12 Video Coding Standards

Application Areas of Video Compression

Application area	Typical data rate	Standard employed
Digital cameras	1…70 Mbit/s 8…50 Mbit/s	"Motion JPEG", H.264/AVC
Digital cinema	Up to 250 Mbit/s	Motion JPEG2000
Digital television broadcast	SD: 2…6 Mbit/s HD: 10…20 Mbit/s	MPEG-2, (H.264/AVC)
DVD Blue-ray Disk	5…8 Mbit/s Up to 40 Mbit/s	MPEG-2 MPEG-2, VC-1
Internet video streaming	100 kbit/s…2 Mbit/s	MPEG-1, H.264/AVC, VP-9, proprietary codecs
Video telephony	30 kbit/s…2 Mbit/s	H.261, H.263, H.264/AVC
Video over 3G	100…500 kbit/s	MPEG-4, H.264/AVC
4G Video, IPTV	500 kbit/s…10 Mbit/s	H.265/HEVC

Common Digital Video Formats

Spatial resolution

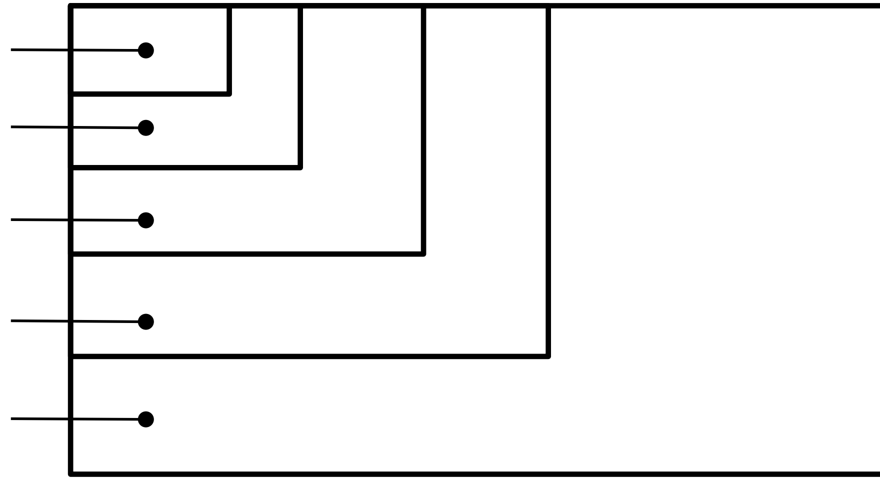
VGA, 640x480, 4:3

ITU-R 601, 720x576, 4:3

ITU-R 709 720p, 1280x720, 16:9

ITU-R 709 1080i, 1920x1080, 16:9

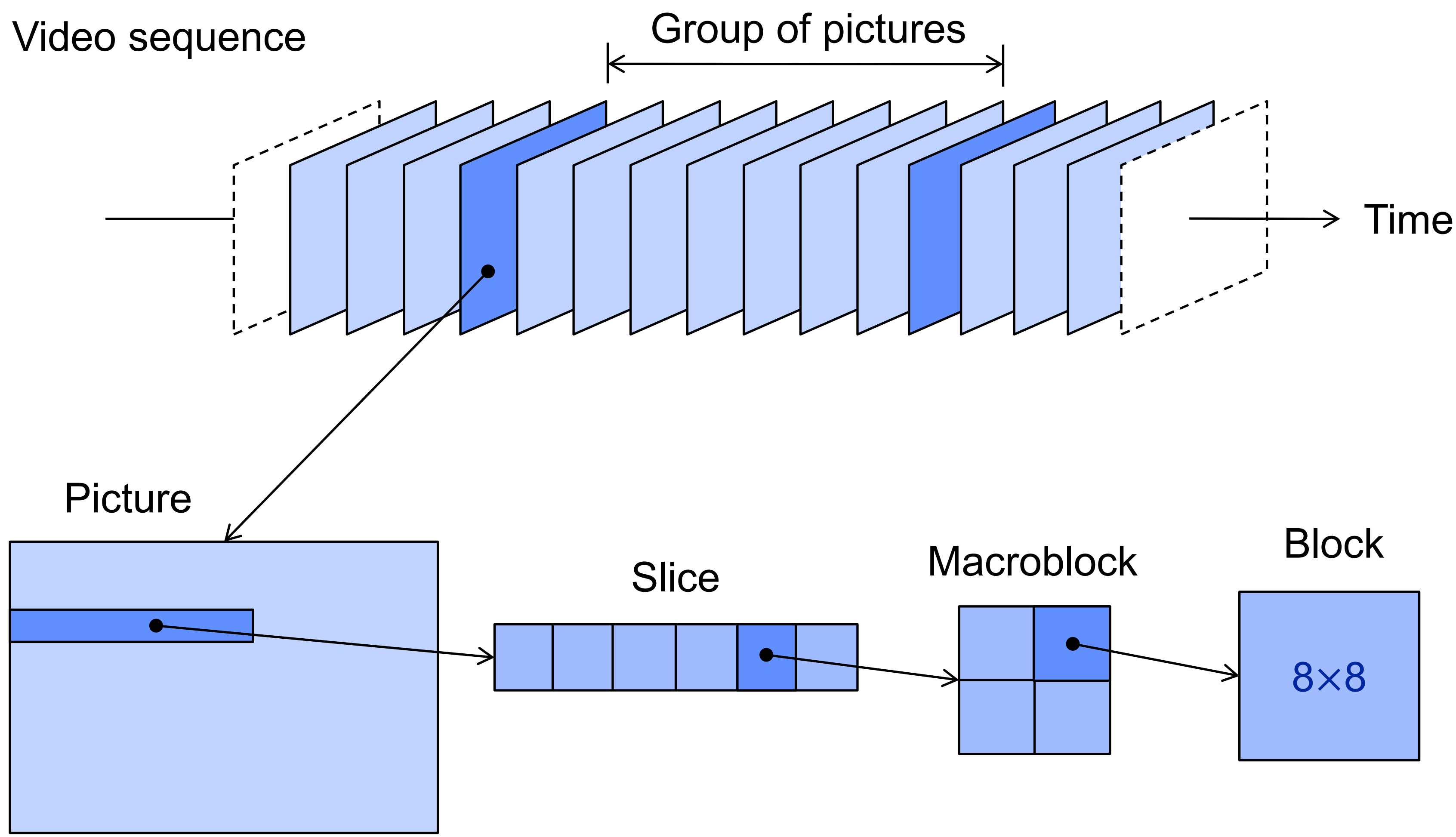
ITU-R 2020, 3840x2160, 16:9



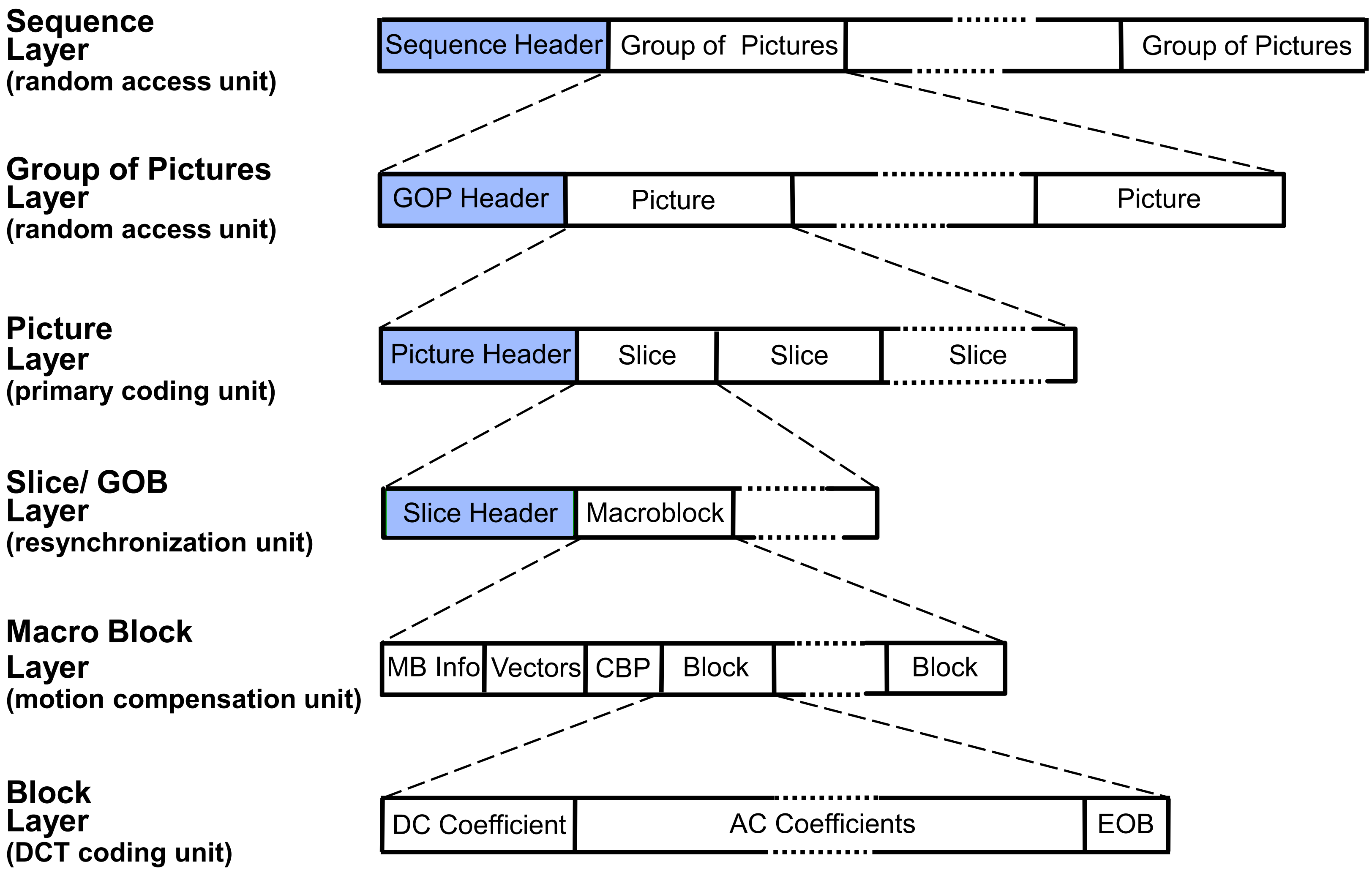
Temporal resolution

Format	Line scanning	Frame rate
VGA	Progressive	24…60 Hz
ITU-R BT.601 “SDTV”	Interlaced	25 Hz
ITU-R BT.709 720p/50 “HD ready”	Progressive	50 Hz
ITU-R BT.709 1080i/25 “Full HD”	Interlaced	25 Hz
ITU-R BT.709 1080p/60 “Full HD”	Progressive	60 Hz
ITU-R BT.2020 “Ultra HD“	Progressive	24…120 Hz

