Digital television
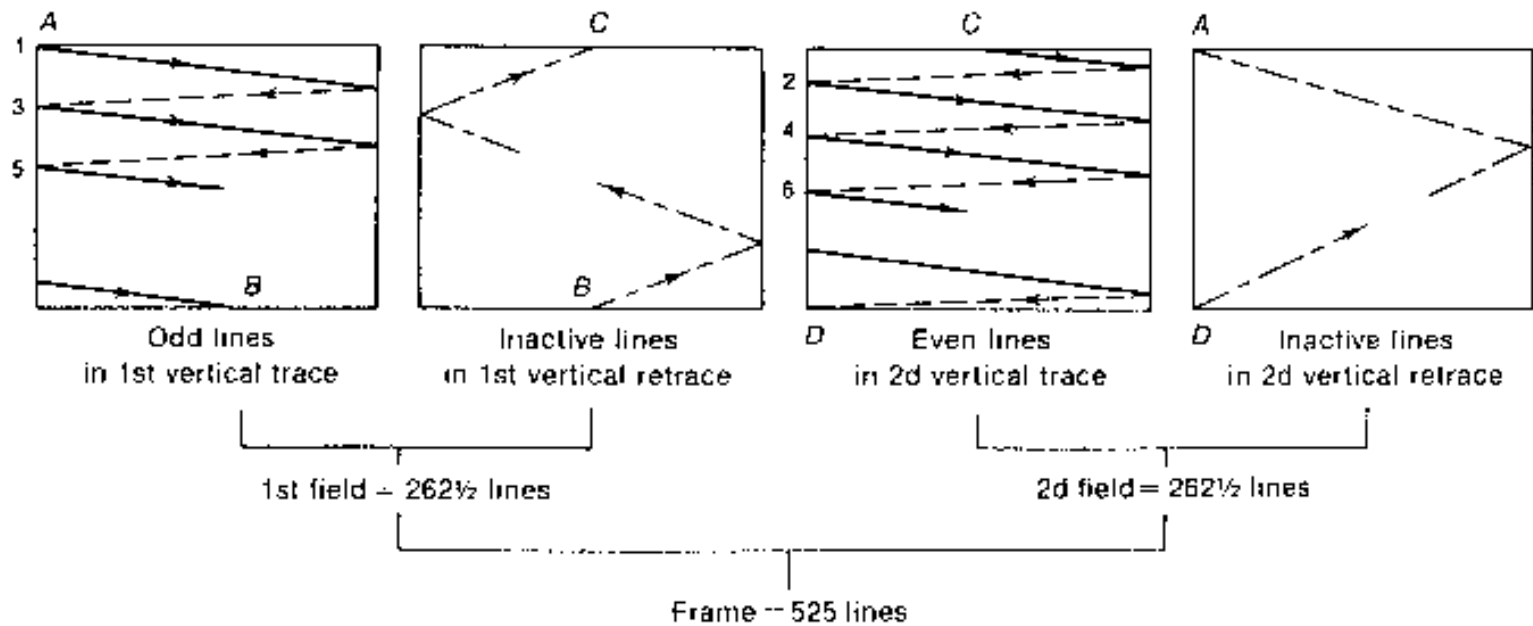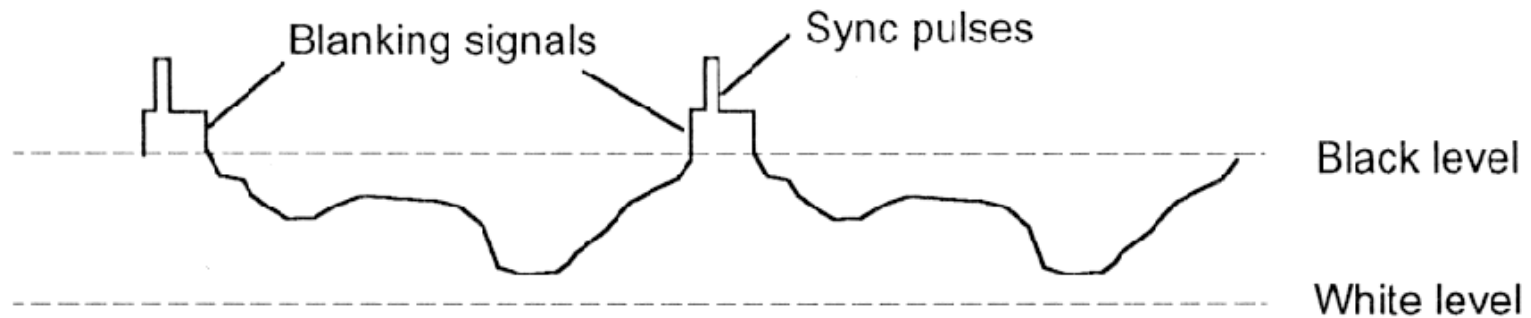
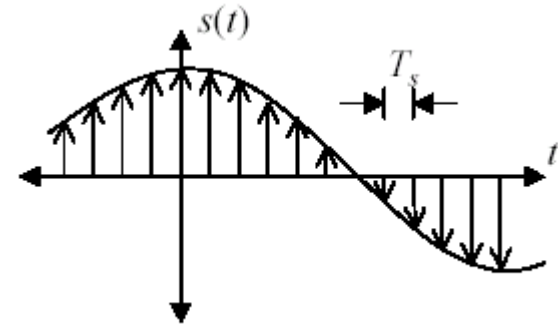# Video and Audio compressing

* Motivation for compression

* Overall structure of codecs

 * Fundamental blocks: DCT, Motion estimation,
        Entropy coding

* MPEG stream

Format of original analog signal

Blanking signals   Sync pulses

Black level

White level

A	C	C	A
1		2	
3		4	
5		6	
B	B	D	D

Odd lines in 1st vertical trace   Inactive lines in 1st vertical retrace   Even lines in 2d vertical trace   Inactive lines in 2d vertical retrace

1st field – 262½ lines   2d field = 262½ lines

Frame -- 525 lines

Sampling of analog signal

Standard sampling

$s[k] = s(kT_s)$

$k \in \{ \; 0 \; 1, 2, 3,...\}$

$1/f_s = T_s$ sampling interval

Nyqvist rate $2 f_{max}$

*Table source: U.Reimers: Digital Video Broadcasting

Digital TV: Sampling

Sampling of video signal (composite video):

     Normally "4fsc" 4 * colorsubcarrier frequency

         e.g. -> 14.318 MHz at 12 bits = 171 Mbit/s

Sampling components (Y / Cb / Cr):

