MPEG-1 & MPEG-2 Compression

6th of March, 2002, Mauri Kangas

Contents:

- MPEG Background
- MPEG-1 Standard (shortly)
- MPEG-2 Standard (shortly)
- MPEG-1 and MPEG-2 Differencies
- MPEG Encoding and Decoding
  - Colour sub-sampling
  - Motion Compensation
  - Slices, macroblocks, blocks, etc.
  - DCT
  - MPEG-2 bit stream syntax
- MPEG-2
  - Supported Features
  - Levels
  - Profiles
- DVB Systems

 © Mauri Kangas 2002

MPEG Background

- MPEG = **M**otion **P**icture **E**xpert **G**roup
- ISO/IEC JTC1/SC29
  - WG11 Motion Picture Experts Group (MPEG)
  - WG10 Joint Photographic Experts Group (JPEG)
  - WG7 Computer Graphics Experts Group (CGEG)
  - WG9 Joint Bi-level Image coding experts Group (JBIG)
  - WG12 Multimedia and Hypermedia information coding Experts Group (MHEG)
- MPEG-1,2 Standardization 1988-
  - Requirement
  - System
  - Video
  - Audio
  - Implementation
  - Testing
- Latest MPEG Standardization: MPEG-4, MPEG-7, MPEG-21

MPEG-1 Standard ISO/IEC 11172-2 (1991)

"Coding of moving pictures and associated audio for digital storage media"

Video

- optimized for bitrates around 1.5 Mbit/s
- originally optimized for SIF picture format, but not limited to it:
  - 352x240 pixels a 30 frames/sec [NTSC based]
  - 352x288 pixels at 25 frames/sec [PAL based]
- progressive frames only - no direct provision for interlaced video applications, such as broadcast television

Audio

- joint stereo audio coding at 192 kbit/s (layer 2)

System

- mainly designed for error-free digital storage media
- multiplexing of audio, video and data

Applications

- CD-I, digital multimedia, and video database (e.g. video-on-demand)

MPEG-2 Standard ISO/IEC 13818-2 (1994)

Video

- 2-15 or 16-80 Mbit/s bit rate (target bit rate: 4...9 Mbit/sec)
- TV and HDTV picture formats
- Supports interlaced material
- MPEG-2 consists of *profiles* and *levels*
  - Main Profile, Main Level (MP@ML) refers to 720x480 resolution video at 30 frames/sec, at bit rates up to 15 Mbit/sec for NTSC video (typical ~4 Mbit/sec)
  - Main Profile, High Level (MP@HL) refers to HDTV resolution of 1920x1152 pixels at 30 frames/sec, at a bit rate up to 80 Mbit/sec (typical ~15 Mbit/sec)

Audio:

- compatible multichannel extension of MPEG-1 audio

System:

- video, audio and data multiplexing defines tow presentations:
  - ***Program Stream*** for applications using near error free media
  - ***Transport Stream*** for more error prone channels

Applications:

- satellite, cable, and terrestrial broadcasting, digital networks, and digital VCR

MPEG-2 and MPEG-1 Differencies

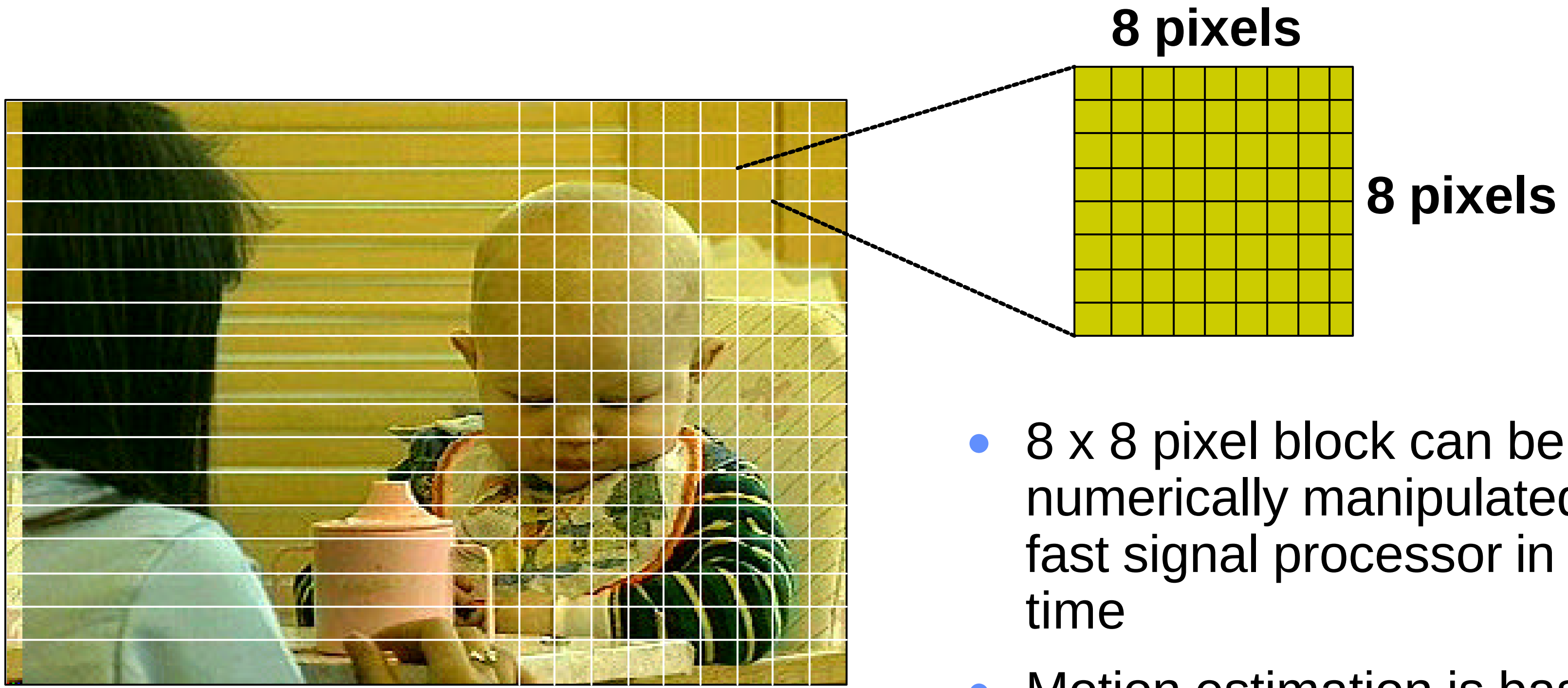
- All MPEG-2 decoders that comply with currently defined profiles and levels are required to decode MPEG-1 constrained bit streams:
- List of differencies
  - IDCT Mismatch Control
  - Macroblock stuffing
  - Run-level escape syntax
  - Chrominance samples horizontal position (co-locate with luminance in MPEG-2, half the way between luminance samples in MPEG-1
  - Slices (in MPEG-2 slices start on the same horizontal row of macroblocks, in MPEG-1 its possible to have all macroblocks of a picture in one slice, for example
  - D-pictures (not permitted in MPEG-2; in MPEG-1 only Intra-DC-coefficient, special end_of_macroblock code)
  - Full-pel Motion Vectors (in MPEG-1 full-pel motion vectors possible, in MPEG-2 always half-pel motion vectors)
  - Aspect Ratio Information (MPEG-1 specifies pel aspect ratio, MPEG-2 specifies display aspect ratio and pel aspect ratio can be calculated from this and from frame size and display size)
  - Forward_f_code and backward_f_code (differencies in parameter location and contents)
  - Constrained_parameter_flag and maximum horizontal_size (MPEG-2 has profile and level mechanism)
  - Bit_rate and vbv_delay (fixed values are reserved for variable bit rate in MPEG-1, other values are for constant bit rate; in MPEG-2 semantics for bit_rate are changed, etc.)
  - VBV (in MPEG-1 VBV is only defined for constant bit rate operation; in MPEG-2 VBV is only defined for variable bit rate and constant bit rate is assumed to be a special case of variable bit rate)
  - temporal_reference (a small difference between MPEG-1 and MPEG-2)
- MPEG-2 syntax can be made to be very close to MPEG-1, by using particular values for the various MPEG-2 syntax elements that do not exist in MPEG-1 syntax

