Master: INFORMATIQUE
Parcours: VICO Visual Computing

UE: Multimedia Communication

video coding

vincent.ricordel@univ-nantes.fr

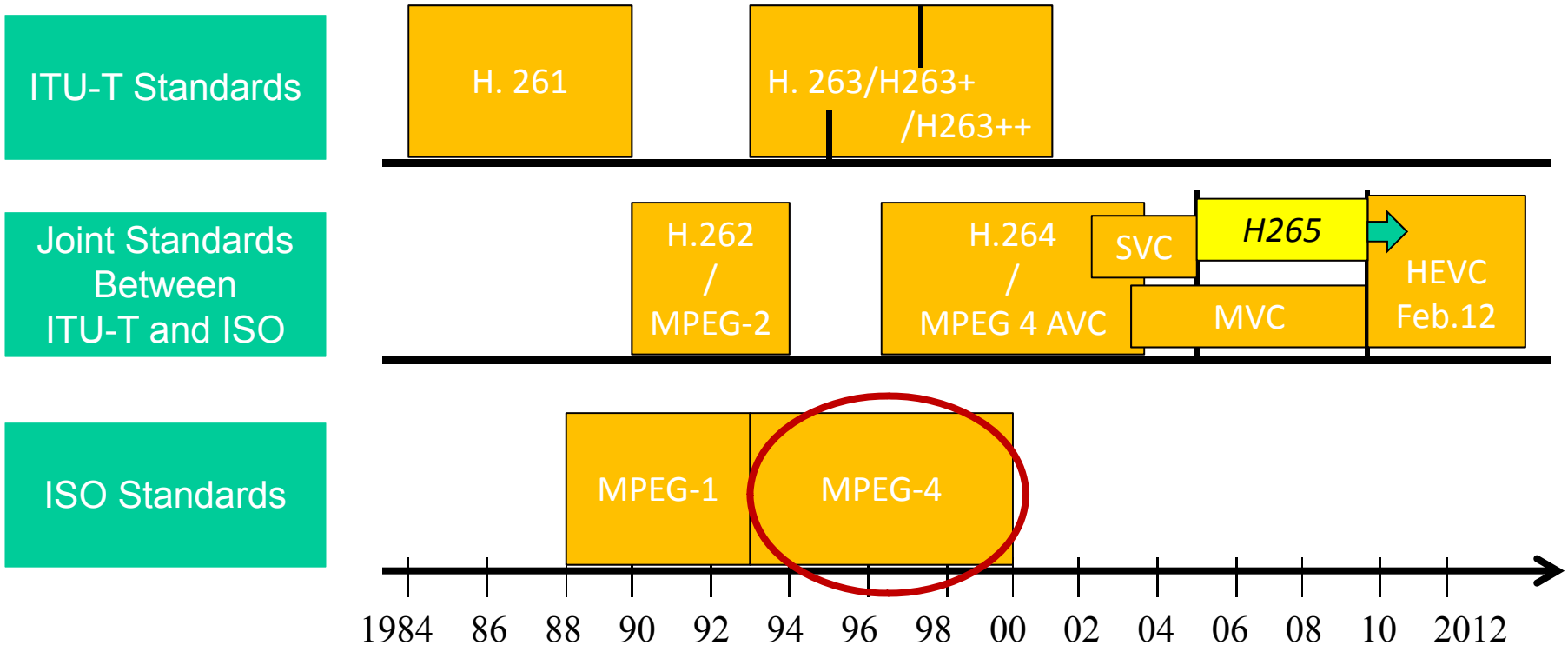
Video coding

Outlines:

- MPEG 4
- H.264/AVC

Video Standard History

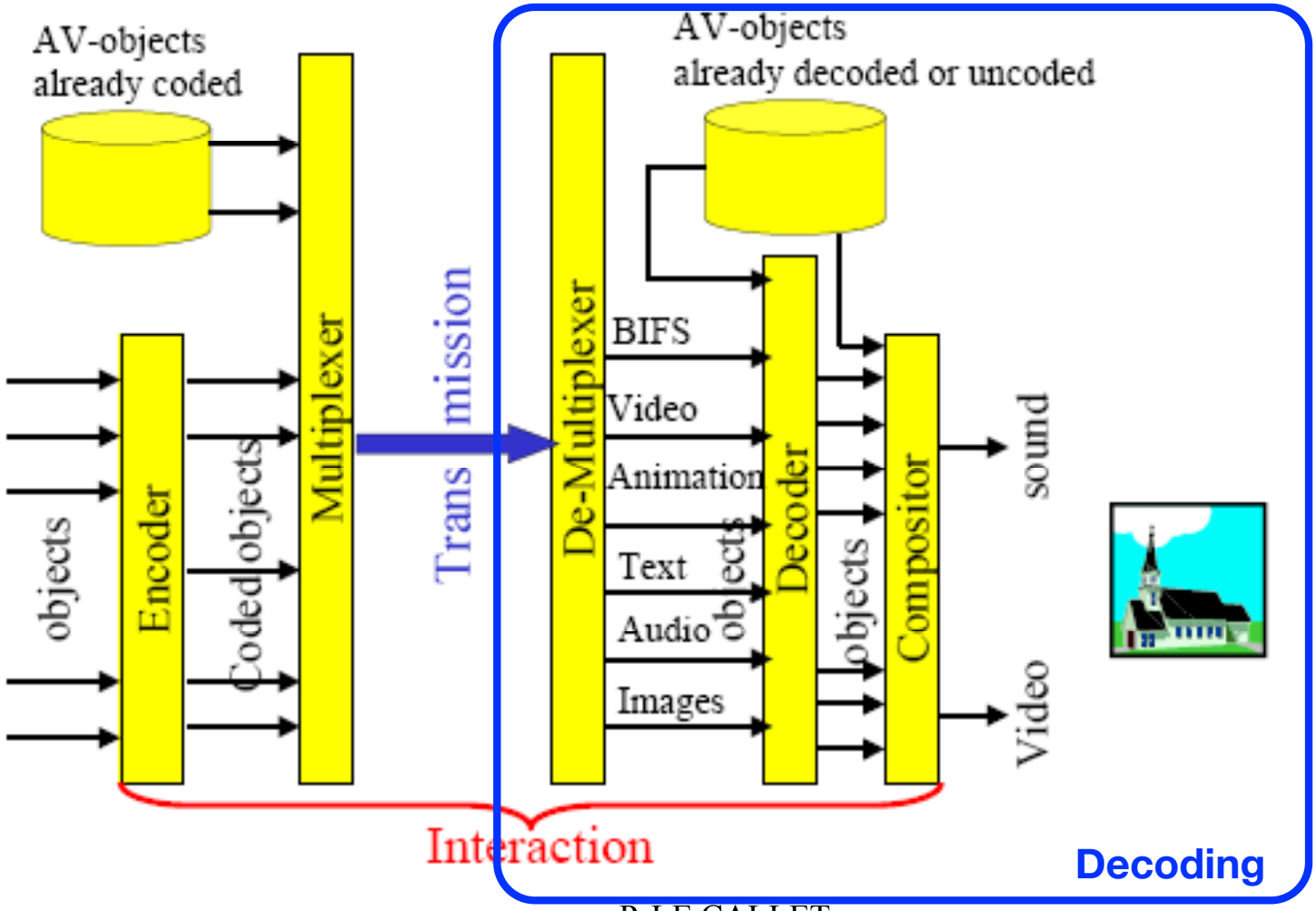
- Two organisations:
  - **ITU-T** (International Telecom Union)
  - **ISO** (International Standard Organisation)



