Master: INFORMATIQUE
Parcours: VICO Visual Computing

UE: Multimedia Communication

Video coding

vincent.ricordel@univ-nantes.fr

Video coding

Outlines:

- Standards History & functionalities
- MPEG (1-2) principles

Standard ?

- What for ?
  - Interoperability

- Standard vs de facto standard

- Video Standard

  A toolbox

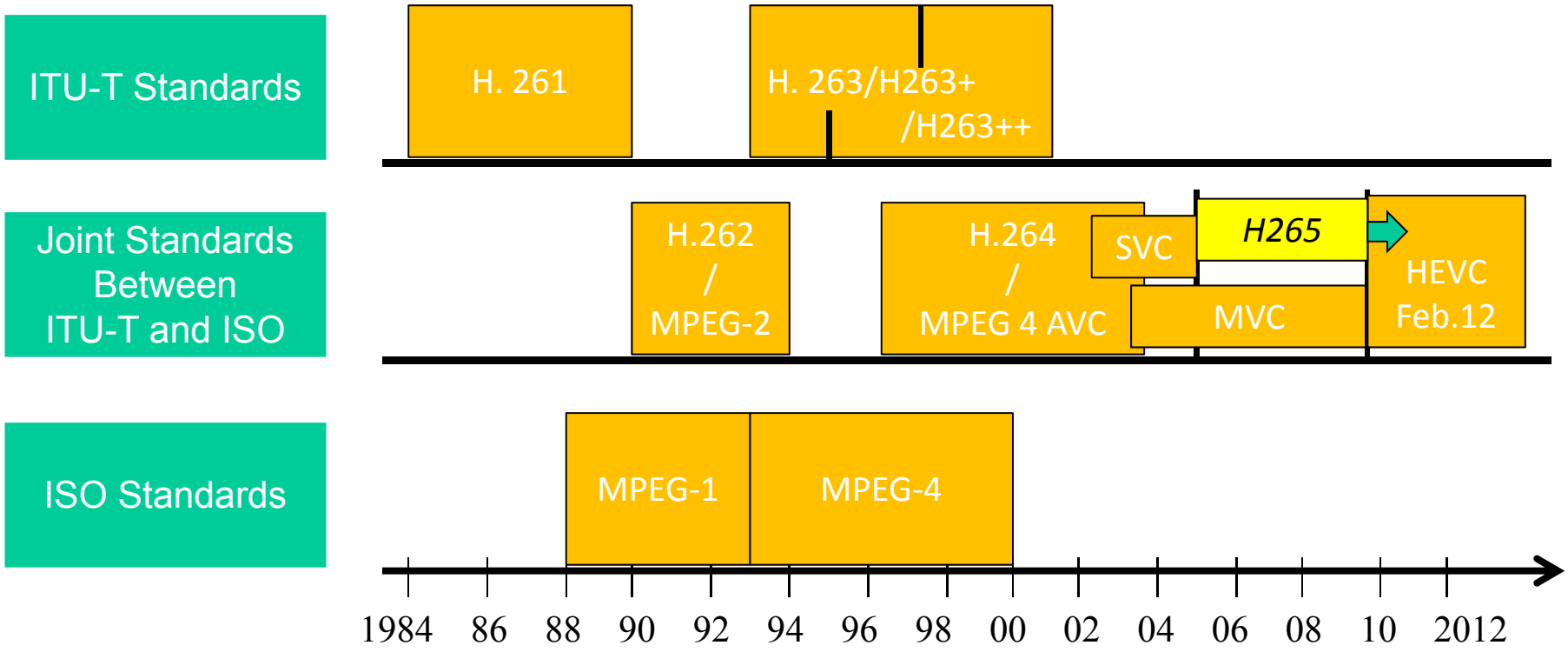
  A Definition of bitstream syntax:

  different CODERs at same bit rate, compliant with the standard might lead to different performances (visual quality) !

- Part 1 : system
- **Part 2 : visual**
- Part 3 : Audio
- Part 4 : conformance
- Part 5 : reference software
- Part 6 : Delivery multimedia Integration framework
- Part 7 : Optimised software for MPEG 4 tools
- Part 8 : MPEG 4 on IP framework
- Part 9 : reference hardware
- **Part 10 : Advanced video coding**

Video Standard History

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



ITU standards

- H261 : first **largely USED** standard for visiophony and visioconf services multiple of 64 kbit/s (64 => 384)
- H263 (version 1, 1995): visio (phony & conference) below 20 kbit/s, better performance & more flexible than H261
- H263+ (version 2, 1998): new modes, basis for MPEG4 part 2
- H263++ (version 3, 2001) : more modes, basis for MPEG-4 (Visual part 10)
- H264 (2003/04) : largely deployed AVC (advanced video coding) .H264-SVC (2005), H264-MVC (2009)
  Target of AVC : HD resolution video
- HEVC (2012)

Jointly With ISO

Standards video MPEG (ISO)

- MPEG 1 (1992) ISO/IEC 11172-2
  - First compression standard for audio and video for storing et reading on CD
  - Bit rate 1.4 Mb/s (1,15 for video), qualité =VCR
- MPEG 2 et H 262 (1994) ISO/IEC 13818-2
  - for TV broadcast
  - Bit rate from 2 to 6 Mb/s (SD) & 15 to 40 MB/s (HDTV)
- MPEG 4 Part 2 (1999) ISO/IEC 14996-2
  - Multimedia applications, new functionalities
  - Low bit rate
  - Support object coding (mesh, texture)

Then joint Standards with ITU

! MPEG-7, MPEG-21, MPEG DASH for other purposes

Since 2012 … by ISO & ITU-T

- HEVC/H.265 “High Efficiency Video Coding”
  - By the JCT-VC “Joint Collaborative Team on Video Coding” created in 2010
    - Target : 2K resolution video
  - Versions
    - April 2013 : 1st version
    - October 2014 : 2nd vers. (MV-HEVC & SHVC)
    - February 2015 : 3rd vers. (3D-HEVC)
- VVC/H.266 “Versatile Video Coding”
  - By the JVET “Joint Video Exploration Team” created in October 2015 (which became the “Joint Video Experts Team” in 2018)
    - Target : 4K resolution video, HDR “High Dynamic Range”, and 360° (omnidirectional immersive media)
  - 1st version : July 2020
- MPEG-IPCC “Immersive Media & Point Cloud Compression”
  - Target & methods : V-PCC (Video-based PCC) and G-PCC (Geometry-based PCC)
  - 2017 : 1st Call
  - 2020 : 1st version

Since 2012 … by AOM

- AOM “Alliance for Open Media”
  - 2015 foundation of the consortium by Amazon, Cisco, Google, Intel Corp., Microsoft, Netflix, Mozilla foundation
    - joined latter by ARM, Nvidia, Facebook, VideoLAN, Adobe, Apple …
  - For Video transmissions over the Internet
  - Open source, royalty free licensing model
- AV1 “AOMedia Video Codec”
  - Successor of VP9 (Google)
  - September 2018 : 1st use on YouTube
  - 2019 : supported by the internet browsers (software decoding)
  - 2020 : supported by processors (hardware decoding)
  - Avril 2019 : SVT-AV1 “Scalable Video Technology for AV1”

Image & video representation

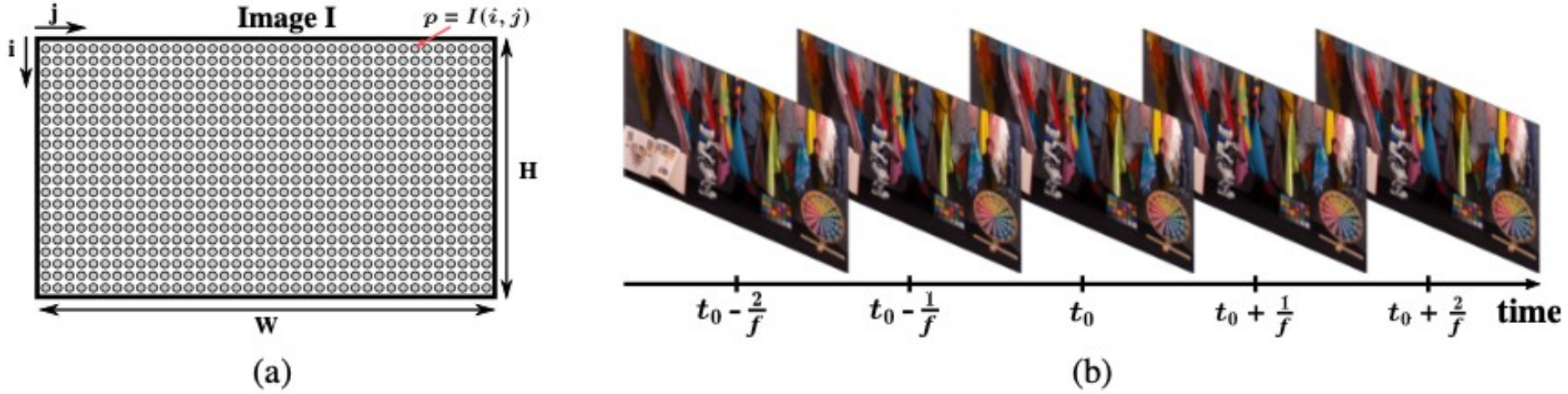


Fig. 2.1 Representation of (a) an image with a spatial resolution $W \times H$ and (b) a video sequence with a frame-rate f.

Typical input formats for compression

	Frame Format	Sampling	Frame size		Frame rate
			Luminance	Chrominance	(frame/second)
•	SQCIF	4:2:0	128x96	64x48	15/10/7.5
•	QCIF	4:2:0 Progres..	176x144	88x72	30/15/10/7.5
•	CIF	4:2:0 Progres..	352x288	176x144	30/15/10/7.5
•	SIF	4:2:0 Progres..	360x288	180x144	25
•	4CIF	4:2:0 Progres..	704x576	352x288	25
•	16CIF	4:2:0 Prog./Interl.	1408x1152	704x576	50/25
•	CCIR601	4:2:2 Interlaced	720x576	360x576	25
•	720 25p	4:2:2 progressive	1280x720	640x720	25
•	720 30p	4:2:2 Progressive	1280x720	640x720	29.97
•	720 50p	4:2:2 Progressive	1280x720	640x720	50
•	720 60p	4:2:2 Progressive	1280x720	640x720	59.94
•	1080 50i	4:2:2 Interlaced	1920x1080	960x1080	50field/s;25 fps
•	1080 50p	4:2:2 Progressive	1920x1080	960x1080	50 fps
•	1080 60i	4:2:2 Interlaced	1920x1080	960x1080	59.94field/s;29.97fps
•	1080 60p	4:2:2 Progressive	1920x1080	960x1080	59.94fps

SD

HD

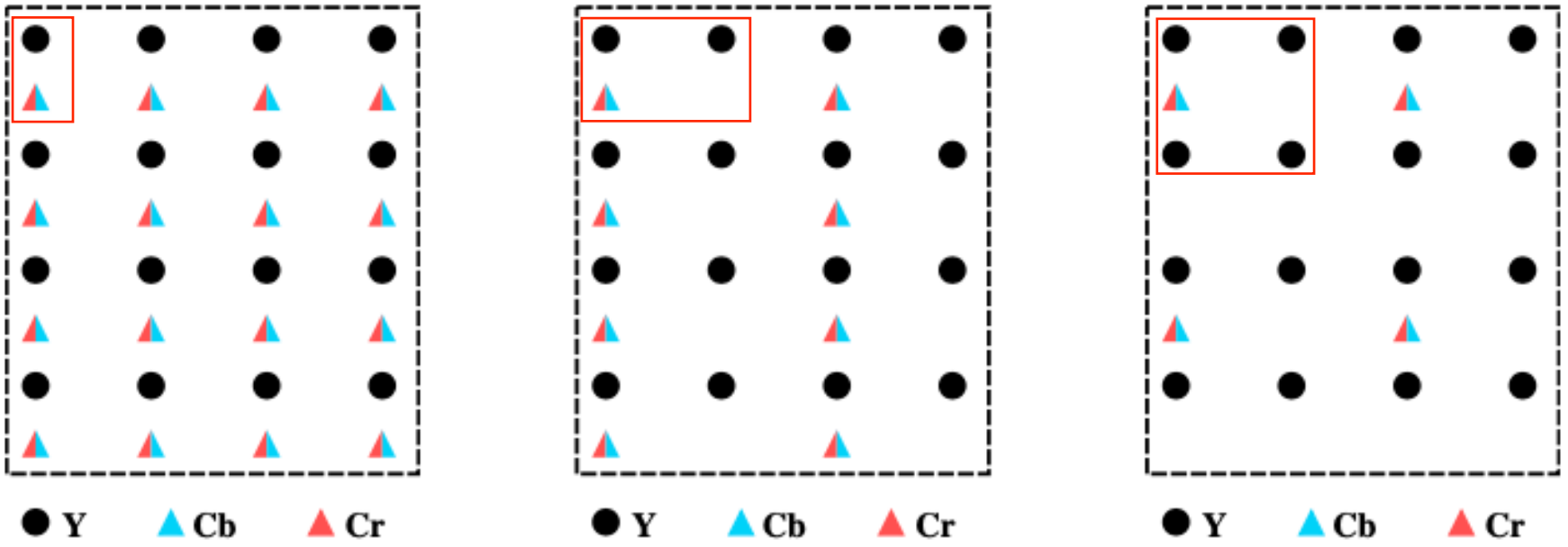
HDTV & UHDTV formats

	HDTV	UHDTV
Aspect ratio	16/9	16/9
Spatial resolution	1920x1080 (2K, 1080p)	3840x2160, 7680x4320 (4K, 8K, 2160p, 4320p)
Frame-rate	60, 60/1.001, 50, 30, 30/1.001, 25, 24, 24/1.001	120, 60, 60/1.001, 50, 30, 30/1.001, 25, 24, 24/1.001
Color gamut	BT.709 color primaries	BT.2020 color primaries (WCG)
Dynamic range	SDR	SDR, HDR

Table 2.1 Parameter values for HDTV and UHDTV video systems.

Video subsampling formats

Only for transmission purpose.
All the pixels have to be coloured when displayed
=> the missing chrominances are interpolated



(a) 4:4:4 24bpp (b) 4:2:2 16bpp (c) 4:2:0 12bpp

Fig. 2.2 Representation of luminance (Y) and chrominance (Cb and Cr) pixel components for common chroma subsampling formats.