Hierarchical Syntax in Video Compression Standards



Hierarchical Picture Syntax in Video Bitstreams

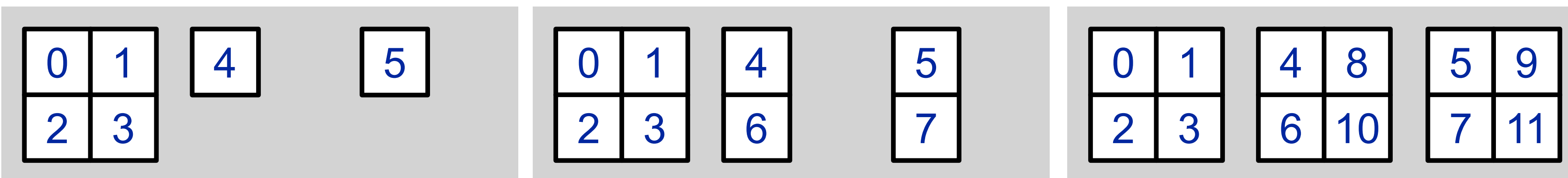


Macroblocks

Basic coding unit macroblock (MB) comprises 16×16 pixels in luminance image plus associated chrominance samples

⇨ Chrominance components within MB depend on sampling format

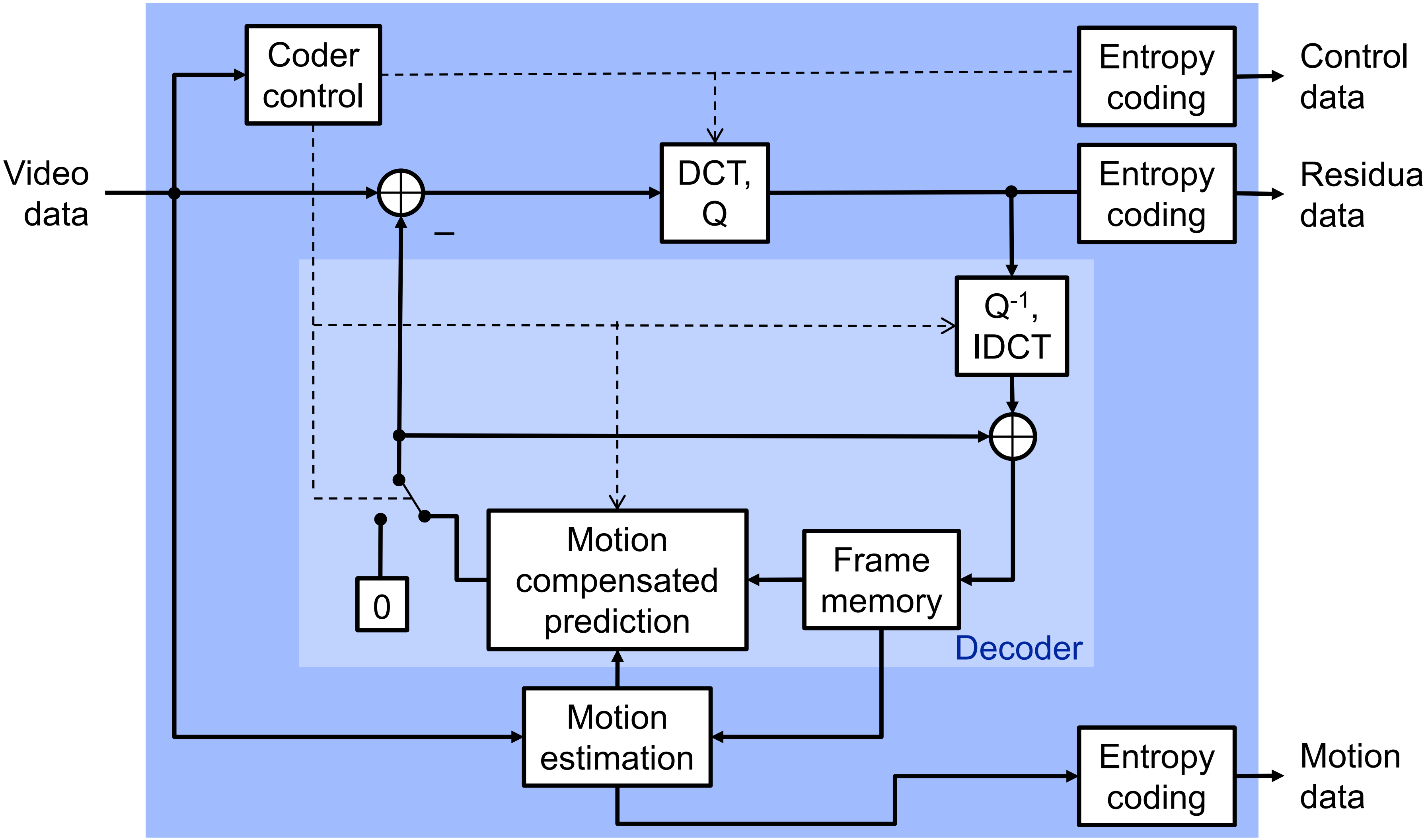
Sampling format 4:2:0	Sampling format 4:2:2	Sampling format 4:4:4



Y	C_b	C_r	Y	C_b	C_r	Y	C_b	C_r
16×16	8×8	8×8	16×16	8×16	8×16	16×16	16×16	16×16

⇨ MB consists of 4 luminance plus 2, 4 or 8 chrominance blocks respectively

12.1 ITU-T Rec. H.261 / H.263 Encoder



ITU-T Rec. H.261

Purpose: video conferencing over ISDN networks, standardized 1990

- Image format CIF or QCIF
- Frame rate 7.5…30 Hz
- Constant bit rate at p x 64 kbit/s, p = 1...30, typically 128 kbit/s

Motion compensated prediction

- Works on 16x16 macroblocks
- Integer-pixel accuracy
- Maximum displacement +/- 15 pixels
- Adaptive loop filter for noise suppression with separable impulse response

Residual coding

- 8x8 DCT, uniform quantization
- Zigzag scanning of coefficients
- Run-level coding using (zero-run, value) symbols

Macroblock Types in H.261

Prediction	MQUANT	MVD	CBP	TCOEFF	VLC
Intra				x	0001
Intra	x			x	0000 001
Inter			x	x	1
Inter	x		x	x	0000 1
Inter+MC		x			0000 0000 1
Inter+MC		x	x	x	0000 0001
Inter+MC	x	x	x	x	0000 0000 01
Inter+MC+FIL		x			001
Inter+MC+FIL		x	x	x	01
Inter+MC+FIL	x	x	x	x	0000 01

'x' means that the item is present in the coded macroblock

Abbreviations:

MC	Motion compensation flag
FIL	Filter-in-the-Loop flag
MQUANT	Quantization step size
MVD	Motion vector
CBP	Coded Block Pattern, indicates coding status of Y, C_r, C_b blocks of a MB
TCOEFF	DCT coefficients

ITU-T Rec. H.263

Purpose: video conferencing over PSTN networks, standardized in 1995 / 1999

- Image formats: sub-QCIF, QCIF, CIF, 4CIF, 16CIF
- Frame rate 5…30 Hz
- Variable bit rate ~20 kbit/s up to several Mbit/s

Motion compensated prediction

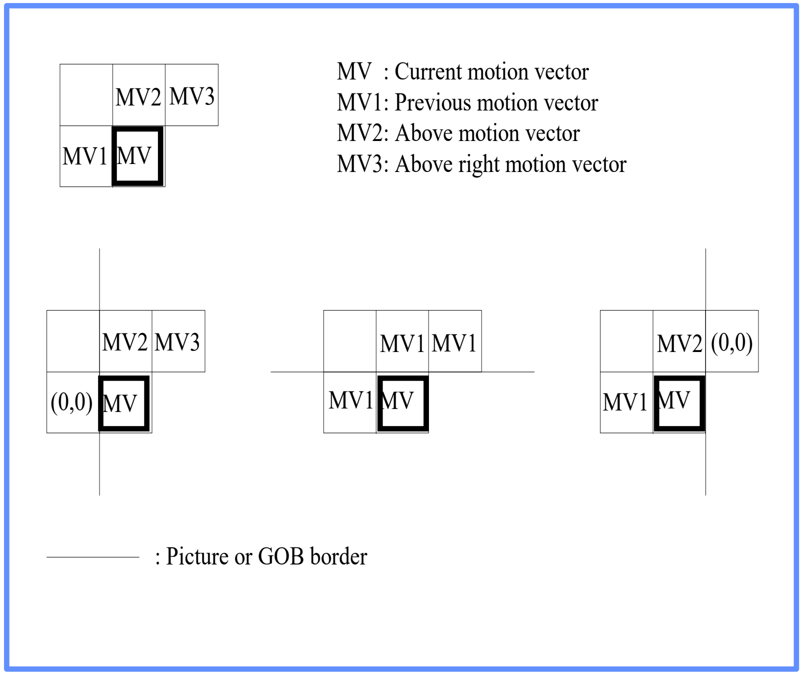
- Works on 16×16 macroblocks or 8×8 blocks
- Half-pixel accuracy
- Maximum displacement +/- 15 pixels
- No loop filter