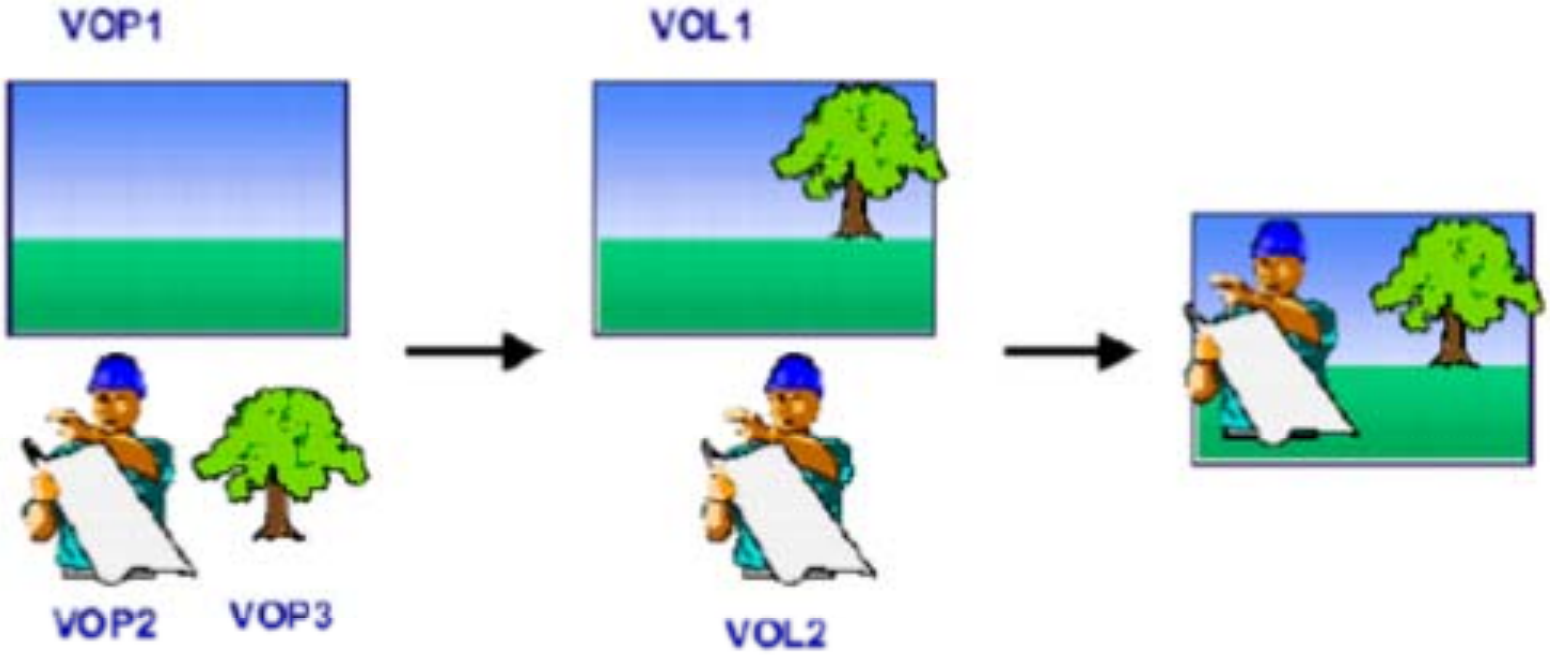
MPEG 4 ISo14496

- 1998-1999
- Coding audio visual object

Decomposition of the scene in Audio Visual objects



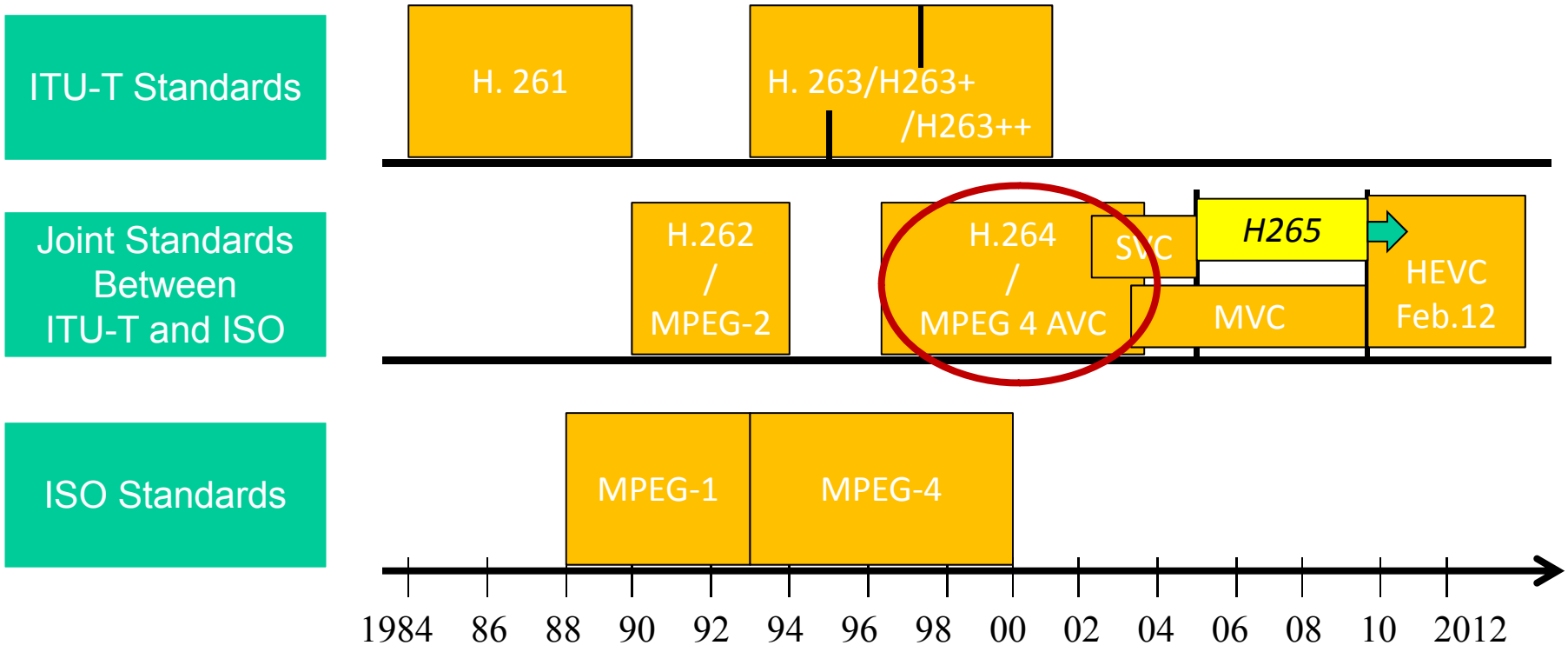
Composition video



- Object-based coding requires the decoder to compose different Video-Object-Planes (VOP) into a scene
- Video-Object-Layers (VOL) enable content-based scalability

Video Standard History

- Two organisations:
  - **ITU-T** (International Telecom Union)
  - **ISO** (International Standard Organisation)



H264 : objectives

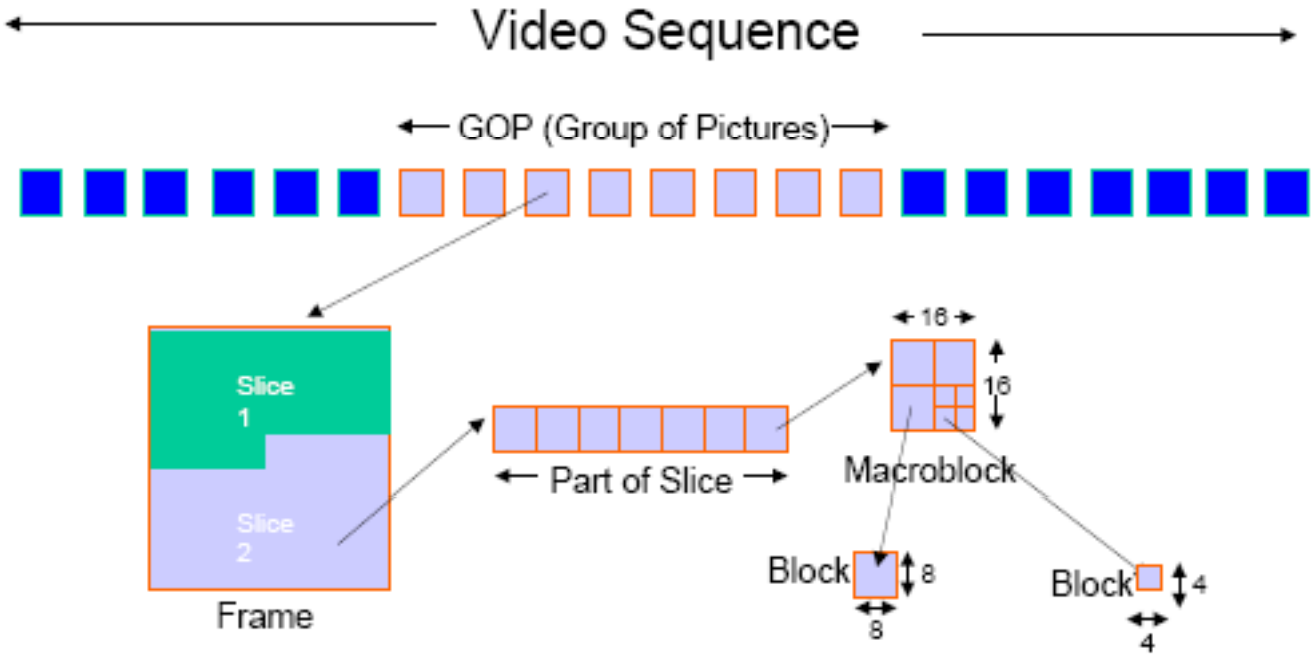
- To lower the bit-rate by about 50% while offering a better Quality of Service (QoS)
- To improve the coding efficiency
- To improve error resiliency
- Transmission over narrow bandwidth channels
- To reduce the overhead syntax
- To deal with 4:2:2 & 4:4:4 sampling formats
- To deal with most of the frame sizes
  - QCIF (176x144)
  - ...
  - TVHD (1920x1080)
  - D-Cinema (4096x2048)

2K video format

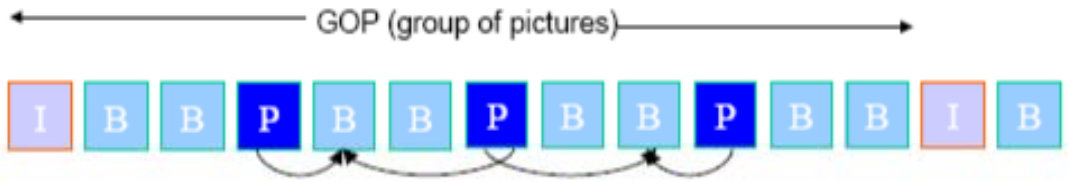
Applications

- For TV distribution (1 à 10 Mb/s)
  - Satellite & Cable for broadcasting applications
  - DVD-Video for entertaintment
  - XDSL for VoD (Video On Demand) applications
- For H.32X services
  - Multimedia Services < 1Mb/s
  - Use of G.72X standards for audio
  - Use of H.26X standards for video
- For Video Streaming
  - With the 2G or 3G networks
  - With the RTP or RTCP protocols