Wide Color Gamut

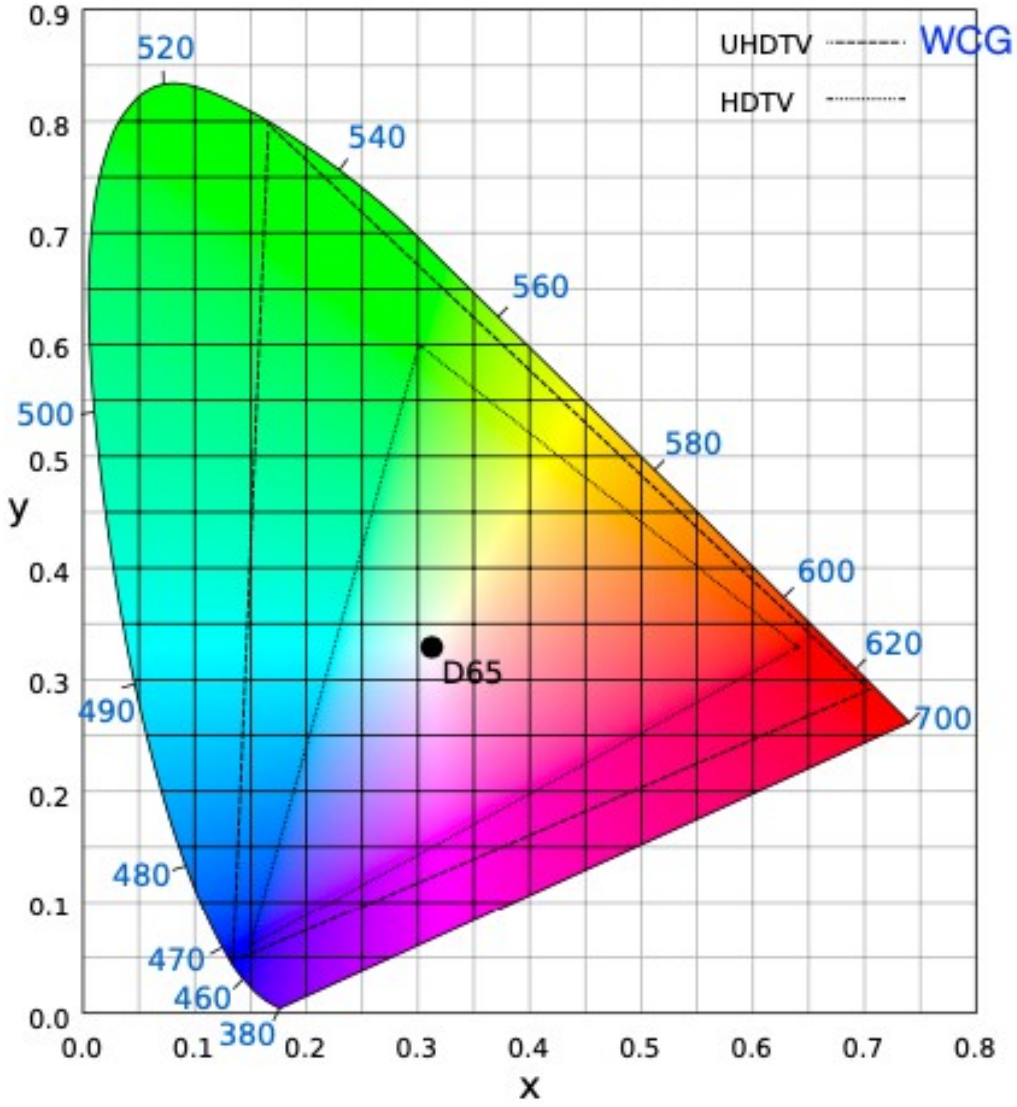


Fig. 2.3 Representation of the Rec. BT.709 (HDTV) and BT.2020 (UHDTV) using the CIE 1931 diagram.

WCG (Wide Color Gamut)

UHDTV deployment

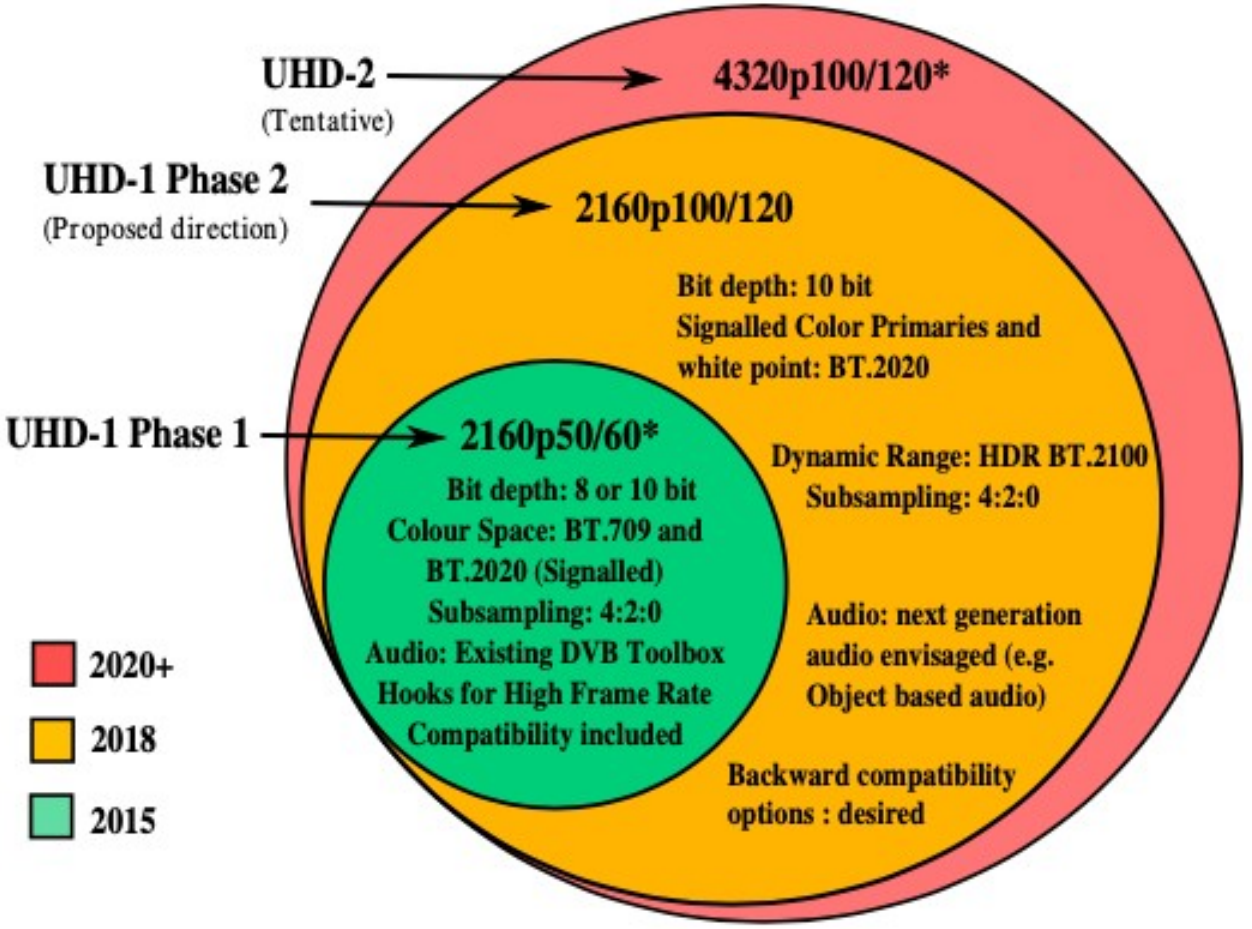


Fig. 1.1 DVB phased UHDTV deployment

Different scalability types

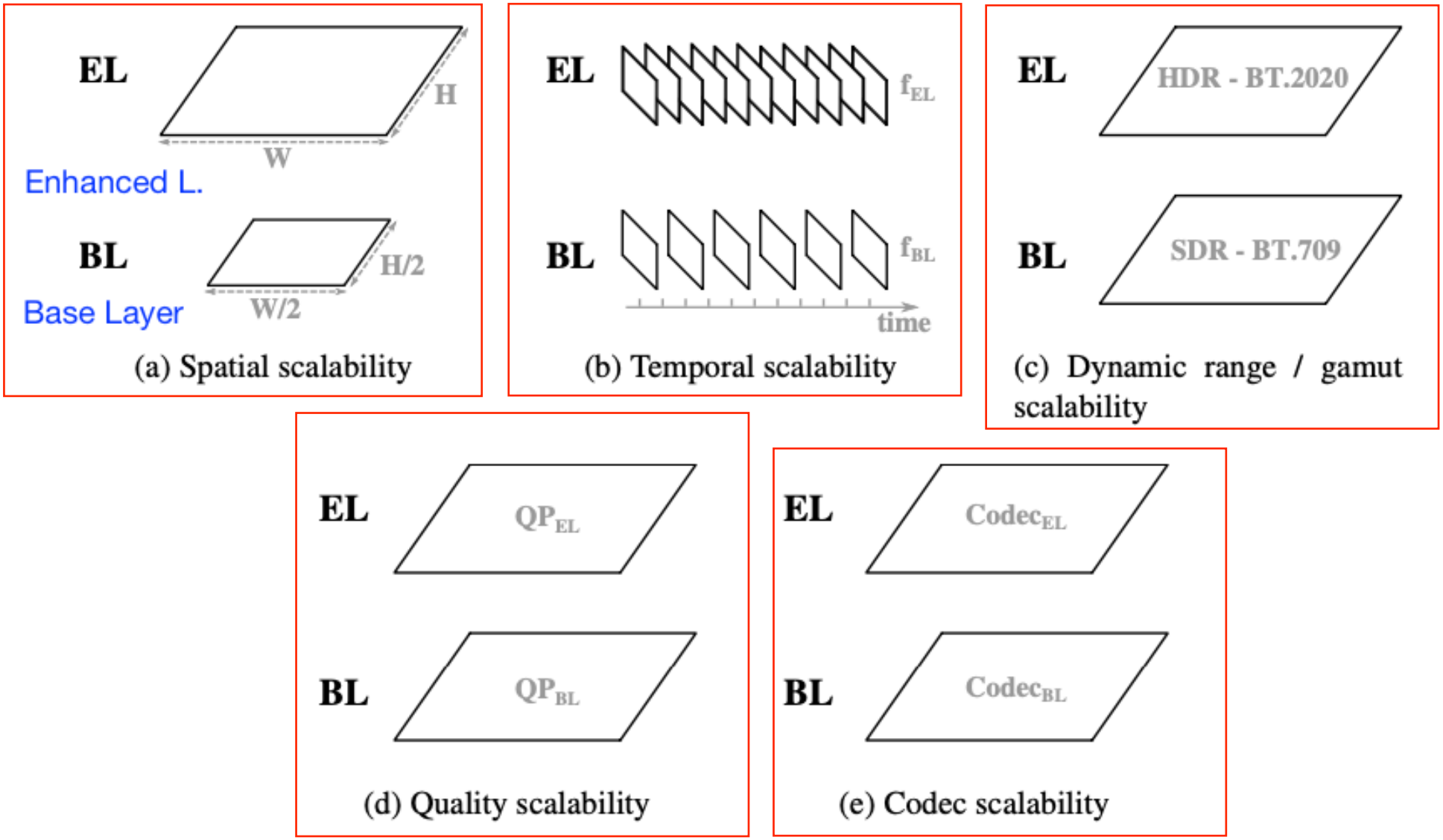


Fig. 2.5 Different types of scalability.

MSE, Quantization, PSNR

$$MSE(\mathbf{I},\mathbf{I}_d) = \frac{1}{W \cdot H} \times \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} (\mathbf{I}(i,j) - \mathbf{I}_d(i,j))^2,$$

$$\mathbf{B}_{n,quant}(u,v) = \Delta \cdot \left\lfloor \frac{|\mathbf{B}_{n,trans}(u,v)|}{\Delta} + \frac{1}{2} \right\rfloor \cdot sgn(\mathbf{B}_{n,trans}(u,v)) \; \forall (u,v) \in [0,H[\times[0,W[,$$

$$PSNR(\mathbf{I},\mathbf{I}_d) = 10 \times \log_{10}\left(\frac{2^b - 1}{MSE(\mathbf{I},\mathbf{I}_d)}\right),$$

For a YUV frame

$$PSNR_{YUV} = \frac{6 \times PSNR_Y + PSNR_U + PSNR_V}{8}.$$

BD-Rate

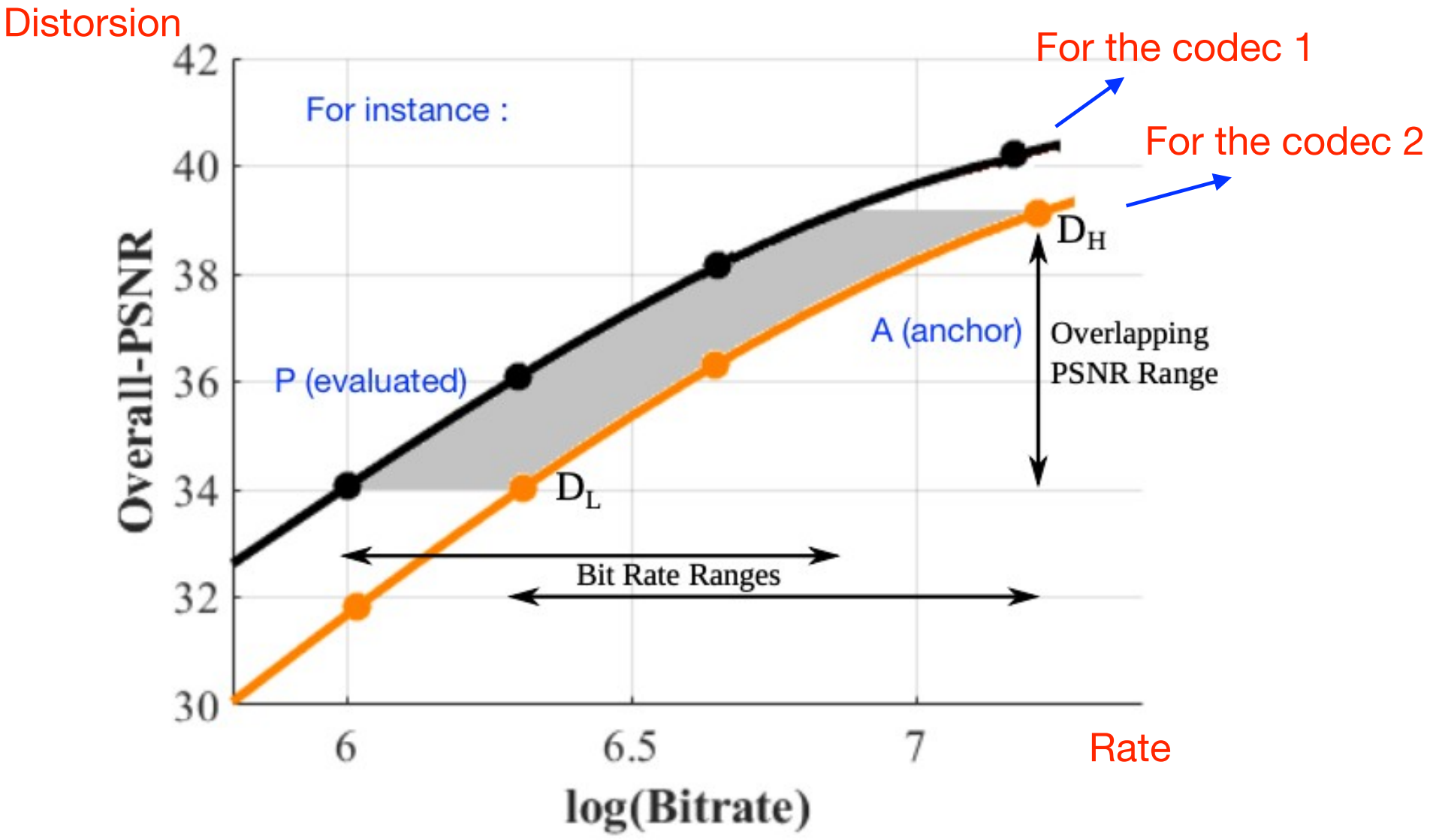


Fig. 2.6 Illustration of Bjøntegaard Delta (BD-Rate) computation.