Residual coding

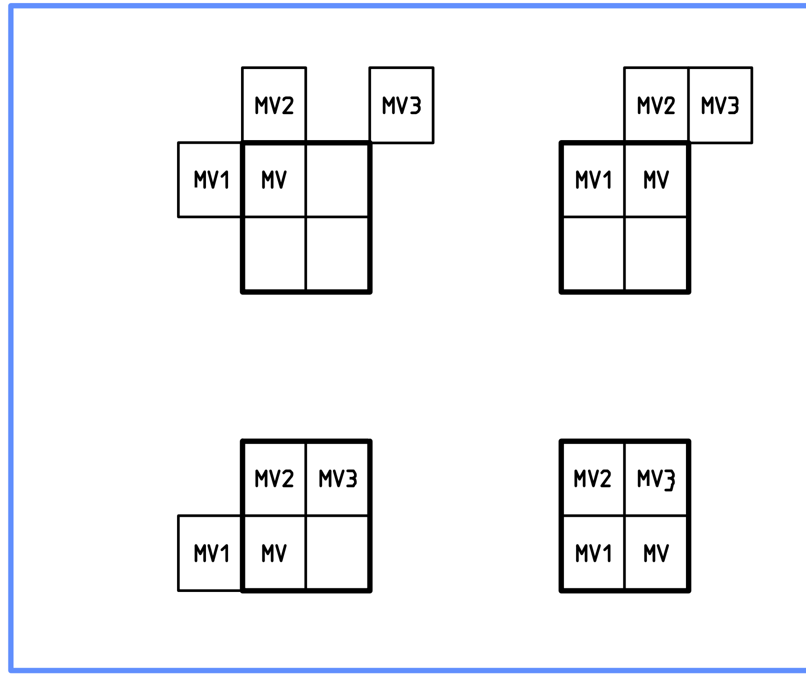
- 8x8 DCT, uniform quantization
- Zigzag scanning of coefficients
- Run-level coding using (zero-run, value) symbols

Motion Vector Prediction in H.263

Core prediction mode
(1 vector / macroblock)



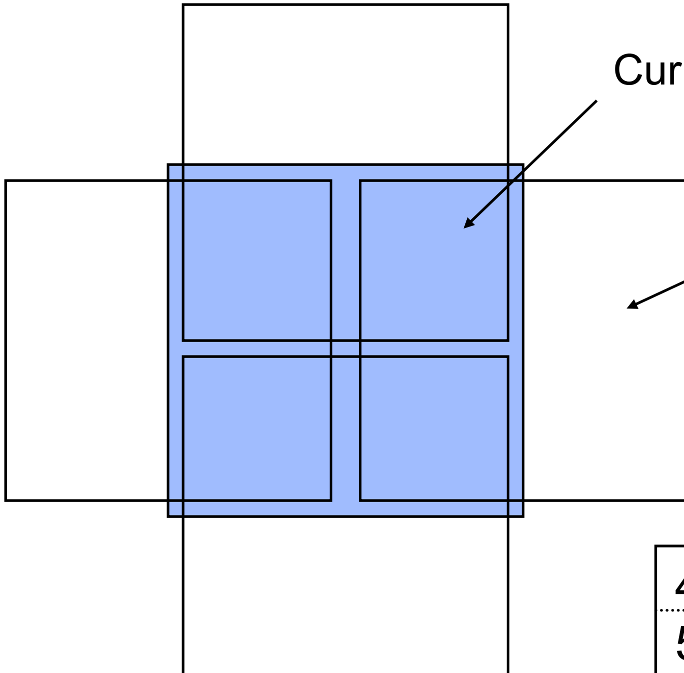
Advanced prediction mode
(4 vectors / macroblock)



☐ = Macroblock

Predictor (MV) = median (MV1, MV2, MV3)

Overlapped Block Motion Compensation in H.263



Weights

Current block

4	5	5	5	5	5	5	4
5	5	5	5	5	5	5	5
5	5	6	6	6	6	5	5
5	5	6	6	6	6	5	5
5	5	6	6	6	6	5	5
5	5	6	6	6	6	5	5
5	5	5	5	5	5	5	5
4	5	5	5	5	5	5	4

Top / bottom block

2	2	2	2	2	2	2	2
1	1	2	2	2	2	1	1
1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
1	1	2	2	2	2	1	1
2	2	2	2	2	2	2	2

Left / right block

2	1	1	1	1	1	1	2
2	2	1	1	1	1	2	2
2	2	1	1	1	1	2	2
2	2	1	1	1	1	2	2
2	2	1	1	1	1	2	2
2	2	1	1	1	1	2	2
2	2	1	1	1	1	2	2
2	1	1	1	1	1	1	2

Principle

- Each pixel is weighted sum of three prediction values
- Three neighboring motion vectors are used
- Sum is divided by 8

Mismatch Errors in Hybrid Video Codecs

Mismatch: numerical discrepancy between the reconstructed output produced by the hypothetical decoder within an encoder and the reconstructed outputs produced by a decoder. With the exception of IDCT, ITU and ISO standards define the decoding process unambiguously.

To minimize:

- IDCT accuracy must conform to IEEE P1180
- Saturation of IDCT inputs (+2047/ -2048) and outputs (+255/ -256)
- Oddification of IDCT inputs (different methods defined in H.261/H.263 and MPEG decoders)

Drift: temporal accumulation of mismatch

To minimize:

- Apply forced intra coding of blocks, e.g. in H.261/H.263 a macroblock at any spatial location may by coded at most 132 times using motion compensated prediction

Subjective Video Quality of H.261 versus H.263

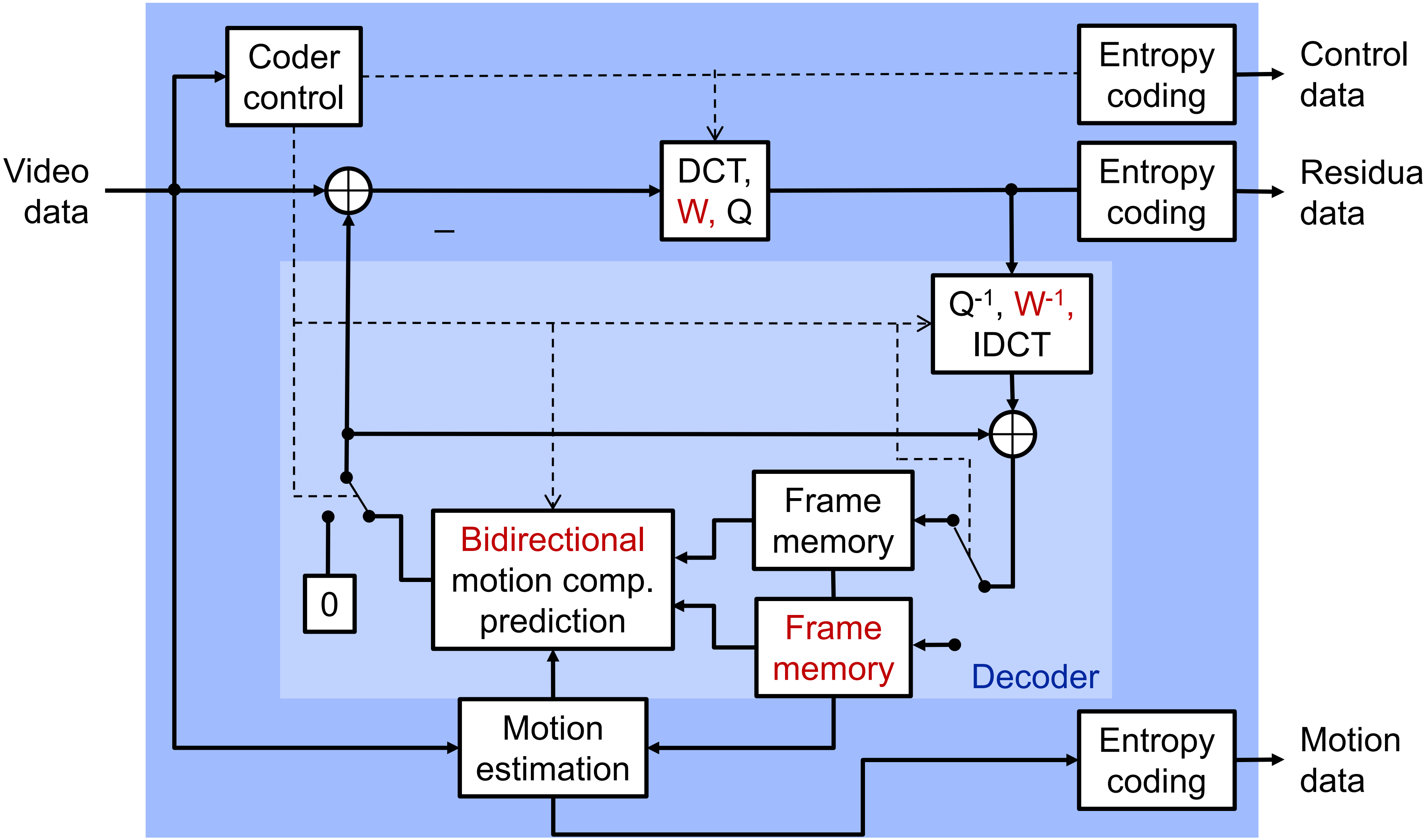
Mean opinion score (MOS) to evaluate video quality in subjective tests

MOS	5	4	3	2	1
Degradation	imperceptible	perceptible but not annoying	slightly annoying	annoying	very annoying