     Normally 13.5 MHz luminance frequency

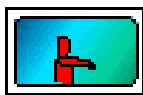   At 4:2:2 sampling totally 27 MHz, 8 bit ->

 PAL TV ⟹ **216 Mbps**

 HDTV TV ⟹ **1 Gbps**

Available: 5 Mbit/s
Compression "40:1"
Compression "250:1" (HEVC)

Table 2.3. Digitisation characteristics and data rates of video signals with irrelevance reduction

Signals	Clock [MHz]	Values/ [line]	Lines	H_o [Mbit/s]	H_{oTotal} [Mbit/s]	Format
R	13.5	864	625	108		4:4:4
G	13.5	864	625	108		ITU 601
B	13.5	864	625	108	324	
Y	13.5	864	625	108		4:2:2
C_B	6.75	432	625	54		ITU 601
C_R	6.75	432	625	54	216	
Y	13.5	720	576	83		4:2:2
C_B	6.75	360	576	41.5		only active image
C_R	6.75	360	576	41.5	166	
Y	13.5	720	576	83		4:2:0
$C_B C_R$	6.75	360	576	41.5	124.5	only active image
Y	6.75	360	288	20.7		4:2:0, SIF
$C_B C_R$	3.375	180	288	10.4	31.1	only active image

*Table source: U.Reimers: Digital Video Broadcasting

MPEG: History

JPEG – during the 80s – first standard 1992

MPEG started in 1988

H.120 – 2Mbit/s – DCPM - videoconferencing

H.261, 1989, – 384 kbit/s-1Mbit/s – DCT - videoconferencing

MPEG-1 (ISO/IEC-11172), 1989-1996 – VHS quality at 1,5 Mbit/s

MPEG-2 (ISO/IEC-13818), 1991-2000 – video broadcasting

MPEG-4 (ISO/IEC-14496), 1995-2002 – objects, wavelets

MPEG-7 2002 – "content-based coding"

MPEG-4 AVC 2004

H.265 HEVC 2013

What is MPEG in general?

A MPEG have several parts; for instance MPEG-4:

- **Systems – Synchronizing and multiplexing of A / V**
- **Visual – Compression of visual objects**
- **Audio – Compression of audio**
- Conformance testing
- Delivery Multimedia Integration Framework
- Reference software
- MPEG-4 delivery over IP networks
- Reference hardware
- **Advanced Video Coding - ITU-T H.264 standard**
- Scene description / app engine
- ISO base media format
- Intellectual property handlign
- MP4 file format
- AVC file format
- ….

MPEG-1

•This standard is named after the ISO committee that created it. It has a bandwidth of about 188 KBps, and consists of three layers:

   •MPEG video, encoded at 138 KBps

   •MPEG system (responsible, among other things, for A/V sync), encoded at 34 KBps

   •MPEG audio, the Third Layer, hence MP3, encoded at <u>128 kilobits per second</u> (16 KBps)

MPEG-2

Improvements for broadcasting
Possibilty to perform parallel decoding

MPEG-4

- MPEG-4 standard: 30 parts
- Most known: MPEG-4 Part 10, or MPEG-4 AVC/H.264
  - Video encoder (Advanced Video Coder)

- HEVC (High Efficiency Video Coder)

Comparison of video coding standards based on equal PSNR[66]				
Video coding standard	Average bit rate reduction compared to			
	H.264/MPEG-4 AVC HP	MPEG-4 ASP	H.263 HLP	H.262/MPEG-2 MP
HEVC MP	35.4%	63.7%	65.1%	70.8%
H.264/MPEG-4 AVC HP	-	44.5%	46.6%	55.4%
MPEG-4 ASP	-	-	3.9%	19.7%
H.263 HLP	-	-	-	16.2%

Source: G.J. Sullivan; Heiko Schwarz; Thiow Keng Tan; Thomas Wiegand (2012-08-22)."Comparison of the Coding Efficiency of Video Coding

Compression of data in general

Lossless

decompressed stream
identical to original

Ex. Run-length encoding
Huffman-encoding
(ZIP, GIF)

Lossy

there is loss in data,
taking advt. of limits
of eye and ear

Ex. Interpolation,
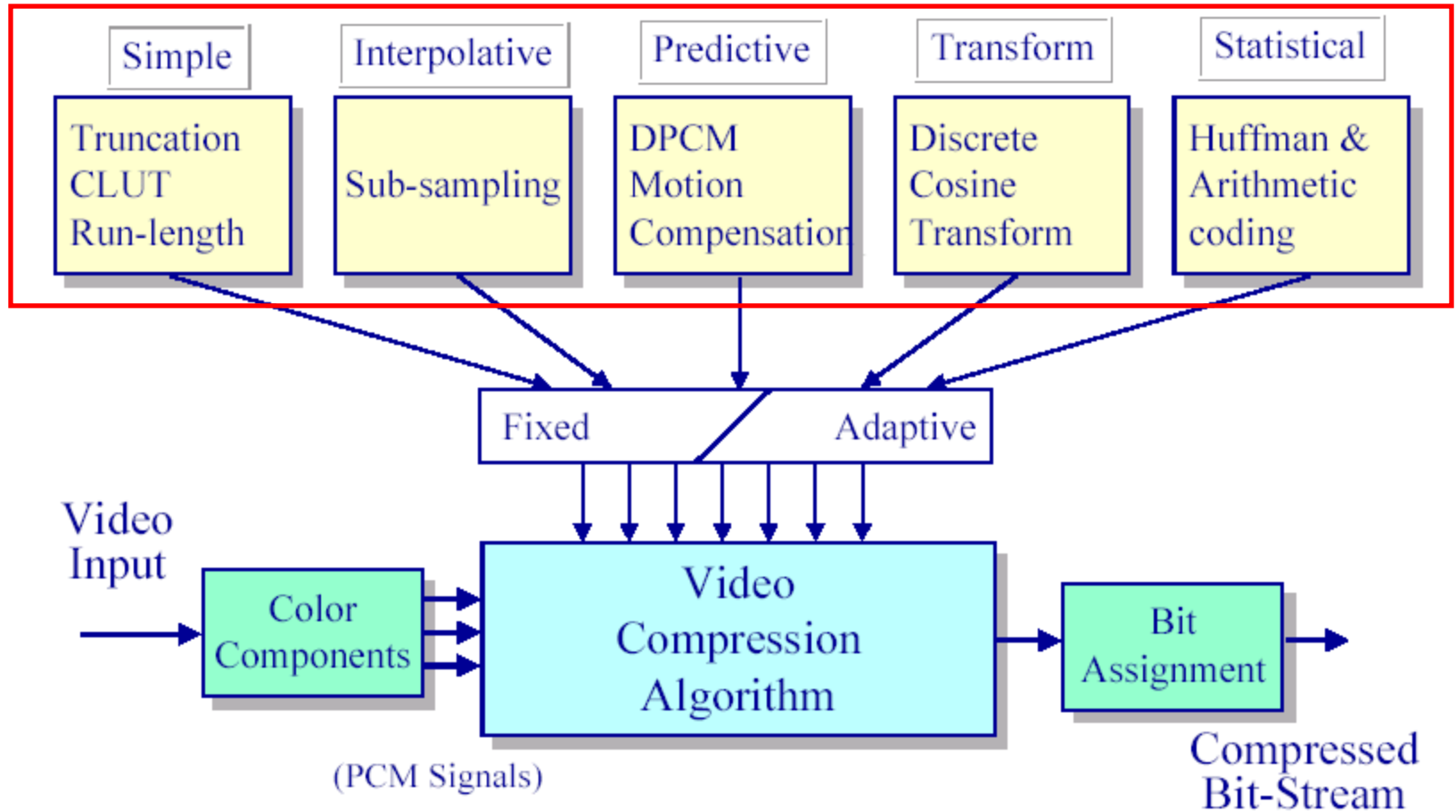subsampling, quantization
(YUV, Wavelets, Fractals)

