.
  - The compression algorithms, and thus the encoders, are completely up to the manufacturers.

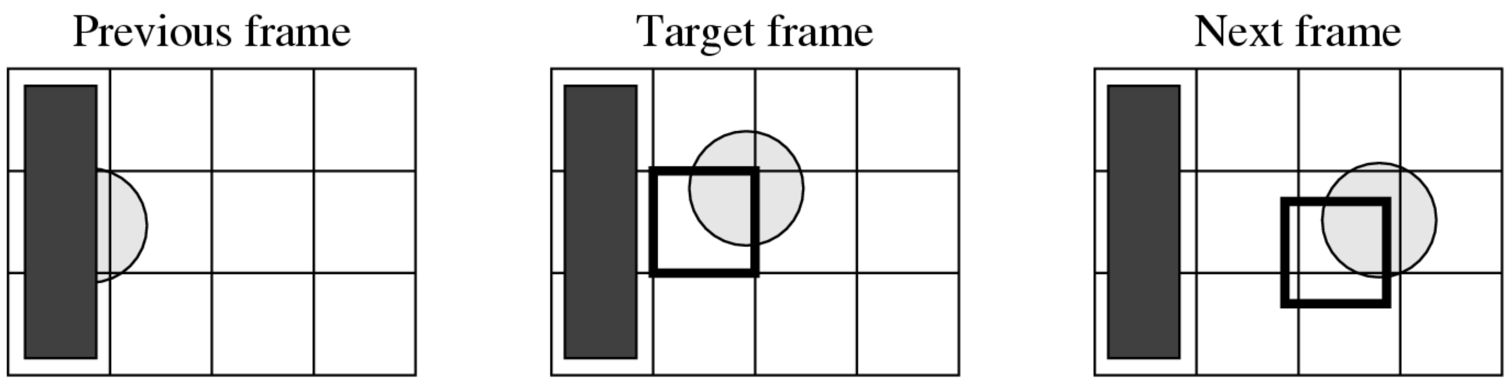
MPEG 1 – storage and retrieval

MPEG 2 – digital TV

MPEG4/H.264 – storage and retrieval

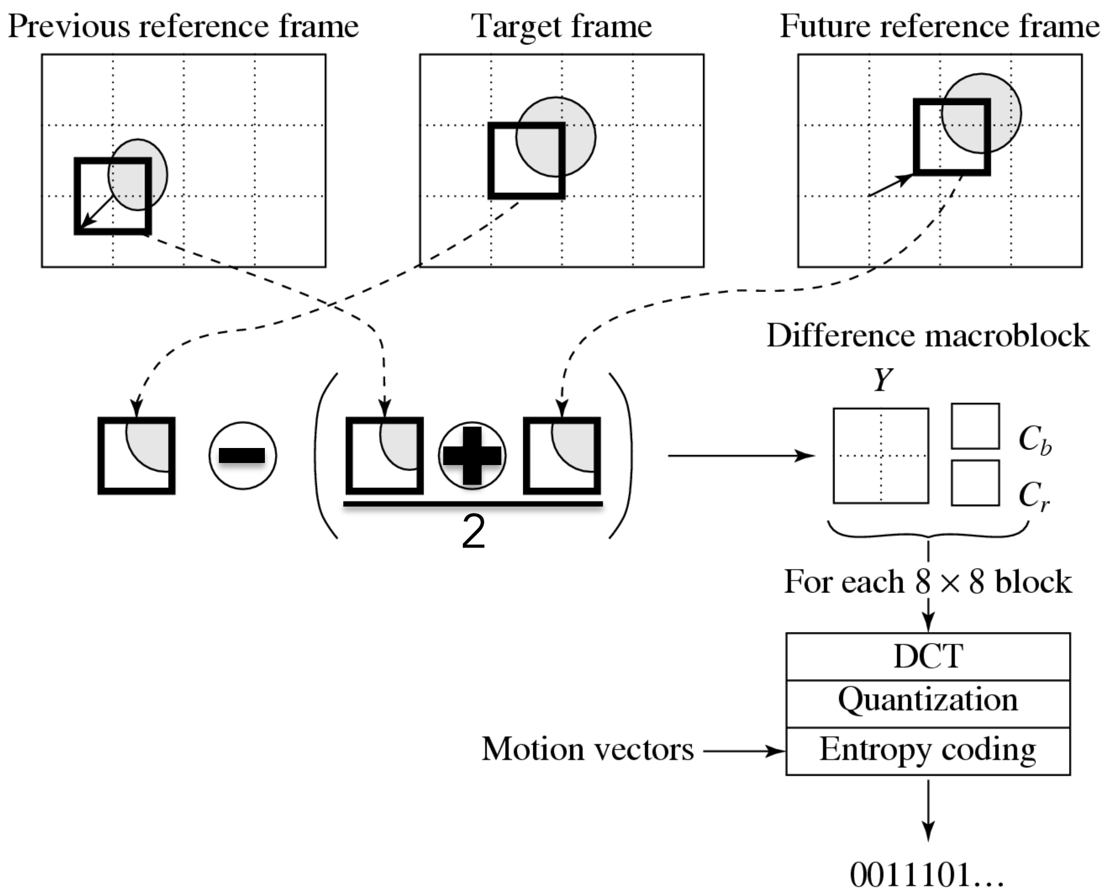
Motion Compensation in MPEG-1

