Chapter 5. Image/Video Compression Standards

- JPEG
- H.261
- MPEG-1
- MPEG-2
- H.263 & H.263+
- MPEG-4
- MPEG-7
- MPEG-4 AVC/H.264

JPEG

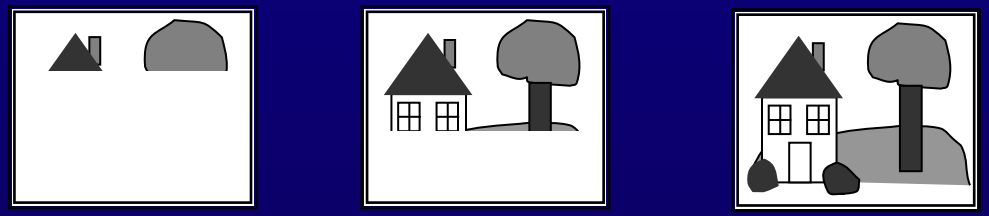
JPEG : Background

- JPEG = Joint Photographic Expert Group, Joint standards committee of ITU-T and ISO
- Flexible standard for monochrome and color image compression
- Intraframe coding scheme, optimized for still images
- Flexible picture size
- Coding of color components separately, arbitrary color spaces possible, best compression for Y/R-Y/B-Y
- Variable compression ratio
- Compression 24:1 for ITU-R 601 images without loss of quality visually

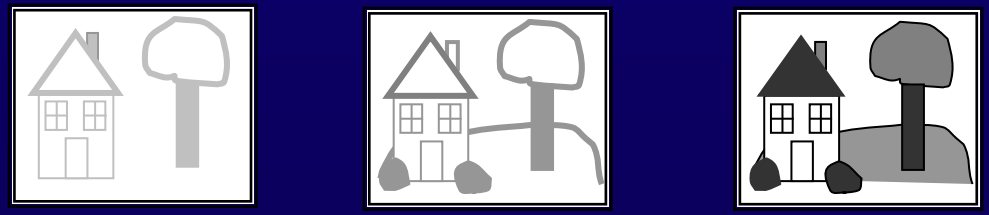
JPEG

Mode of operations:

- Sequential DCT
- Progressive DCT
- Sequential Lossless
- Hierarchical



Sequential coding

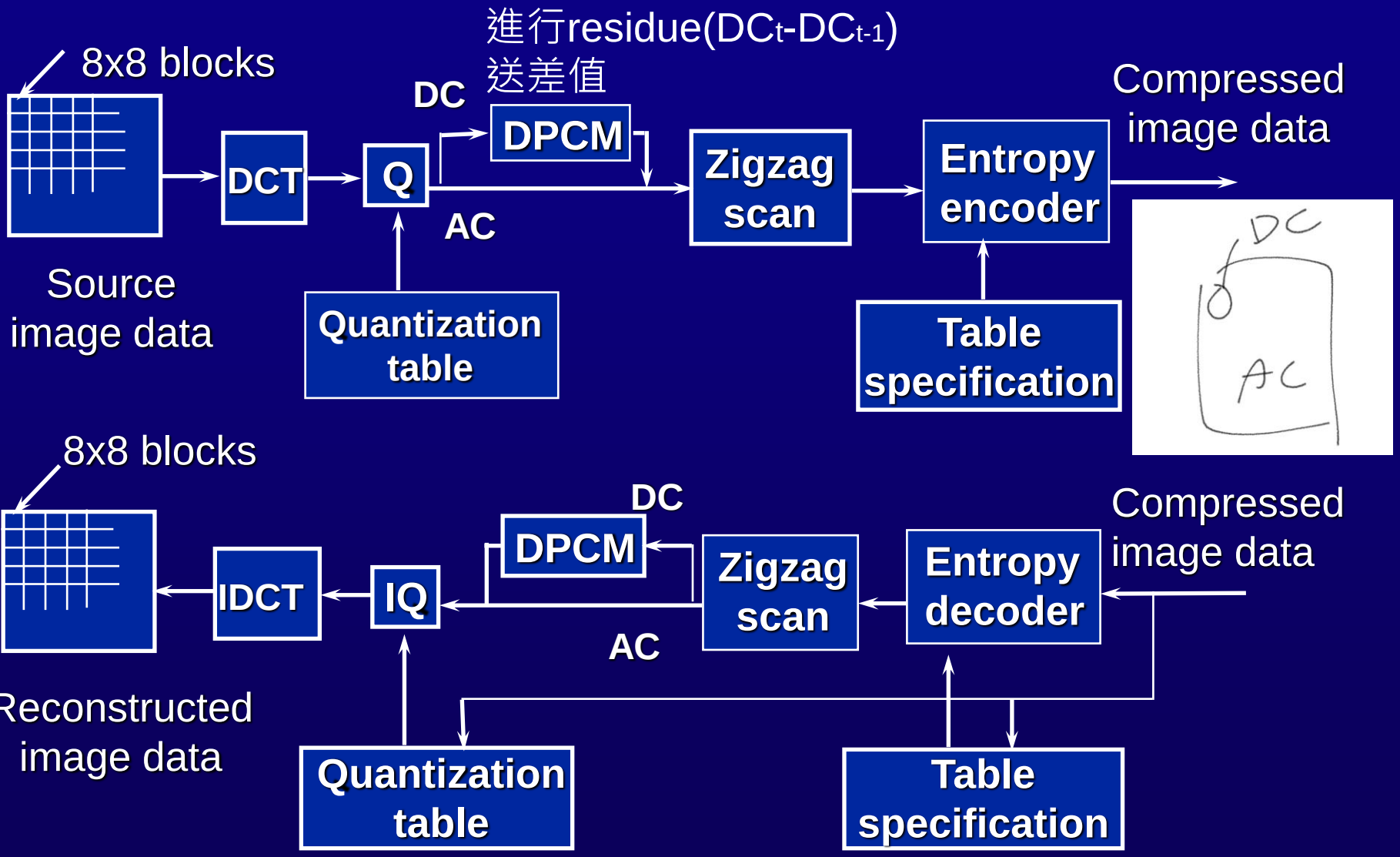


Progressive coding

JPEG : Goal

- Allow applications to tradeoff easily between desired compression ratio and image quality
- Work independently of image types
- Have modest computational complexity that would allow software-only implementation even in low-end computers
- Allow both sequential (single scan) and progressive (multiple scan) coding
- Offers the option for hierarchical coding
- Compress video frame by frame (MJPEG)

Sequential DCT-based Systems



Sequential DCT-based Systems

- Baseline System: minimum capability that must be present in all JPEG systems (source sample precision limited to 8 bits; only Huffman coding for entropy coding; up to 2 sets of entropy coding tables)

- Extended Sequential Systems: capabilities beyond the baseline requirement (source sample precision of 8 or 12 bits; can use Arithmetic coding; up to 4 sets of entropy coding tables)

An Example of JPEG Baseline System

- Original:

52	55	61	66	70	61	64	73
63	59	66	90	109	85	69	72
62	59	68	113	144	104	66	73
63	58	71	122	154	106	70	69
67	61	68	104	126	88	68	70
79	65	60	70	77	68	58	75
85	71	64	59	55	61	65	83
87	79	69	68	65	76	78	94

- After level shift by -128:

- 76	- 73	- 67	- 62	- 58	- 67	- 64	- 55
- 65	- 69	- 62	- 38	- 19	- 43	- 59	- 56
- 66	- 69	- 60	- 15	16	- 24	- 62	- 55
- 65	- 70	- 57	- 6	26	- 22	- 58	- 59
- 61	- 67	- 60	- 24	- 2	- 40	- 60	- 58
- 49	- 63	- 68	- 58	- 51	- 65	- 70	- 53
- 43	- 57	- 64	- 69	- 73	- 67	- 63	- 45
- 41	- 49	- 59	- 60	- 63	- 52	- 50	- 34

An Example of JPEG Baseline System

- After DCT:

- 415	- 29	- 62	25	55	- 20	- 1	3
7	- 21	- 62	9	11	- 7	- 6	6
- 46	8	77	- 25	- 30	10	7	- 5
- 50	13	35	- 15	- 9	6	0	3
11	- 8	- 13	- 2	- 1	1	- 4	1
- 10	1	3	- 3	- 1	0	2	- 1
- 4	- 1	2	- 1	2	- 3	1	- 2
- 1	- 1	- 1	- 2	- 1	- 1	0	- 1