MPEG-2 vs. MPEG-1 Decoding Process

MPEG-1 decoding process is 'almost the same' as the MPEG-2 decoding process when:

- progressive_sequence = '1'
- chroma_format = '01' (4:2:0)
- frame_rate_extension_n = 0 and frame_rate_extension_d = 0 (MPEG-1 frame rate)
- intra_dc_precision = '00' (8-bit Intra-DC precision)
- picture_structure = '11' (frame-picture, because progressive_sequence = '1')
- frame_pred_frame_dct = 1 (only frame-based prediction and frame DCT)
- concealment_motion_vectors = '0' (no concealment motion vectors)
- q_scale_type = '0' (linear quantiser_scale)
- intra_vlc_format = '0' (MPEG-1 VLC table for Intra MBs)
- alternate_scan = '0' (MPEG-1 zigzag scanning order)
- repeat_first_field = '0' (because progressive_sequence = '1')
- chroma_420_type = '1' (chrominance is "frame-based", because progressive_sequence = '1'
- progressive_frame = '1' (progressive_sequence = '1')

MPEG Compression is Based on Processing 8 x 8 Pixel Blocks



- 8 x 8 pixel block can be numerically manipulated by fast signal processor in real time
- Motion estimation is based on comparing the blocks between series of pictures

Only Moving Areas Have to Be Coded

new picture

-

previous picture

=

difference

Encoder

Decoder

difference

+

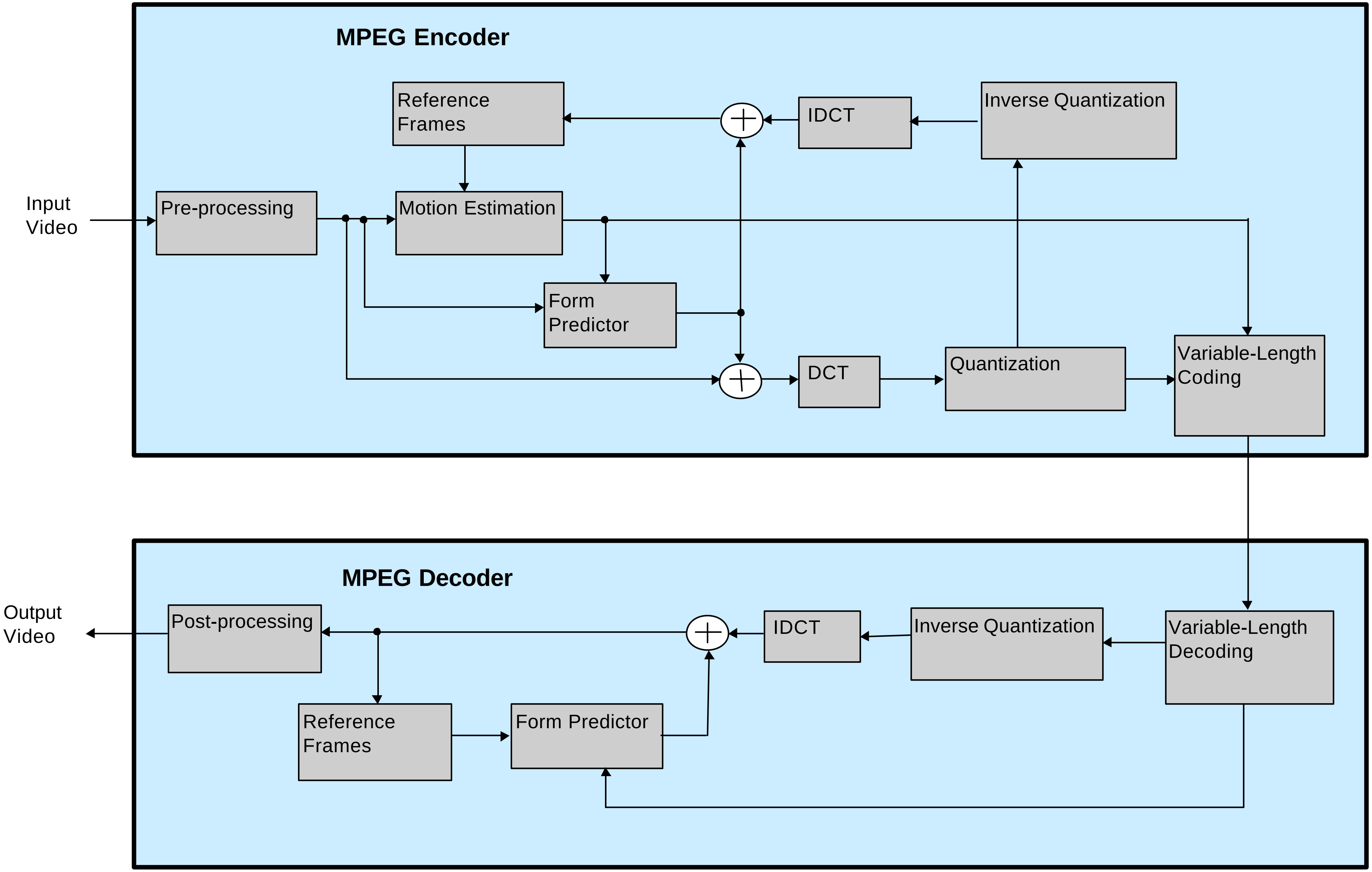
previous picture

=

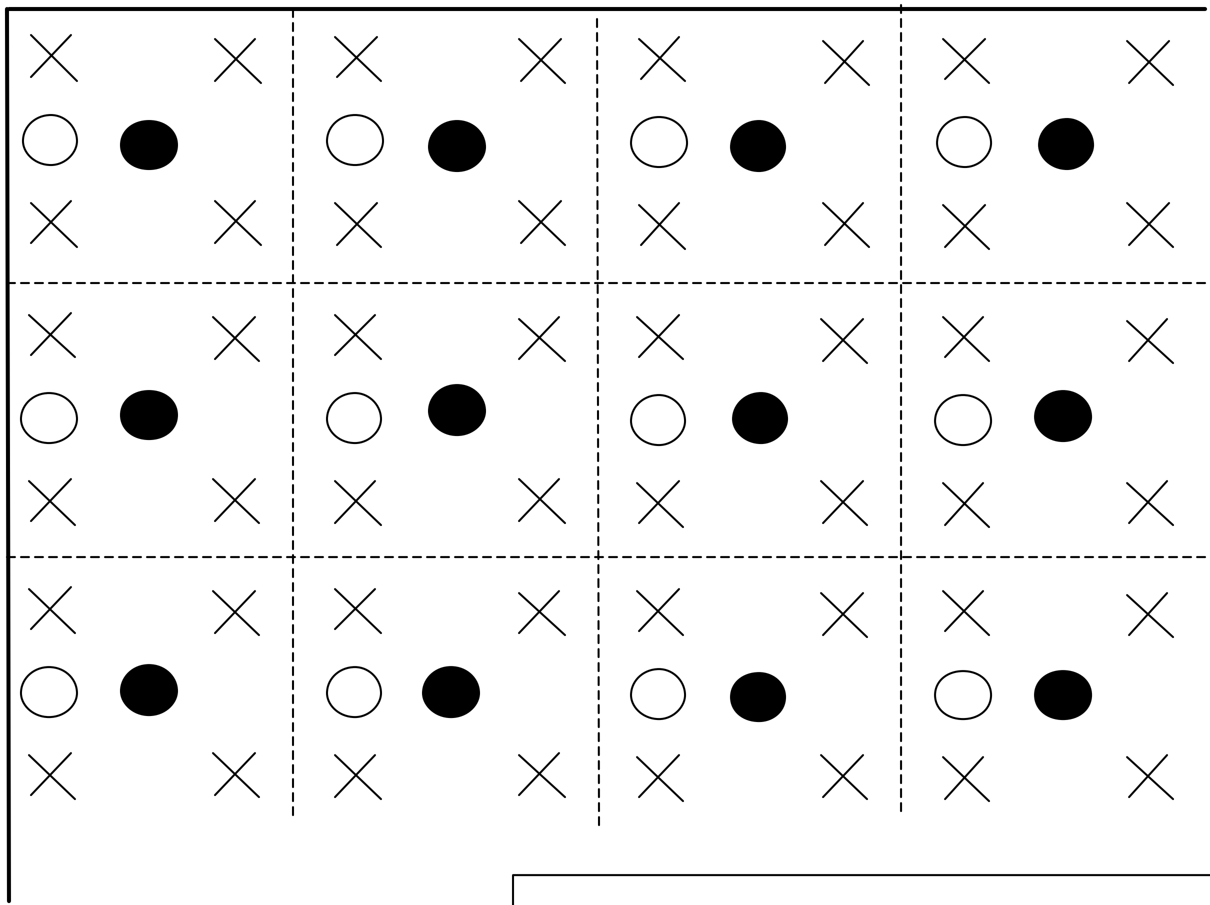
new picture

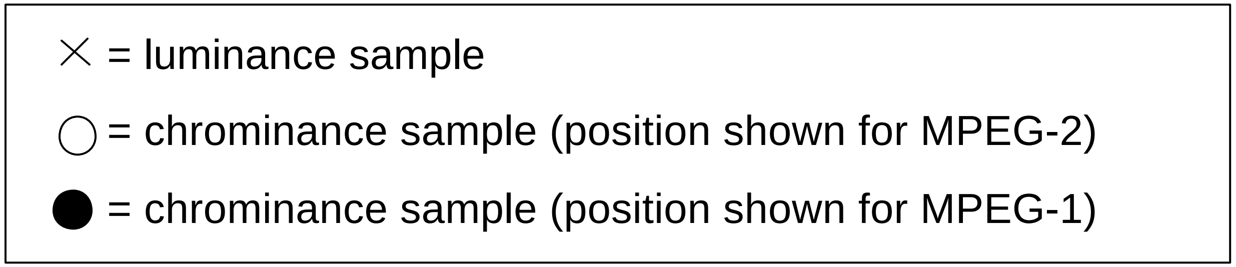
© Mauri Kangas 2002

MPEG Encoding and Decoding

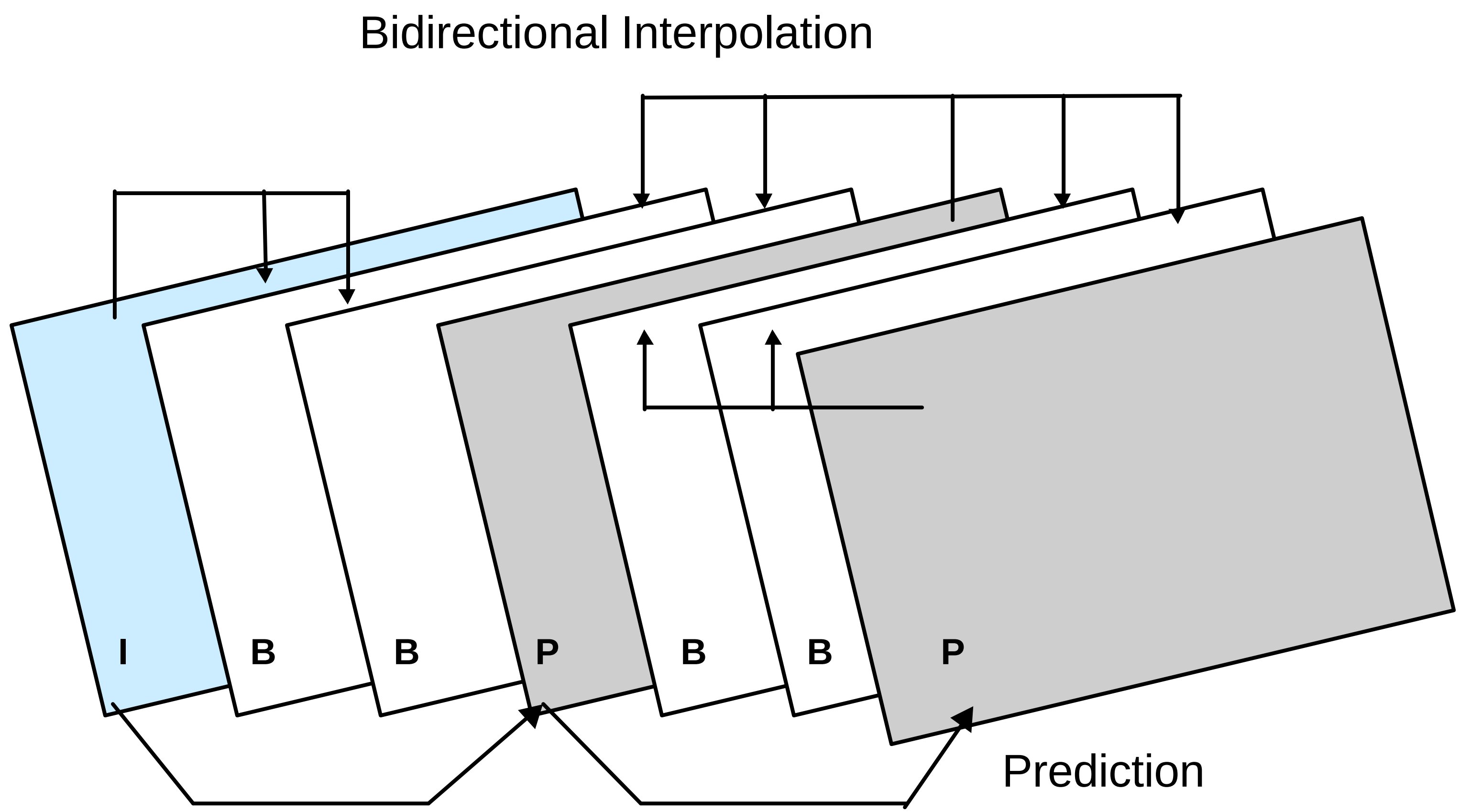


MPEG Colour Sub-sampling 4:2:0





Motion Compensation

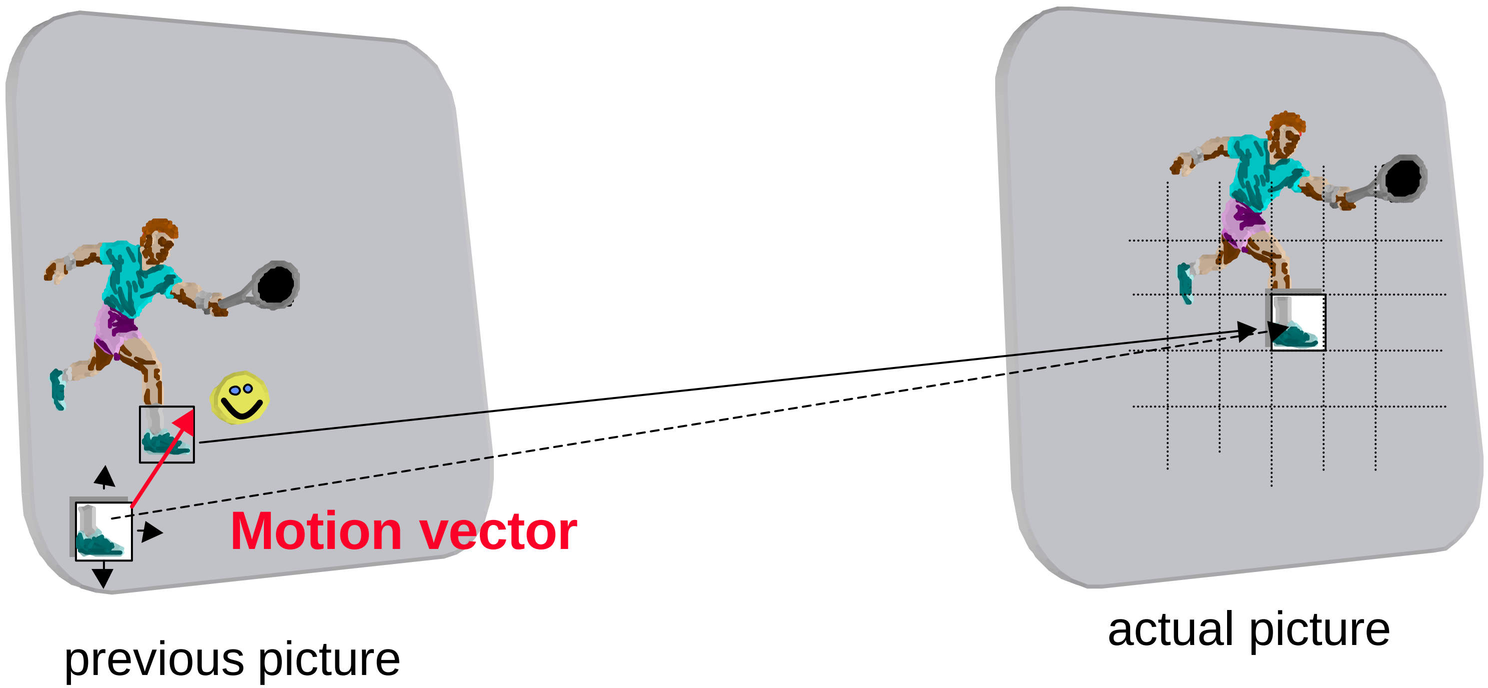


I = Intra-Frame
B = Bi-directionally interpolated frame
P = Predicted frame

- From a video signal (stream of pictures) it is not necessary to send every picture
- Whole picture is needed only when all the content is changed!
- Several pictures has to be buffered to memory to make prediction forward and bacward