Hybrid coding

Combination of lossy / lossless etc.
Jpeg, Mpeg,
Last step entropy

MPEG compression techiniques

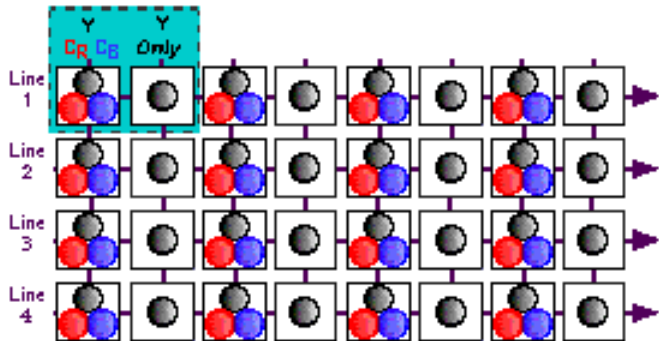I-frame   P-frame   MPEG-2 Video

- Temporal, spatial redundancy
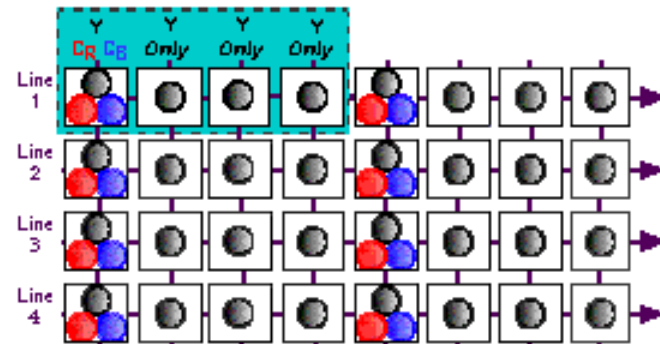- Impercitible details
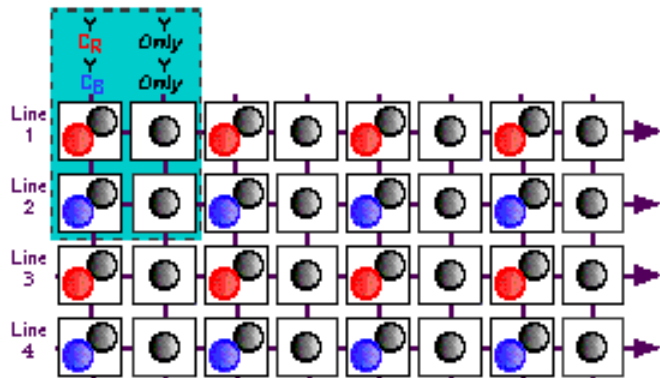
Video compression - methods

Video Sampling Formats

4:2:2 Sampling (Digital Betacam)
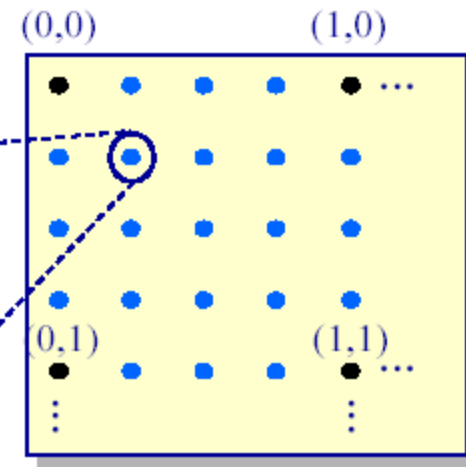High profile MPEG-2

4:1:1 Sampling (NTSC DV)

4:2:0 Co-sited (PAL DV)
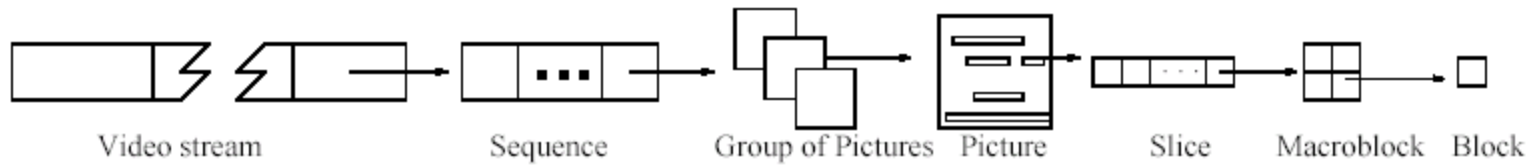Normal profile MPEG-2

Sample interpolation
Chrominance

◆ Bi-linear interpolation:

$$U(1, 1) = U(0,0)\times 0.75 + U(1,0)\times 0.25 + U(0,1)\times 0.75 + U(1,1)\times 0.25$$

(0,0)   (1,0)

(0,1)   (1,1)

Sub-sampled
U or *V* component

MPEG fundamental blocks

Video stream   Sequence   Group of Pictures   Picture   Slice   Macroblock   Block

Slice and Macroblock

GOP

Picture

Slice

Macroblock (MB)