The Need for Bidirectional Search

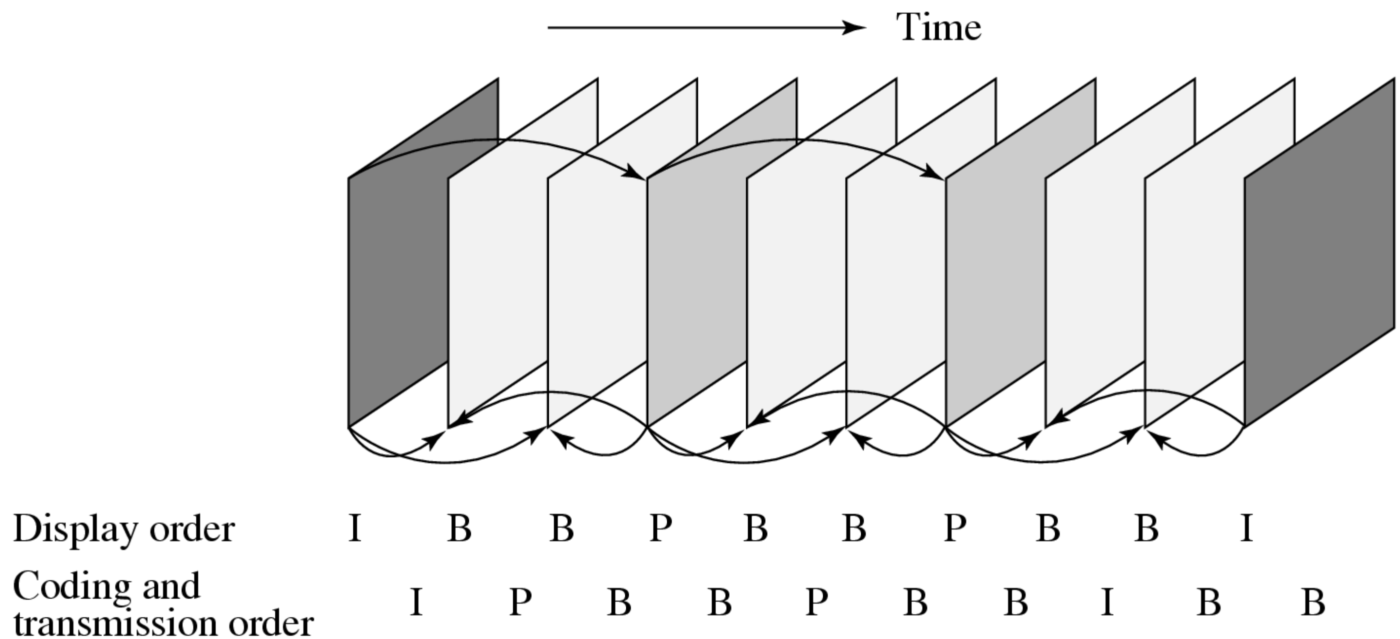


- The MB containing part of a ball in the Target frame cannot find a good matching MB in the previous frame because half of the ball was occluded by another object. A match however can readily be obtained from the next frame.