Motion Compensation

- Try to match each block in the actual picture to content in the previous picture. Matching is made by shifting each of the 8 x 8 blocks of the two successive pictures pixel by pixel each direction -> Motion vector
- Substract the two blocks -> Difference block
- Transmit the motion vector and the difference block



Transmission Order of the Frames

At the encoder input:

1	2	3	4	5	6	7	8	9	10	11	12	13
I	B	B	P	B	B	P	B	B	I	B	B	P

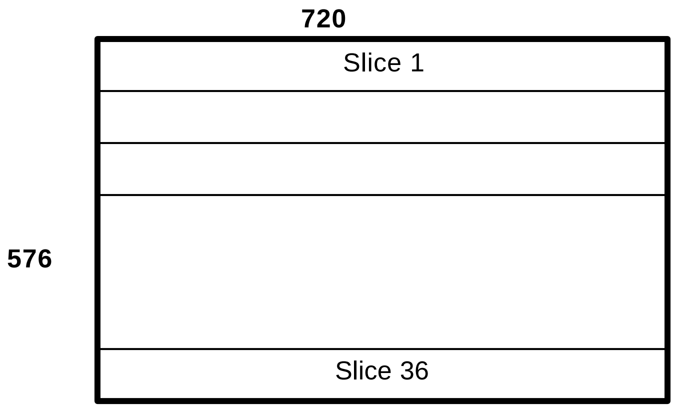
At the encoder output, in the coded bit stream and at the decoder input:

1	4	2	3	7	5	6	10	8	9	13	11	12
I	P	B	B	P	B	B	I	B	B	P	B	B

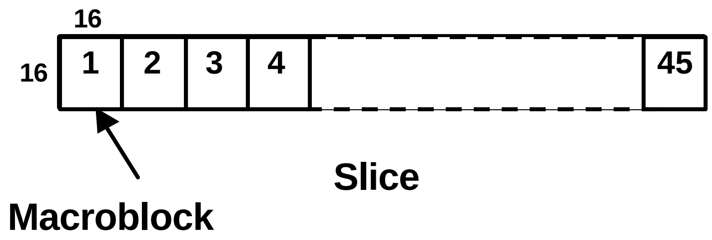
At the decoder output:

1	2	3	4	5	6	7	8	9	10	11	12	13
I	B	B	P	B	B	P	B	B	I	B	B	P

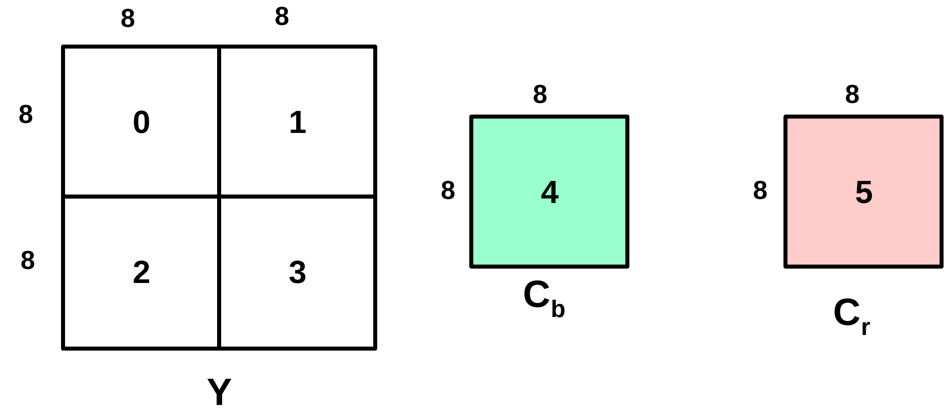
Slice and Macroblock Structure



Frame



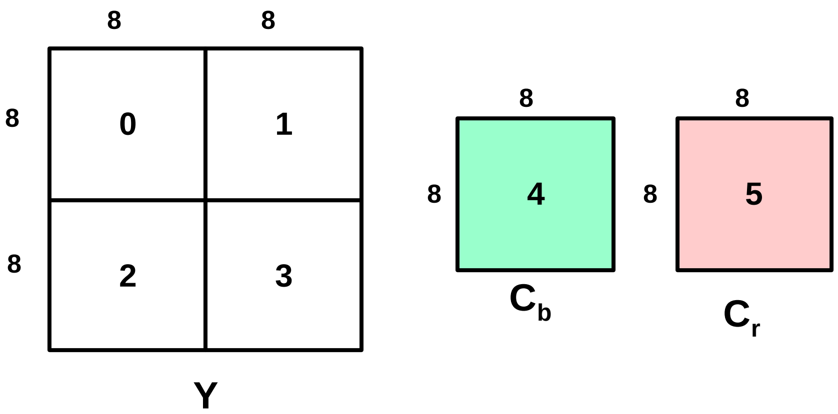
Colour subsampling 4:2:0



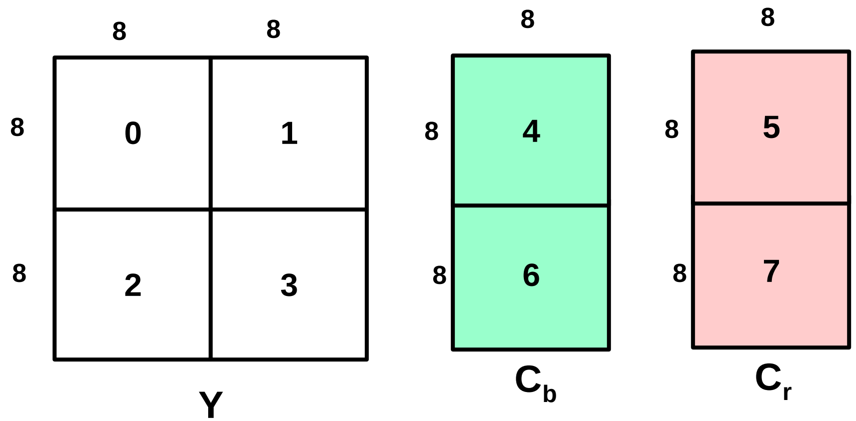
DCT blocks

Colour Subsampling

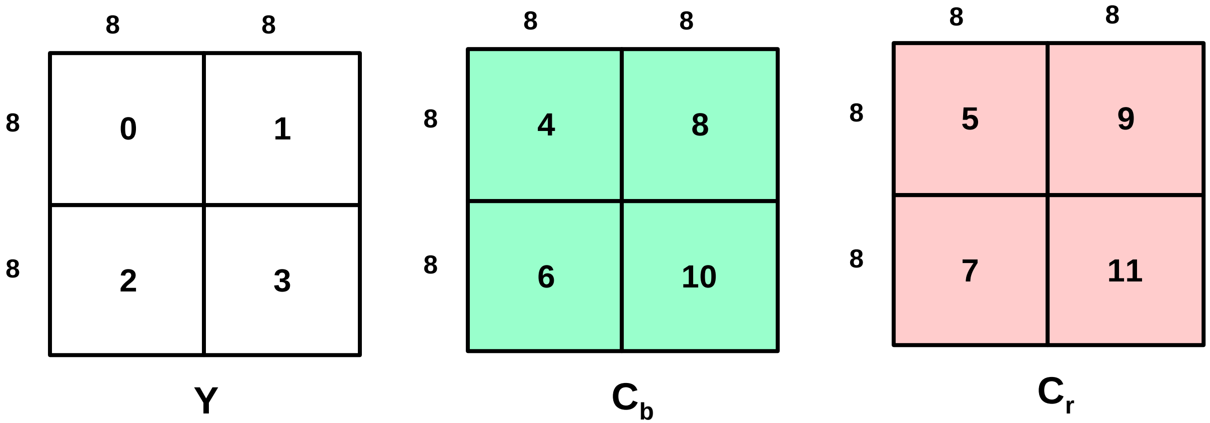
Colour subsampling 4:2:0



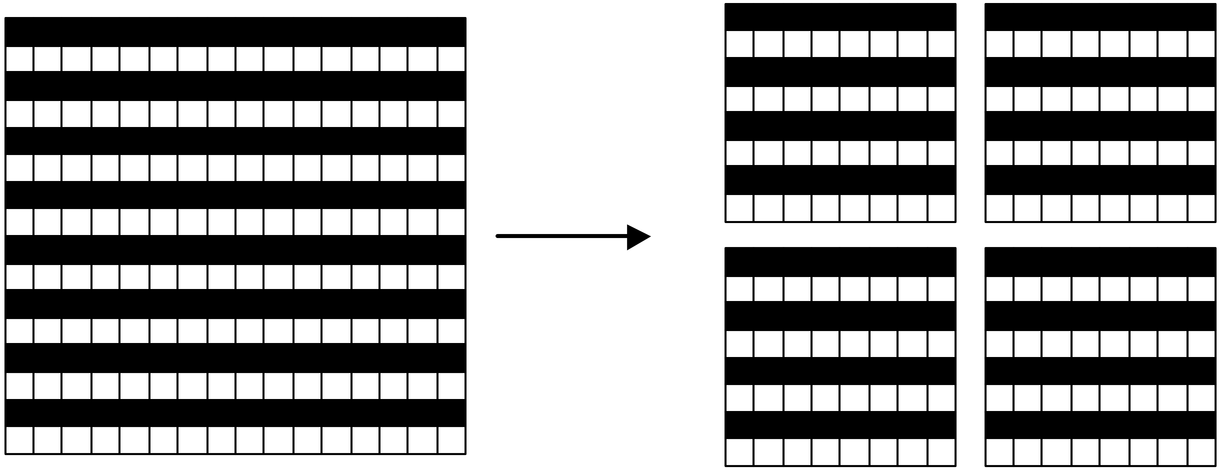
Colour subsampling 4:2:2



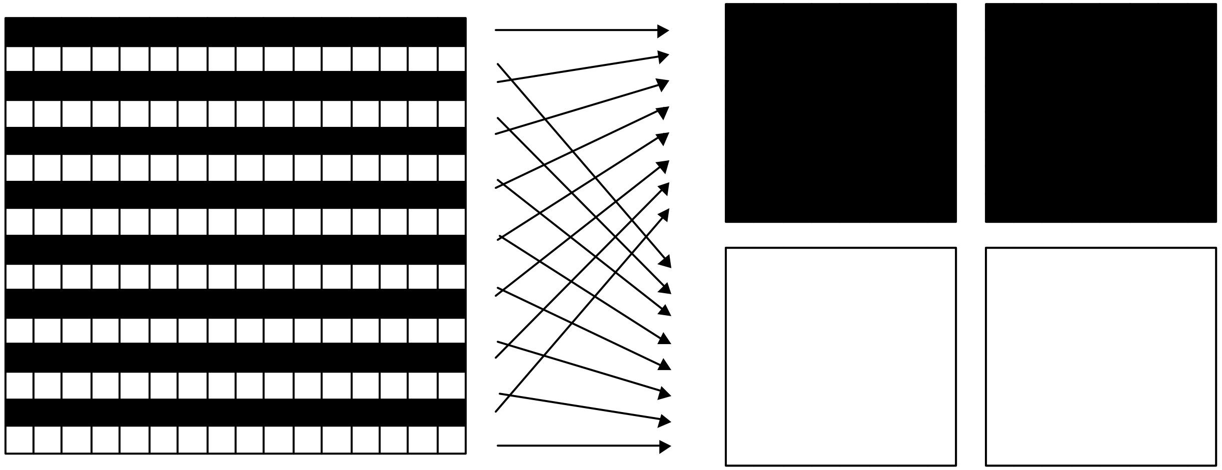
Colour 'subsampling' 4:4:4



Field vs. Frame DCT



Mapping from 16 x 16 block to 8 x 8 blocks for <u>frame-organized</u> data

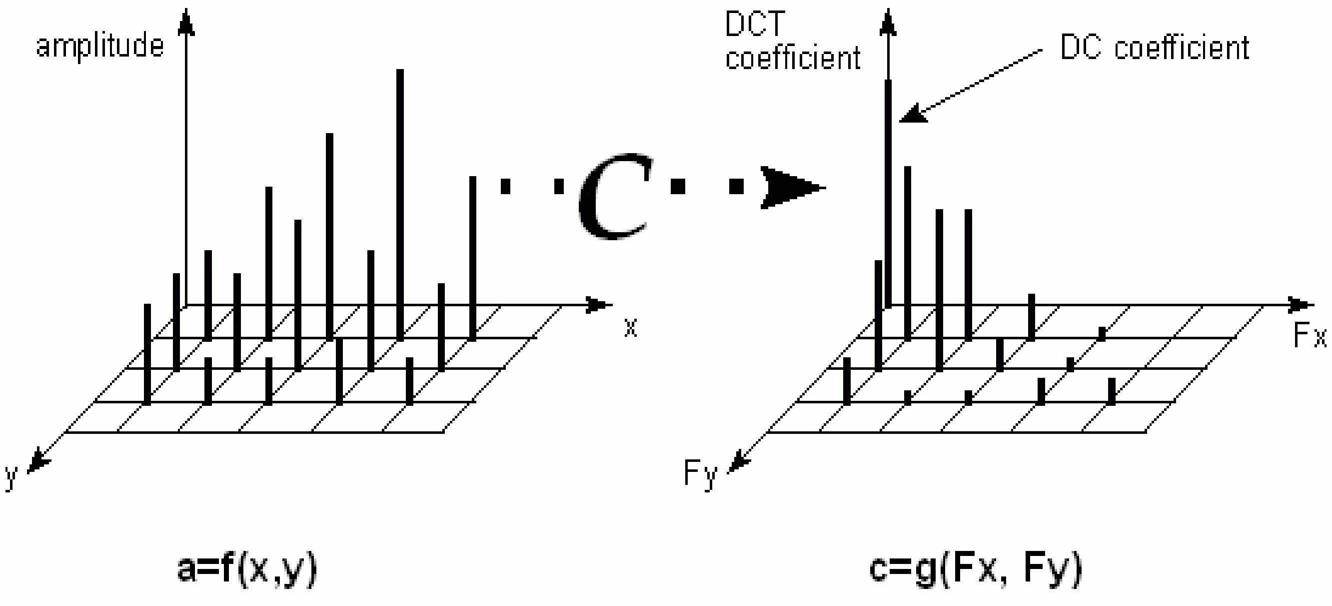


Mapping from 16 x 16 block to 8 x 8 blocks for <u>field-organized</u> data

© Mauri Kangas 2002

Discrete (Digital) Cosine Transform is A Way to Arrange Information

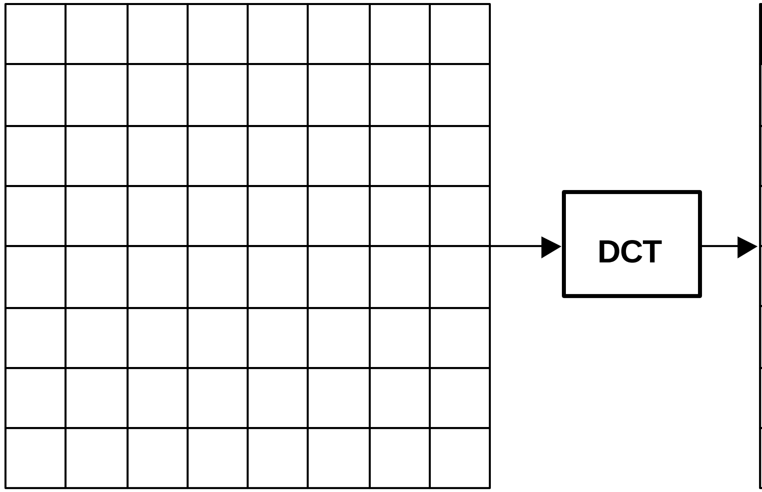
(Luminance, chrominance)