Hierarchy



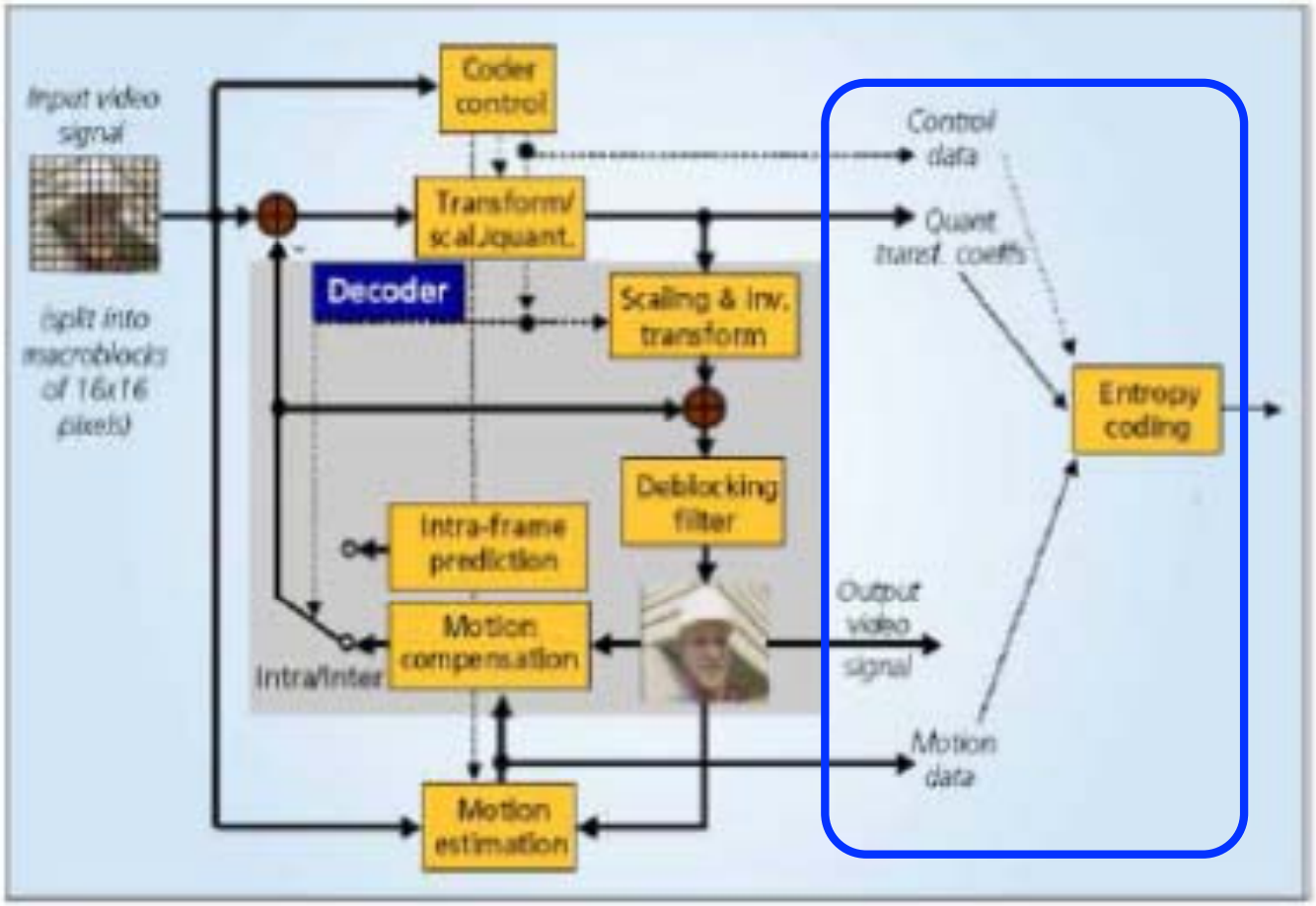
Relationships between different frame types (recall)



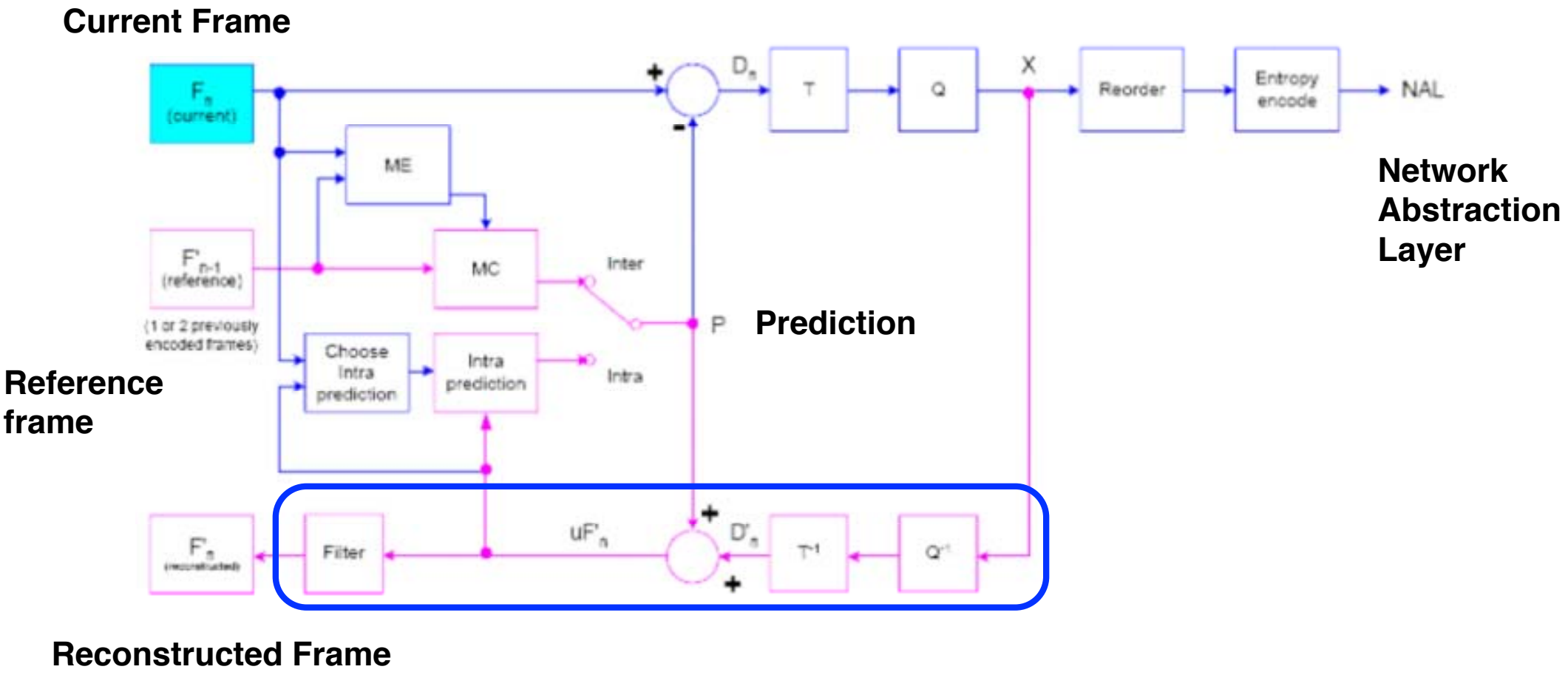
Display order : I B B P B B P

Coding & Decoding order : I P B B P B B

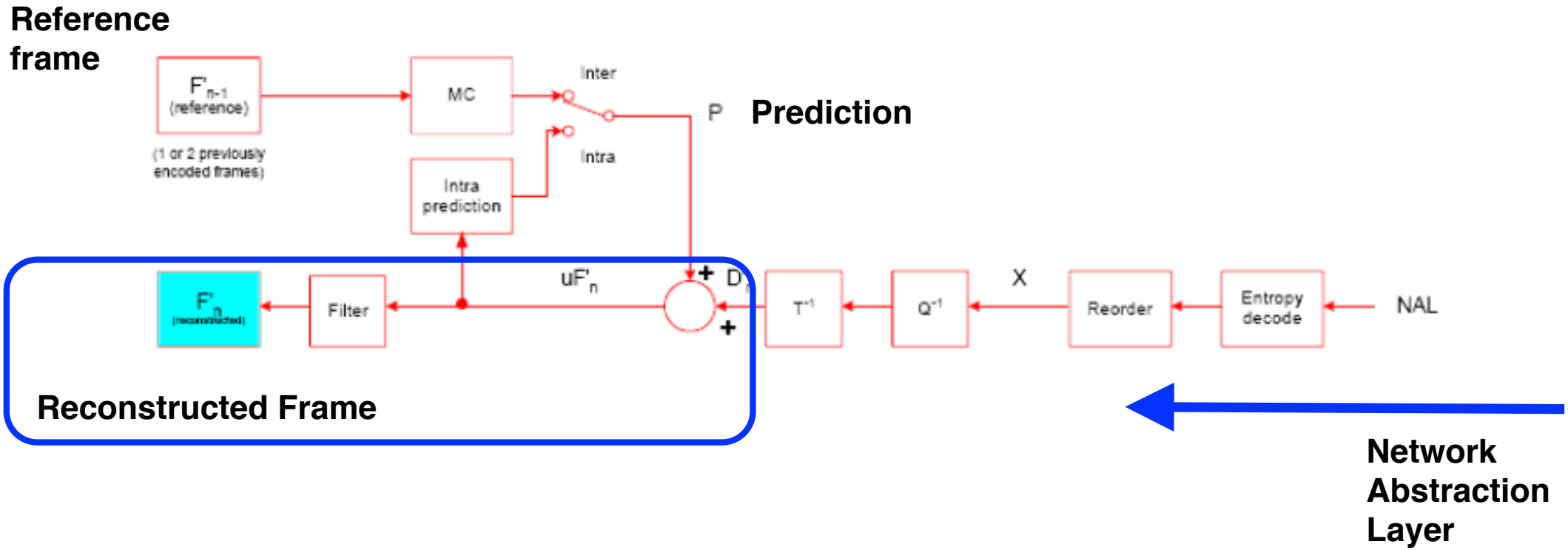
Coder:Decoder block diagram



Coder block diagram



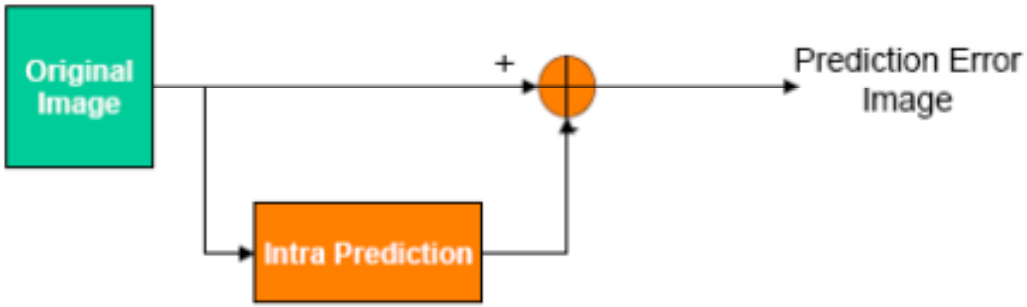
Decoder block diagram



Intra prediction

Intra coding:

- No needs of Reference Frame
- Use spatial Prediction
  - Needs the use of adjacent blocks
  - Produces an error predicted frame easy to encode



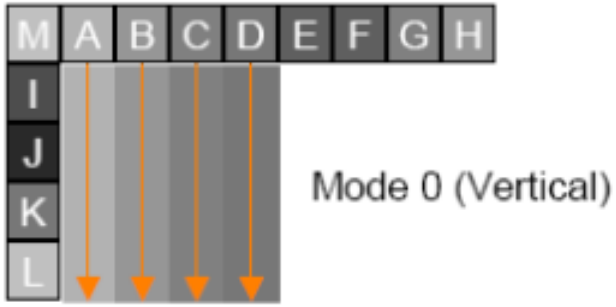
3 Types of Intra Prediction

- Luminance:
  - Macroblock 16x16
  - Block 4x4
- Chrominance:
  - Macroblock 8x8

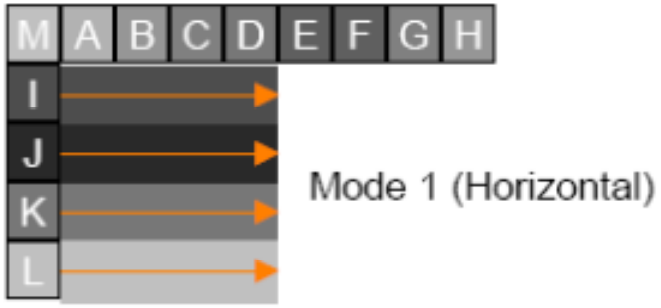
Intra prediction

Prediction of 4x4 blocks (luminance)

- 9 modes
- Use of pixels from neighborhood blocks



Mode 0 (Vertical)

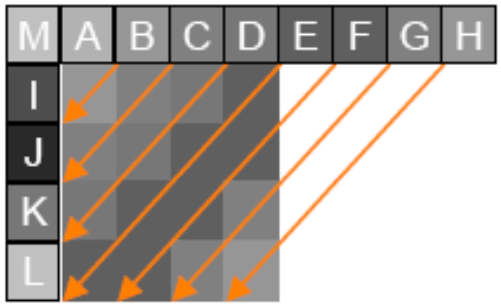


Mode 1 (Horizontal)

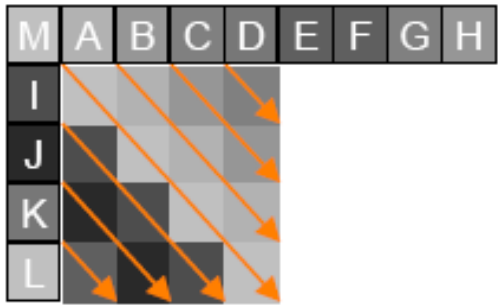
Mode 2 (DC)

Intra prediction

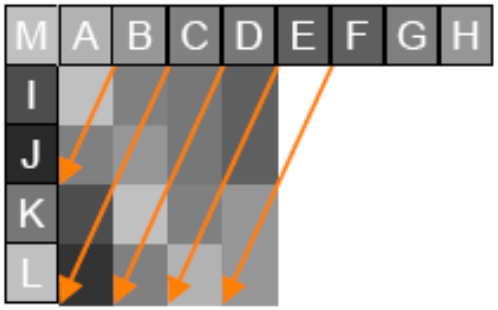
Prediction of 4x4 blocks



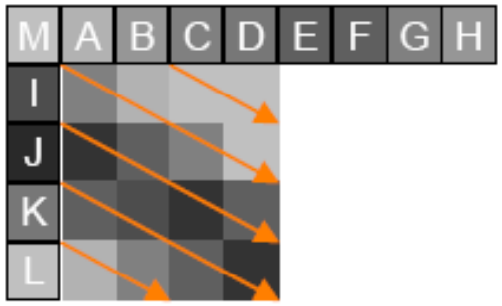
Mode 3 (Diagonal left)



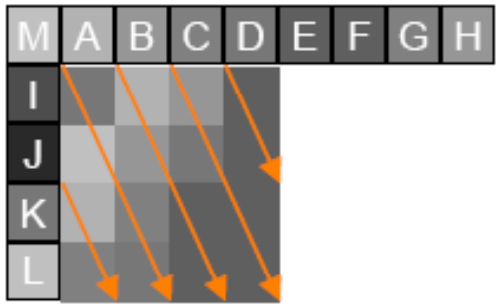
Mode 4 (Diagonal right)



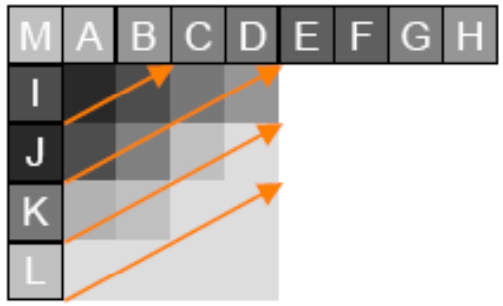
Mode 5 (vertical left)



Mode 6 (Horizontal down)



Mode 7 (Vertical right)



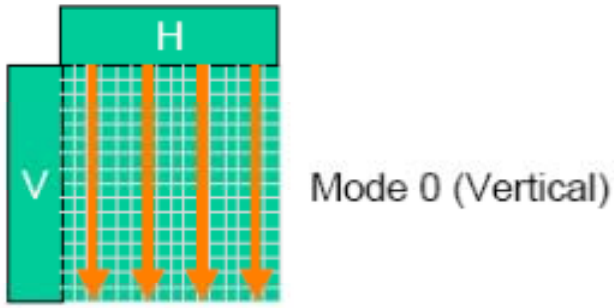
Mode 8 (Horizontal up)

Same types of Intra prediction for 8x8 blocks in FREXt profiles)
(chrominance)

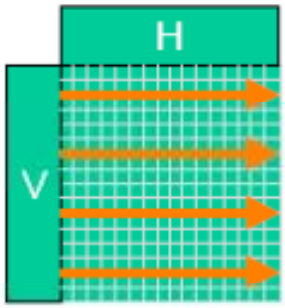
Intra prediction

Prediction of 16 x 16 Macroblocks (luminance)

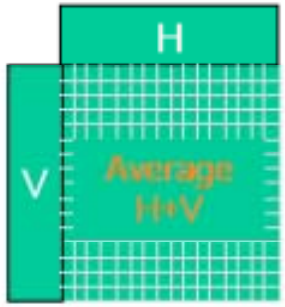
- 4 modes
- Same Modes as for luminance (16x16) and chrominance (8x8)



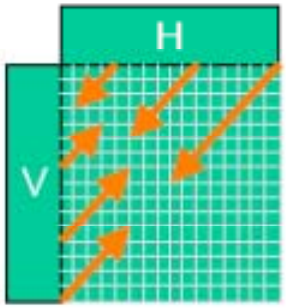
Mode 0 (Vertical)



Mode 1 (Horizontal)



Mode 2 (DC)

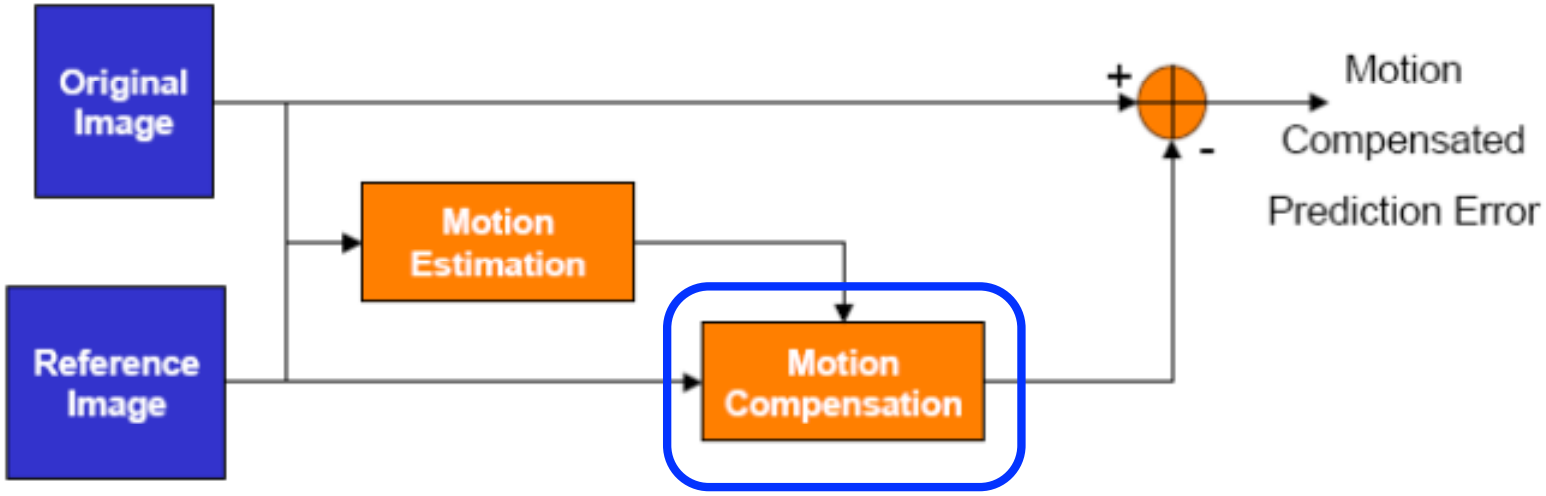


Mode 3 (Plane)