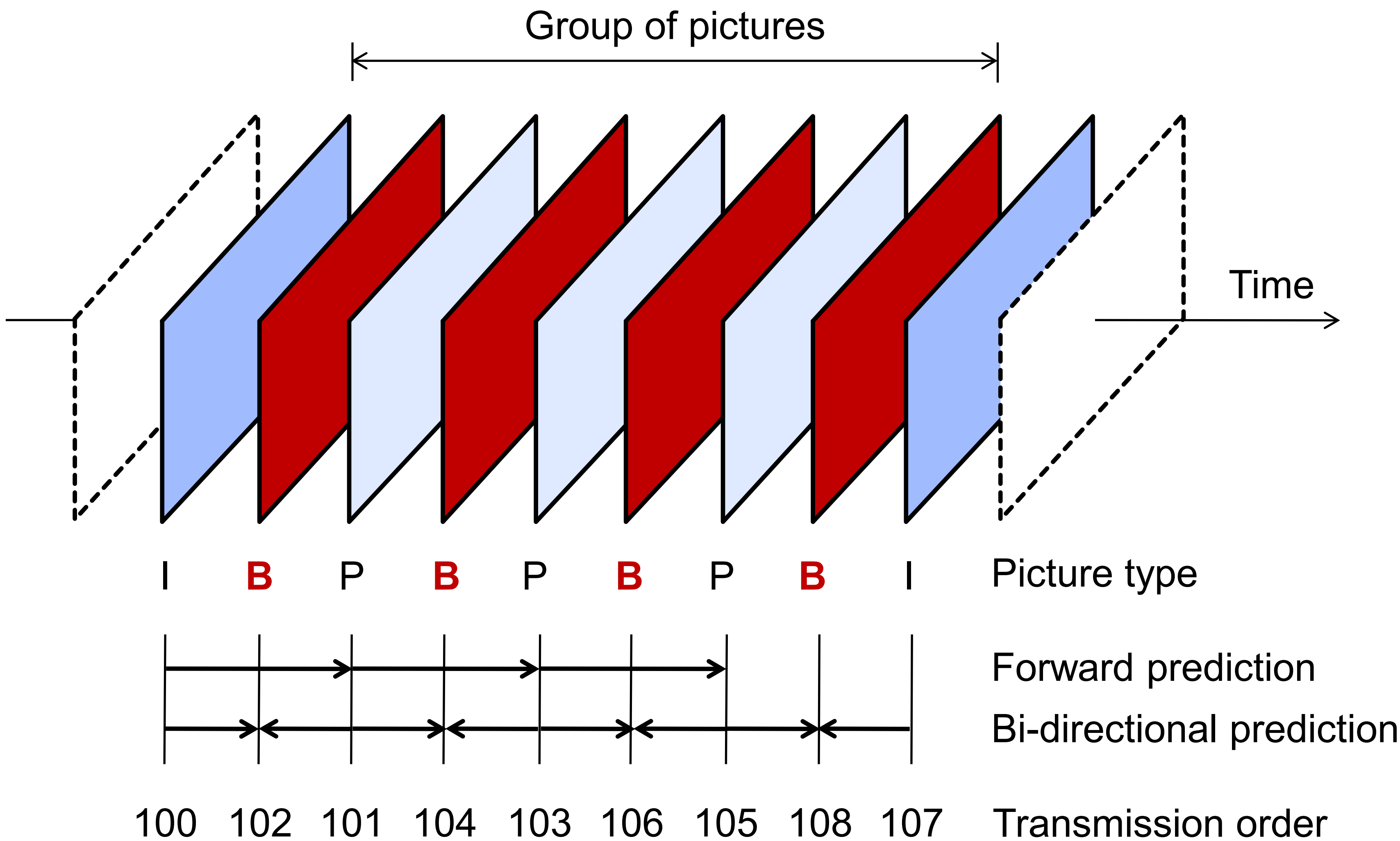
Result for subjective tests on a set of typical conversational video sequences

Connection	V.34 Modem	ISDN 1B	ISDN 2B	ISDN 6B	ISDN 24B
Data rate	22.5 kbit/s	48 kbit/s	112 kbit/s	320 kbits/s	1408 kbit/s
Frame rate	5 Hz	8.33 / 10 Hz	12.5 / 15 Hz	25 / 30 Hz	25 / 30 Hz
Image size	QCIF	QCIF / CIF	CIF	CIF	CIF
MOS H.261	1.1	1.3	2.0	3.3	4.5
MOS H.263	1.7	2.8	3.9	4.8	4.9

12.2 ISO/IEC MPEG-1 / MPEG-2 Encoder



MPEG-1 / MPEG-2 Picture Types



ISO/IEC MPEG-1

Purpose: video on CD-ROM at VHS like quality, ISO/IEC 11172 [1993]

- Image format ITU-R 601 at frame rates of 25 or 30 Hz
- Bit rate typically 1.5 Mbit/s

Coding of I-pictures

- Intraframe coded using 8x8 DCT
- Arbitrary weighting matrix for coefficients
- Uniform quantization
- Differential coding of DC coefficients
- Zigzag scan, run-level coding for AC coefficients

Coding of P-pictures

- Motion compensated prediction from already encoded I- or P-picture
- Half-pixel accuracy, maximum range +/- 512 pixels, bilinear interpolation
- Differential coding of motion vectors
- Prediction error is coded with 8x8 DCT as in I-pictures

ISO/IEC MPEG-1 (Cont.)

Coding of B-pictures: motion compensated prediction from two consecutive P-pictures (or I-picture respectively) and mode switching

Forward prediction from previous I- or P-picture
- As for P-Pictures using one motion vector per macroblock

Backward prediction from subsequent I- or P-picture
- Using one motion vector per macroblock
- Imposes one frame delay and picture re-ordering

Bi-directional prediction from two neighboring images:
- Average forward and backward prediction by interpolation
- Requires two motion vectors per macroblock, delay and re-ordering

All B-picture prediction modes have in common:

- Half-pixel accuracy, maximum range +/- 512 pixels, bilinear interpolation
- Differential coding of motion vectors
- Prediction error is coded with 8x8 DCT as in I-pictures

Demo 12 „B-Prediction“

ISO/IEC MPEG-2

Purpose: TV broadcast of interlaced video, ISO/IEC 13818 / ITU-T H.262 [1995]

- Variable image formats up to HDTV 1080p and 60 Hz frame rate
- Supports 4:2:0, 4:2:2 and 4:4:4 color format
- Bit rate typically 4…8 Mbit/s for SD, 20 Mbit/s for HD

Same coding principle as MPEG-1 with the following extensions:

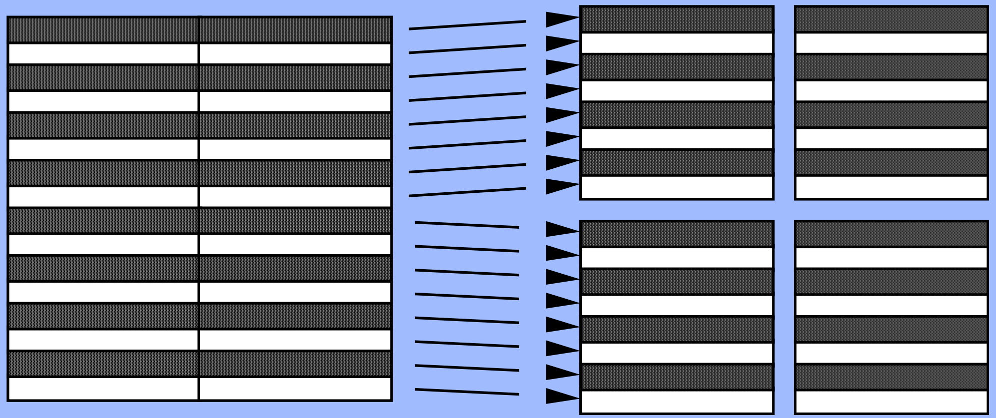
Compression of interlaced digital video
- using frame pictures and field pictures
- adaptive frame/field prediction and frame/field DCT

Improved compression efficiency
- using different quantization and VLC tables
- additional alternating coefficient scan patterns

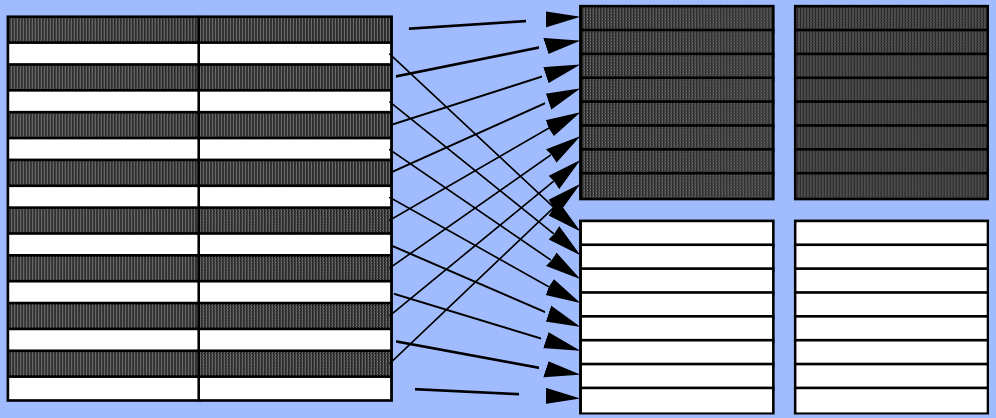
Support for spatial, temporal and SNR scalability
- not very efficient and rarely implemented