Complexity

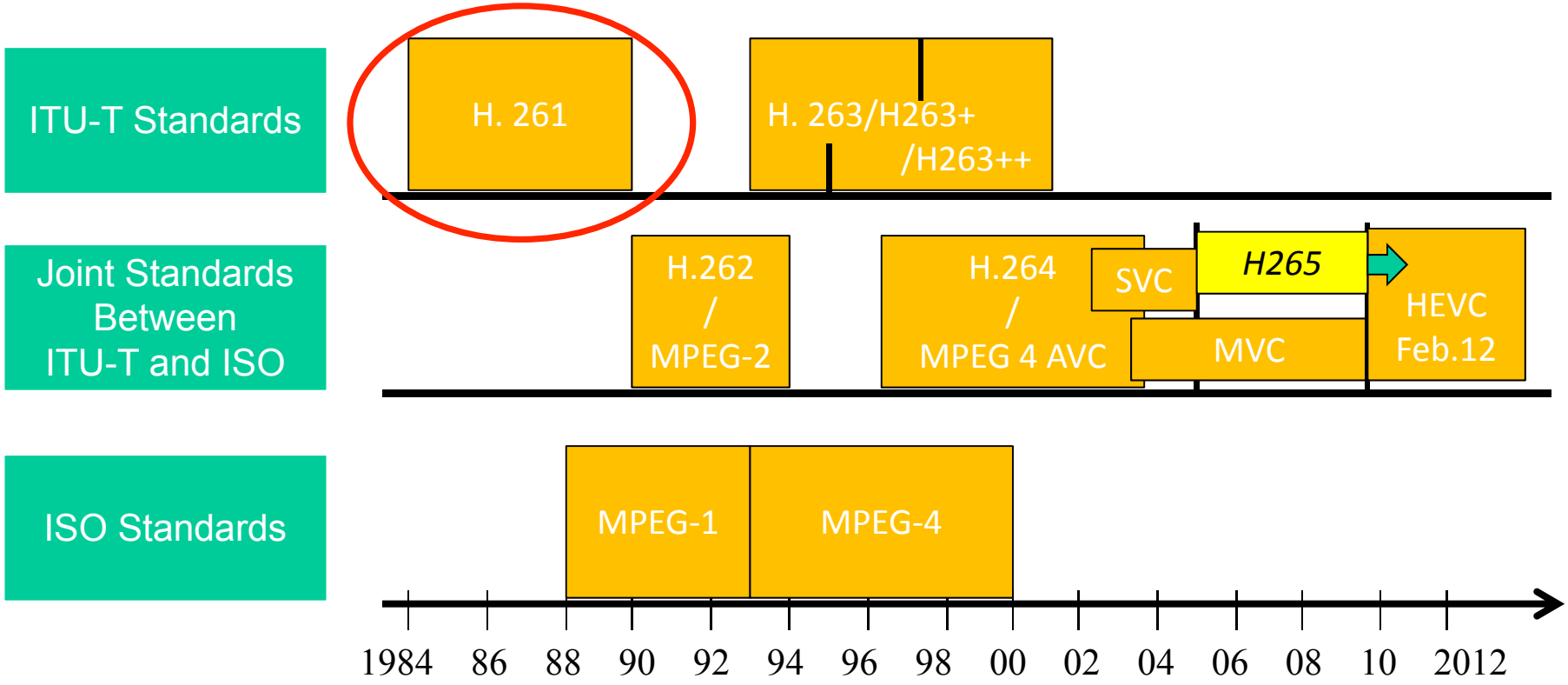
$$TR_{\%}(T_P, T_A) = 100 \times (1 - \frac{T_P}{T_A})$$

Encoding time with Codec 1

Encoding time with Codec 2

Video Standard History

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



H261 (part visual of H320)

- 1990 : visio conference
- Input formats:

Input image format	CIF	QCIF
Number of active pels/line		
Lum (Y)	360 (352)	180 (176)
Chroma (C_R,C_B)	180 (176)	90 (88)
Number of active lines/frame		
Lum (Y)	288	144
Chroma (C_R,C_B)	144	72
Interlacing	1:1	1:1
Temporal rate (f/s)	30; 15; 10 or 7.5	30; 15; 10 or 7.5
Aspect ratio	4:3	4:3

H261 bitrate

- Bit rate: Multiple of 64 kbps (=canal ISDN, Integrated Services Digital Networks)
  - Typically 384 Kbps with audio
- Quality : acceptable @ 128 kbps if poor motion
- maximum delay: 150 ms
  - Interactive service !

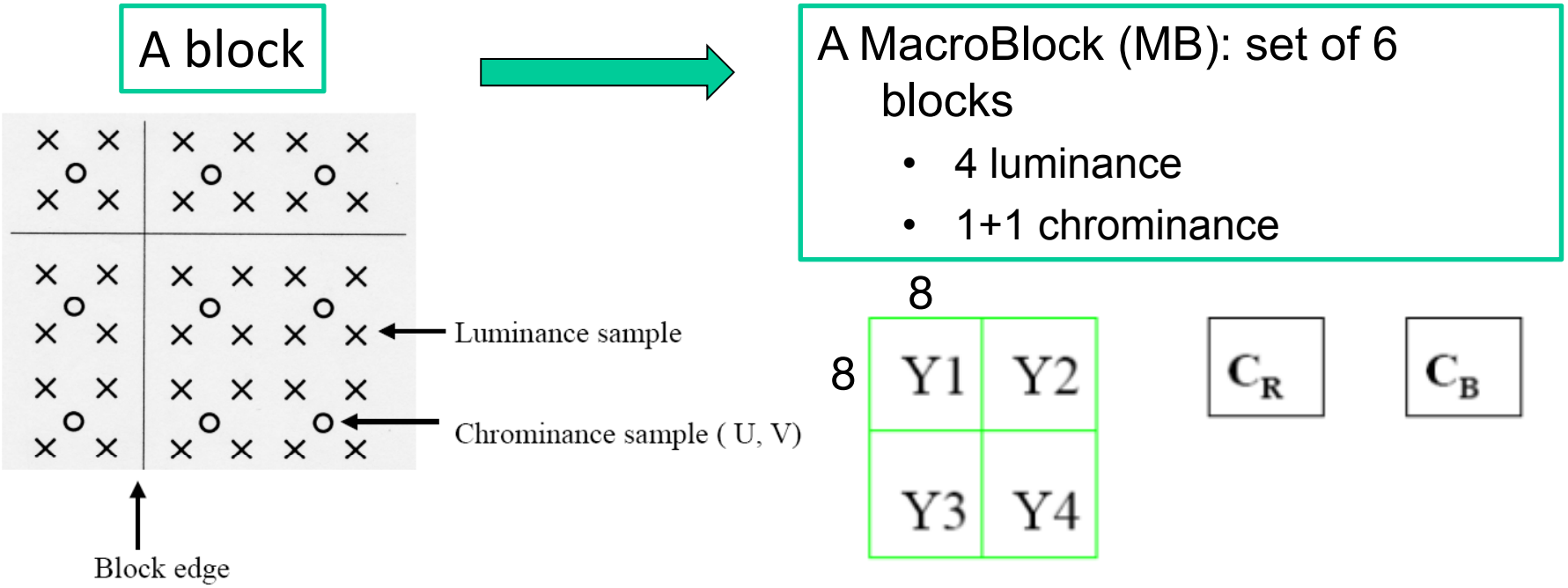
- Original Bandwidth:
  - CIF 30 fps : 37.3 Mbit/s
  - QCIF 30 fps : 9.35 Mbit/s
- Compression rate :
  - QCIF 10 fps => 64 kb/s =>tc=48
  - CIF 15 fps => 384 kb/s => tc=50

H261 : principles

- Block based Hybrid coding: DCT & DPCM with motion compensation

H261 : principles

- Block based Hybrid coding: DCT & DPCM with motion compensation



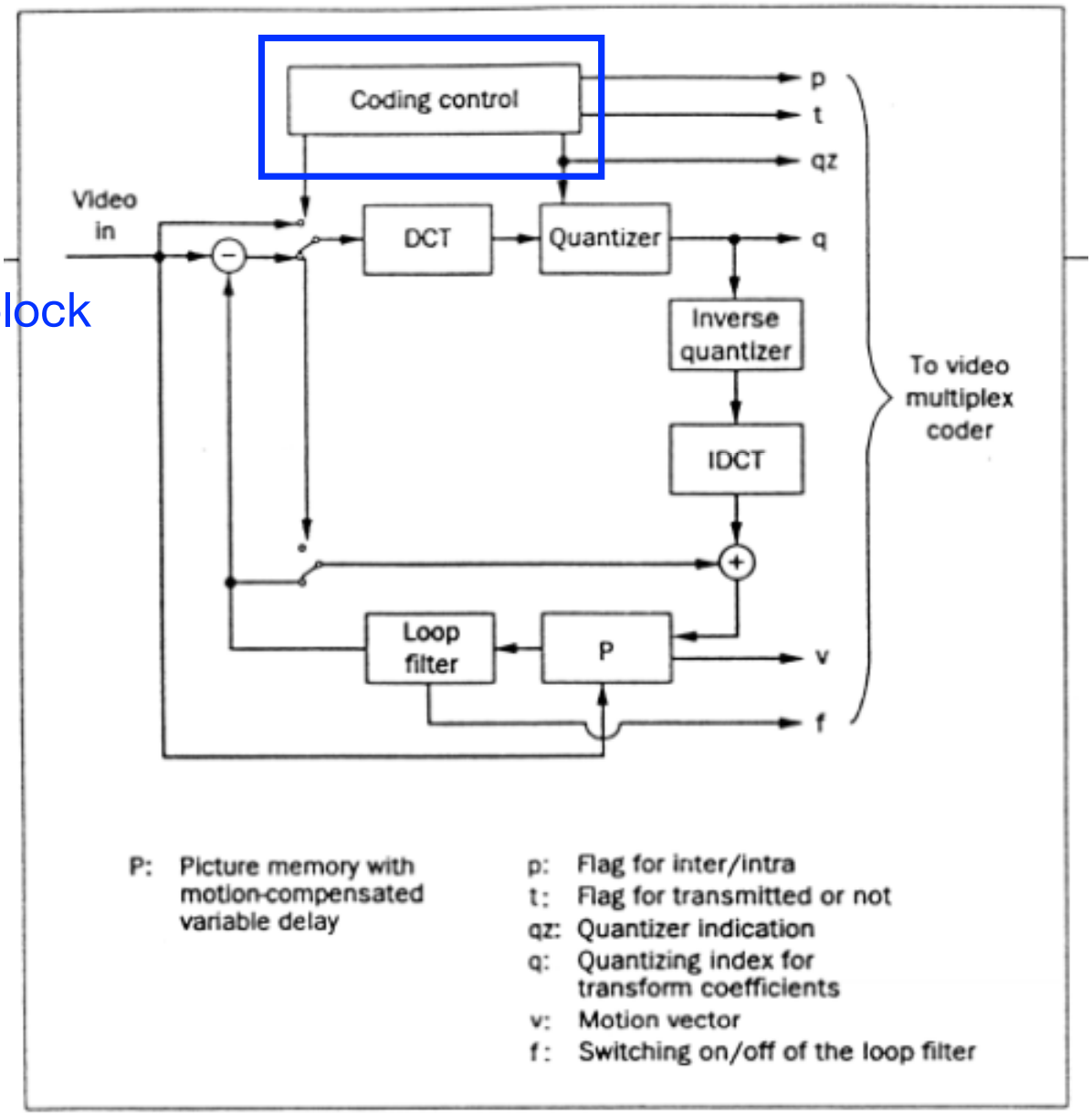
H261 : principles

- Block based Hybrid coding: DCT & DPCM with motion compensation

- Two main modes: Intra & Inter
  - Intra : a block is transformed (DCT), quantized & coded VLUC (Variable Length Universal code)
  - Inter : a block is predicted from previous decoded frame by Frame difference **or** motion estimation and compensation

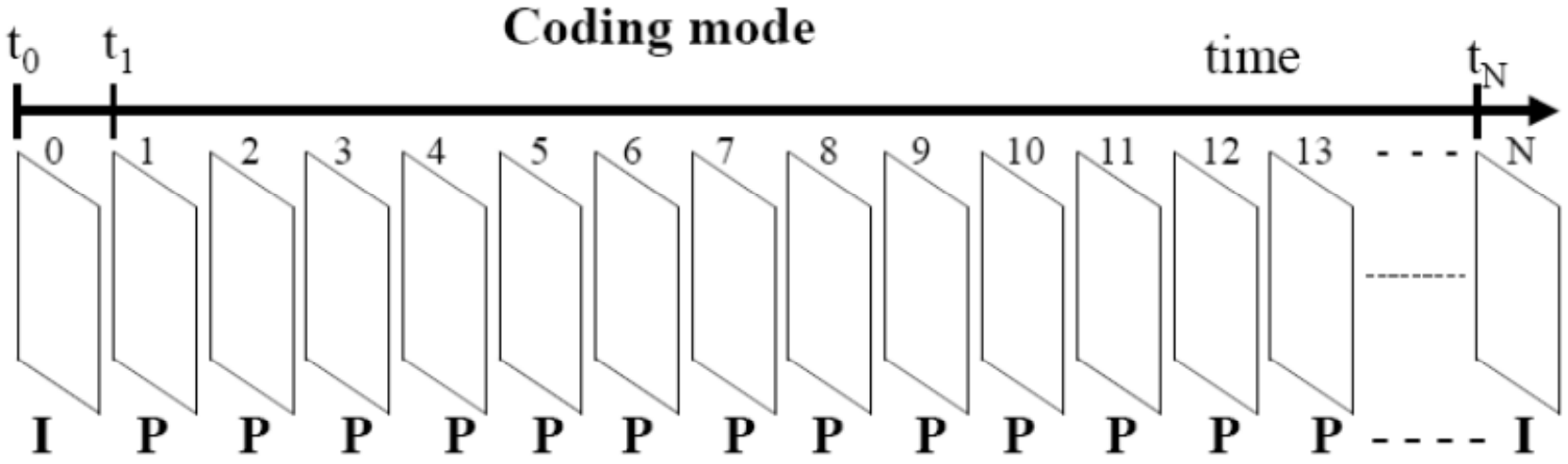
    code(quantification(DCT(prediction error)
  - Bit rate controlled by adjusting quantization steps => BR constant in mean