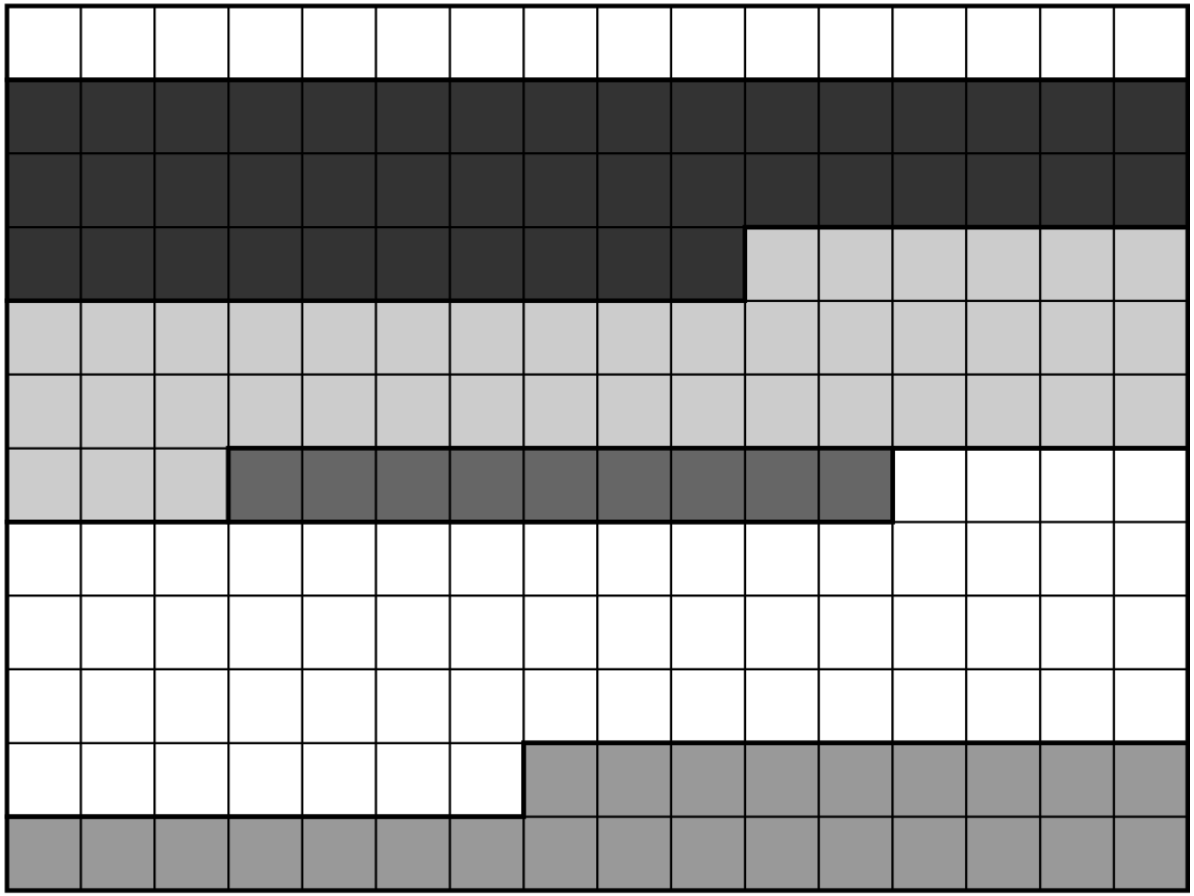
B-frame Coding Based on Bidirectional Motion Compensation



MPEG Frame Sequence



GOBs are replaced by Slices



MPEG-1 Slices

- May contain variable numbers of macroblocks in a single picture.
- May also start and end anywhere as long as they fill the whole picture.
- Each slice is coded independently — additional flexibility in bit-rate control.
- Slice concept is important for error recovery.

Different quantization tables for its Intra and Inter coding

- For DCT coefficients in Intra mode:

$$QDCT[i,j] = round\left(\frac{8\times DCT[i,j]}{step\_size[i,j]}\right) = round\left(\frac{8\times DCT[i,j]}{Q_1[i,j]*scale}\right)$$

- For DCT coefficients in Inter mode:

$$QDCT[i,j] = \left\lfloor\frac{8\times DCT[i,j]}{step\_size[i,j]}\right\rfloor = \left\lfloor\frac{8\times DCT[i,j]}{Q_2[i,j]*scale}\right\rfloor$$