Inter prediction

Inter Frame Coding:

- Need of Reference Images
- Use of Temporal Prediction
  - Macroblock-based decomposition
  - Macroblock-based Motion estimation (displacement vector)
  - Pixels Interpolation according to the resolution
  - 2 displacement vectors are possible per pixel



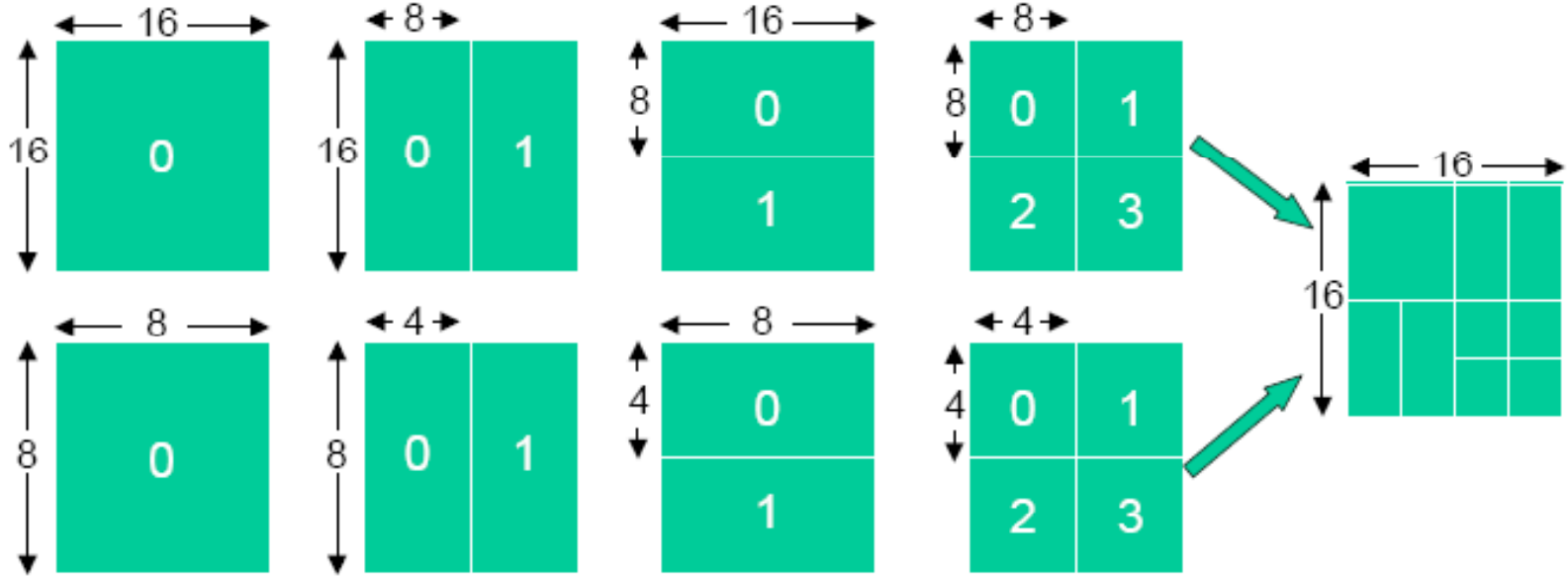
Inter prediction

Reference Frames:

- Two lists of Frame buffers can be used
  - List 0: for slices P, SP or B
  - List 1: for slices B
- 2 types of Reference Frames can be used
  - Long term reference frames
  - Short term reference frames

(SP: Switch-Predictive)

Macroblock Partition:

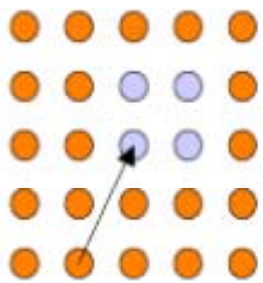


Inter prediction

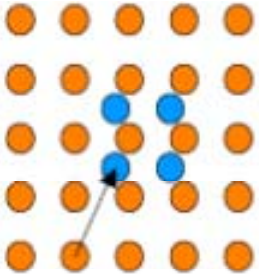
Motion vectors (1):

– Resolution going down to 1/4 pixel width/height

Motion vector with pixel precision

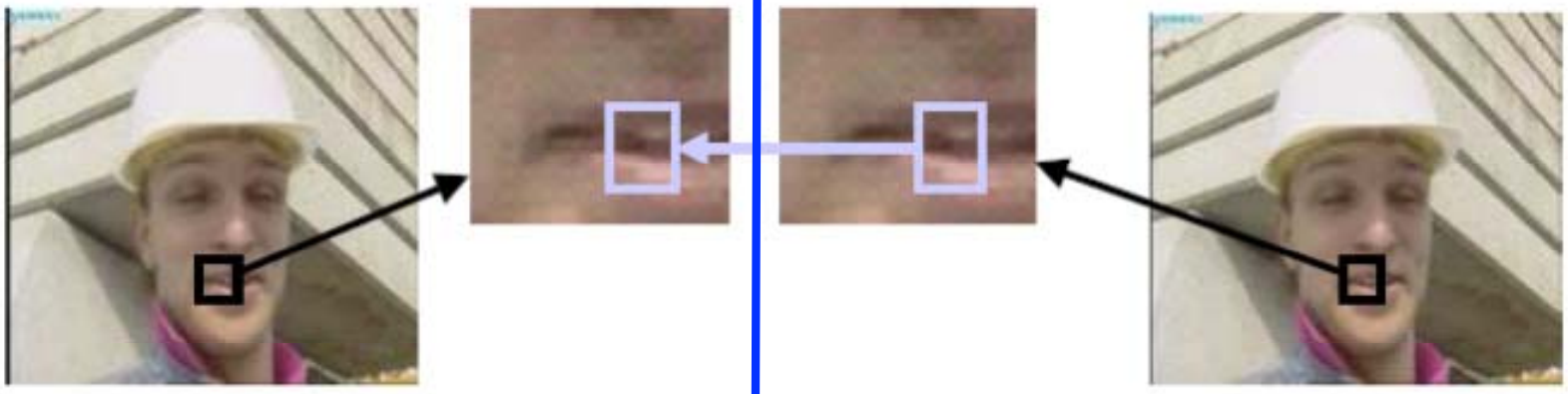


Vector (1,-2)



Vector (0.75,-1.5)

Motion vector with quater pixel precision



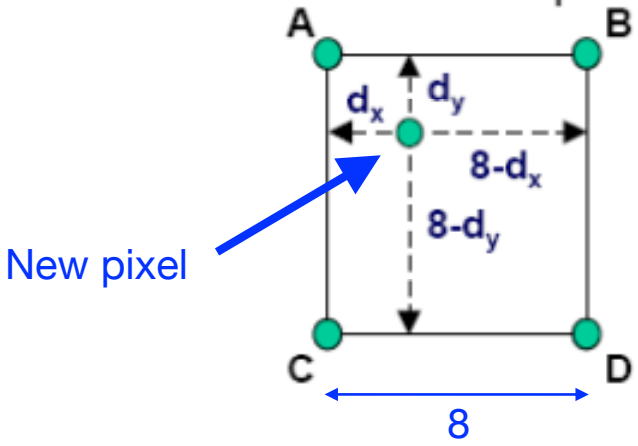
Reference frame (decoded)

Current frame

Inter prediction

Motion vectors (2)

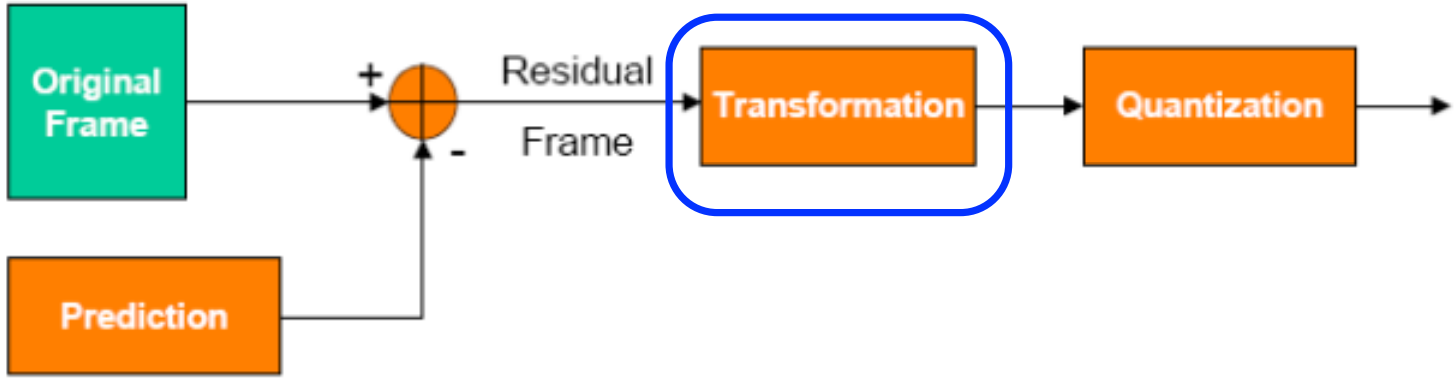
- Pixels interpolation for non-integer motion vector
  - Luminance component:
    - 6-tap FIR filter for resolution ½
    - Average of the 2 adjacent values for resolution ¼
  - Chrominance components:
    - Linear interpolation



$$v=((8-d_x)(8-d_y)A+d_x(8-d_y)+(8-d_x)d_yC+d_xd_yD+32)/64$$

Transformation and quantization

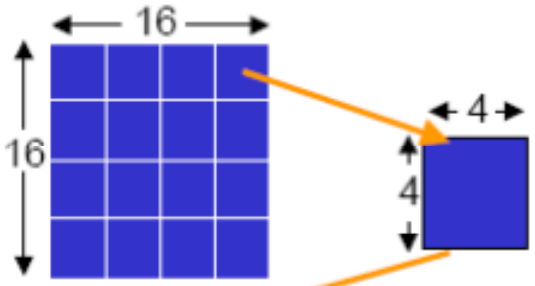
General block-diagram:



- 4x4 Integer DCT for all the macro-blocks
- Hadamard Transformation for DC elements in Intra 16x16 mode

Transformation and quantization

Integer DCT Transformation :



$$Y = \left(CXC^T\right) \otimes E_{forw} =$$

$$a = 1/2 \text{ and } b = \sqrt{\frac{2}{5}}$$

$$\left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \begin{bmatrix} x_{00} & x_{01} & x_{02} & x_{03} \\ x_{10} & x_{11} & x_{12} & x_{13} \\ x_{20} & x_{21} & x_{22} & x_{23} \\ x_{30} & x_{31} & x_{32} & x_{33} \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 1 \\ 1 & 1 & -1 & -2 \\ 1 & -1 & -1 & 2 \\ 1 & -2 & 1 & -1 \end{bmatrix} \right) \otimes \begin{bmatrix} a^2 & \frac{ab}{2} & a^2 & \frac{ab}{2} \\ \frac{ab}{2} & \frac{b^2}{4} & \frac{ab}{2} & \frac{b^2}{4} \\ a^2 & \frac{ab}{2} & a^2 & \frac{ab}{2} \\ \frac{ab}{2} & \frac{b^2}{4} & \frac{ab}{2} & \frac{b^2}{4} \end{bmatrix}$$

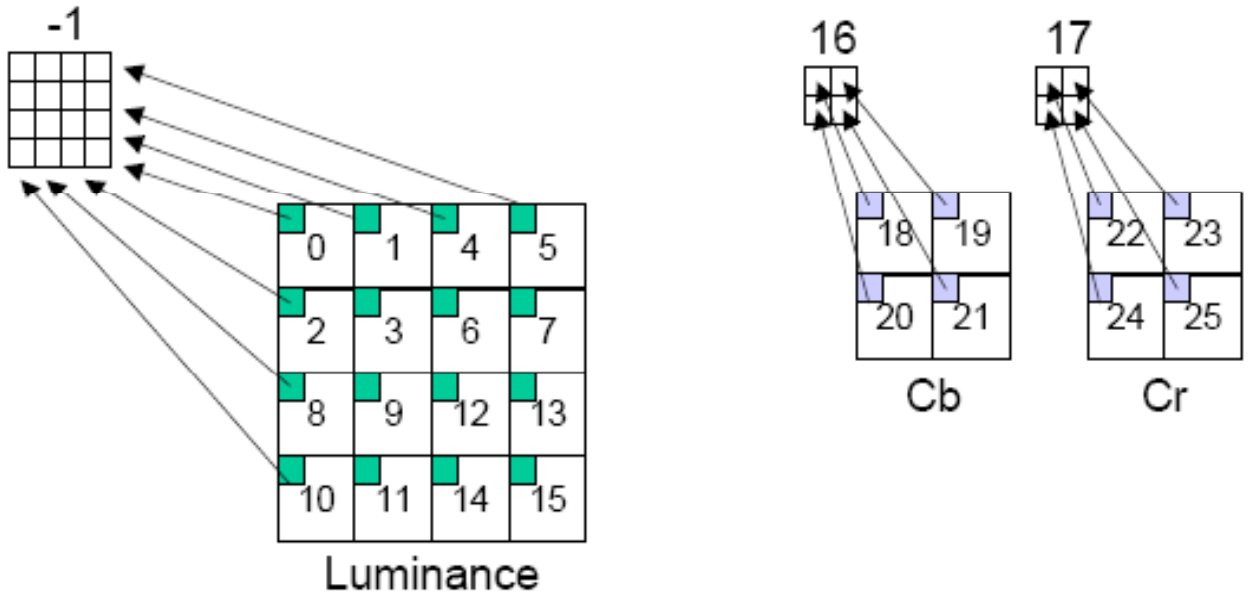
- Quantization performed by using E_{forw}:

$$Y_Q(i,j) = \left[Y(i,j) \cdot Q(QP\%6,i,j) + f\right] / 2^{17+QP/6}, \quad i,j = 0,\ldots,3$$

Transformation and quantization

Hadamard Transformation:

– Construction of DC blocks



Transformation and quantization

Hadamard Transformation :

- Transformation of DC 4x4 blocks

$$Y_D = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_{D00} & x_{D01} & x_{D02} & x_{D03} \\ x_{D10} & x_{D11} & x_{D12} & x_{D13} \\ x_{D20} & x_{D21} & x_{D22} & x_{D23} \\ x_{D30} & x_{D31} & x_{D32} & x_{D33} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \right)$$

$$Y_{QD}(i,j) = \left[Y_D(i,j) \cdot Q(QP\%6,0,0) + 2 \cdot f \right] / 2^{18+QP/6}, \quad i,j = 0,\ldots,3$$

- Transformation of DC 2x2 blocks

$$Y_D = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_{D00} & x_{D01} \\ x_{D10} & x_{D11} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$Y_{QD}(i,j) = \left[Y_D(i,j) \cdot Q(QP\%6,0,0) + 2 \cdot f \right] / 2^{18+QP/6}, \quad i,j = 0,1$$

Example

- Original block

$$\begin{pmatrix} 43 & 216 & 255 & 249 \\ 49 & 198 & 193 & 211 \\ 48 & 194 & 177 & 171 \\ 46 & 214 & 225 & 169 \end{pmatrix}$$

- Encoder chooses 4×4 DC prediction mode: Prediction

$$\begin{pmatrix} 128 & 128 & 128 & 128 \\ 128 & 128 & 128 & 128 \\ 128 & 128 & 128 & 128 \\ 128 & 128 & 128 & 128 \end{pmatrix}$$

- Prediction error

$$\begin{pmatrix} -85 & 88 & 127 & 121 \\ -79 & 70 & 65 & 83 \\ -80 & 66 & 49 & 43 \\ -82 & 86 & 97 & 41 \end{pmatrix}$$

- Block after transform

$$\begin{pmatrix} 610 & -1256 & -686 & -558 \\ 279 & -478 & 111 & -69 \\ 176 & -160 & -120 & 100 \\ -13 & -14 & 3 & 3 \end{pmatrix}$$

- Quantized coefficients

$$\begin{pmatrix} 7 & -9 & -8 & -4 \\ 2 & -2 & 1 & 0 \\ 2 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

- (run,level) pairs: (0,7), (0,-9), (0,2), (0,2), (0,-2), (0,-8), (0,-4), (0,1), (0,-1), (2,-1), (1,1)

- Reconstructed block

$$\begin{pmatrix} 57 & 204 & 246 & 238 \\ 52 & 194 & 189 & 204 \\ 48 & 195 & 174 & 169 \\ 50 & 207 & 217 & 167 \end{pmatrix}$$

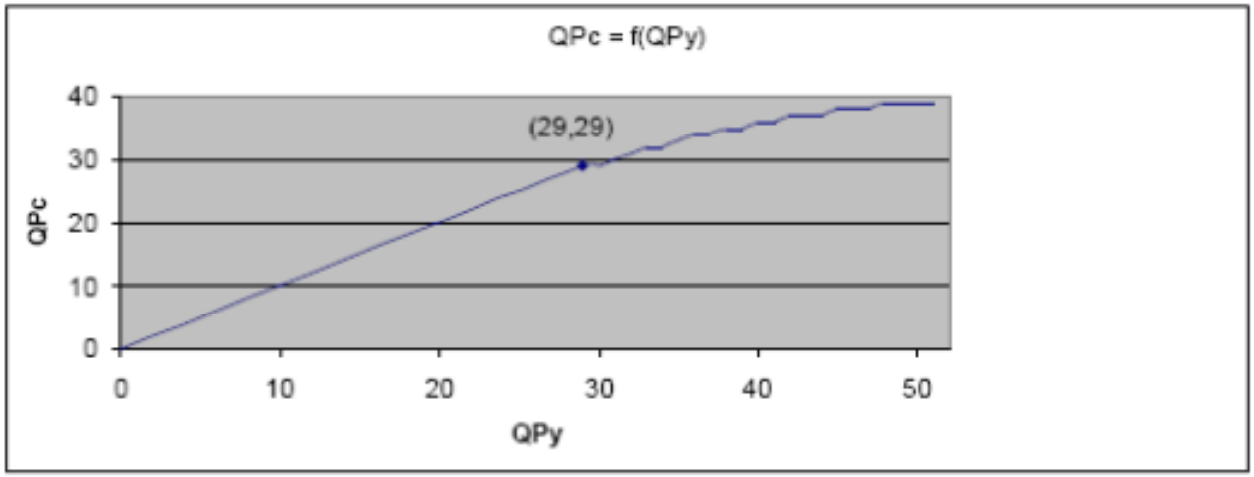
Quantized-coeff. =
Round(coeff. / quantization-step)