Samples from Picture **”Arranged” samples**

DCT - Discrete Cosine Transform

Input pixel block 8x8, 8-9 bits/sample



Coefficient block 8x8, ≥ 12 bits/sample; default quantization matrix coefficients

DC	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	35	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

Increasing vertical frequency

Increasing horizontal frequency

8 x 8 DCT:

$$F(u,v) = \frac{C(u)C(v)}{4}\left[\sum_{x=0}^{7}\sum_{y=0}^{7} f(x,y)\cos\left(\frac{(2x+1)up}{16}\right)\cos\left(\frac{(2y+1)vp}{16}\right)\right], \qquad u,v = 0,\ldots,7$$

8 x 8 IDCT:

$$f(x,y) = \frac{1}{4}\left[\sum_{u=0}^{7}\sum_{v=0}^{7} C(u)C(v)F(u,v)\cos\left(\frac{(2x+1)up}{16}\right)\cos\left(\frac{(2y+1)vp}{16}\right)\right], \qquad x,y = 0,\ldots,7$$

 © Mauri Kangas 2002

Quantization

- After quantization DCT coefficients are 12 bits or more, while the sarting data was 8 (-9) bits/pixel
- Quantizer step size in the decoder:

  SS = QF[m, n] x QS

  where

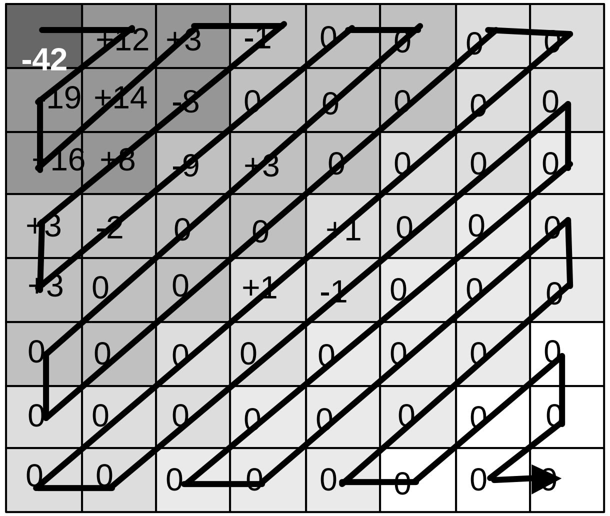
  * QF[m, n] is dependent from the location of the coefficient within the DCT block; in the default case QF[m,n] is within the range (16...83) for the intra frames and 16 for other frames
  * QS is the base quantizer step size
- standard gives one default quantisation matrix (same for luminance and chrominance) for intra frames and one quantisation matrix (same for luminance and chrominance)
  - with colour sub-sampling 4:2:0 only two matrices are used (one for luminance and one for chrominance)
  - with colour sub-sampling 4:2:2 and 4:4:4 four matrices are used (intra and non-intra matrix for luminance and (intra and non-intra matrix for chrominance)
  - when donloading these matrices the same matrix can be speified for both luminance and chrominance

Scanning the Quantized DCT Coefficients

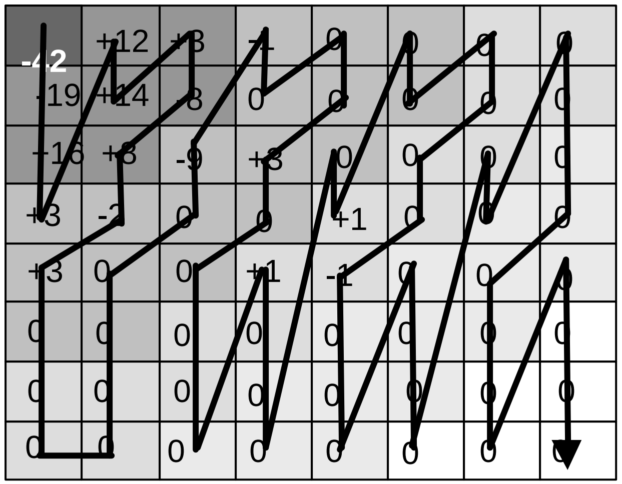
-42	+12	+3	-1	0	0	0	0
-19	+14	-8	0	0	0	0	0
+16	+8	-9	+3	0	0	0	0
+3	-2	0	0	+1	0	0	0
+3	0	0	+1	-1	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Increasing vertical frequency