H261


Coding control
p
t
qz
Video in
DCT
Quantizer
q
Inverse quantizer
IDCT
To video multiplex coder
Loop filter
P
v
f
P: Picture memory with motion-compensated variable delay
p: Flag for inter/intra
t: Flag for transmitted or not
qz: Quantizer indication
q: Quantizing index for transform coefficients
v: Motion vector
f: Switching on/off of the loop filter

block

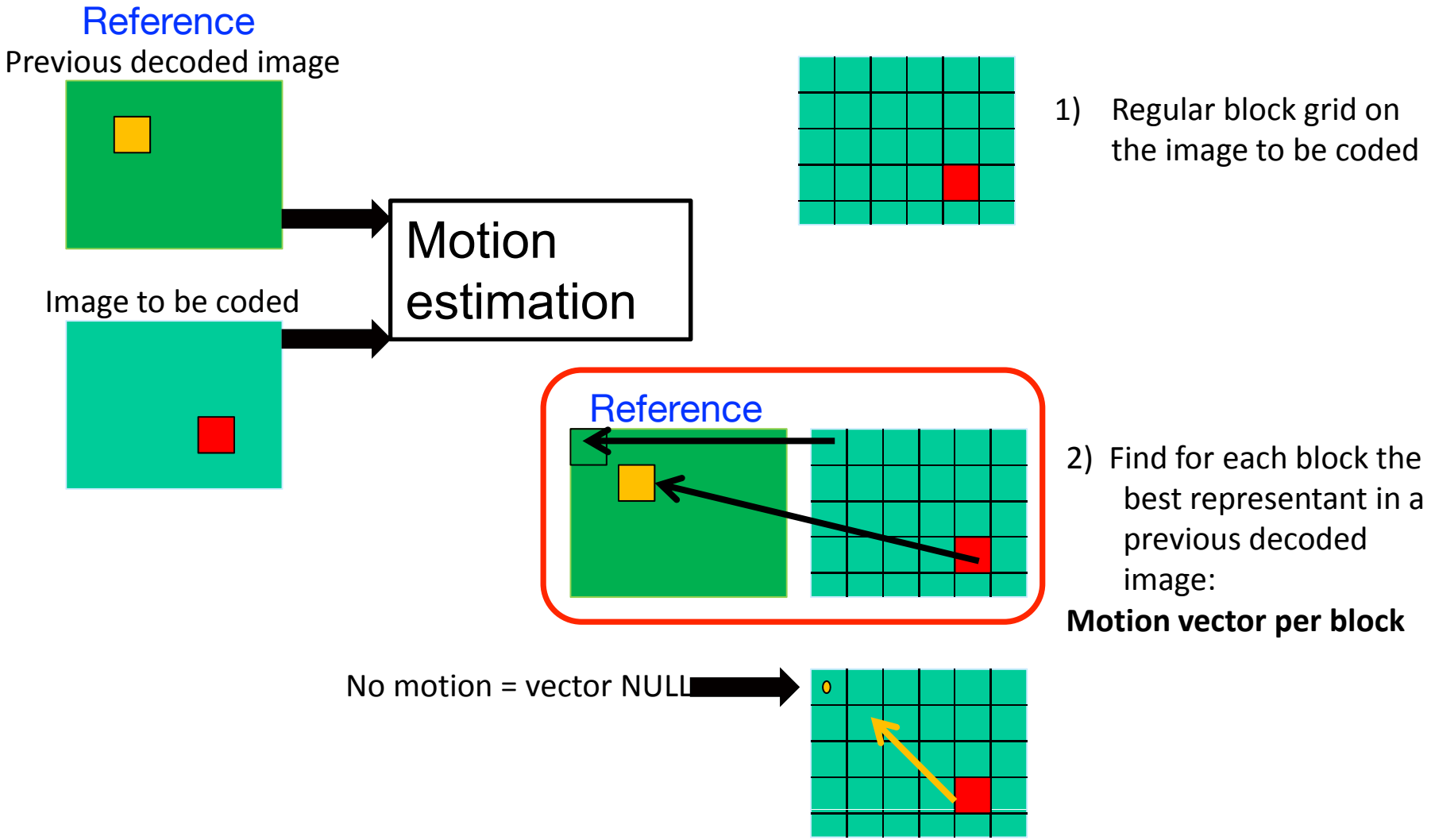
Coding sequence



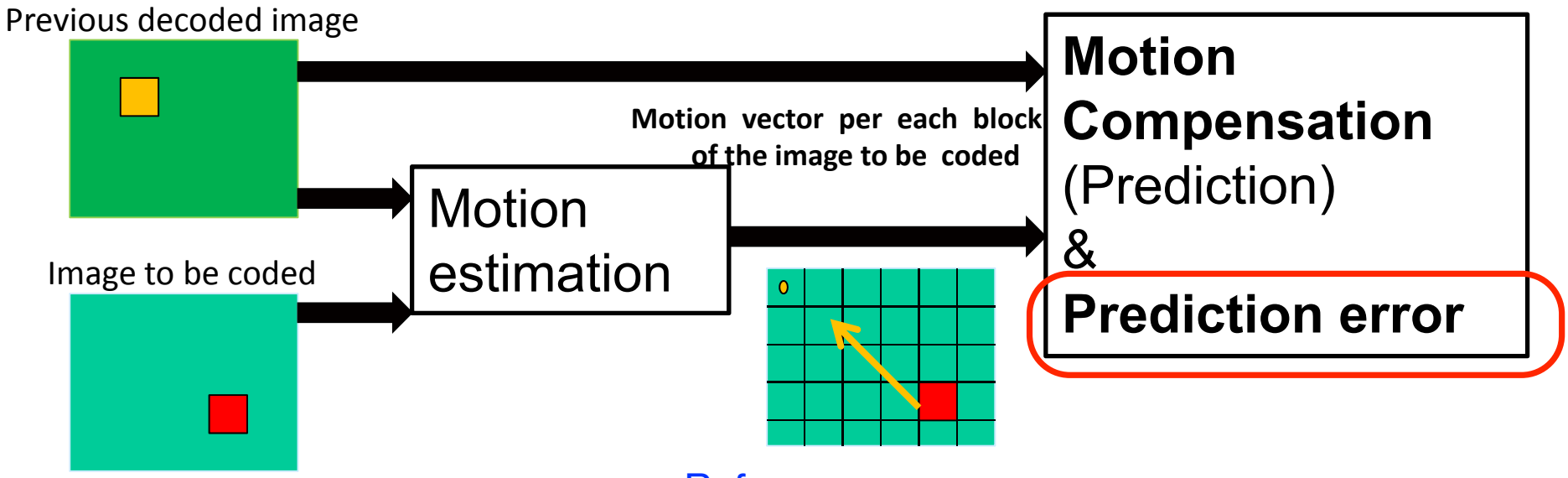
I: Intra frame coding *(first frame & every 132 frames (at least))*

P: Inter frame coding *(by motion compensation of the previous decoded frame)*

Inter coding

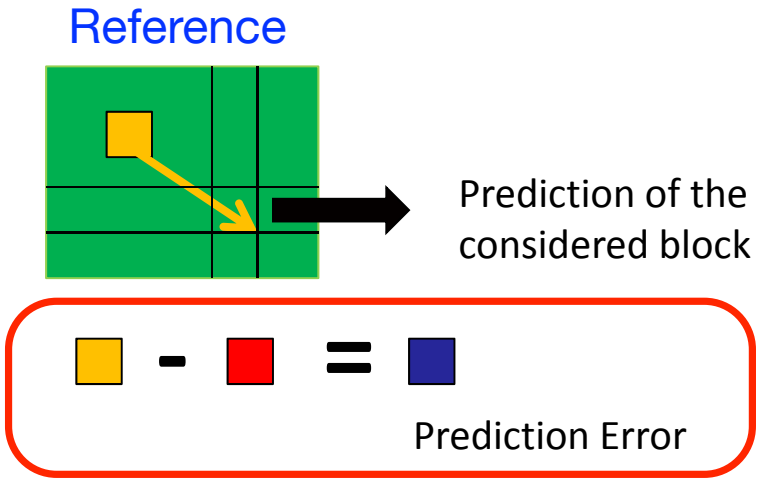


Inter coding (2)

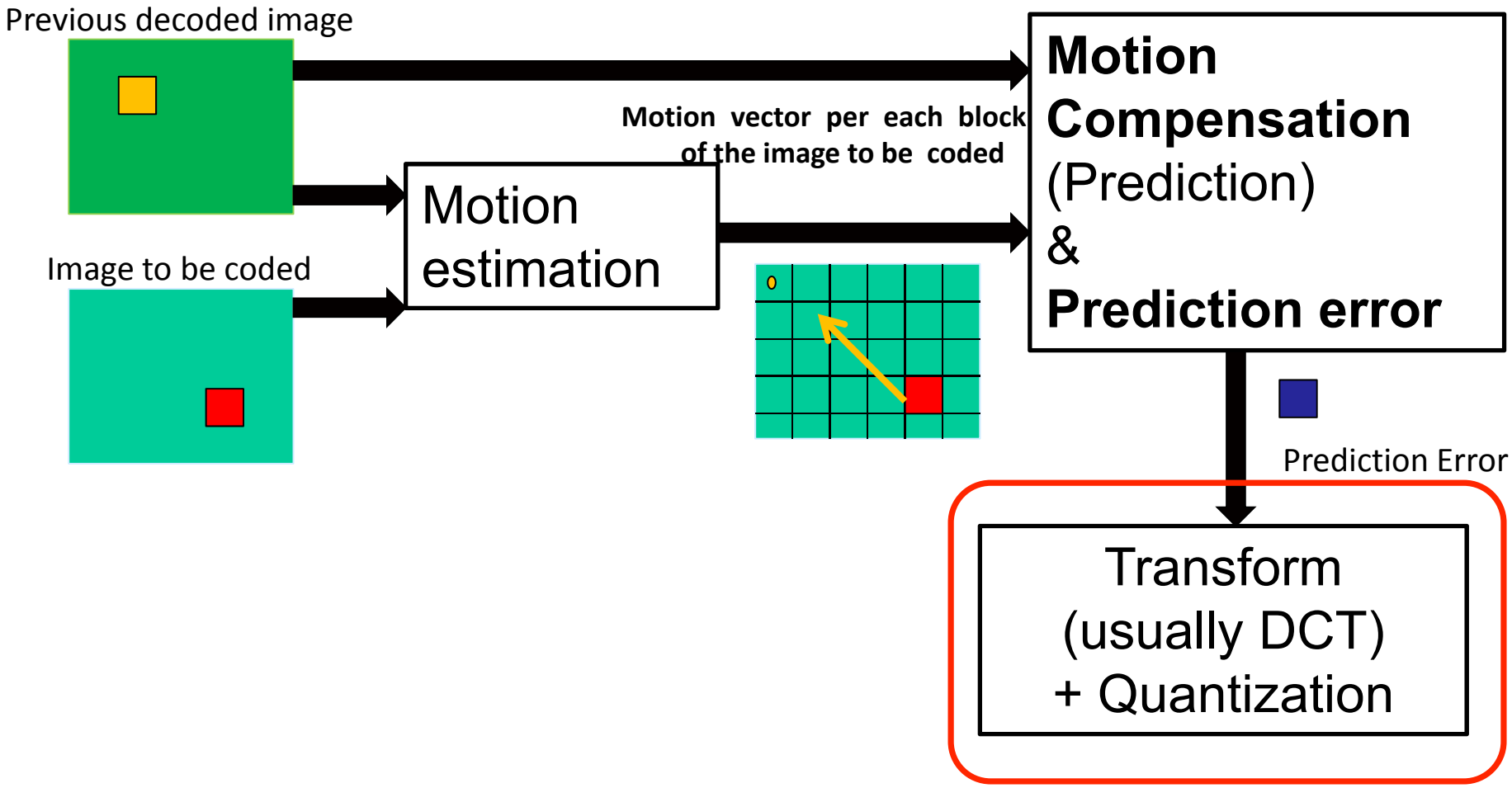


1) For each block : compensate the motion (from previous decoded image) using the motion vector => prediction
2) Compute the difference between the two blocks:

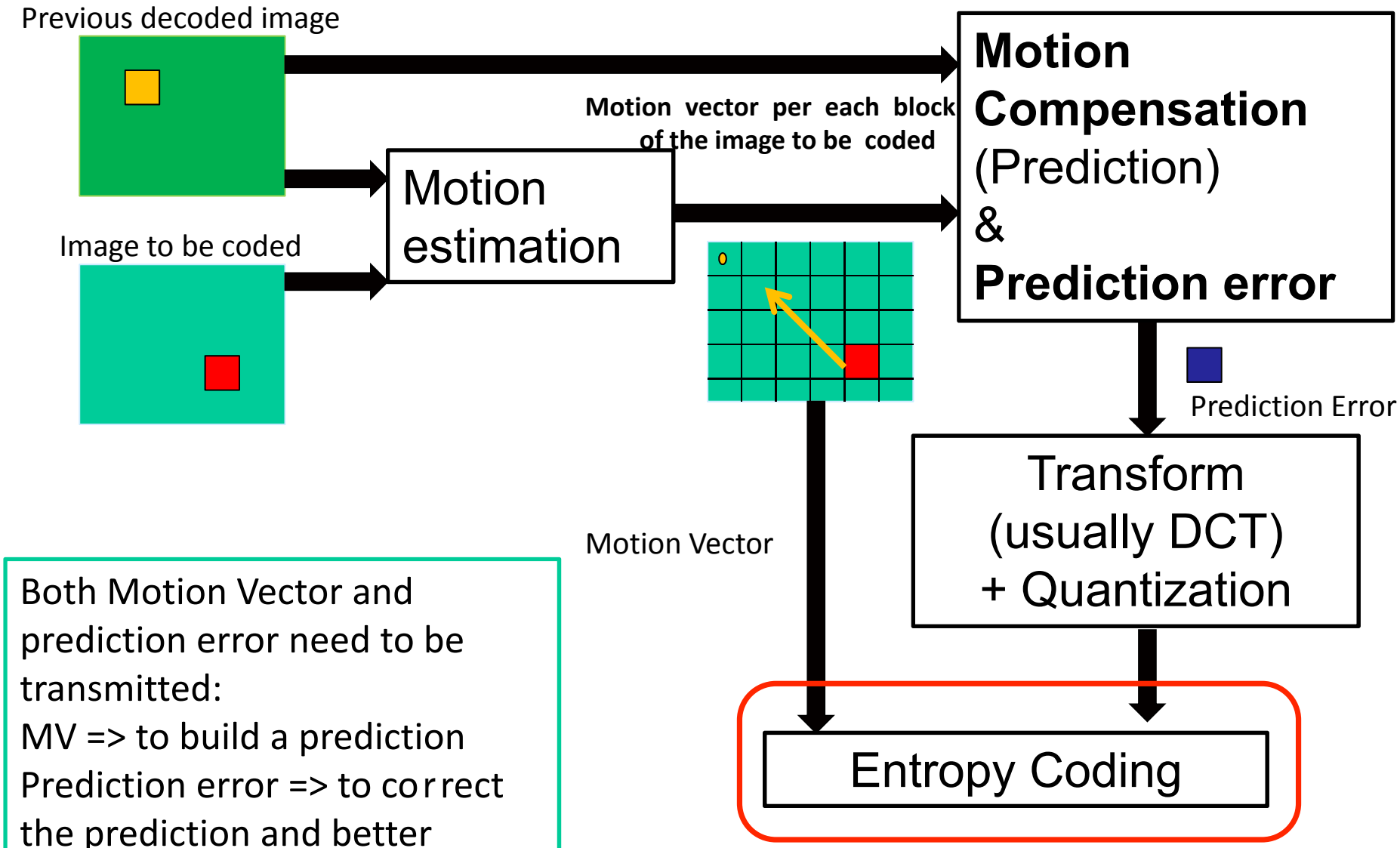
Block to be coded and its prediction



Inter coding (3)



Inter coding (4)



Both Motion Vector and prediction error need to be transmitted:
MV => to build a prediction
Prediction error => to correct the prediction and better decoding result

Coding modes

- For each MB, we have the choice !