Y Y
Y Y

C_B C_R

JPEG / Still coding

Intra-frame coding: Overall

4	7	10	13	16	19	22	25
7	10	13	16	19	22	25	28
10	13	16	19	22	25	28	31
13	16	19	22	25	28	31	34
16	19	22	25	28	31	34	37
19	22	25	28	31	34	37	40
22	25	28	31	34	37	40	43
25	28	31	34	37	40	43	46

Image Store

Step-size Table

Huffman Code-book

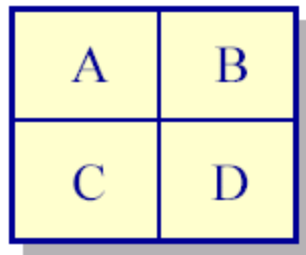Color-Space Convertor → Level Shifting → DCT Encoder → Quantizer → Entropy Encoder → Compressed Image bit-stream (1 component)

◆ Compression comes from:

» Chrominance subsampling
» DCT coefficient quantization
» Difference coding DC coefficients
» Statistical & run-length coding of AC coefficients

◆ Qualitative results:

» 0.25 - 0.5 bpp — ok for some applications
» 0.5 - 0.75 bpp — ok for many
» 0.75 - 1.5 bpp — excellent
» 1.5 - 2.0 — indistinguishable

Transformation based compression

2 x 2 array of pixels

1-D array of differences

A	B
C	D

A	B–A	C–A	D–A

Encode differences with less precision!!

Common transformations: Fourier, DCT, Wavelets
MPEG2 uses DCT

Discrete cosinus transform

◆ The 1-dimensional transform:

$$F(\mu) = \frac{C(\mu)}{2} \sum_{x=1}^{7} f(x) \cos\frac{(2x+1)\mu\pi}{16}$$
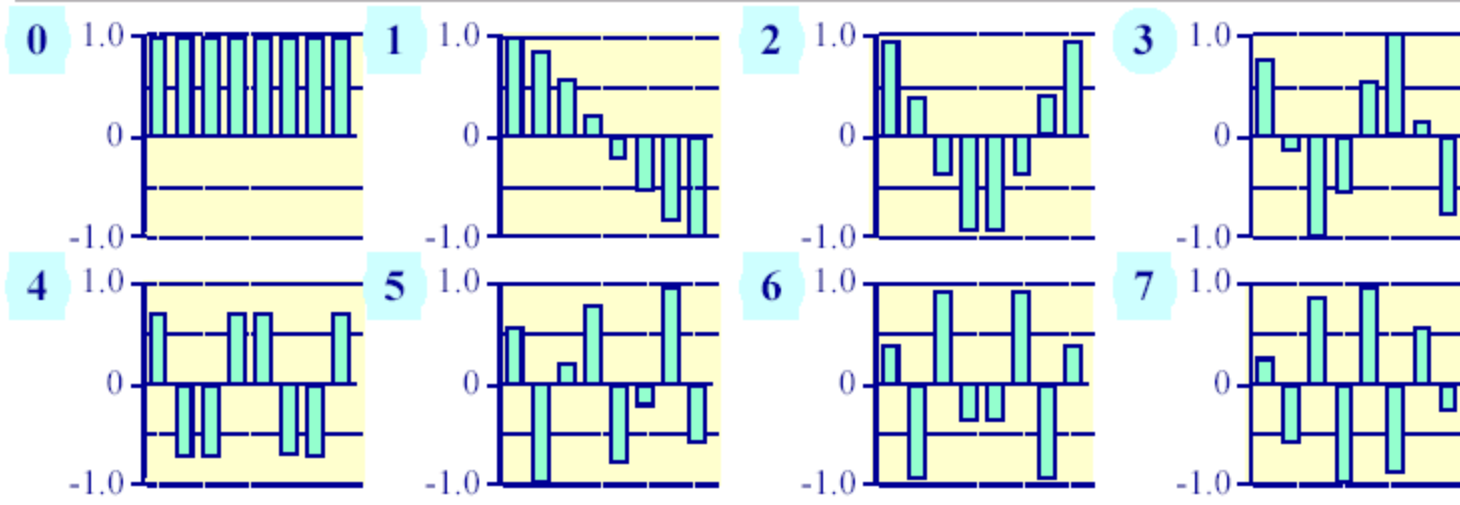
» $F(\mu)$ is the DCT coefficient for $\mu = 0..7$
» $f(x)$ is the xth input sample fot $x = 0..7$
» $C(\mu)$ is a constant (equal to $2^{-0.5}$ if $\mu = 0$ and 1 otherwise)

◆ The 2-dimensional (spatial) transform:

$$F(\mu,v) = \frac{C(\mu)C(v)}{2\cdot 2} \sum_{y=1}^{7} \sum_{x=1}^{7} f(x,y) \cos\frac{(2x+1)\mu\pi}{16} \cos\frac{(2y+1)v\pi}{16}$$

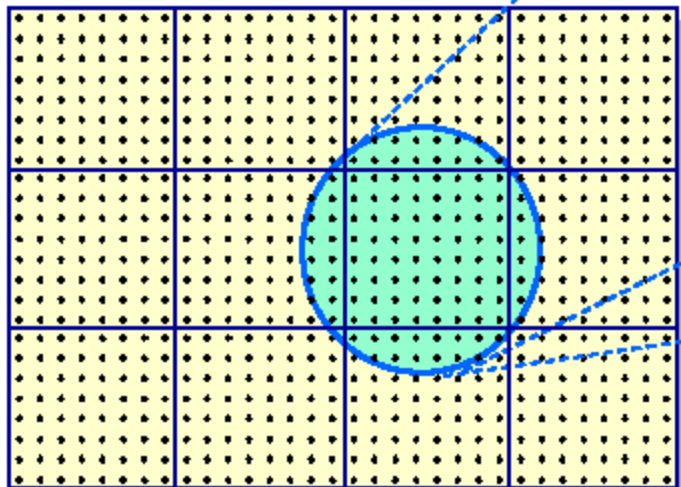# Discrete cosinus transform
Represents input as sum of basic functions

DCT on frames

◆ Apply the DCT in x and y dimensions simultaneously to 8x8 pixel blocks

  » Code coefficients individually with fewer bits

Video Frame

172	-18	15	-8	23	-9	-14	19
21	-34	24	-8	-10	11	14	7
-9	-8	-4	6	-5	4	3	-1
-10	6	-5	4	-4	4	2	1
-8	-2	-3	5	-3	3	4	6
4	-2	-4	6	-4	4	2	-1
4	-3	-4	5	6	3	1	1
0	-8	-4	3	2	1	4	0