MPEG-2 Adaptive Frame/Field DCT Coding

Frame DCT Coding



Field DCT Coding



MPEG-2 Video Profiles and Levels

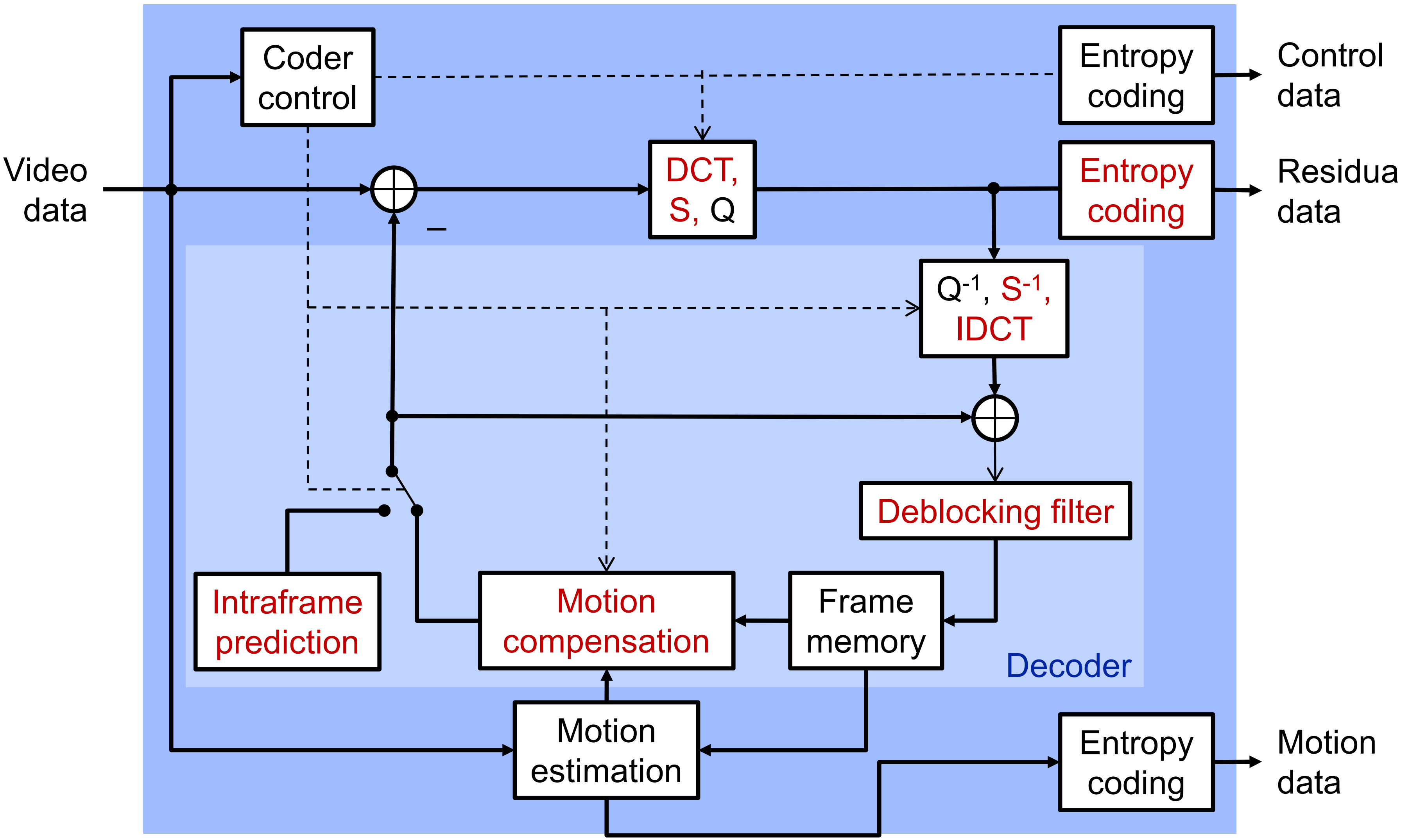
Profile defines constraints on syntax elements to limit implementation complexity
- Only subset of coding modes is allowed
- Example: forbid B-pictures or scalable coding

Level defines constraints on parameter values to limit required resources
- Image resolution or video frame rate must not exceed maximum values

PROFILE / LEVEL	SIMPLE No B-Pictures 4:2:0	MAIN 4:2:0	SCALABLE 4:2:0	HIGH 4:2:0 & 4:2:2	4:2:2
HIGH < 1920x1152x60 H D T V		1 layer rate < 80 Mb/s		3 layers # rate < 100 Mb/s	
HIGH-1440 < 1440x1152x60 H D T V		1 layer rate < 60 Mb/s	<u>SPATIAL</u> 3 layers # rate < 60 Mb/s	3 layers # rate < 80 Mb/s	
MAIN < 720x576x30 S D T V	1 layer rate < 15 Mb/s	1 layer rate < 15 Mb/s	<u>SNR</u> 2 layers # rate < 15 Mb/s	3 layers # rate < 20 Mb/s	1 layer rate < 50 Mb/s
LOW < 352x288x30 L D T V		1 layer rate < 4 Mb/s	<u>SNR</u> 2 layers # rate < 4 Mb/s		

= not defined

12.3 ITU-T Rec. H.264 / MPEG-4 AVC



Basic H.264 Coding Elements

H.264/MPEG-4 AVC “Advanced Video Coding”, standardized in 2003

Common with previous standards are

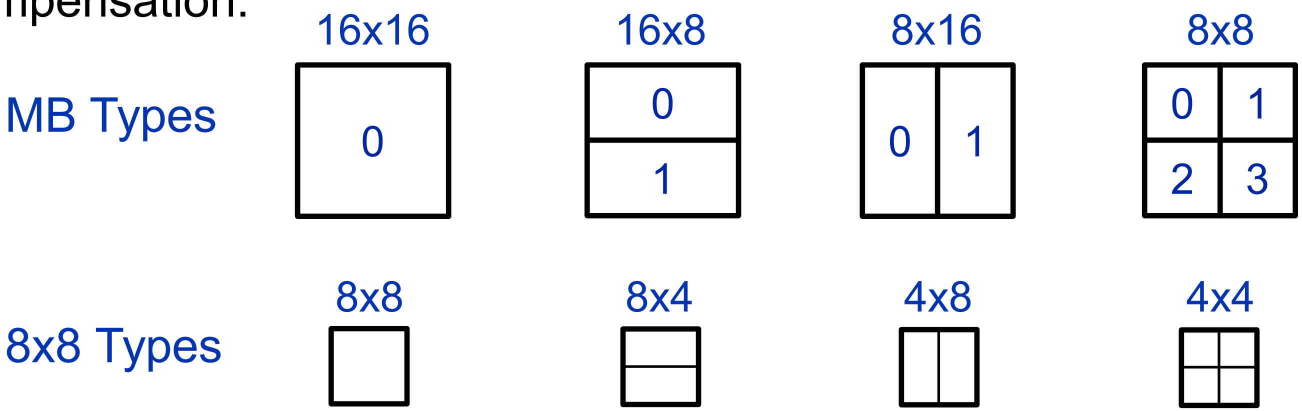
- Macroblocks-based processing
- Motion compensated prediction
- Block transform of prediction error
- Scalar quantization
- Entropy coding using VLC tables
- I-, P-, and B-pictures

Important new elements are

- Variable block-size motion compensation
- Quarter-pixel resolution of motion vectors
- Multiple reference frames for temporal prediction
- Intraframe prediction
- 4x4 integer transform for coding of prediction error
- Deblocking filter

H.264 Motion Compensation

Temporal prediction in H.264 is based on variable blocksize motion compensation:



Motion vectors have quarter-pixel accuracy obtained by two step interpolation:

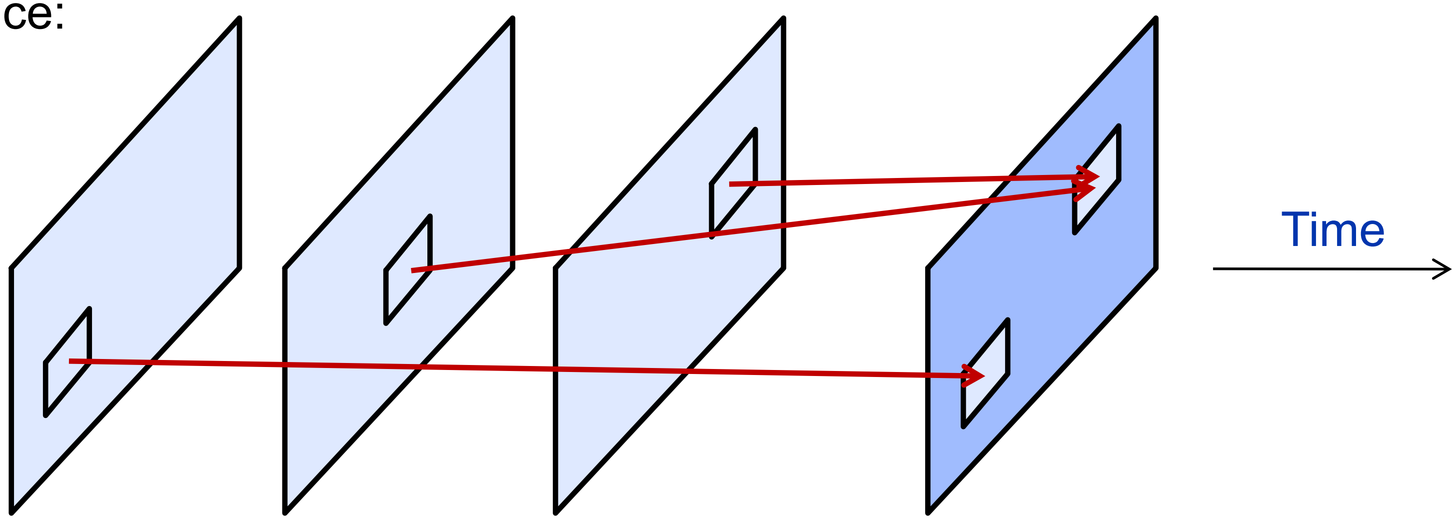
- Half-pixel positions are interpolated using impulse response function

$$h[n] = \frac{1}{32}\{1;\ -5;\ 20;\ 20;\ -5;\ 1\}$$

- Quarter-pixel positions are obtained from half-pixel positions using bi-linear interpolation

H.264 Multiple Reference Frames

Temporal prediction in H.264 can use macroblocks from multiple frames as reference:



Weighted prediction: each MB in a frame may have an individual weight assigned before overlapping (signaled per slice / or distance weighted)

Generalized B frames with two modifications with respect to MPEG-2:

- B-pictures may serve as reference for further prediction
- Reference frames for B-pictures may be both in the past

H.264 Intraframe Prediction

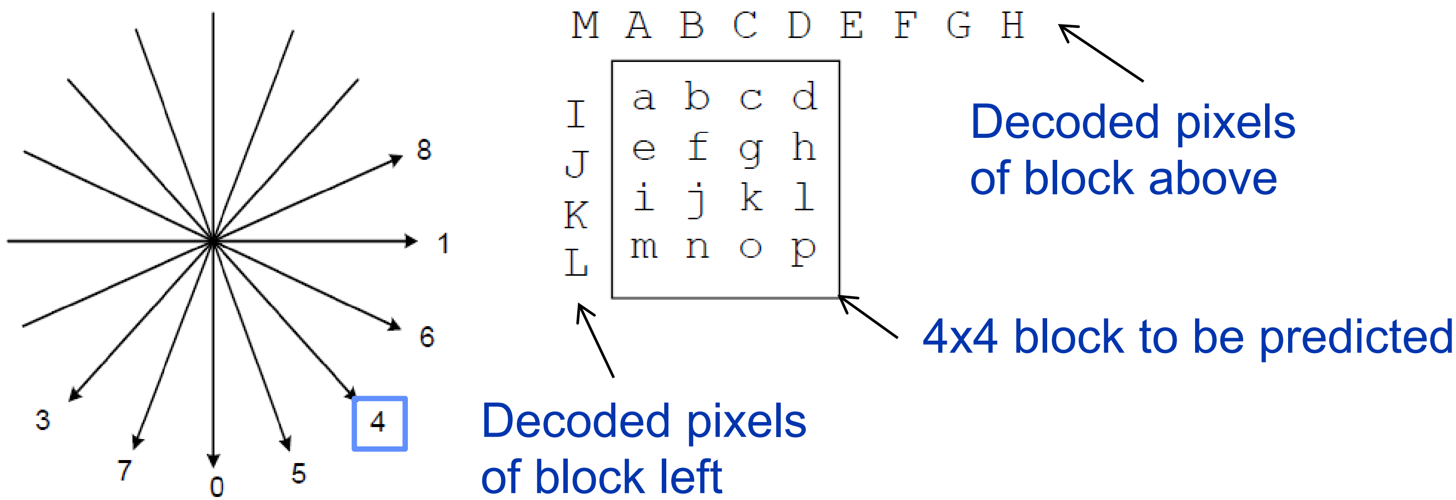
Directional spatial prediction of 4x4 blocks in intraframe coding using nine different types

- Mode 2: DC prediction for uniform image areas

  `a= … =p=(A+B+C+D+I+J+K+L+4)>>3`

E.g. Mode 4: Diagonal down/right prediction

`a=f=k=p=M,        e=j=o=I,` etc.



Additional intra prediction on 16x16 macroblocks using 4 direction modes

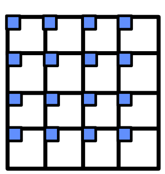
H.264 Integer 4x4 Transform

4x4 Integer transform to decorrelate prediction error image (after temporal or intra prediction) using integer approximation of 4x4 DCT

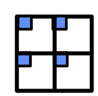
- Columns are orthogonal but not normalized, normalization included as scaling in subsequent quantization process

$$\boldsymbol{A}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \qquad \widetilde{\boldsymbol{A}}_1^{-1} = \begin{bmatrix} 1 & 1 & 1 & 1/2 \\ 2 & 1/2 & -1 & -1 \\ 1 & -1/2 & -1 & 1 \\ 1 & -1 & 1 & -1/2 \end{bmatrix}$$

Second transform of all 4x4 DC coefficients with a macroblock using Walsh-Hadamard transform



Luma: $\boldsymbol{A}_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$

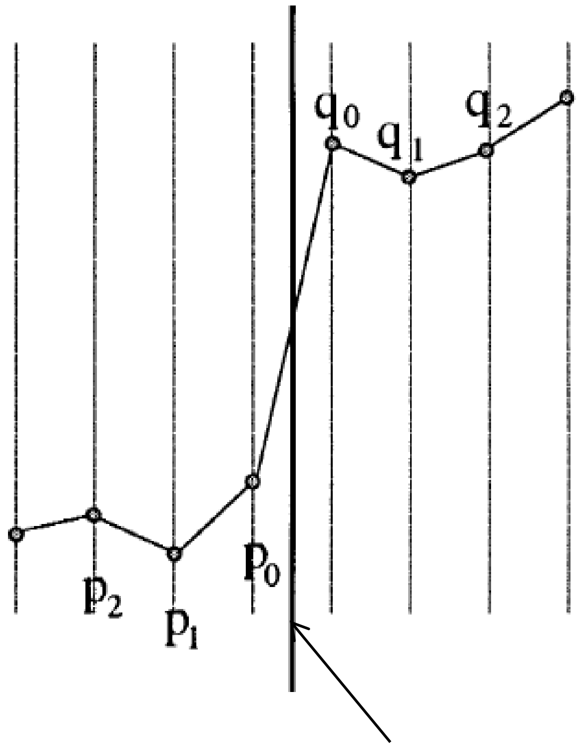


Chroma: $\boldsymbol{A}_3 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

H.264 Deblocking Filter

Filtering along block boundaries to reduce blocking artifacts, filter in the loop

⇨ Applied separately in horizontal and vertical direction



4x4 Block edge

Idea: filter only if edge at block boundary is caused by coarse quantization

⇨ Filter p_0 and q_0 only if following three conditions are met:

$$|p_0 - q_0| < \alpha(\mathrm{QP})$$
$$|p_1 - p_0| < \beta(\mathrm{QP})$$
$$|q_1 - q_0| < \beta(\mathrm{QP})$$

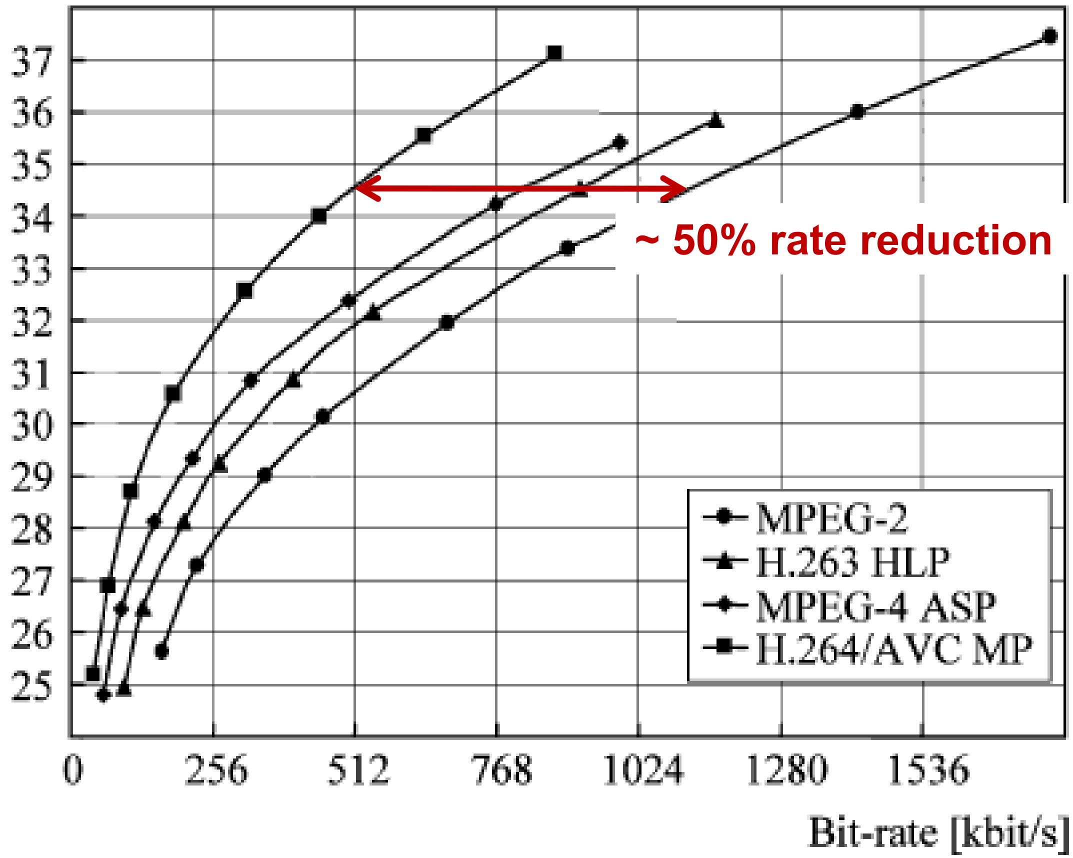
$$\text{with} \quad \beta(\mathrm{QP}) << \alpha(\mathrm{QP})$$

⇨ Filter p_1 and q_1 additionally if

$$|p_2 - p_0| < \beta(\mathrm{QP}) \quad \text{or} \quad |q_2 - q_0| < \beta(\mathrm{QP})$$

(QP = quantization parameter)

Performance of H.264



PSNR Curves for video sequence “Tempete”, CIF, 15 Hz [Sullivan, Wiegand, 2005]