- Several options:
  - Intra or FD or MCFD
  - Modification of the quantization step
  - With or without transmission of quantized prediction error

FD : Frame Difference
MCFD : Motion Compensated Frame Difference

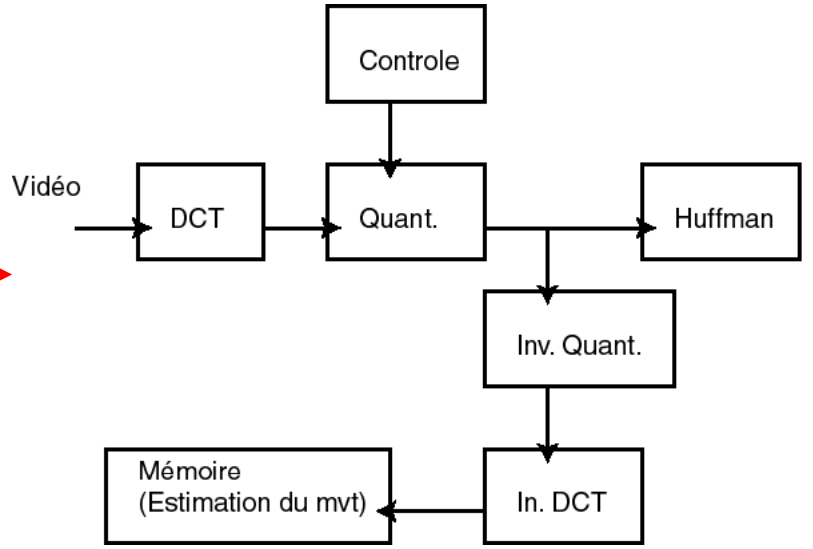
Prediction	MQUANT	MVD	CBP	TCOEFF	VLC
Intra				x	0001
Intra	x			x	0000 001
Inter			x	x	1
Inter	x		x	x	0000 1
Inter+MC		x			0000 0000 1
Inter+MC		x	x	x	0000 0001
Inter+MC	x	x	x	x	0000 0000 01
Inter+MC+FIL		x			001
Inter+MC+FIL		x	x	x	01
Inter+MC+FIL	x	x	x	x	0000 01

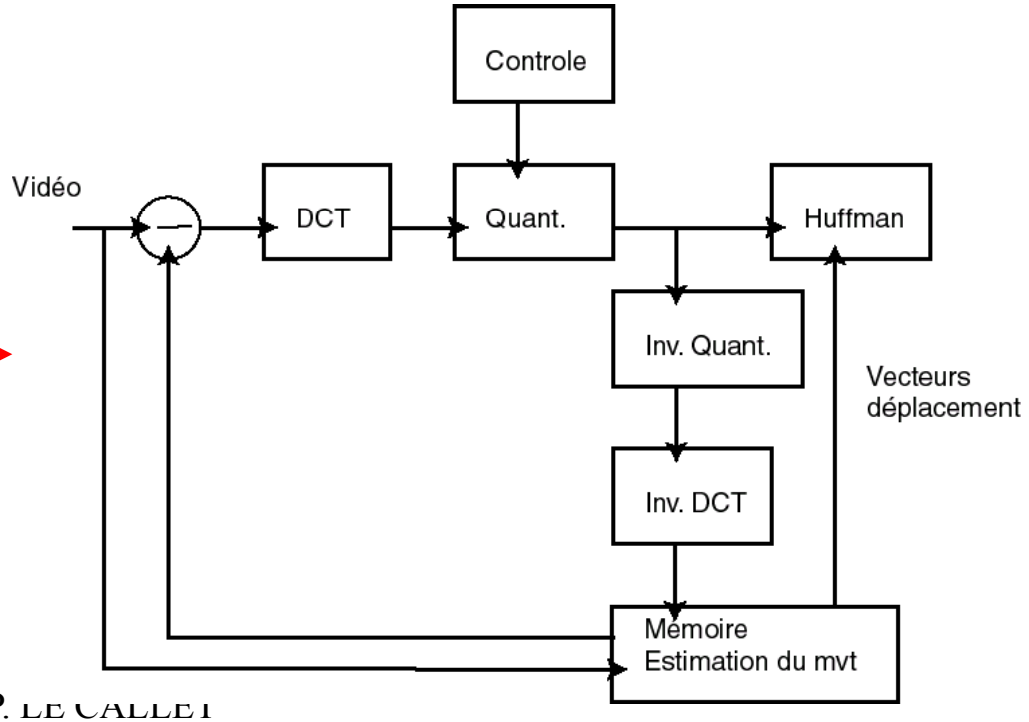
Indication of the coding mode in the bitstream

Summary

- « I »
- « P »

(2 possibilities: FD or MCFD)



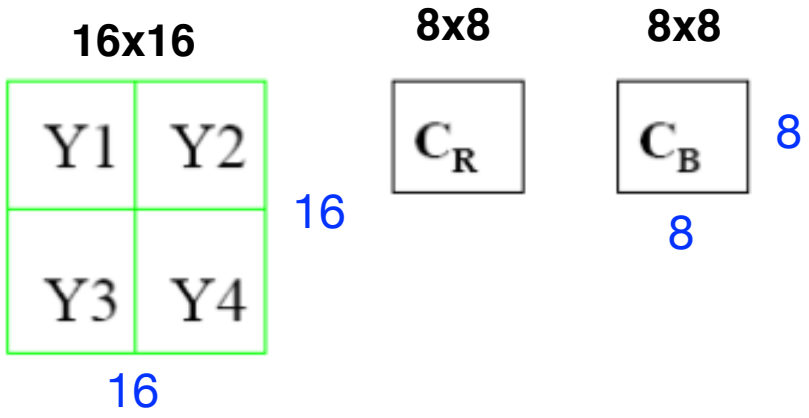


Structuration

- H261 divides frames into 4 hierarchical structures:
  - Picture layer
  - Group of Blocks Layer (GOB)
  - Macroblock layer (MB)
  - Block layer (B)

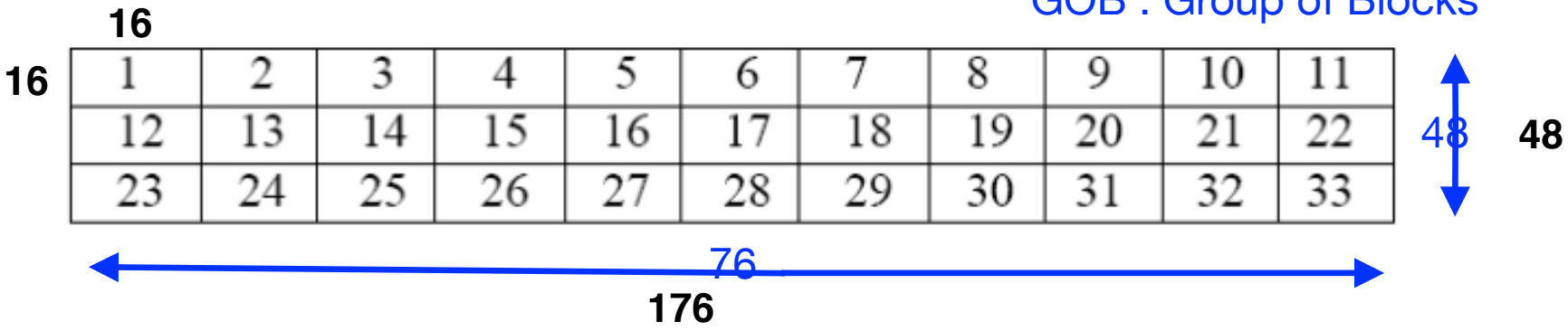
H261 : Structuration MB et GOB

- MB layer : set de 6 blocks
  - 4 luminance
  - 1+1 chrominance



- One GOB contains 33 MB

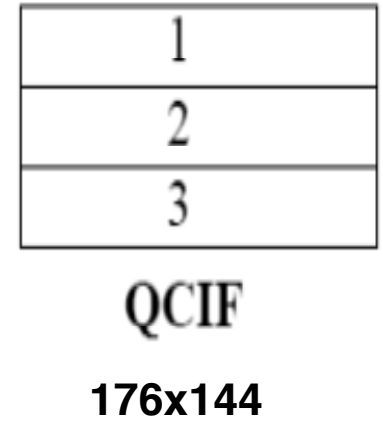
GOB : Group of Blocks



Structuration : Picture layer

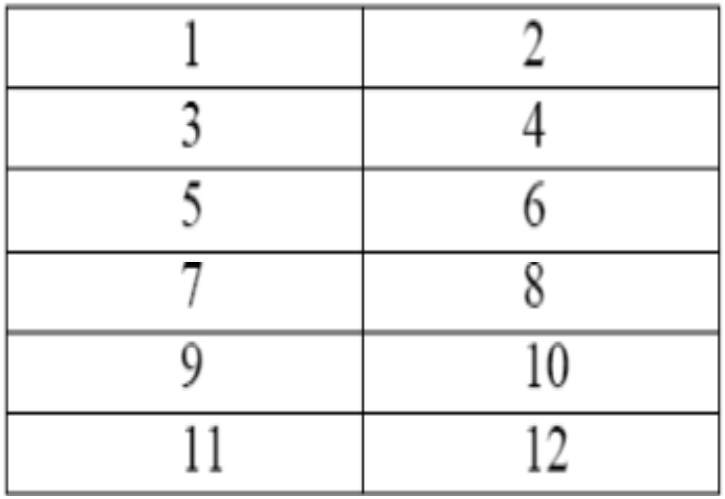
Composition of the Picture layer

Format	Number of GOBs in a frame	Number of MBs in a GOB	Total number of MBs in a frame
CIF	12	33	396
QCIF	3	33	99



QCIF

176x144



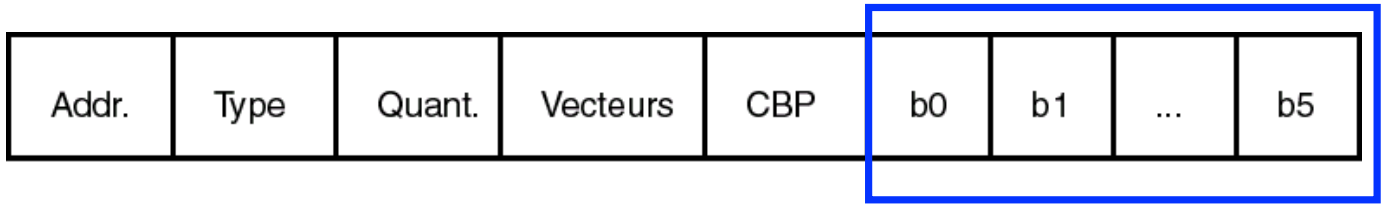
CIF

352x288

Bitstream

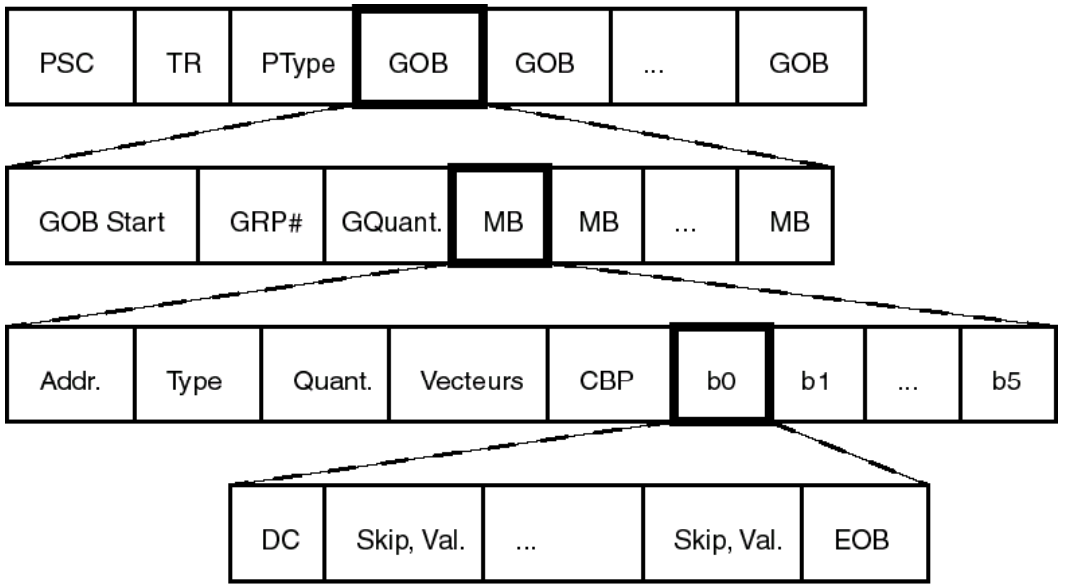
- For each MB

Header



  - Addr : position in Frame
  - Type : coding mode
  - Quant. : quantization step **(if needed)**
  - Vecteurs **(if needed)**
  - CBP: transmitted DCT coefficient for each bloc (from 0 to 5) only if needed (for each of them)
    - CBP « Coded Block Pattern » : list of blocs for which it is needed => to interpret following information
  - b0, b1, , b5: code(quantized(DCT coeff)) **if needed**

« H261 Bitstream » structure



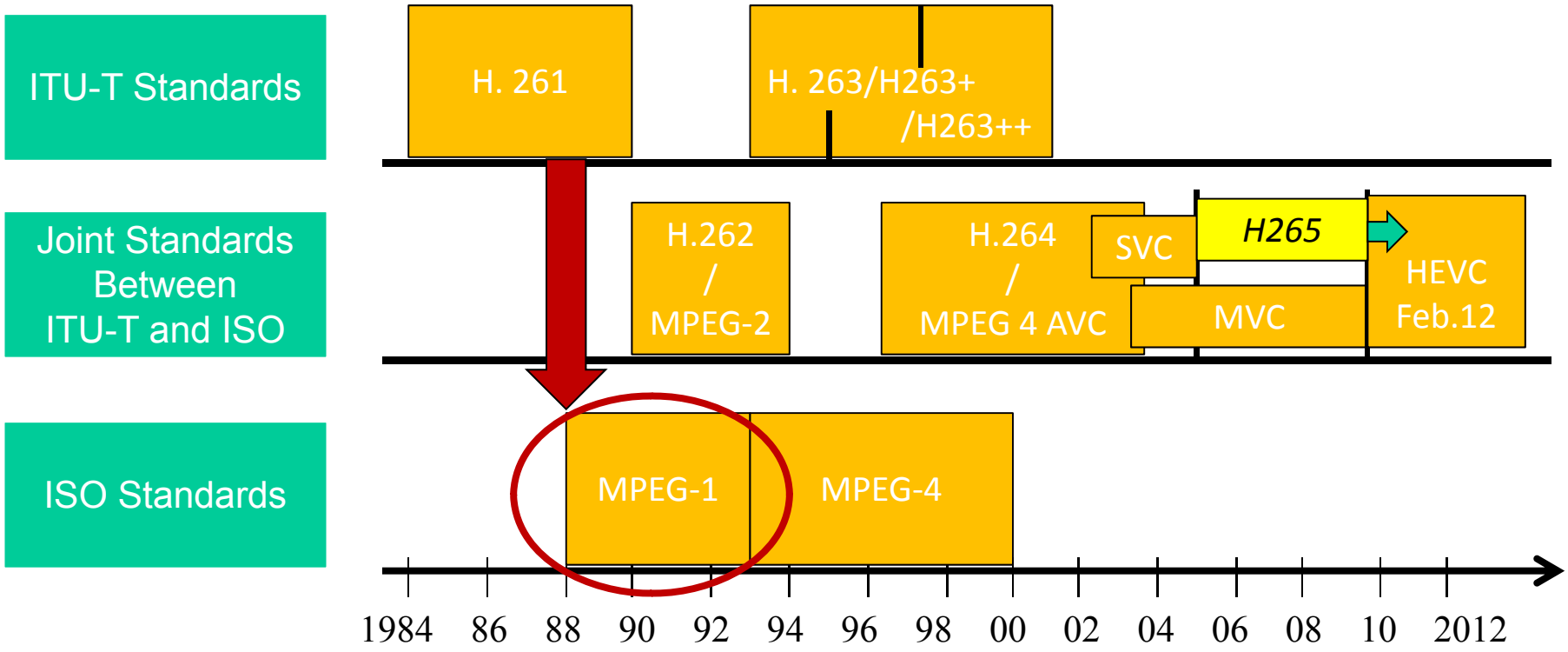
- PSC

« Picture Start Code » délimitation des images

- TR « Temporal Reference » pour synchro. avec audio
- PType « Picture Type » : image « P » ou « I »
- GOB « Group of Block » : 3x11 macroblocs
- Grp# « Group number # » : cas où des GOB sont passés
- GQuant «Group Quantization Value »
  pour quantification particulière d ’un GOB