Reconstructed-coeff. =
Quantized-coeff. x quantization-step

Transformation and quantization

Quantization step:

- 52 different quantization steps (QPy or QPc)
- The quantization step is double every 6 increments of QPy
- QPy can be changed by a value Delta transmitted within the macroblock layer
- QPc is deduced from QPy

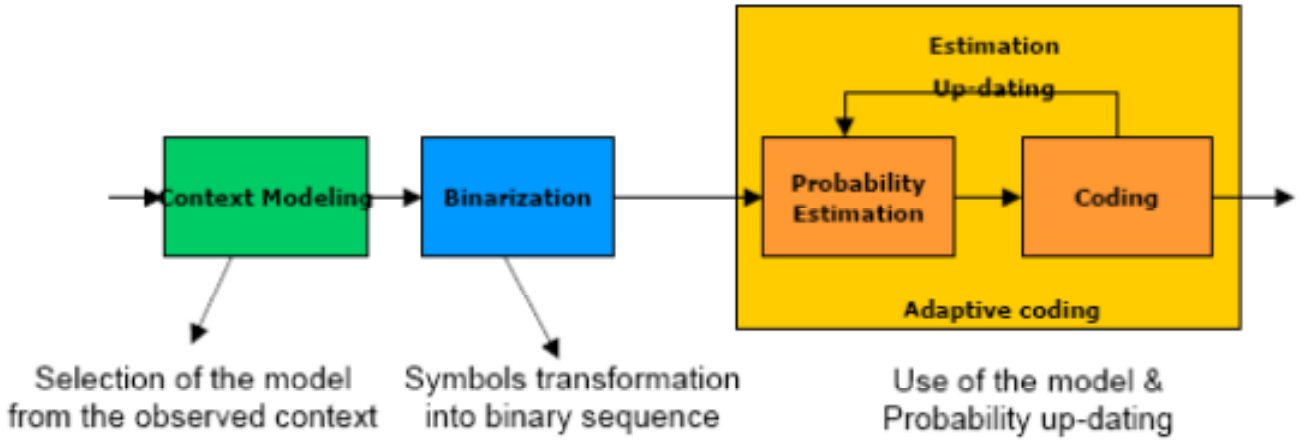


Entropy coding

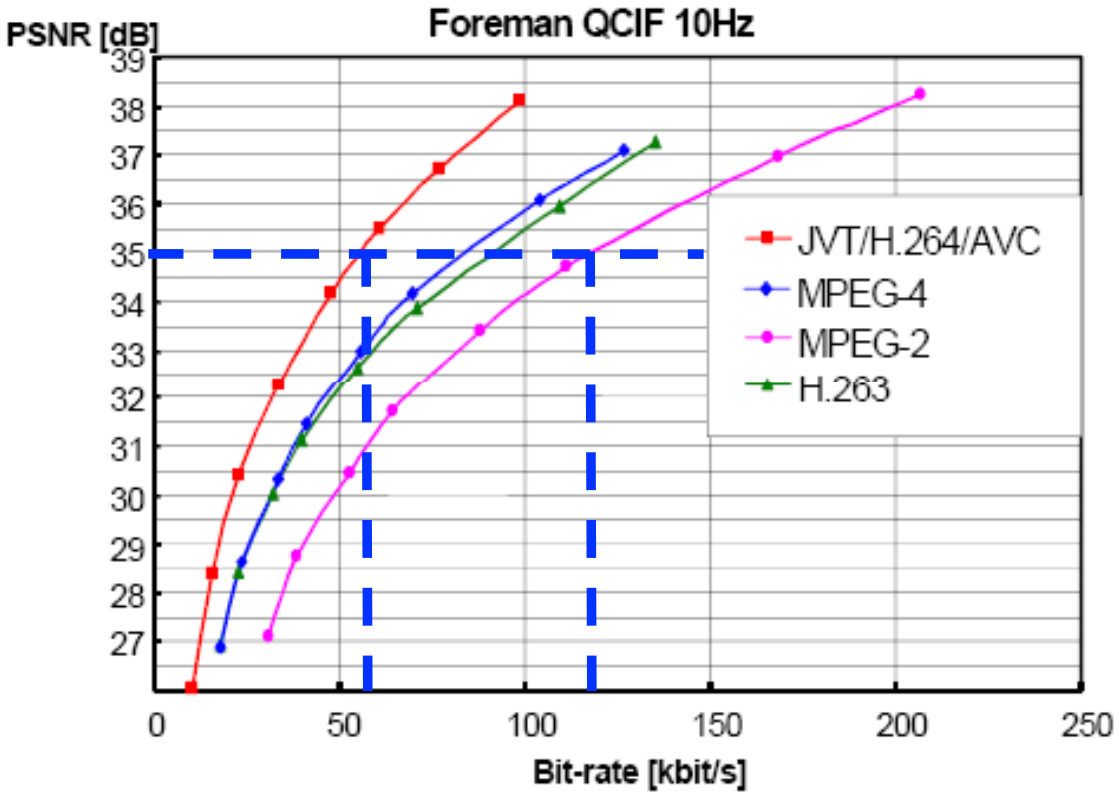
- VLC :
  - **Residual block data (quantized transform data) is coded using context-adaptive variable length coding (CAVLC)**
  - **Other elements (header data, motion vectors, et.c.) are coded using Exp-Golomb codes, either directly or via table lookup.**

CABAC (Context-Adaptive Arithmetic Coding)

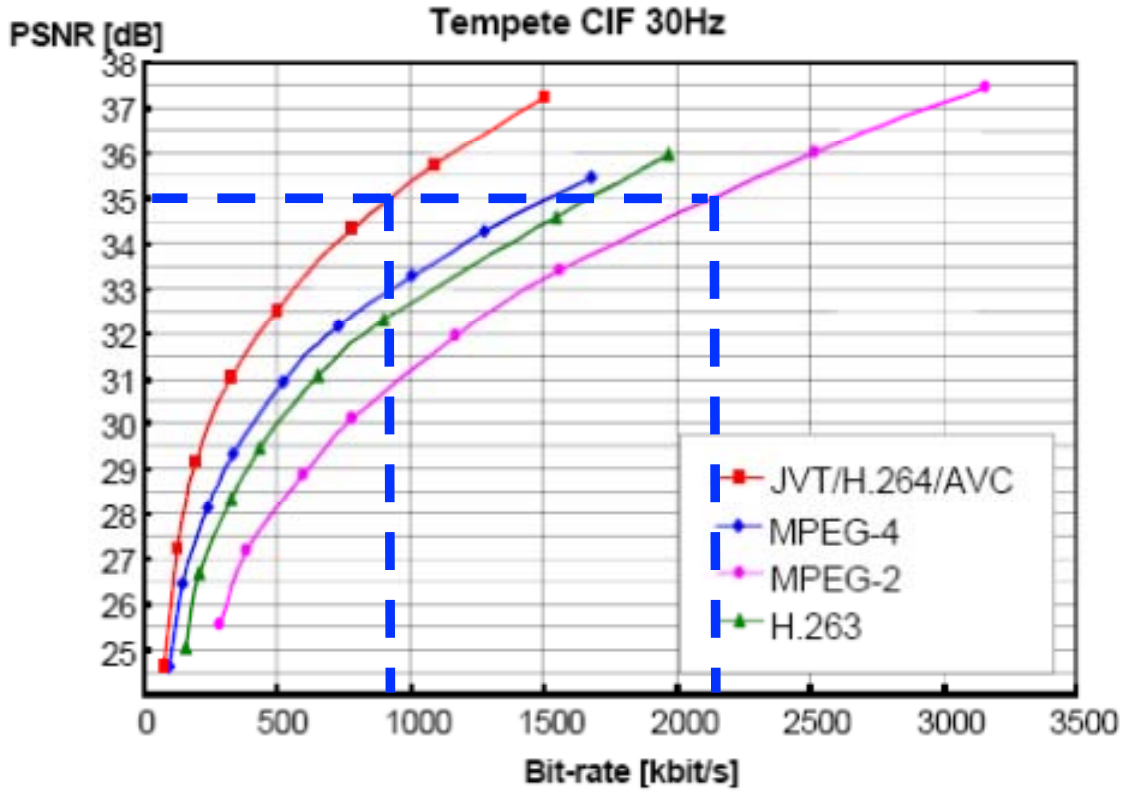
- Use of adaptive probability modes
- Use of inter-symbol correlation trough the different contexts
- Use of arithmetic coding for representing non integer value symbols
- Need to binarize the non-binary elements



Performances



Performances



Performances

MIPS: Millions of Instructions per second

Functions	Arithmetic		Controlling		Data transfer		
	MIPS	%	MIPS	%	MIPS	Mbyte/s	%
Integer-pel motion estimation	95491.9	78.31	21 915.1	55.37	116 830.8	365 380.7	77.53
Fractional-pel motion estimation	21396.6	17.55	14 093.2	35.61	30 084.9	85 045.7	18.04
Fractional-pel interpolation	558.0	0.46	586.6	1.48	729.7	1067.6	0.23
Lagrangian mode decision	674.6	0.55	431.4	1.09	690.7	2642.6	0.56
Intra prediction	538.0	0.44	288.2	0.73	585.8	2141.8	0.45
Variable lenght coding	35.4	0.03	36.8	0.09	44.2	154.9	0.03
Transform and quantization	3223.9	2.64	2178.6	5.50	4269.0	14 753.4	3.13
Deblocking	29.5	0.02	47.4	0.12	44.2	112.6	0.02
Total	121 948.1	100.00	39 577.3	100.00	153 469.3	471 299.3	100.00
Base of the estimated computing load : Baseline profile ; 30 CIF frame/s ; 5 reference frames ; + - 16-Pel search range ; QP = 20							