12.4 ITU-T Rec. H.265 / MPEG-H HEVC

High efficiency video coding (HEVC), most recent standard released in 2013

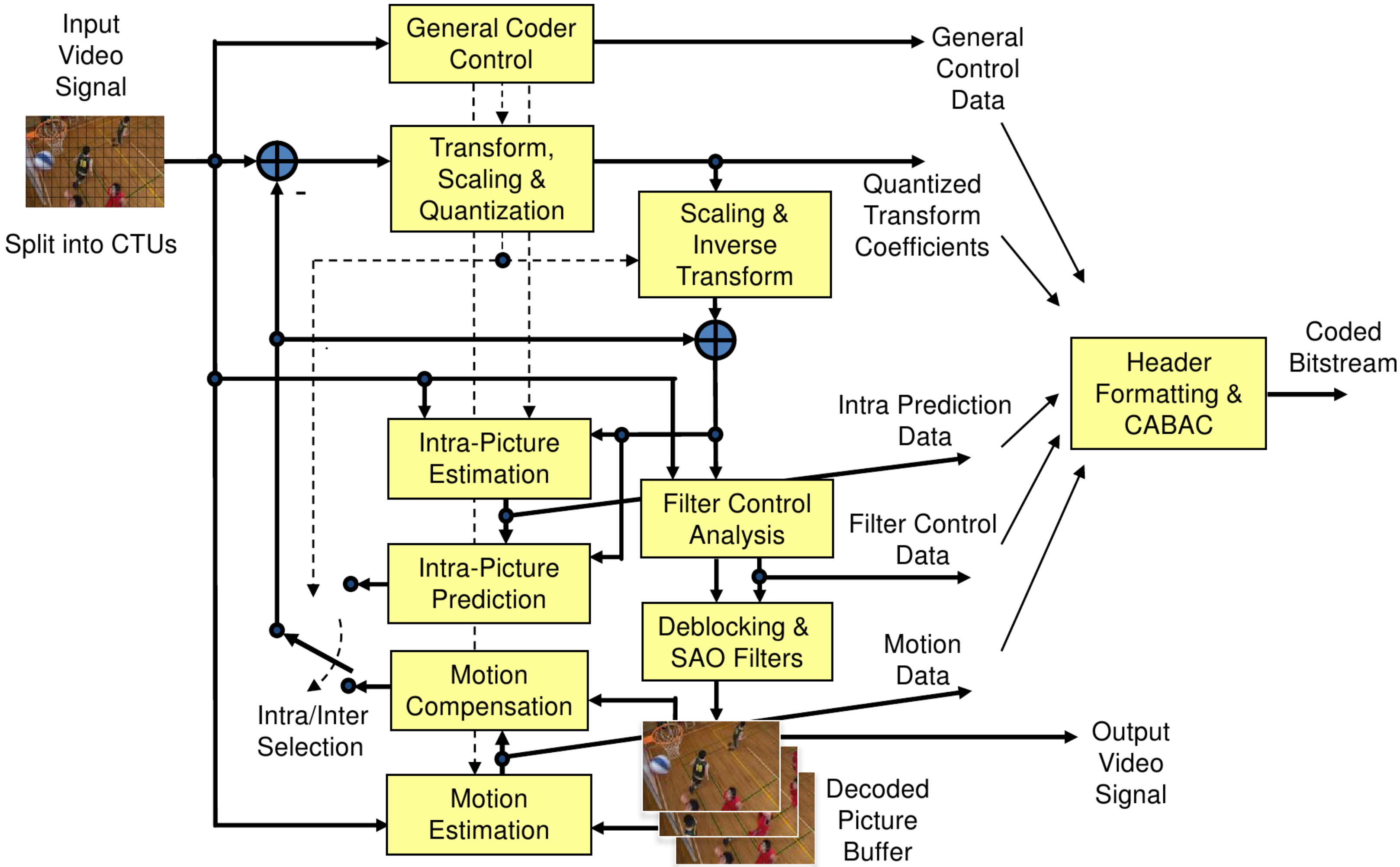
Common elements with predecessor H.264/AVC

- Variable block size motion compensation
- Intraframe spatial prediction
- Scalar quantization
- Transform of prediction error using DCT
- Arithmetic coding of all syntax elements

New elements yielding higher compression performance

- Larger block structures with quad-tree partitioning for motion compensation as well as transform coding (coding tree unit, CTU)
- Enhanced intraframe prediction modes
- Discrete sine transform as alternative transform (DST)
- Additional sample adaptive offset loop filter (SAO)

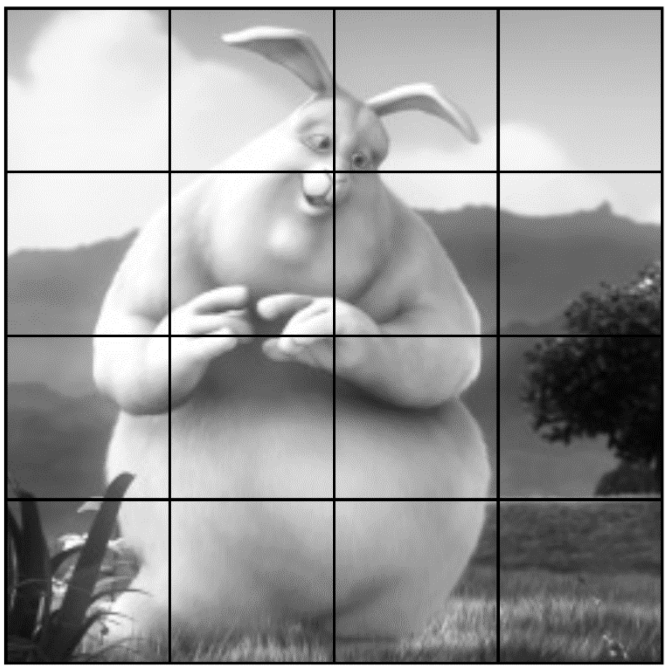
Overview of the HEVC Encoder



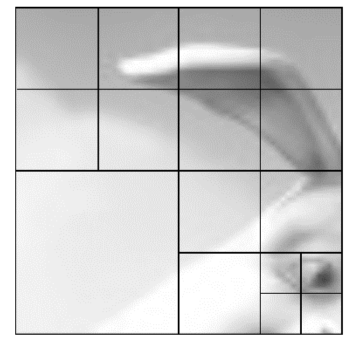
[Source: Sullivan et al., "Overview of the High Efficiency Video Coding (HEVC) Standard," IEEE Transactions on Circuits and Systems for Video Technology, 2012]

HEVC Block Partitioning

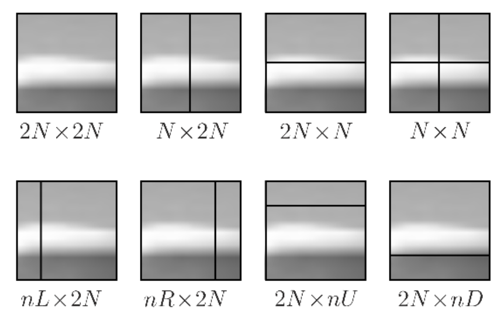
- Partitioning of a picture into Coding Tree Units (CTUs)



- Partitioning of a CTU into Coding Units (CUs)



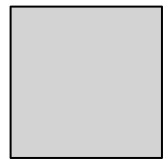
- Partitioning of CUs into Prediction Units (PUs)



Enhanced Intraframe Prediction in HEVC

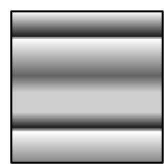
Basic principle: spatial prediction using up to $4N + 1$ decoded boundary samples of adjacent blocks

- DC mode: mean value of boundary samples



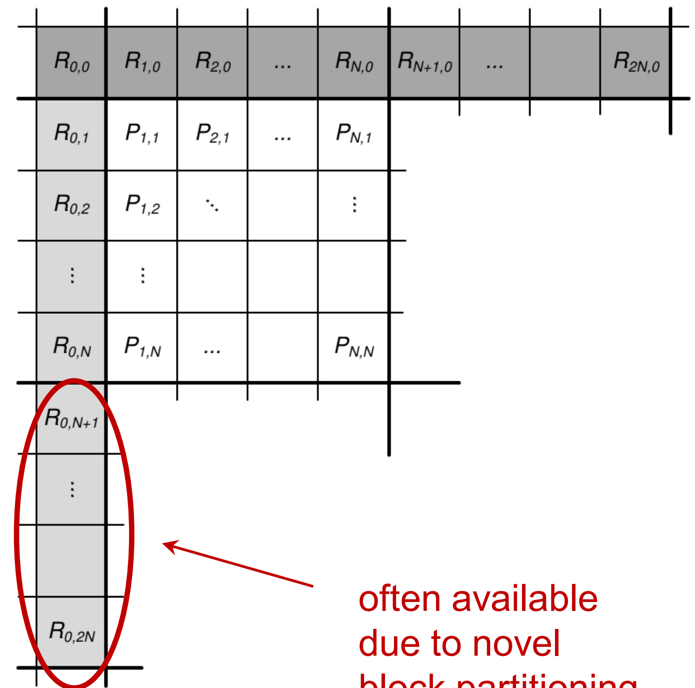
- Planar mode for smooth surfaces with slope

- 33 directional modes for angular prediction



Coding: selected mode encoded by prediction using 3 most probable modes from previously decoded blocks