DCT Coefficients

29

DCT on frames: Example

Quantization

DCT coefficients

172	-18	15	-8	23	-9	-14	19
21	-34	24	-8	-10	11	14	7
-9	-8	-4	6	-5	4	3	-1
-10	6	-5	4	-4	4	2	1
-8	-2	-3	5	-3	3	4	6
4	-2	-4	6	-4	4	2	-1
4	-3	-4	5	6	3	1	1
0	-8	-4	3	2	1	4	0

Step-siz table

4	7	10	13	16	19	22	25
7	10	13	16	19	22	25	28
10	13	16	19	22	25	28	31
13	16	19	22	25	28	31	34
16	19	22	25	28	31	34	37
19	22	25	28	31	34	37	40
22	25	28	31	34	37	40	43
25	28	31	34	37	40	43	46

Image Store → ... → DCT Encoder → Quantizer →

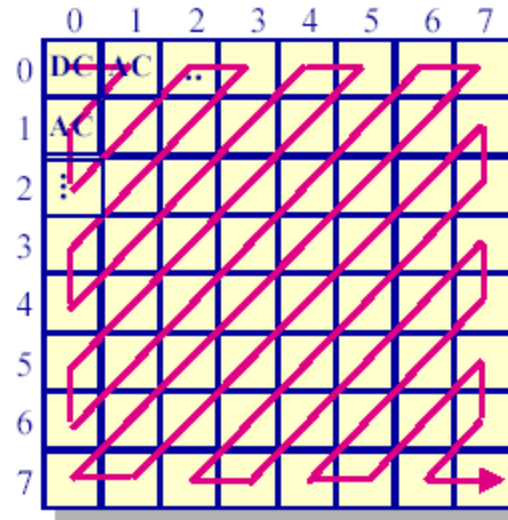| 43	3	2	1	1	0	0	0
3	3	2	1	0	0	0	0
1	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

◆ Step-size table can be scaled to control degree of compression
  » Scaling factor called the "*q-factor*"

* Variable in MPEG stream

Zig-zag walking

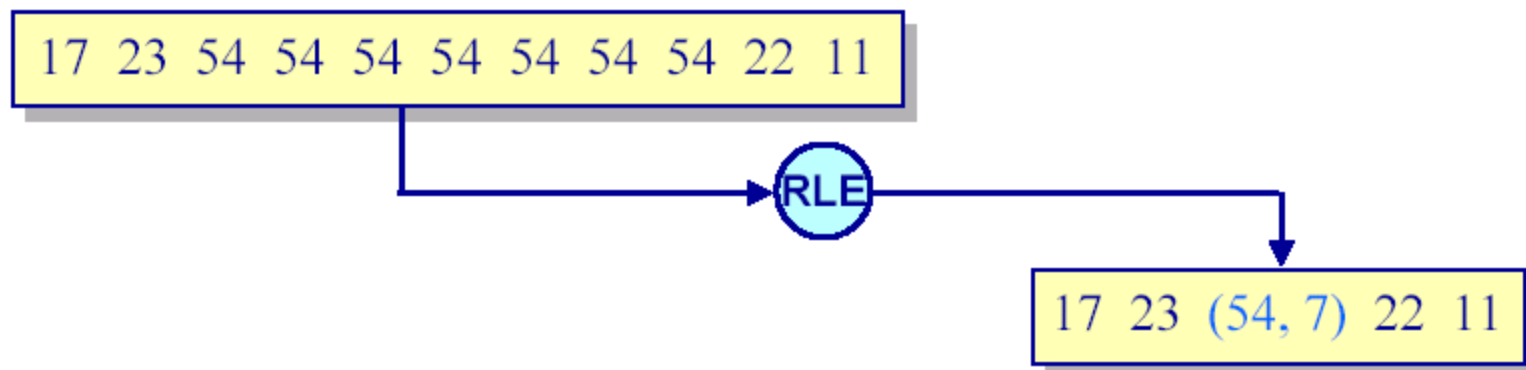In order to increas effectiveness of run-length encoding

43,3,3,1,3,2,1,2,1,1,1,0,0,1,1,0,0,....
After RLE:
  43, (3,2), 1, 3, 2, 1, 2, (1,3), (0,2), (1,2),(0,49)