Increasing horizontal frequency



Zigzag Scan

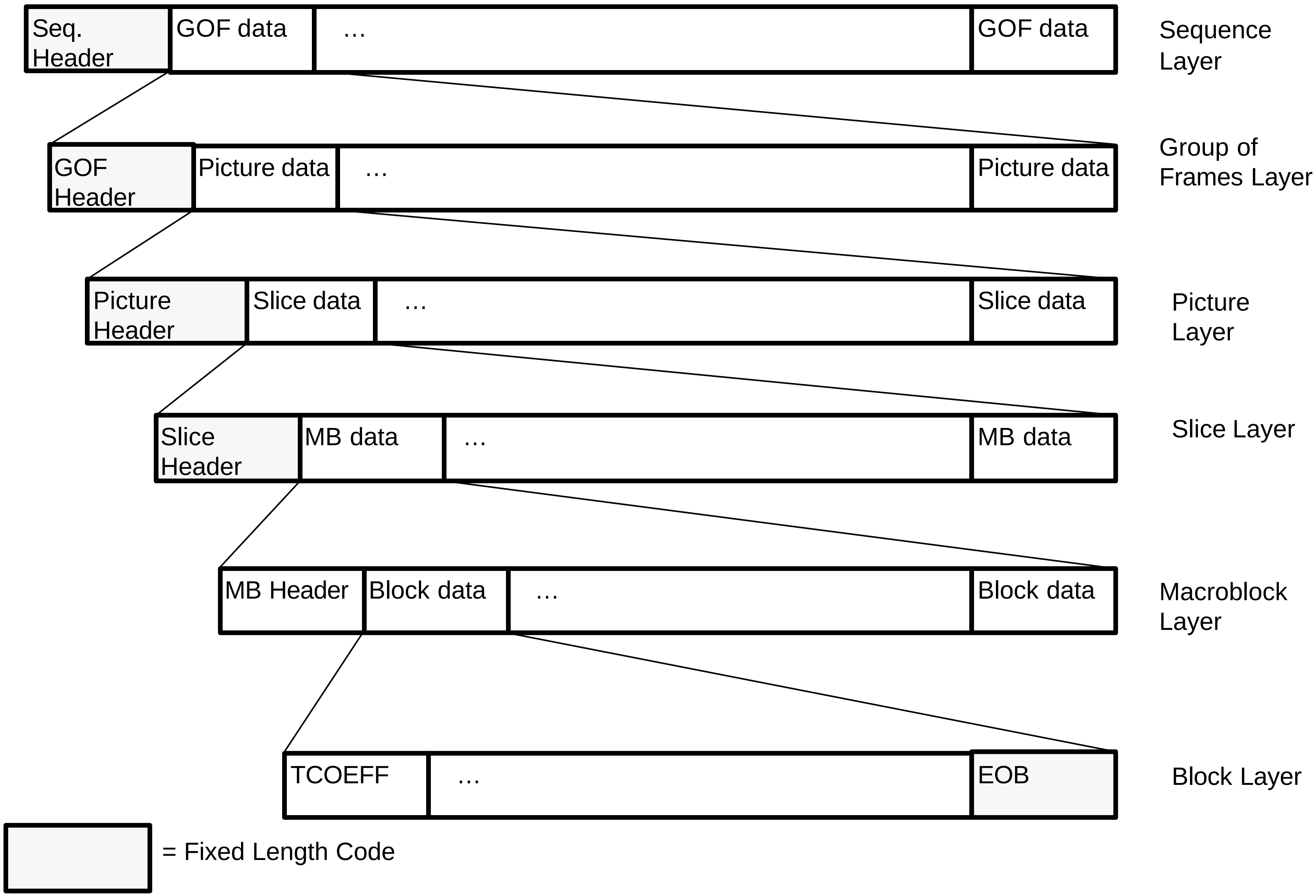


Vertical Scan

Variable-Length Coding

- Variable length coding is used for (example):
  - quantized coefficients
  - macroblock prediction type
  - motion vectors
  - etc.
- A Modified Huffman code is used throughout MPEG
  - optimal variable-length coding for the chosen alphabet
  - an "ESCAPE" code is added to the code; when some input symbol is recognized, which doesn't belong to a high-probability symbol set, the input symbol is inserted as such preceded by the "ESCAPE" code
- The quantized DCT coefficients are taken into a 64-symbol vector by reading the 8 x 8 DCT block in zigzag or vertical scan order
- Then run-length amplitude coding is used
  - first the DC quantized coefficient receives its own Huffman code
  - because there is a edundancy between adjacent DC quantized coefficients in non-predicted blocks, only the difference between these is Huffman coded
  - the remaining quantized coefficients are parsed into a sequence of runs, where a run is defined as zero or more zeroes followed by a single nonzero value.

MPEG-2 Bitstream Syntax



Features Supported by the MPEG-2 Algorithm

- Different chrominance sampling formats (i.e., 4:2:0, 4:2:2, and 4:4:4) can be represented
- Video in both the progressive and interlaced scan formats can be encoded
- The decoder can use 3:2 pull down to represent a ~24 fps film as ~30 fps video
- The displayed video can be selected by a movable pan-scan window within a larger raster
- A wide range fo picture qualities can be used
- Both constant an variable bit rate channels are supported
- ISO/IEC 11172-2 bit streams are decodable
- Bit streams for high and low (hardware) complexity decoders can be generated
- Editing of encoded video is supported
- The encoded bit stream is resilient to errors

MPEG-2 Profiles

The profiles are specific subset of the bit stream syntax in the MPEG-2 standard (profile-P@level-L):

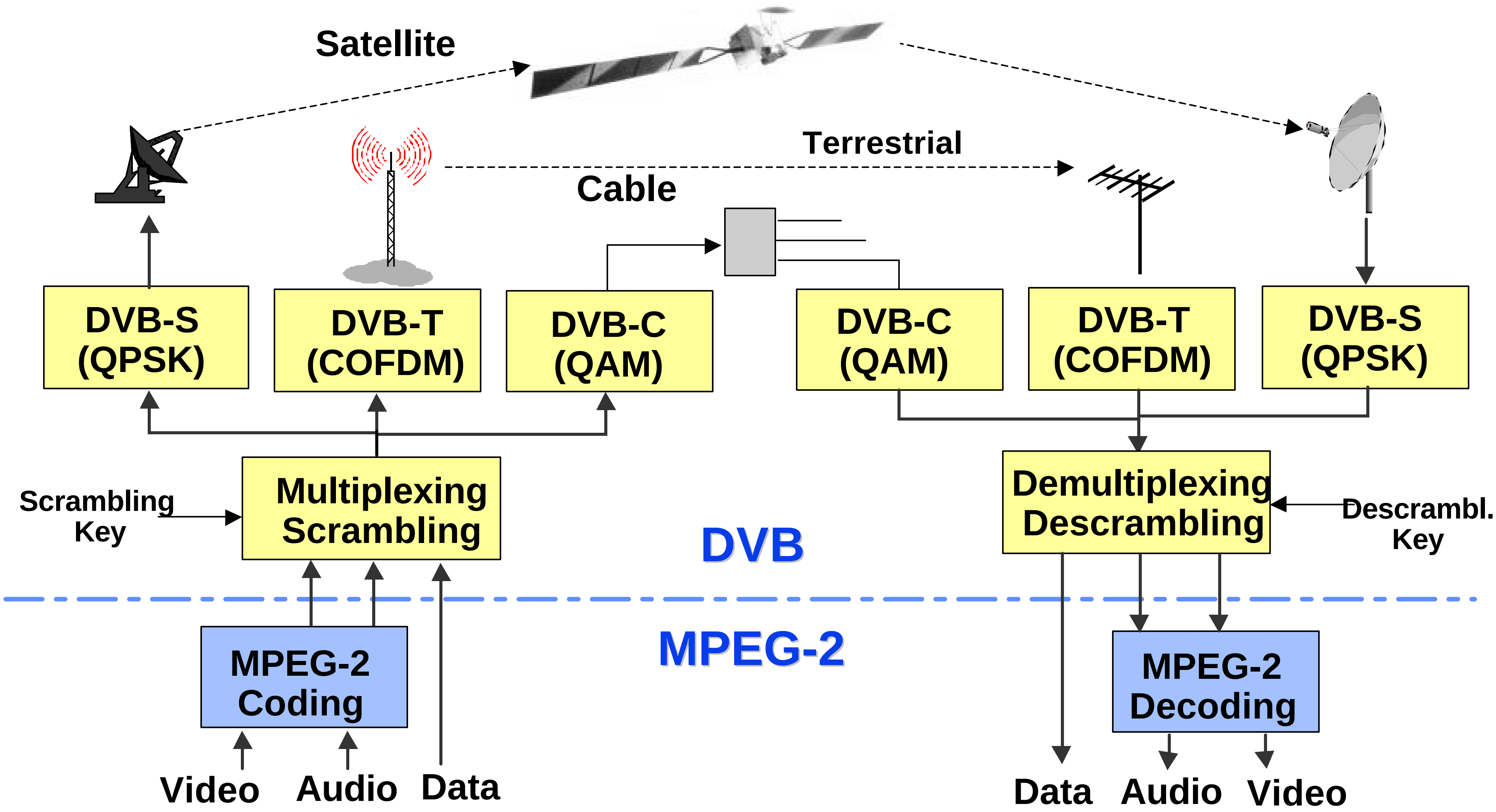
- **Main Profile** was designed to accommodate most initial applications of MPEG-2, in terms of both functionality requirements and cost constraints
- **High Profile** has more functionalities thn Main Profile; allows SNR, spatial and an additional type of scalability giving high quality picture quality when all features are utilized in the decoder
- **Simple profile** is intended for low cost applications; no B-pictures (8 Mbits of memory required)
- **Spatial Scalable Profile** can provide two layer coding with different resolutions on layers (low resolution reproduction and combination gives full-resolution reproduction)
- **SNR Scalable Profile** provides layers with the same pixel resolution by different picture quality (quantization level); the first stream gives a reasonable picture quality and the other stream gives a refinement to the fisrt stream reproduction

MPEG-2 Levels

A level is a defined set of constraints imposed on the parameters of the MPEG-2 bit stream (profile-P@level-L):

- **Main Level** is to be used by initial applications of MPEG-2. Upper bounds of the sampling density correspond to CCIR601 picture format: 720 x 576 (PAL, 25 Hz) or 720 x 480 (NTSC, 30 Hz)
- **High Levels** are intended for HDTV systems. The High Level supports 1920 pixels per line (1920 x 1152), and the High-1440 Level 1440 pixels per line respectively (1440 x 1152)
- **Low Level** corresponds to the quarter-CCIR601 picture format (SIF)

DVB Systems


Satellite
Terrestrial
Cable
DVB-S
(QPSK)
DVB-T
(COFDM)
DVB-C
(QAM)
DVB-C
(QAM)
DVB-T
(COFDM)
DVB-S
(QPSK)
Scrambling
Key
Multiplexing
Scrambling
Demultiplexing
Descrambling
Descrambl.
Key
DVB
MPEG-2
MPEG-2
Coding
MPEG-2
Decoding
Video
Audio
Data
Data
Audio
Video

Compression Efficiency

- **Uncompressed studio quality TV Transmission (Standard definition):**

  576 lines, 720 samples/line

  8 bits/sample

  50 pictures/second (non interlaced)
  => 165 888 000 bits/second (black and white) or
  => 331 776 000 color (color with 50% of luminance resolution)

- **MPEG-1 coding 1,15 Mbit/second (quality: VHS)**

  (25 pict/s, 288 x 360 pixels, non-interlaced)