Example: block with size $N = 4$

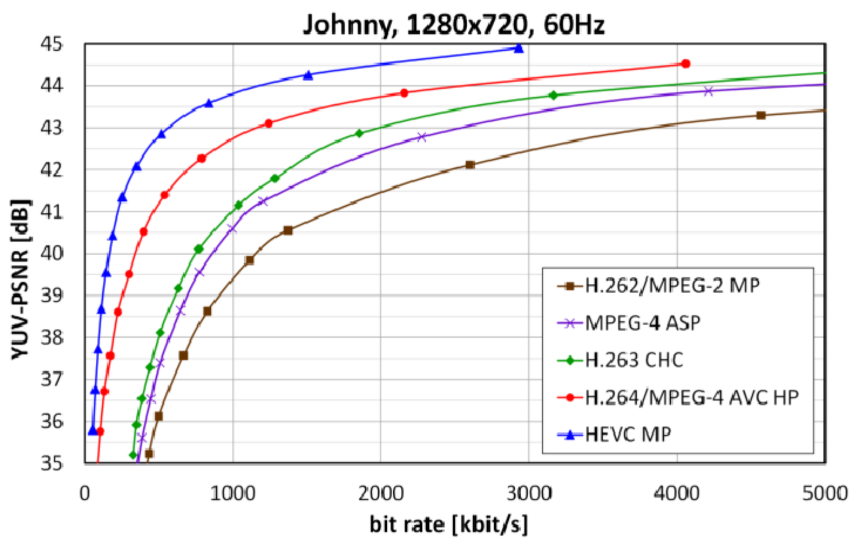


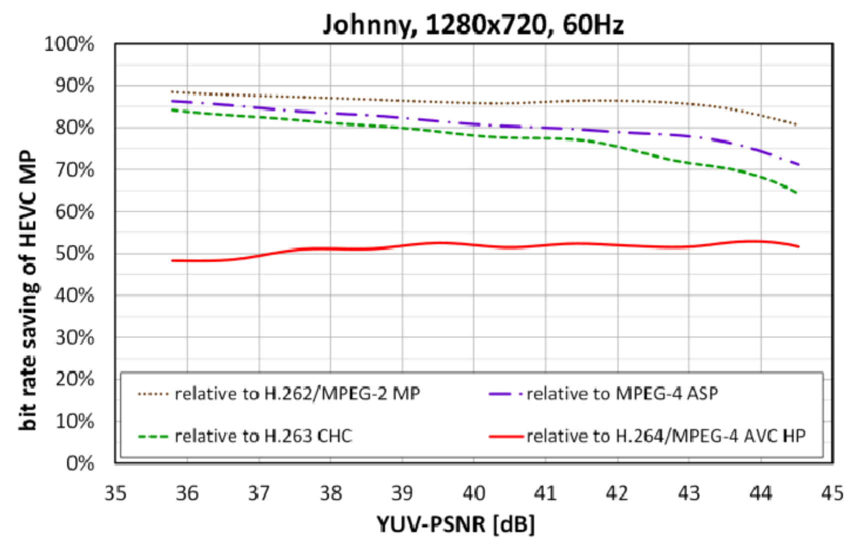
often available due to novel block partitioning scheme!

Performance of HEVC

Rate-distortion curves and rate savings of HEVC Main Profile compared with

- H.264/MPEG-4 AVC High Profile (HP)
- H.263 Conversational High Compression (CHC)
- MPEG-4 Advanced Simple Profile (ASP)
- H.262/MPEG-2 Main Profile (MP)



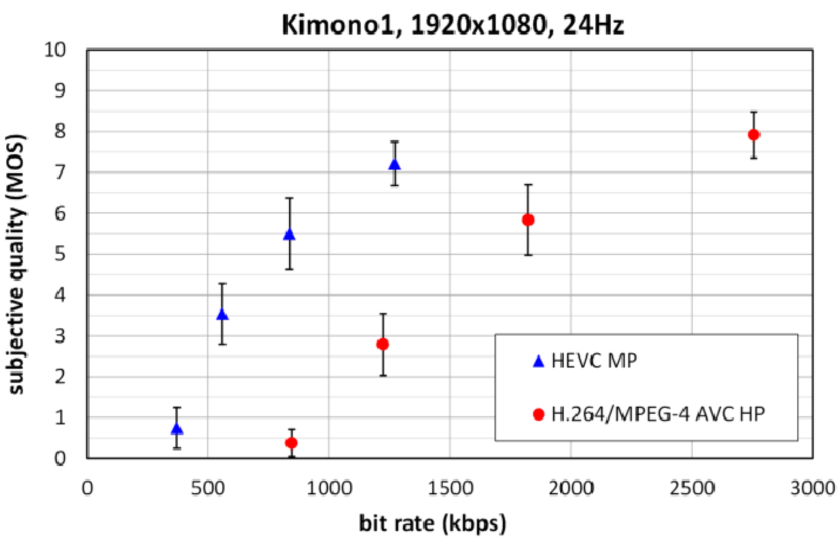


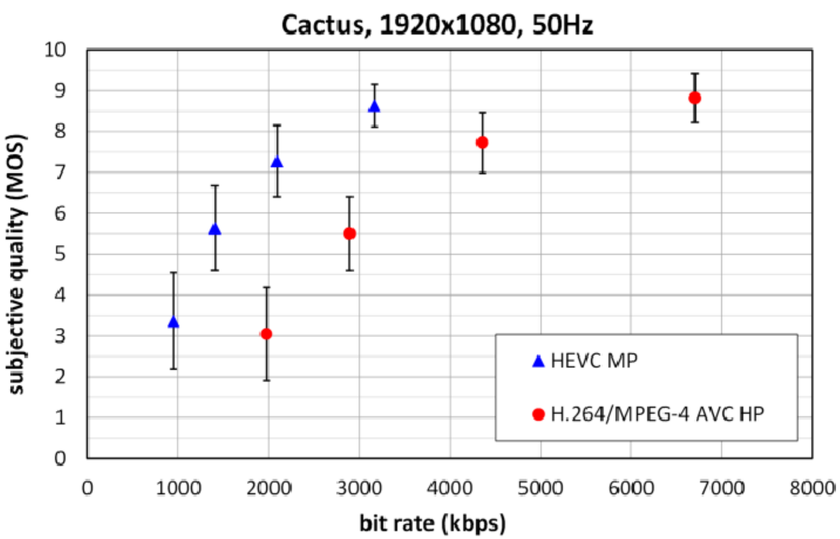
[Source: J.-R. Ohm et al., "Comparison of the Coding Efficiency of Video Coding Standards – Including High Efficiency Video Coding (HEVC)", IEEE Transactions on Circuits and Systems for Video Technology, 2012]

Subjective Evaluation of HEVC

Results of subjective test for HEVC Main Profile compared with H.264/AVC High Profile

- Average bit rate savings of about 53% for same visual quality
- Subjective tests confirm results of rate distortion curves based on PSNR





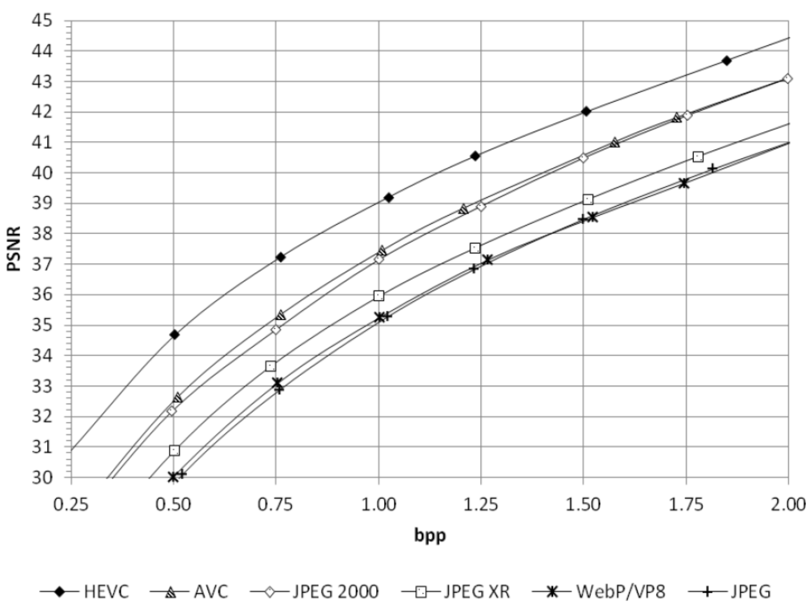
[Source: J.-R. Ohm et al., "Comparison of the Coding Efficiency of Video Coding Standards – Including High Efficiency Video Coding (HEVC)", IEEE Transactions on Circuits and Systems for Video Technology, 2012]

Performance of HEVC for Still Image Coding

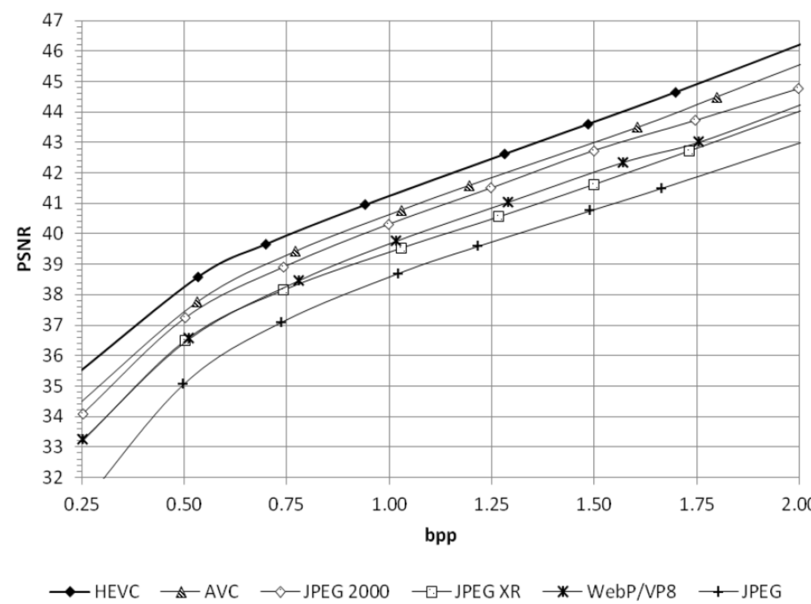
Intraframe mode of HEVC can also be used to code still images

- Outperforms H.264/AVC, JPEG 2000, and JPEG
- Bit rate savings of 44% (32%) compared to JPEG (H.264/AVC intra)

Barbara, 512x512



Lena, 512x512



[Source: T. Nguyen, D. Marpe, "Performance Analysis of HEVC-Based Intra Coding for Still Image Compression," Picture Coding Symposium, Krakow, Poland, May 2012]

Summary Video Coding Standards

- Current video coding standards are based on motion compensated prediction and transform coding of prediction error
- Intraframe prediction for random access and higher coding efficiency
- Basic coding unit is macroblock with 16x16 luminance pixels plus chroma
- Motion accuracy typically half-pixel resolution, quarter-pixel in H.264/AVC
- H.263 for conversational video services
- MPEG-2 for TV broadcast distribution of primarily interlaced video
- H.264/AVC for advanced video compression in multimedia systems
- H.265/HEVC with optimized tools is state-of-the-art in video compression