Run-length encoding
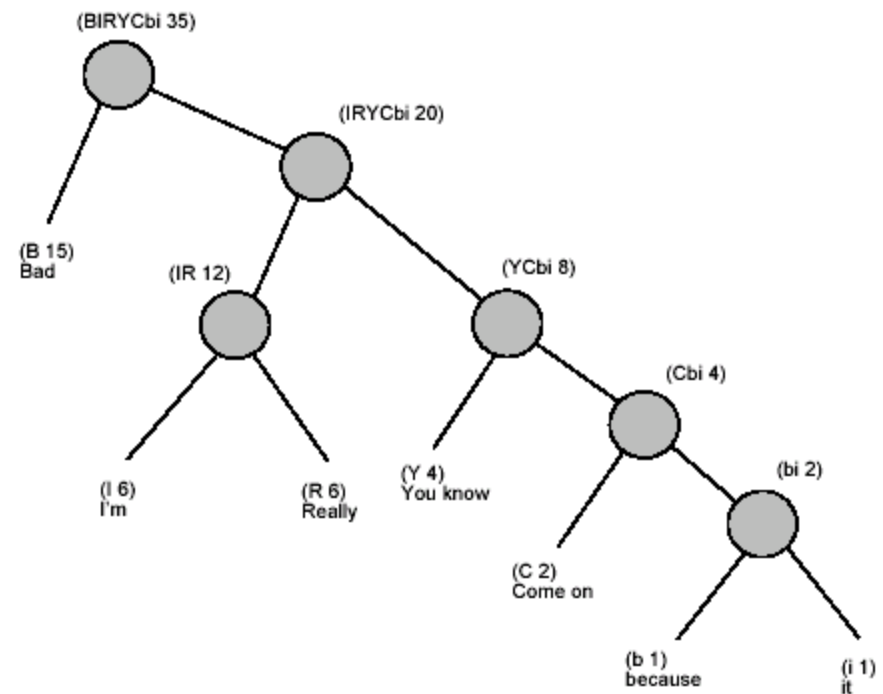

17 23 54 54 54 54 54 54 54 22 11
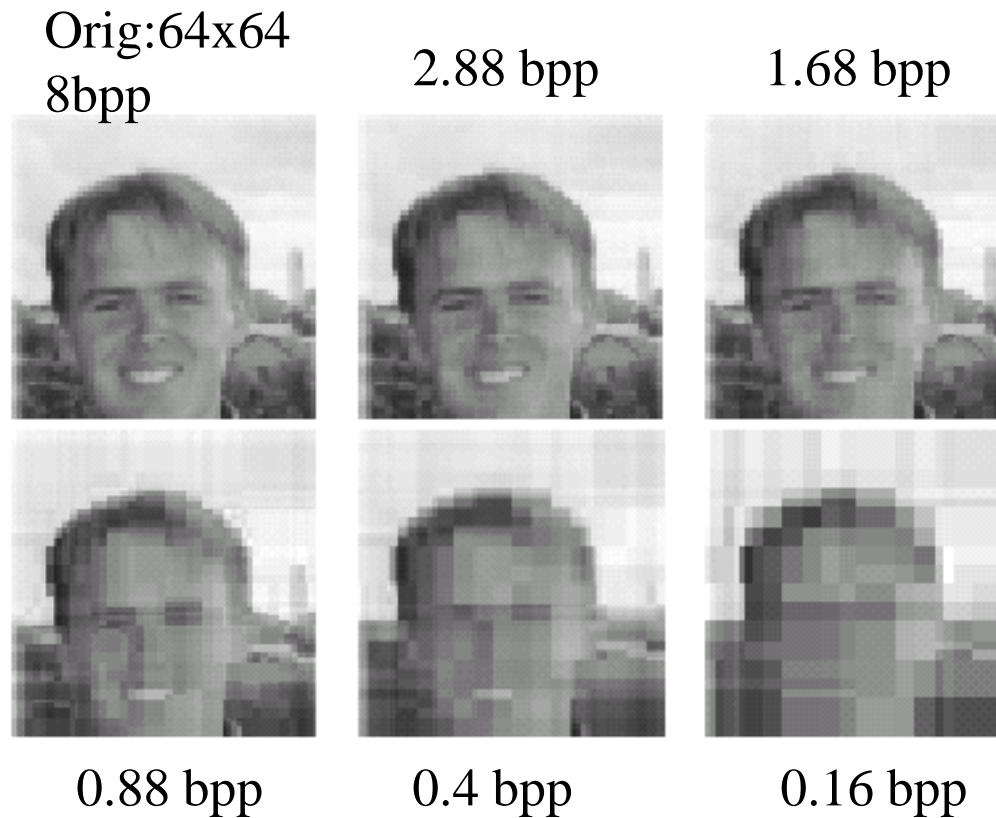
RLE

17 23 (54, 7) 22 11

Huffman encoder - principles

```
Because I'm bad, I'm bad-- come on
Bad, bad-- really, really bad
You know I'm bad, I'm bad--
you know it
Bad, bad-- really, really bad
You know I'm bad, I'm bad-- come on, you kn
Bad, bad-- really, really bad
```
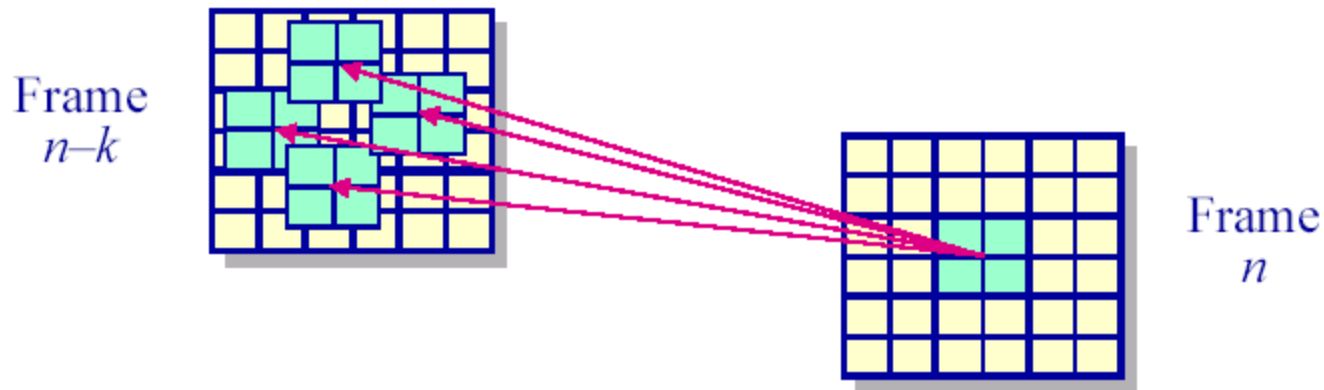
Phrase	Symbol	Frequency	Code Length	Code
Because	b	1	5	00001
I'm	I	6	3	011
Bad	B	15	1	1
Come on	C	2	4	0001
It	i	1	5	00000
Really	R	6	3	010
You know	Y	4	3	001

(BIRYCbi 35)
(IRYCbi 20)
(B 15) Bad
(IR 12)
(YCbi 8)
(Cbi 4)
(I 6) I'm
(R 6) Really
(Y 4) You know
(bi 2)
(C 2) Come on
(b 1) because
(i 1) it

Transformation based compression: Wavelets

Orig:64x64 8bpp   2.88 bpp   1.68 bpp

0.88 bpp   0.4 bpp   0.16 bpp
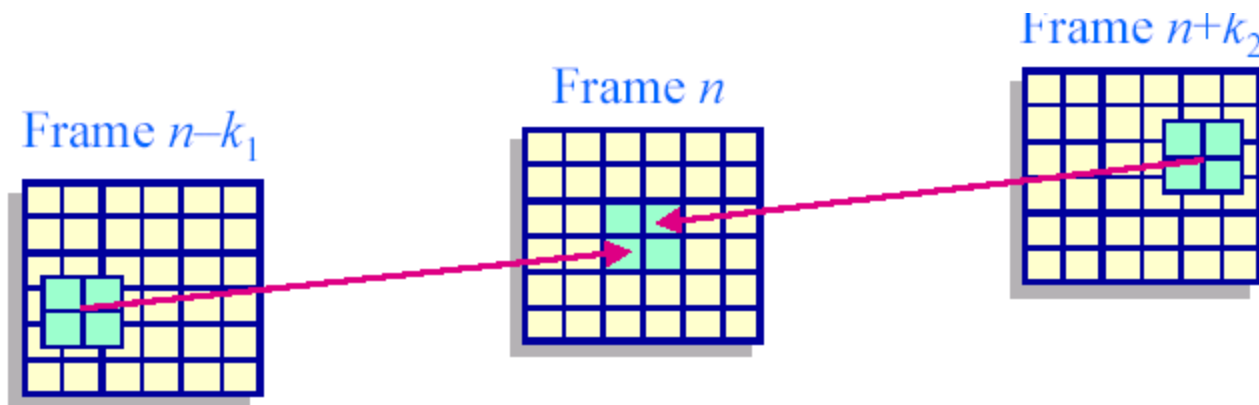
MPEG motion compensation