Default Quantization Table (Q_1) for Intra-Coding

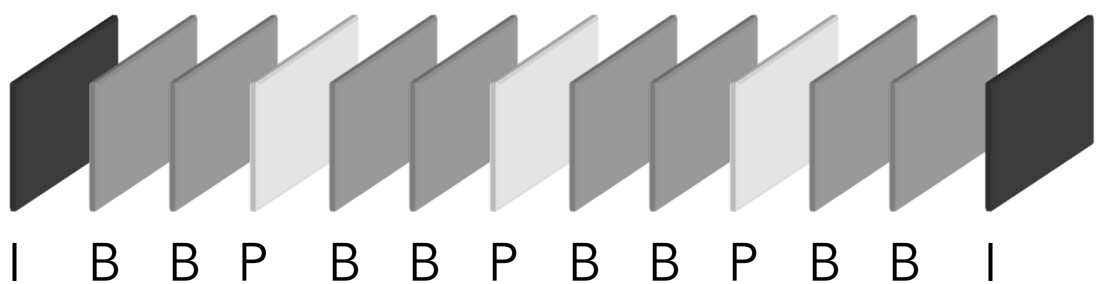
8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	25	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

Default Quantization Table (Q_2) for Inter-Coding

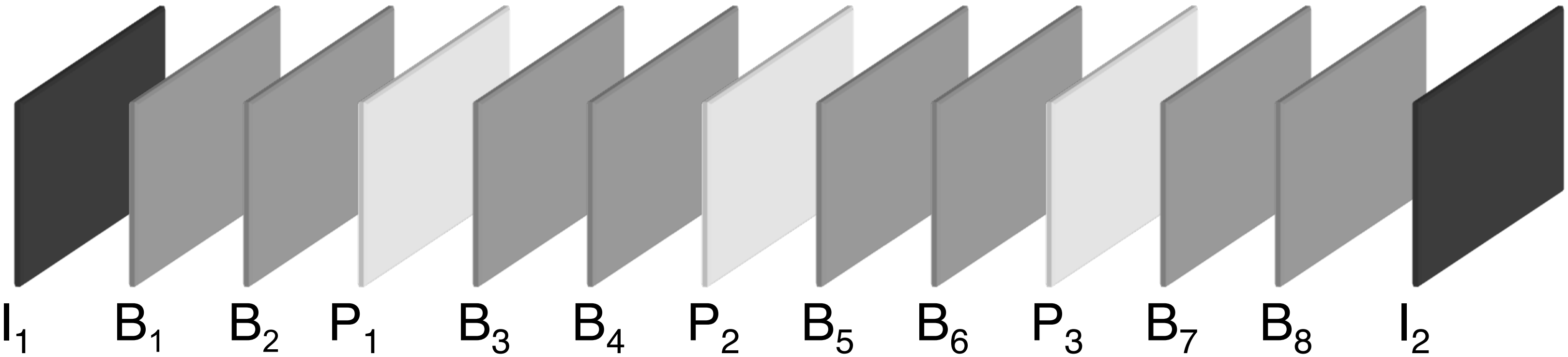
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

Half-pixel precision with an increased range of [−512, 511.5]

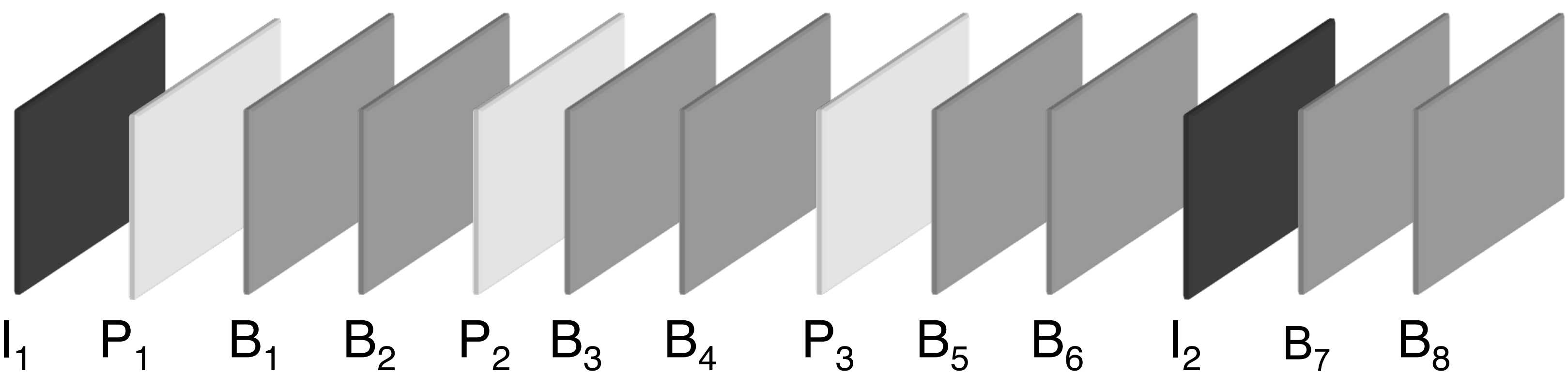
Random access by GOP layer in which each GOP is time coded!