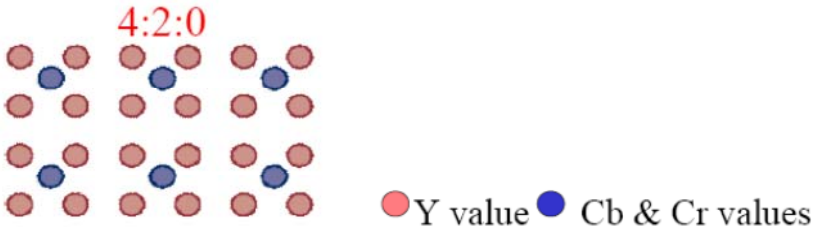
Video Standard History

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



MPEG1

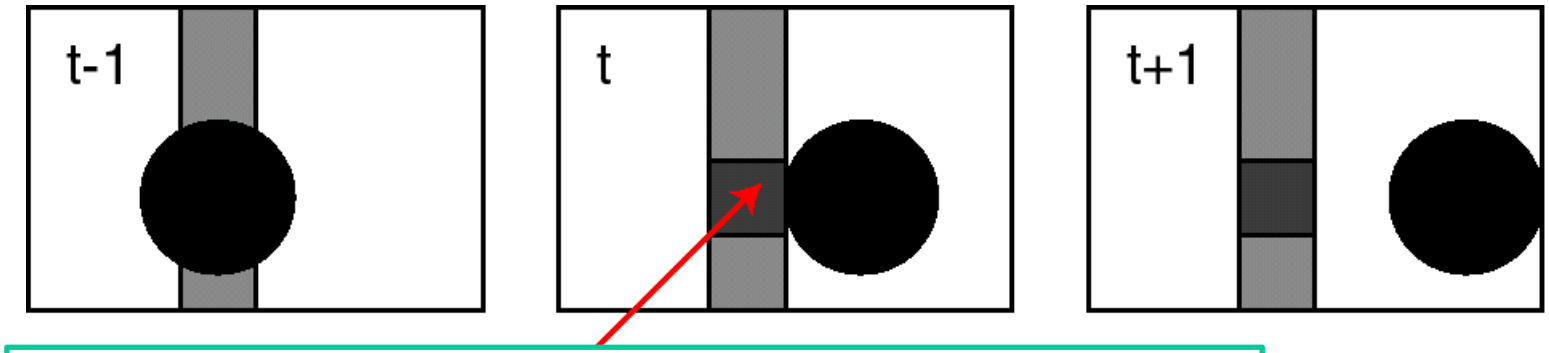
- « Moving Picture Coding Experts Group », 1988
- 3 parts: Vidéo + Audio + System
- Main Application: storage on CD ROM
  - Progressive content
  - Frame rate 25 30 fps
  - Color: 4:2:0 uniquement



- MPEG1 : 1.5 Mbits/s, VCR quality, (CIF 352x288 + CD audio)
-

MPEG1: new inter prediction mode !

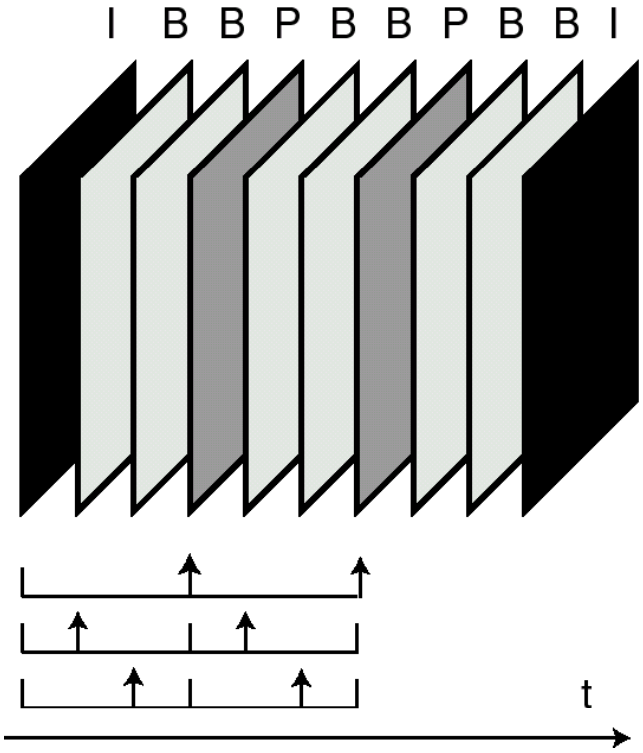
- Based on following observation : prediction from past images, as done in H261, not always efficient



➔ introducing Bidirectionnal prediction

Bi-directionnal prediction: consequences

- 3 types of images:
  - I « intra-frame »
  - P « predicted frame »
  - B « bidirectional frame »
    prediction of macroblocks
    from past and future



- Typical input order (displaying order)

  I1 B1 B2 **P1** B3 B4 **P2** B5 B6 **I2** B7 B8 **P3**

  but different coding & decoding order:

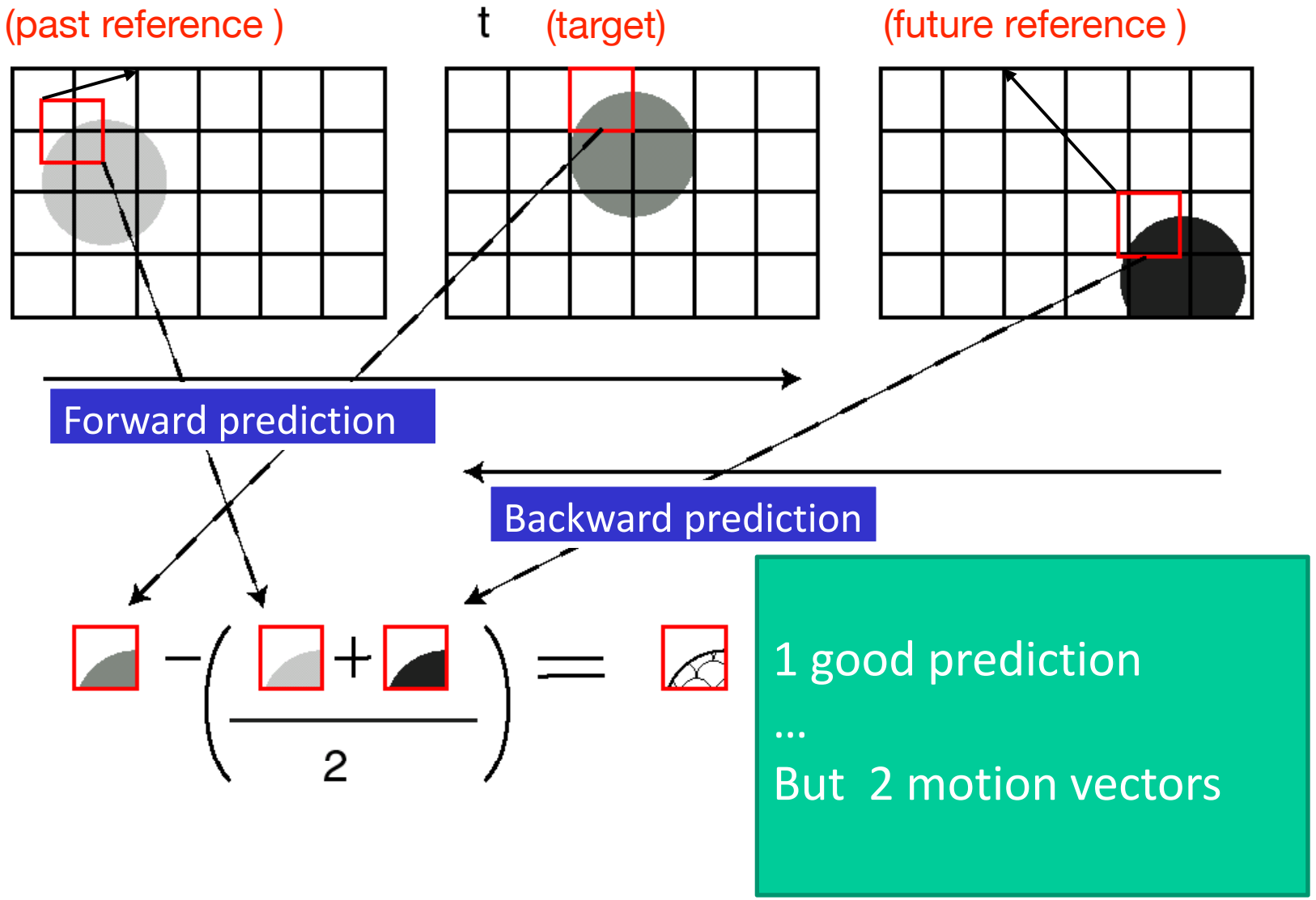
  I1 **P1** B1 B2 **P2** B3 B4 **I2** B5 B6 **P3** B7 B8

  ➔ needs to be reorganized before visualisation

GOP
« Group Of Pictures »

Bi directionnal prediction: Forward, Backward & Both !

- combination of two predictions



Motion estimation (by Block Matching)

- Larger Search Window

WHY ?

- Half pixel estimation !

HOW ?

Quantization

- Quantization step for Intra
  - Depends on the DCT coefficient frequency (Jpeg like)

- Quantization for Inter => of error prediction:
  - Same for all DCT coefficients

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	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

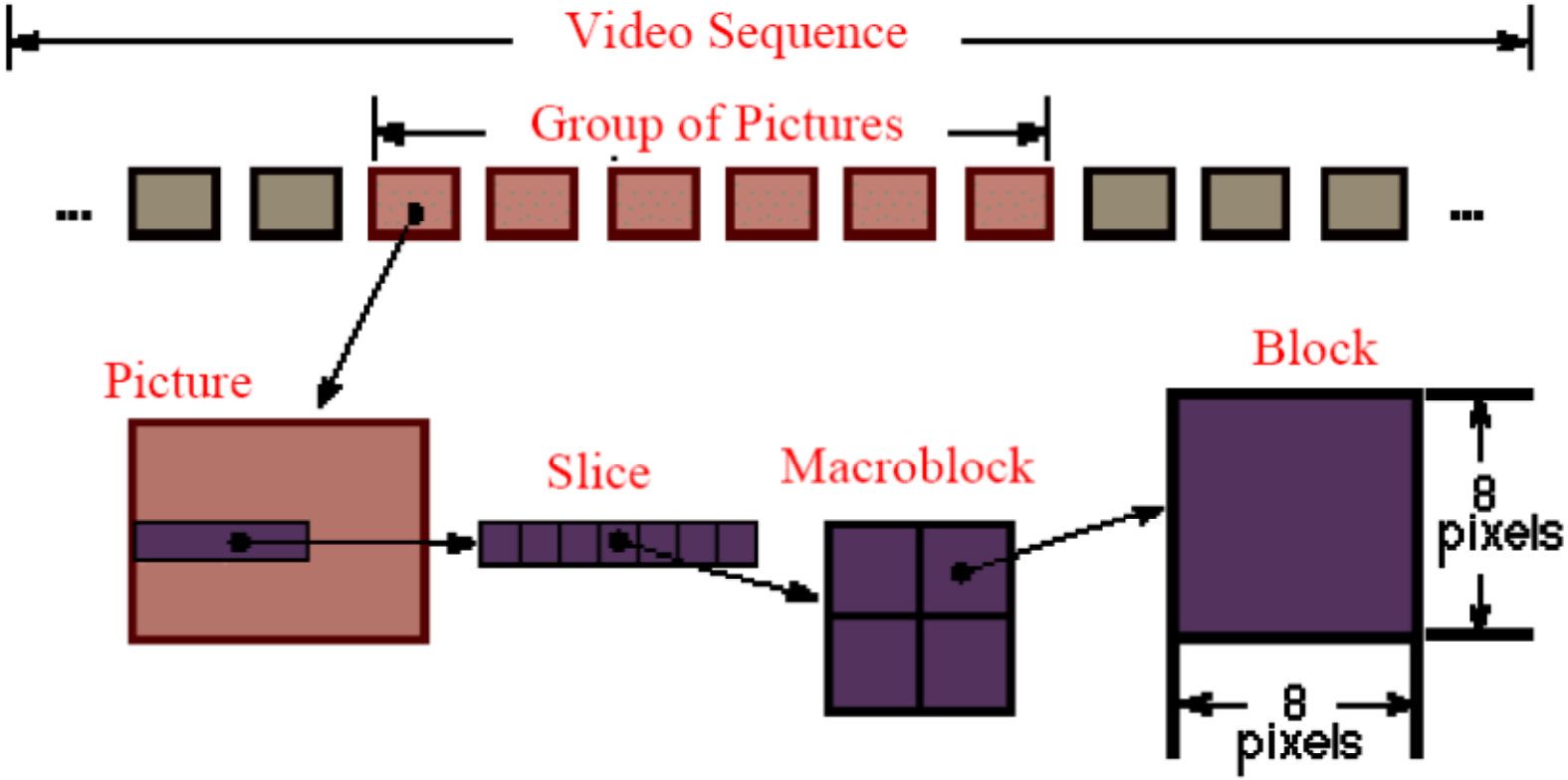
(a) quantization table for intra coding.

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

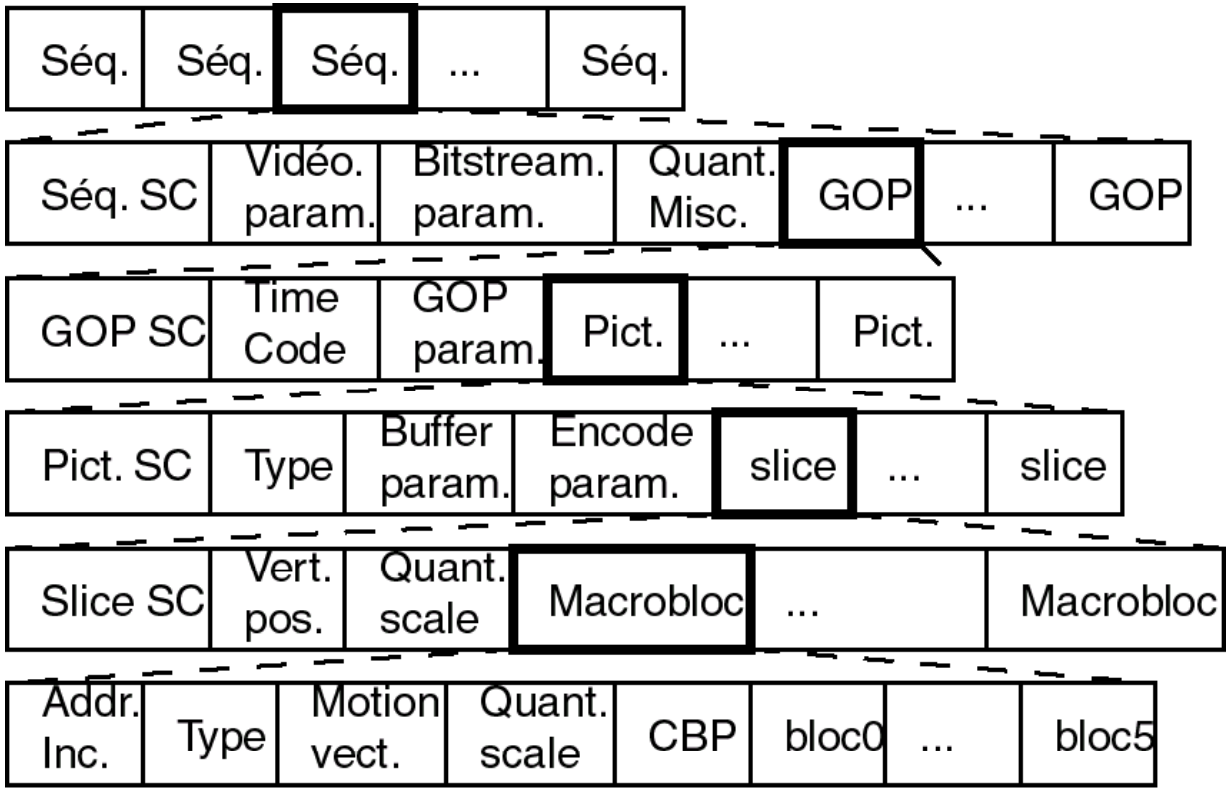
(b) quantization table for non-intra coding.