ME

=> Inter prediction (ME/MC) = more than 95% of the computational cost

Comparison

Modules	STANDARDS		
	MPEG-2	**MPEG-4 ASP**	**H.264 baseline profile**
Motion estimation & compensation			
Block-size	16 x 16	16 x 16 and 8 x 8	16 x 16 ; 16 x 8 ; 8 x 16 ; 8 x 8 ; 8 x 4 ; 4 x 8 ; 4 x 4
Qarter-pel precision	No	Yes	Yes
Multiple reference frame	Up to 2	Up to 2	Yes (5 references frames)
Intraprediction	DC prediction	AC / DC prediction	Yes (9 modes for 4 x 4 blocks and 4 modes for 16 x 16 blocks)
Transform	8 x 8 DCT	8 x 8 DCT	4 x 4 integer transform
Entropy coding	VLC	VLC	VLC and CAVLC
In-loop deblocking filter	No	No	Yes

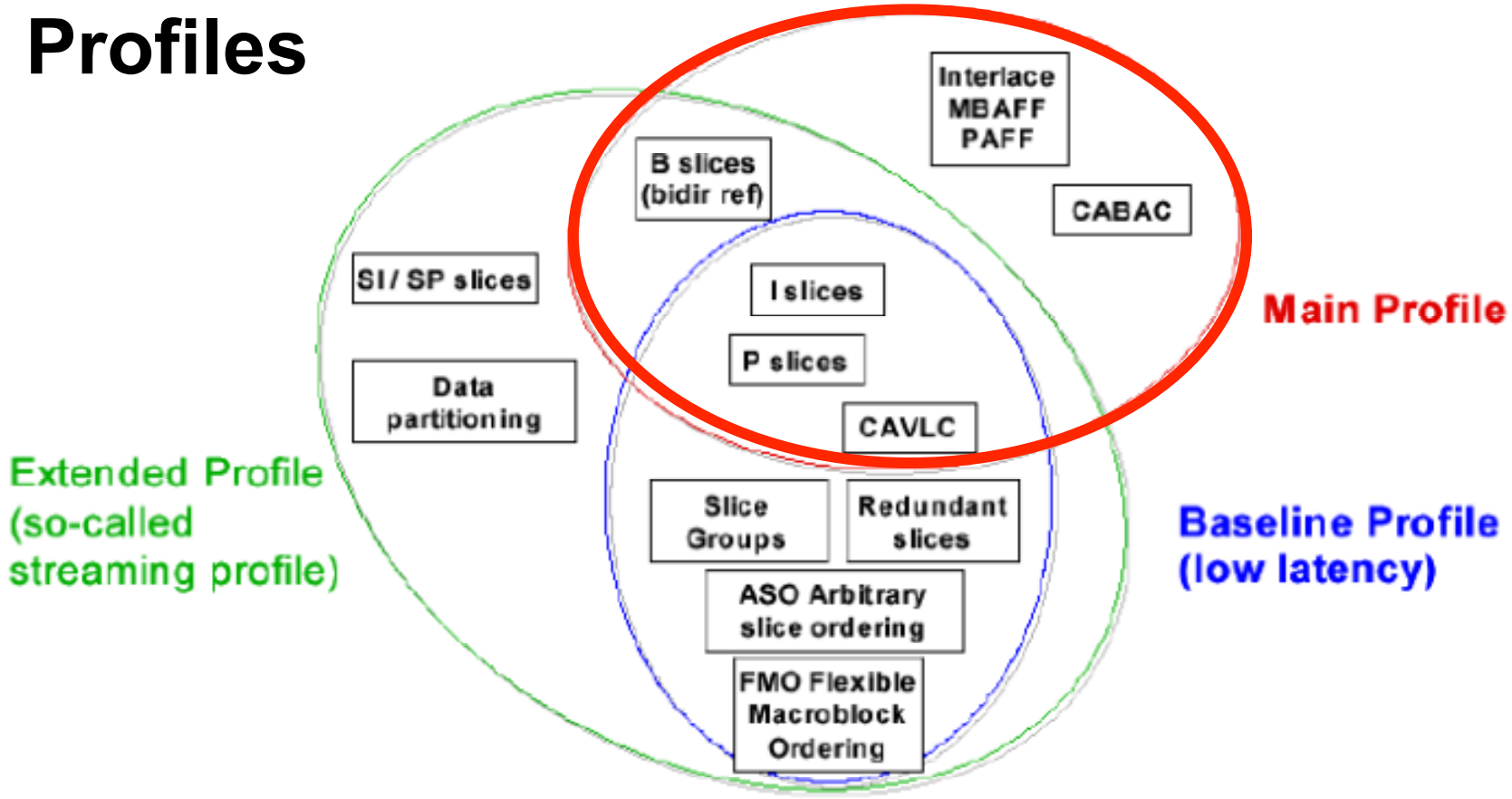
Complexity of AVC …

Comparison with VC1 (WM 9)

	MPEG4-AVC / H.264	VC1
INTRA coding	- **16x16, 8x8, 4x4** blocks - Prediction in spatial domain - 4 modes for 16x16 - 9 modes for 8x8 and 4x4	- **8x8** blocks - Prediction in transform domain -2 modes (H, V)
INTER coding	- **16x16 to 4x4** blocks - ¼ pel precision - Field/Frame predictions	- **16x16, 16x8 or 8x8** blocks - ¼ pel precision - Field/Frame predictions
Transforms	- Integer-DCT transform - Adaptive block transform **8x8 or 4x4** for Inter - Fixed transform for Intra (4x4 TF for 16x16 and 4x4, 8x8 TF for 8x8)	- Integer-DCT transform - Adaptive block transform **8x8, 8x4, 4x8, 4x4** for Inter - Fixed transform for Intra (8x8)
Quantization	- **52** quantization steps - Separate Luma/Chroma - 2 scanning modes	- **31** quantization steps - 8 scanning modes
In-loop filtering	- **4x4** block edges	- **8x8** block edges
Entropy coding	- Context adaptive + Arithmetic coding	- Context adaptive VLC + Bit-plane coding

VC 1 : Microsoft codec
for Window Media 9 (HD DVD, Blu-ray)

Profiles



- **Baseline Profile** – for low-delay end-to-end applications;
- **eXtended Profile** – for mobile applications and e-streaming;
- **Main Profile** – for broadcasting application at SD (Standard Definition) level
- **High Profile**

Level examples (for the main profile)

Level number:					3	3.1	3.2	4	4.1	4.2
Max frame size (macroblocks):					1 620	3 600	5 120	8 192	8 192	8 192
Max macroblocks/second:					40 500	108 000	216 000	245 760	245 760	491 520
Max frame size (samples):					414 720	921 600	1 310 720	2 097 152	2 097 152	2 097 152
Max samples/second:					10 368 000	27 648 000	55 296 000	62 914 560	62 914 560	125 829 120
Format	Luma Width	Luma Height	MBs Total	Luma Samples						
SQCIF	128	96	48	12 288	172.0	172.0	172.0	172.0	172.0	172.0
QCIF	176	144	99	25 344	172.0	172.0	172.0	172.0	172.0	172.0
QVGA	320	240	300	76 800	135.0	172.0	172.0	172.0	172.0	172.0
525 SIF	352	240	330	84 480	122.7	172.0	172.0	172.0	172.0	172.0
CIF	352	288	396	101 376	102.3	172.0	172.0	172.0	172.0	172.0
525 HHR	352	480	660	168 960	61.4	163.6	172.0	172.0	172.0	172.0
625 HHR	352	576	792	202 752	51.1	136.4	172.0	172.0	172.0	172.0
VGA	640	480	1 200	307 200	33.8	90.0	172.0	172.0	172.0	172.0
525 4SIF	704	480	1 320	337 920	30.7	81.8	163.6	172.0	172.0	172.0
525 SD	720	480	1 350	345 600	30.0	80.0	160.0	172.0	172.0	172.0
4CIF	704	576	1 584	405 504	25.6	68.2	136.4	155.2	155.2	172.0
625 SD	720	576	1 620	414 720	25.0	66.7	133.3	151.7	151.7	172.0
SVGA	800	600	1 900	486 400	-	56.8	113.7	129.3	129.3	172.0
XGA	1024	768	3 072	786 432	-	35.2	70.3	80.0	80.0	160.0
720p HD	1280	720	3 600	921 600	-	30.0	60.0	68.3	68.3	136.5
4VGA	1280	960	4 800	1 228 800	-	-	45.0	51.2	51.2	102.4
SXGA	1280	1024	5 120	1 310 720	-	-	42.2	48.0	48.0	96.0
525 16SIF	1408	960	5 280	1 351 680	-	-	-	46.5	46.5	93.1
16CIF	1408	1152	6 336	1 622 016	-	-	-	38.8	38.8	77.6
4SVGA	1600	1200	7 500	1 920 000	-	-	-	32.8	32.8	65.5
1080 HD	1920	1088	8 160	2 088 960	-	-	-	30.1	30.1	60.2
2Kx1K	2048	1024	8 192	2 097 152	-	-	-	30.0	30.0	60.0
4XGA	2048	1536	12 288	3 145 728	-	-	-	-	-	-
16VGA	2560	1920	19 200	4 915 200	-	-	-	-	-	-
3616x1536 (2.35:1)	3616	1536	21 696	5 554 176	-	-	-	-	-	-
3672x1536 (2.39:1)	3680	1536	22 080	5 652 480	-	-	-	-	-	-
4Kx2K	4096	2048	32 768	8 388 608	-	-	-	-	-	-
4096x2304 (16:9)	4096	2304	36 864	9 437 184	-	-	-	-	-	-