Display Order



Decoding Order

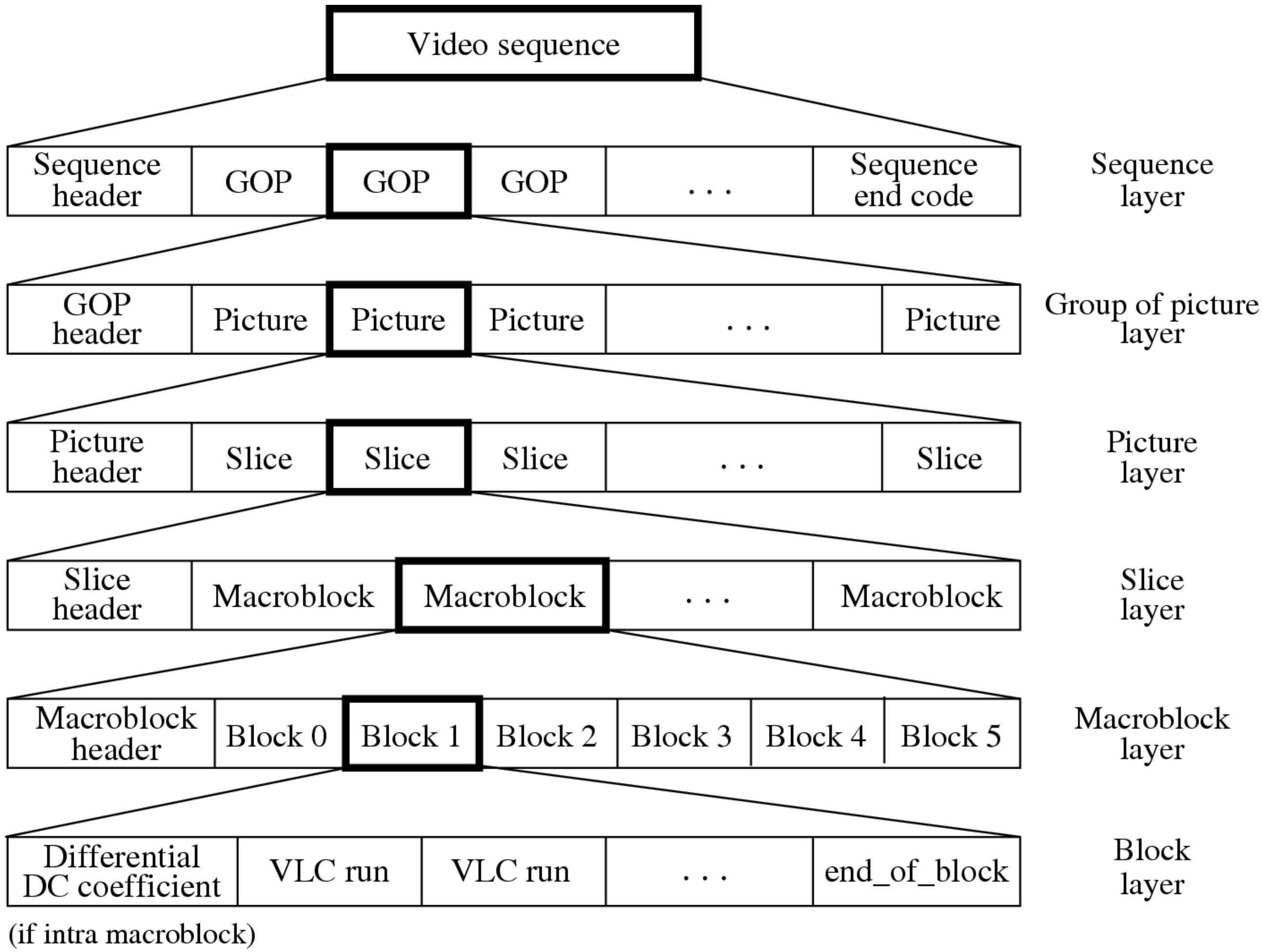


Typical Sizes of MPEG-1 Frames

Type	Size	Compression
I	18kB	7:1
P	6kB	20:1
B	2.5kB	50:1
Avg	4.8kB	27:1

- The typical size of compressed P-frames is significantly smaller than that of I-frames — because temporal redundancy is exploited in inter-frame compression.
- B-frames are even smaller than P-frames — because of (a) the advantage of bi-directional prediction and (b) the lowest priority given to B-frames.