Hierarchy (bitstream)

- 6 levels

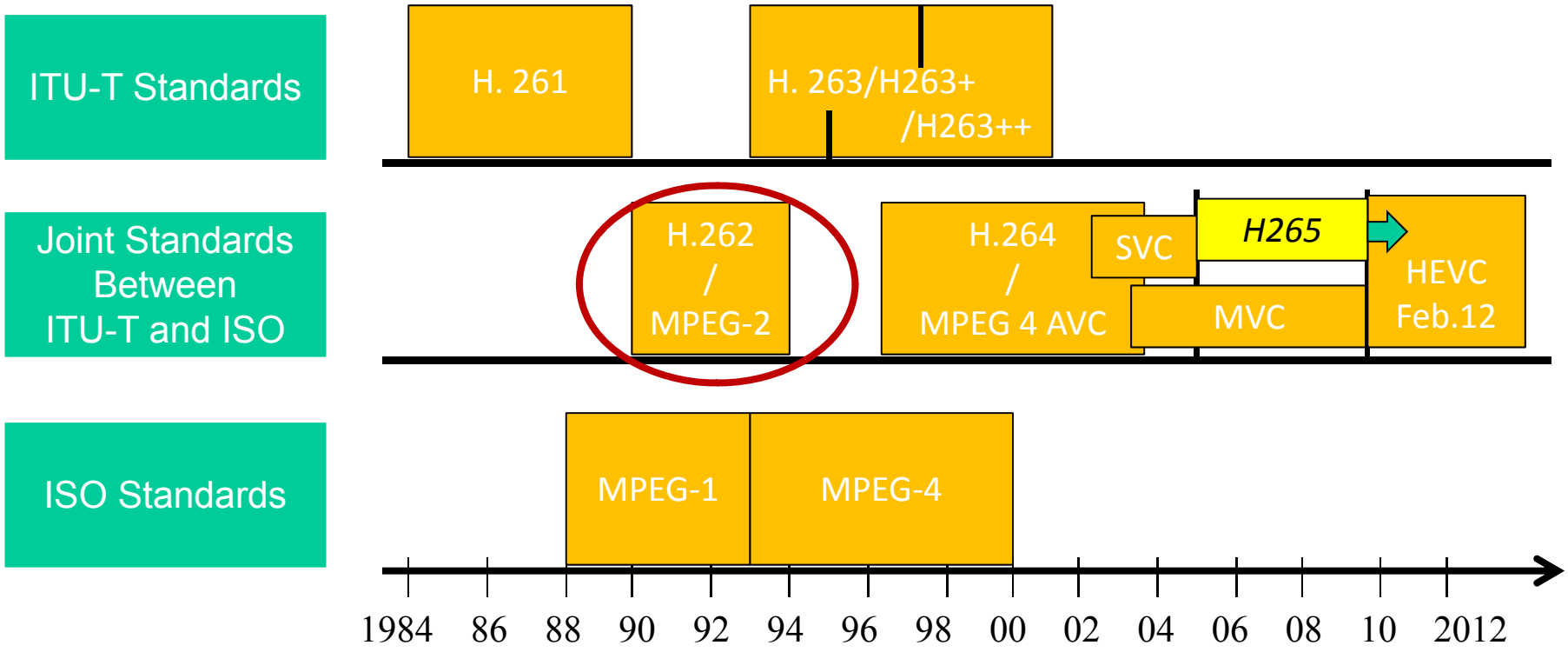


« Bitstream »


Séq.
Séq.
Séq.
...
Séq.
Séq. SC
Vidéo. param.
Bitstream. param.
Quant. Misc.
GOP
...
GOP
GOP SC
Time Code
GOP param.
Pict.
...
Pict.
Pict. SC
Type
Buffer param.
Encode param.
slice
...
slice
Slice SC
Vert. pos.
Quant. scale
Macrobloc
...
Macrobloc
Addr. Inc.
Type
Motion vect.
Quant. scale
CBP
bloc0
...
bloc5

Video Standard History

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



P. LE CALLET

MPEG 2

- MPEG 2 / H 262 (1994) ISO/IEC 13818-2
  - For TV broadcast

Kind of MPEG1 plus:

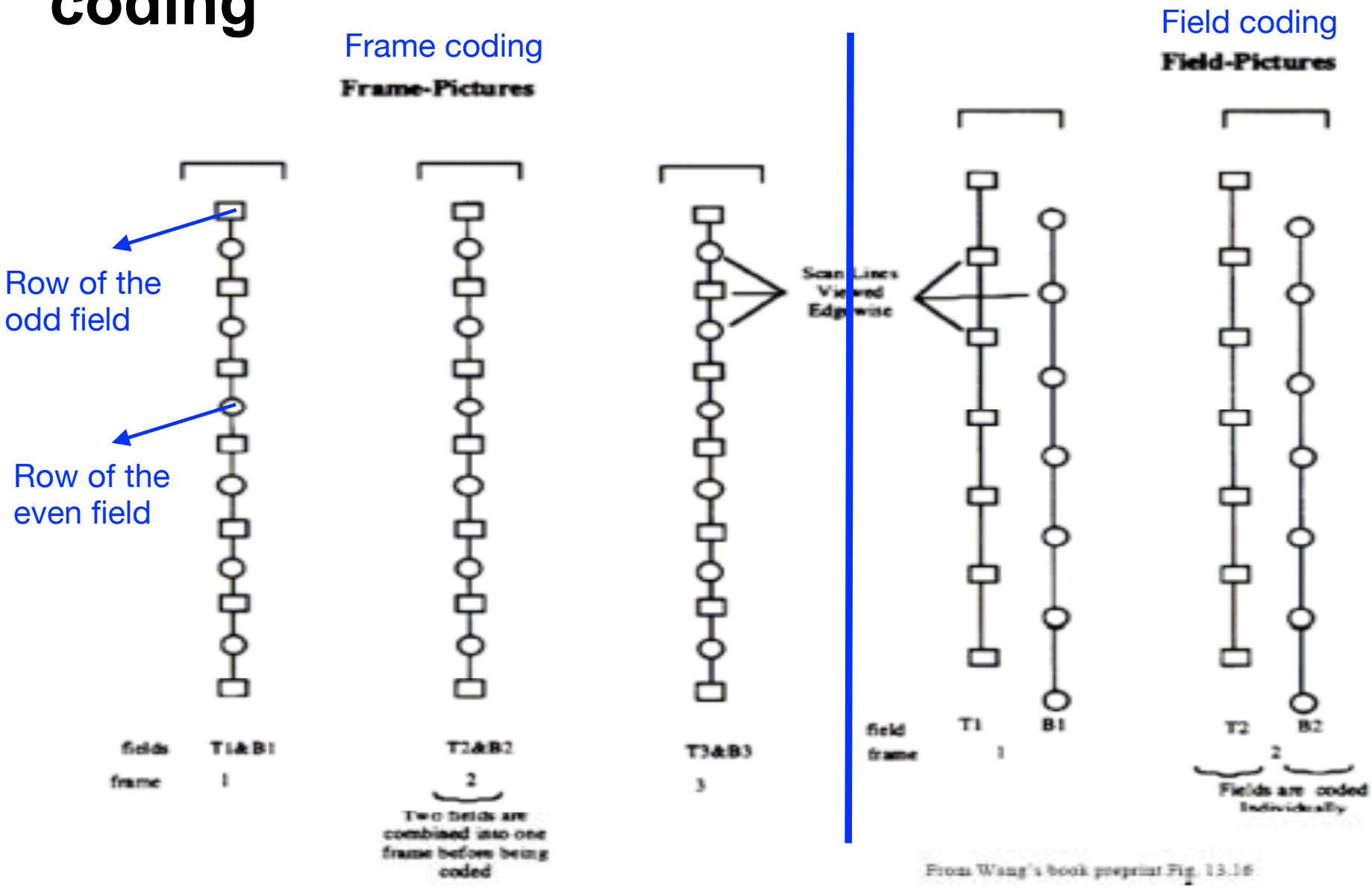
- **progressive or interleaved**
- 4:2:2 et 4:2:0, 25/30 fps
- Formats
  - SD : 720*576 et 720*480
  - HD : 1920*1080i et 1280*720

- Bit rate from 2 to 6 Mb/s (SD) & 15 to 40 MB/s (HDTV)
- Introduction concepts of « prolife » & « level »
- **Motion estimation half pixel accuracy**

Interleaved video & ME issue

Pixels of 2 fields in a frame
=> Can be close spatially but far temporally

Managing Interleaved: Field and Frame coding

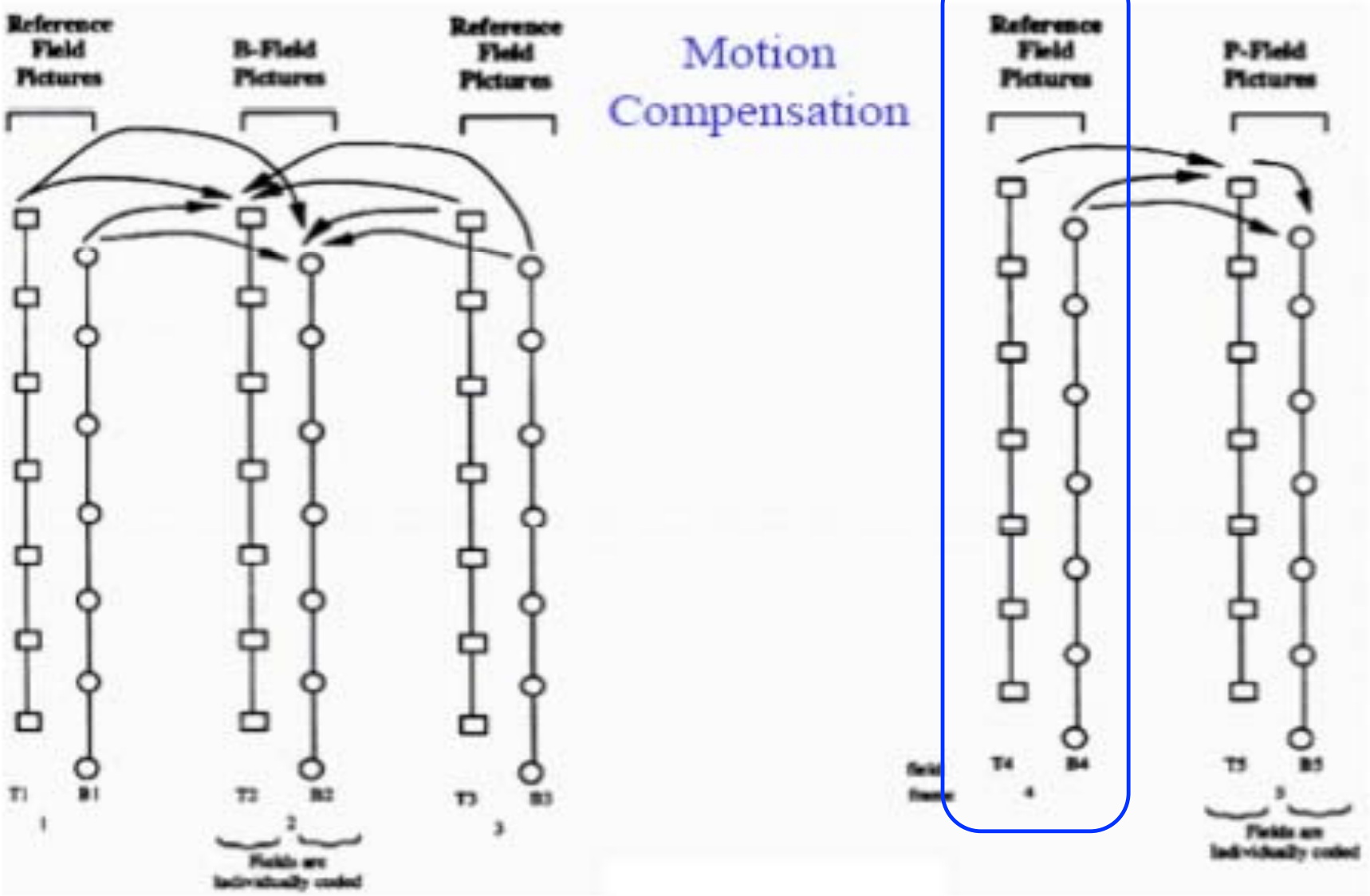


Field coding

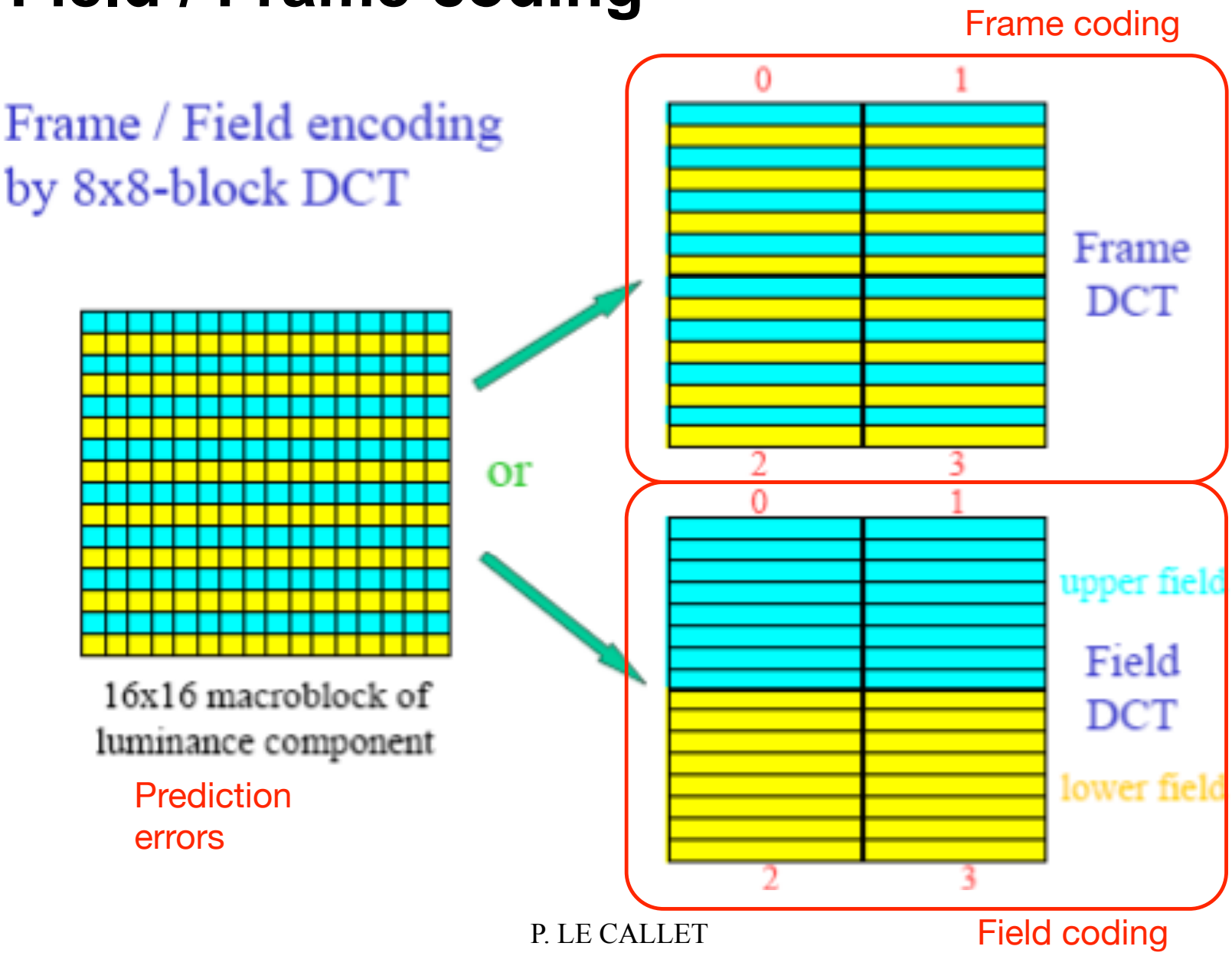
for B Prediction

for P Prediction

References



Field / Frame coding



MPEG 2 : profile et level

Profile => use case

Level => setting

Layers & Profiles	Simple Profile	Main Profile	SNR Scalability Profile	Spatial Scalability Profile	High Profile
Features	No B-frames 4:2:0 No Scalability	B-frames 4:2:0 No Scalability	B-frames 4:2:0 SNR Scalability	B-frames 4:2:0 SNR or Spatial Scalability	B-frames 4:2:0 or 4:2:2 SNR or Spatial Scalability
Low Level (352 x 288)/(352 x 240) 25/30 Fps		<= 4 Mbps	<= 4 Mbps		
Main Level (720 x 576)/(720 x 480) 25/30 Fps	<= 15 Mbps	<= 15 Mbps	<= 15 Mbps		<= 20 Mbps
High Level 720p (1280 x 720)p 24/25/30/50/60 Fps		<= 60 Mbps		<= 60 Mbps	<= 80 Mbps
High Level 1080 i/p (19200 x 1080)i/p 50/59.94//25/29.97 Fps		<= 80 Mbps			<= 100 Mbps