- Performed on macroblock (16x16 pixels)
- The prediction search space is not specified in the standard
- Objective: Find an motion vector *w* that minimizes some cost function, e.g.

$$\sum_{j=0}^{15}\sum_{i=0}^{15} f(frame_n[16b_x+i, 16b_y+j] - frame_{n-1}[16b_x+w_x+i, 16b_x+w_y+j])$$
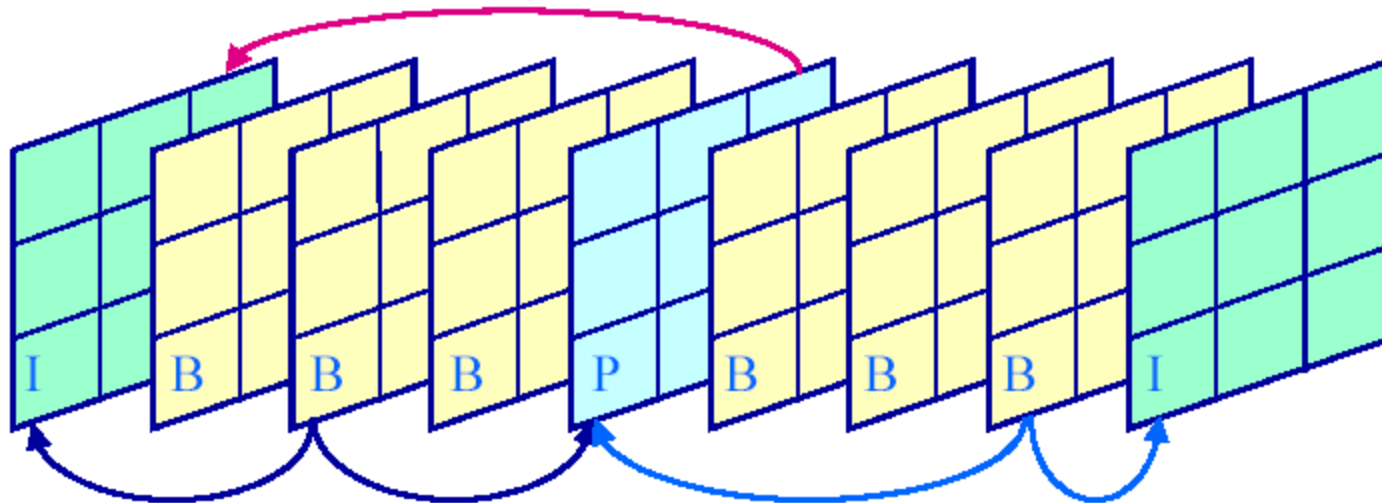
MPEG motion compensation
Bi-directional prediction

Frame $n-k_1$ Frame n Frame $n+k_2$

- Not only prediction, also interpolation
- Both past and future frames are used to predict the current frame
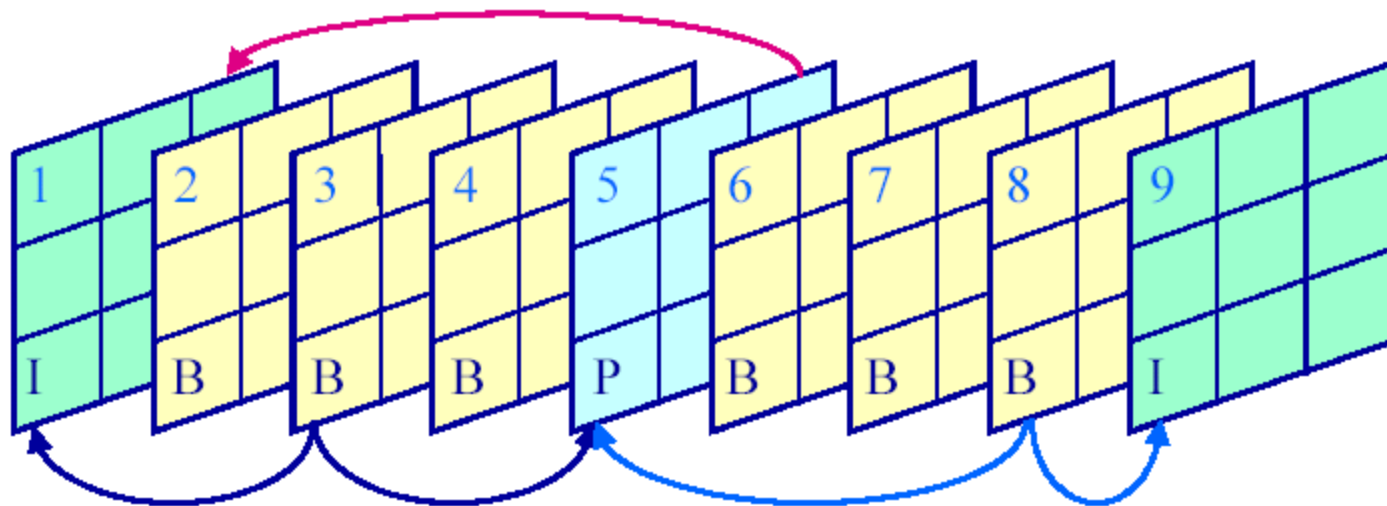
MPEG motion compensation
Bi-directional prediction

- 3 types of pictures in a MPEG stream
  - I – Intracoded pictures (still images)
  - P – Predicted pictures from previous I or P pictures
  - B – Bi-directinally pictures (interpolated pictures)