- **MPEG-2 coding**

  News: 2..3 Mbit/s

  Movies: 2...4 Mbit/s, (24 pict/s)

  Sports: 4...8 Mbit/s

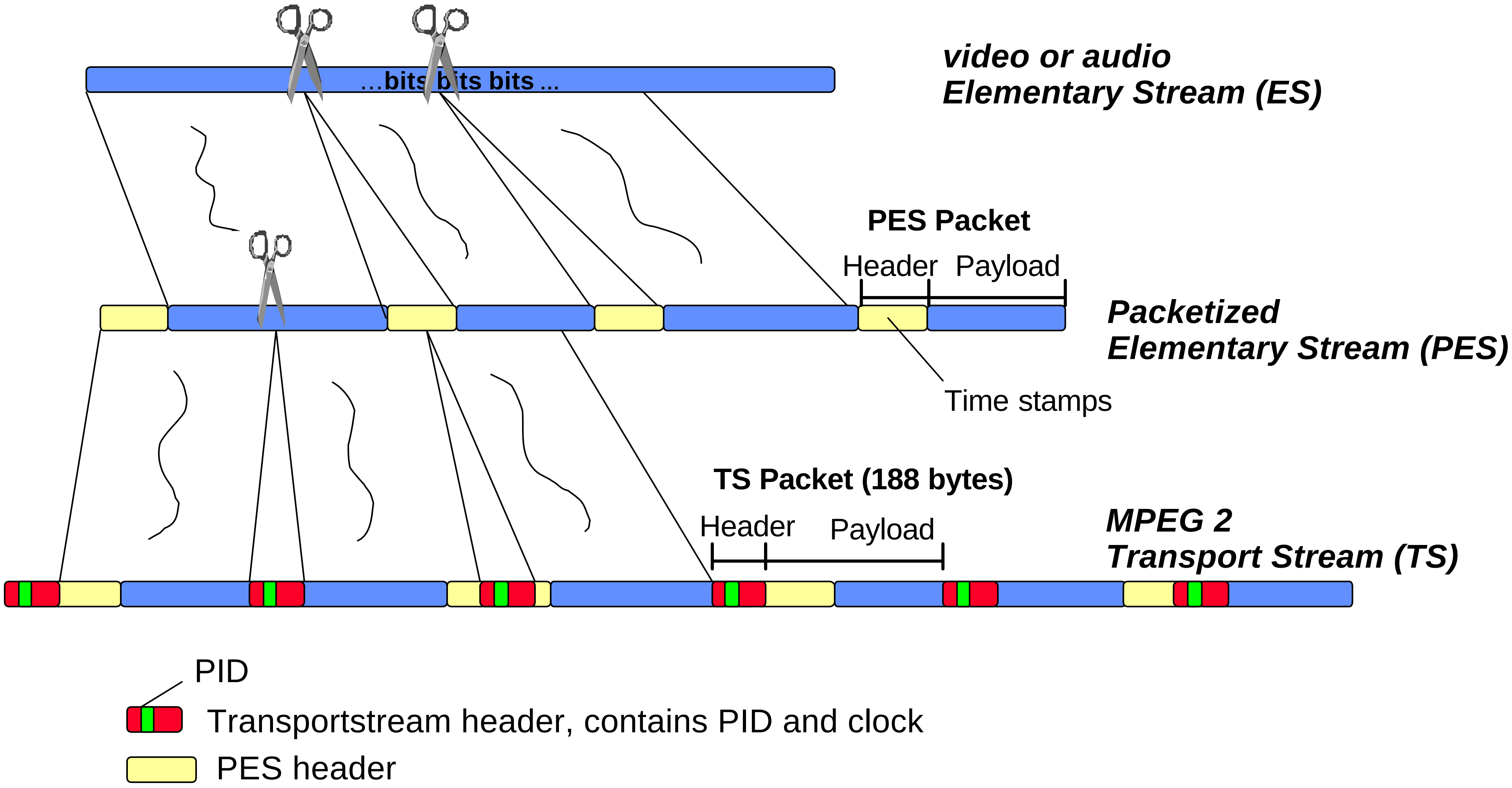
  HDTV: 15...20 Mbit/s

- **MPEG-4 coding**

  VHS quality video: 500...700 kbit/s

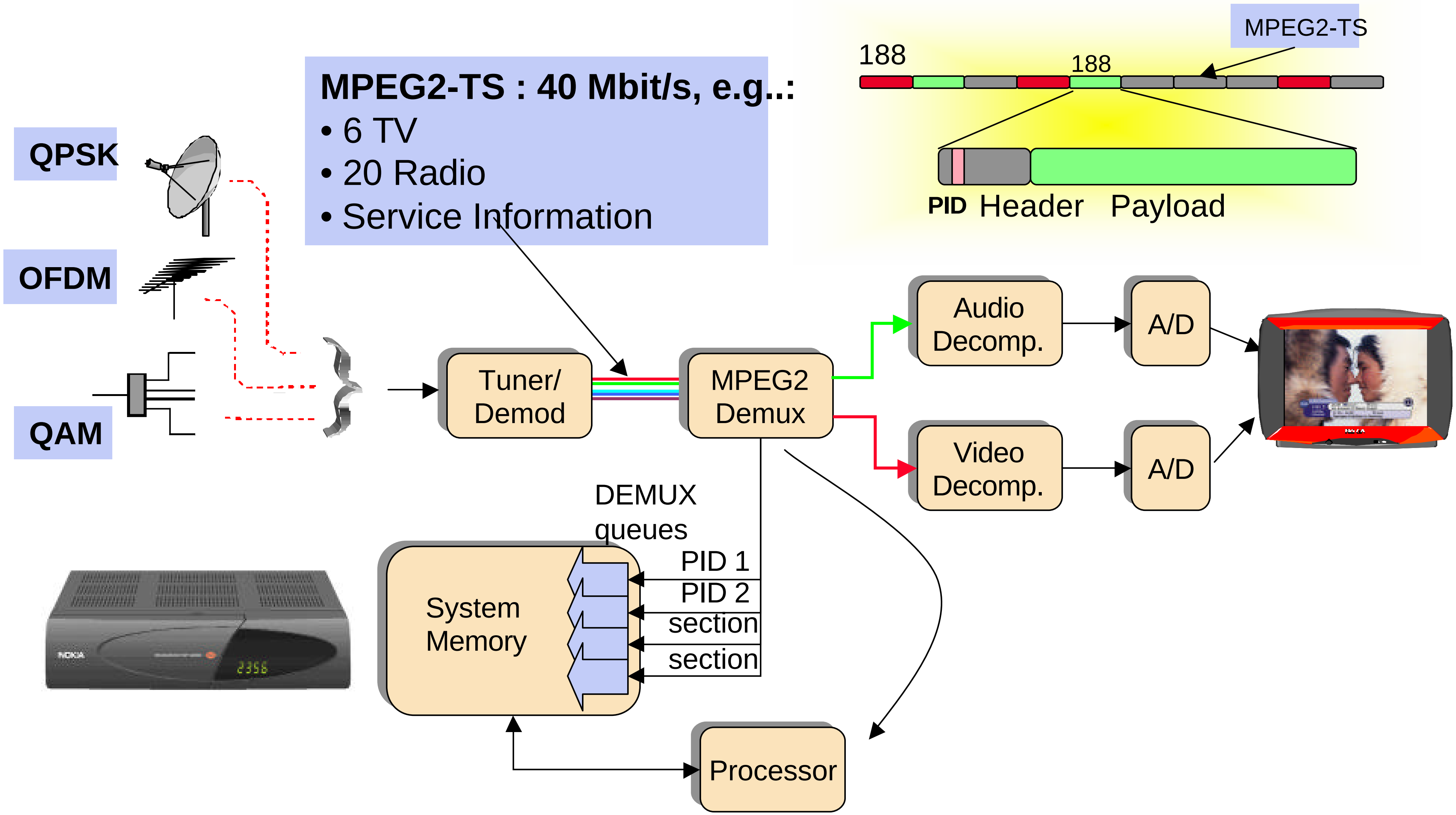
  Movies: 1...2 Mbit/s

The MPEG Transport Stream



Rule: Every elementary stream gets its own (Packet ID) PID

Processing of The Streams in The STB



 © Mauri Kangas 2002