Layers of MPEG-1 Video Bitstream.



MPEG2

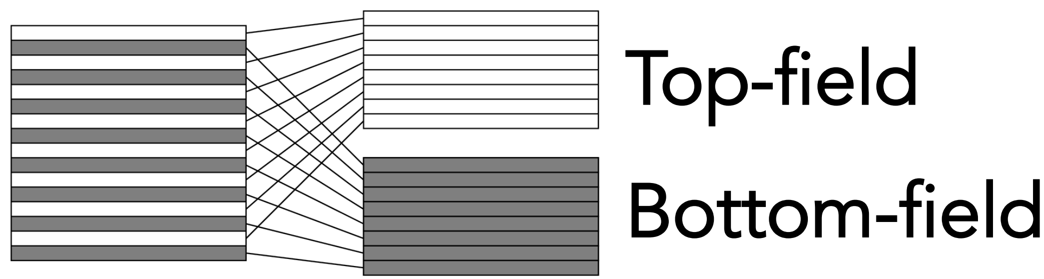
For higher quality video at a bit-rate of more than 4 Mbps!

MPEG-2 7 Profiles

1. Simple
2. Main
3. SNR scalable
4. Spatially scalable
5. High
6. 4:2:2
7. Multiview.

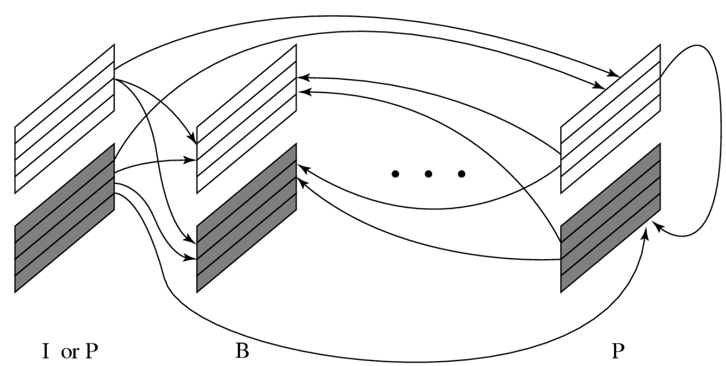
Supporting Interlaced Video

Frame–picture



(a)

Field–pictures



(b)

Supporting Interlaced Video

- MPEG-2 must support interlaced video as well since this is one of the options for digital broadcast TV and HDTV.
- In interlaced video each frame consists of two fields, referred to as the *top-field* and the *bottom-field*.
- In a *Frame-picture*, all scanlines from both fields are interleaved to form a single frame, then divided into 16×16 macroblocks and coded using MC.
- If each field is treated as a separate picture, then it is called *Field-picture*.

Five Modes of Predictions

1. Frame Prediction for Frame-pictures
2. Field Prediction for Field-pictures
3. Field Prediction for Frame-pictures
4. 16×8 MC for Field-pictures
5. Dual-Prime for P-pictures

MPEG-2 Scalabilities

- A base layer and one or more enhancement layers can be defined
- Useful for
  - Networks with very different bit-rates.
  - Networks with variable bit rate (VBR) channels.
  - Networks with noisy connections.

MPEG-2 supports the following scalabilities

1. SNR Scalability—enhancement layer provides higher SNR.
2. Spatial Scalability — enhancement layer provides higher spatial resolution.
3. Temporal Scalability—enhancement layer facilitates higher frame rate.
4. Hybrid Scalability — combination of any two of the above three scalabilities.
5. Data Partitioning — quantized DCT coefficients are split into partitions.

MPEG 2 Support of 4:2:2 and 4:4:4 chroma subsampling

Better resilience to bit-errors: In addition to *Program Stream*, a *Transport Stream* is added to MPEG-2 bit streams!

Two types of scales are allowed

$$scale_i = i \ \& \ scale_i \neq i$$

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$scale_i$	1	2	3	4	5	6	7	8	10	12	14	16	18	20	22	24
i	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
$scale_i$	28	32	36	40	44	48	52	56	64	72	80	88	96	104	112	

MPEG-4

- MPEG-4 departs from its predecessors in adopting a new object-based coding
- The bit-rate for MPEG-4 video now covers a large range between 5 kbps to 10 Mbps.

Composition MPEG-4 Videos (VOP = Video object plane)