MPEG 2 : performances

Bitrate	8,67 % of type I		25 % of type P		66,33 % of type B	
12,2 Mb/s	9,37%	1,08	27,24%	1,09	63,39%	0,96
3,0 Mb/s	16,20%	1,87	36,78%	1,47	47,02%	0,71
1,0 Mb/s	21,38%	2,47	37,14%	1,49	41,48%	0,63
0,54 Mb/s	18,86%	2,18	26,73%	1,07	54,41%	0,82
Sequence *Stefan Edberg* ; Format CIF ; GOP of 12 frames (IBBPB….); % of total bitrate / Representativity						

Typical repartition of the bit budget

1 I

GOP 12 images IBBPBBP … **=> 3 P**

1er col : % du débit total **8 B**

2eme col :débit relatif

Video sequence & GOP structure

8,67% bitrate / I frame

8,33% bitrate / P frame

8,29% bitrate / B frame

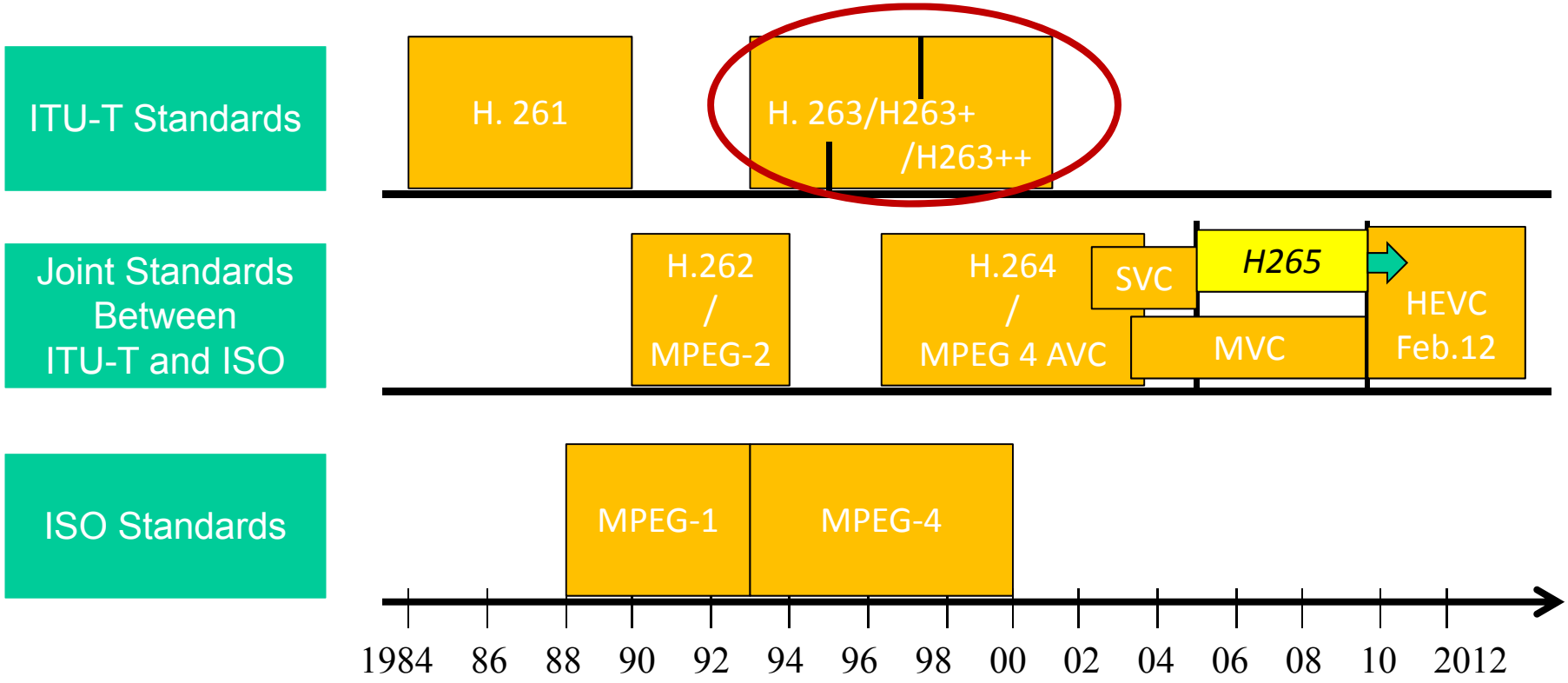
Bitrate		26 "I"	75 "P"	199 "B"	Average
12,2 Mb/s	Y	45,06 dB	48,64 dB	47,94 dB	48,02 dB
	Cb	48,64 dB	49,62 dB	49,20 dB	49,42 dB
	Cr	48,64 dB	49,69 dB	49,22 dB	49,47 dB
3,0 Mb/s	Y	35,88 dB	34,75 dB	36,71 dB	35,26 dB
	Cb	40,88 dB	38,85 dB	40,03 dB	39,27 dB
	Cr	40,78 dB	38,78 dB	39,97 dB	39,20 dB
1,0 Mb/s	Y	29,43 dB	28,08 dB	28,99 dB	28,40 dB
	Cb	36,79 dB	35,25 dB	35,36 dB	35,39 dB
	Cr	36,40 dB	34,74 dB	34,89 dB	34,90 dB
0,54 Mb/s	Y	25,79 dB	25,84 dB	25,55 dB	25,76 dB
	Cb	34,41 dB	33,68 dB	33,43 dB	33,68 dB
	Cr	33,90 dB	33,12 dB	32,88 dB	33,12 dB
Sequence *Stefan Edberg* ; 300 frames ; GOP of 12 frames lenght (IBBPB….)					
Average PSNR for different bitrates					

Si codage intra uniquement

Bitrate		300 "I"
8,0 Mb/s	Y	45,57 dB
	Cb	44,29 dB
	Cr	44,29 dB
3,0 Mb/s	Y	32,05 dB
	Cb	38,16 dB
	Cr	37,93 dB
1,2 Mb/s	Y	25,81 dB
	Cb	34,44 dB
	Cr	33,93 dB
Sequence *Stefan Edberg* ; 300 frames		
GOP of I frames only		
Average PSNR for different bitrates		

Video Standard History

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



P. LE CALLET

H263

- 1996
  - Mainly for streaming, visio,videophony
  - CIF,QCIF, SQCIF, 10 fps

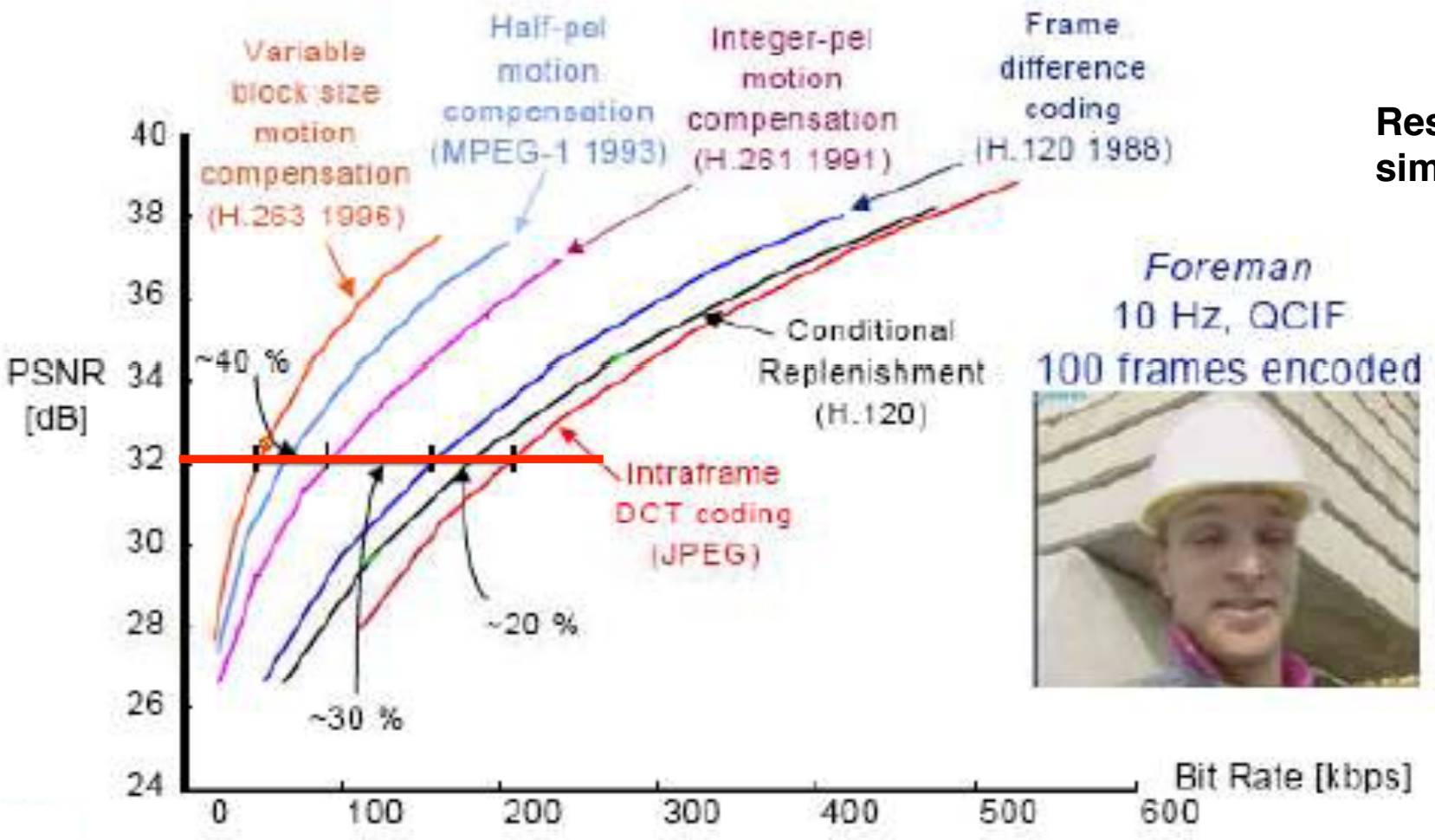
- Kind of H262 plus
  - Advanced modes for motion prediction:

  OBMC : overlapped block motion compensation

  1 to 4 motion per macroblocs !

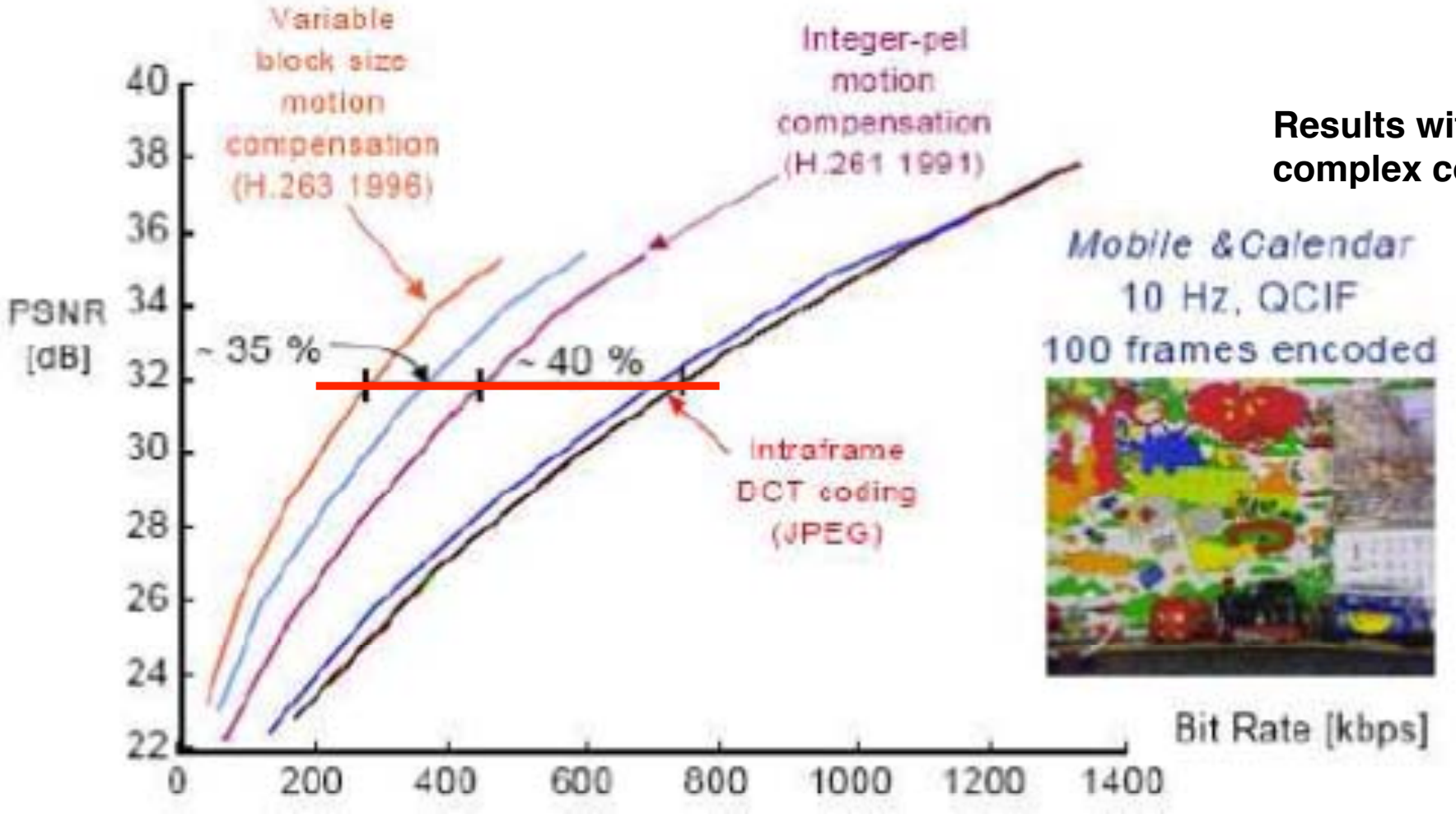
  - VLC (Varable length coding) using new Arithmetic coding (save 5% in bit rate)

Performances



Results with a simple content

Performances



Results with a complex content