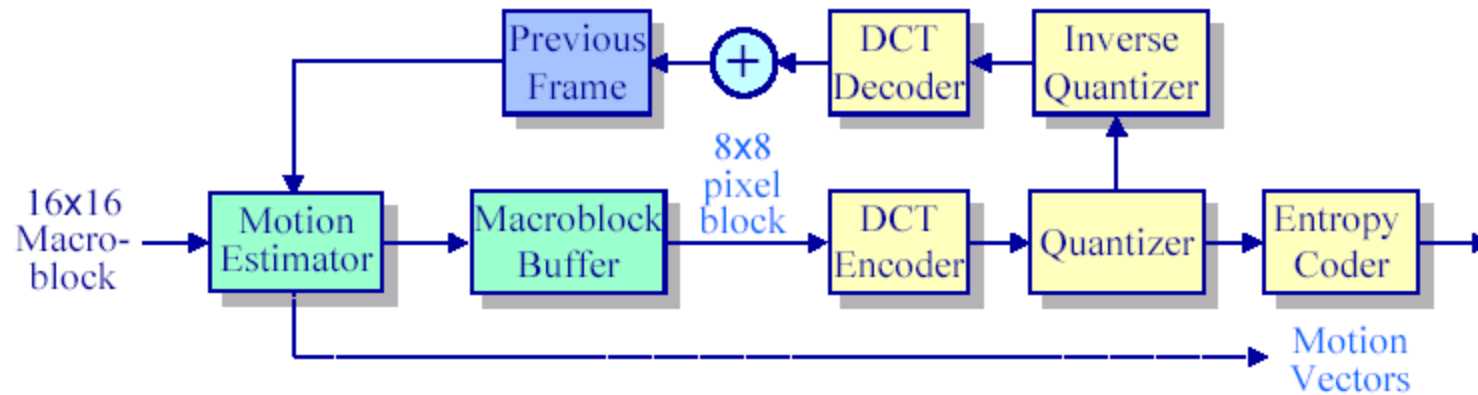# MPEG motion compensation
Bi-directional prediction

- ◆ Directional prediction implies that frames cannot be encoded or transmitted in the order they are scanned
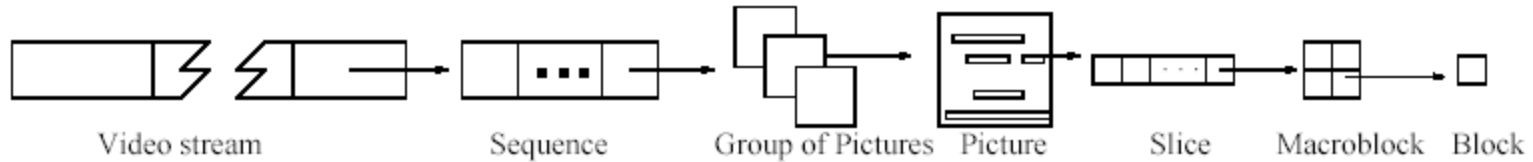
- ◆ Encoding & transmission order:
  - » I_1 P_5 B_2 B_3 B_4 I_9 B_6 B_7 B_8 P_{13} B_{10} B_{11} B_{12} I_{17} B_{14} B_{15} B_{16} ...

Structure of MPEG encoding

MPEG fundamental blocks

Video stream   Sequence   Group of Pictures   Picture   Slice   Macroblock   Block
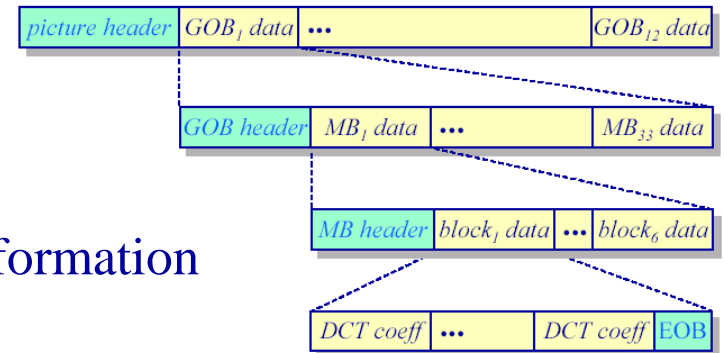
MPEG-2 Stream

» Sequence Layer
 * decoding parameters (bit-rate, buffer size, picture resolution, framerate, ...)
» Group of Pictures Layer
 * a random access point
» Picture Layer
 * picture type and reference picture information
» Slice Layer
 * position and state information for decoder resynchronization
» Macroblock Layer
 * coded motion vectors
» Block Layer
 * coded DCT coefficients, quantizer step size, *etc.*

| picture header | GOB_1 data | ... | | GOB_{12} data |

| GOB header | MB_1 data | ... | | MB_{33} data |

| MB header | $block_1$ data | ... | $block_6$ data |

| DCT coeff | ... | DCT coeff | EOB |

MPEG fundamental blocks

| quantised DCT coefficients for one 8x8 block (variable length coded) | Block layer |

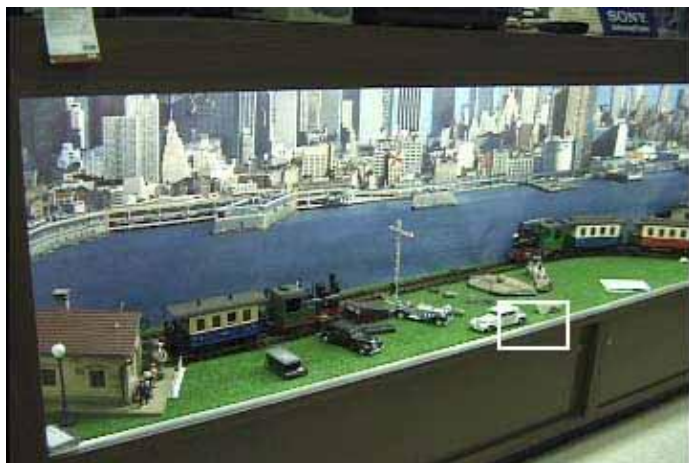| macroblock address | mode | (optional) quantisation value | motion vectors | coded block pattern | luminance blocks | chrom. blocks |

Macroblock layer (containing four luminance and two chrominance blocks for 4:2:0 video)

| start code | slice address | quantisation value | macroblock 0 | macroblock 1 | ...... | macroblock n - 1 |

Slice layer (containing n macroblocks)

| start code | picture flags | slice 0 | slice 1 | ...... | slice m - 1 |

Picture layer (containing m slices)

| start code | sequence parameters | (optional) quantisation weighting matrix | profile and level | picture 0 | picture 1 | ...... | picture p - 1 |

Sequence layer (containing p pictures)

Each picture is divided into m horizontal slices, each comprising n macroblocks. For 4:2:0 video, each macroblock contains four luminance and two chrominance 8x8 blocks of quantised DCT coefficients.

Compression artifacts

MPEG standardization
What about research???

- pre-processing: (or "how to fit a square peg in a round hole?").
- motion estimation (or "how to efficiently find a good prediction.")
- macroblock decision models (efficient, but does it also optimise the ripple effect on subsequent macroblocks ?)
- rate control and buffer management in editing environments (MPEG: video only exists within a sequence. Real world: decoder is displaying picture from previous sequence, while reconstructing a picture from the new sequence)
- implementation complexity reduction (real time encoding on normal PC hardware).