- After quantization:

- 26	- 3	- 6	2	2	0	0	0
1	- 2	- 4	0	0	0	0	0
- 3	1	5	- 1	- 1	0	0	0
- 4	1	2	- 1	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

Default Quantization Tables

- Luminance Table:

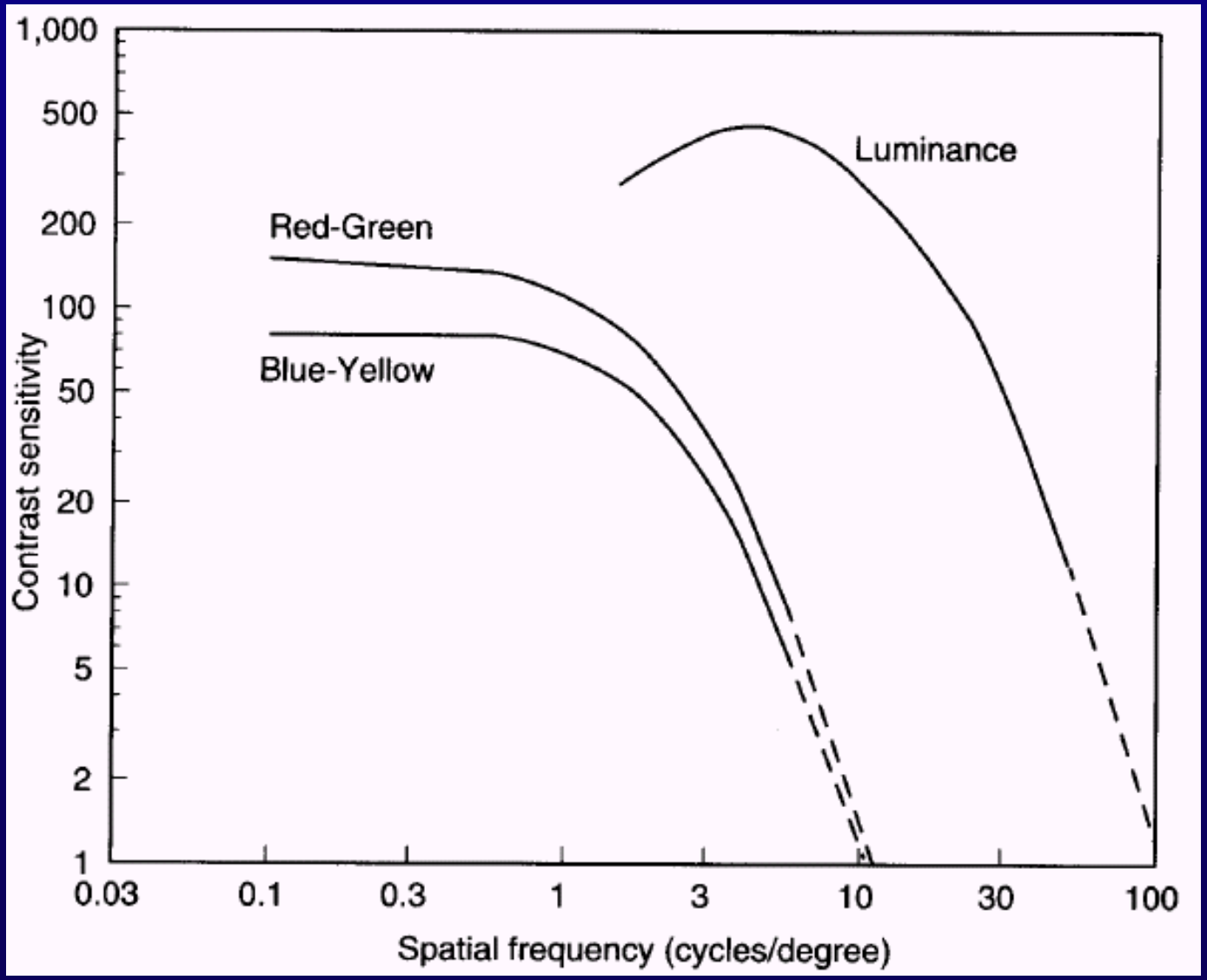
數字為quantization step size，越左上角越重要，也就step要越小(細緻，保留資訊)

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

- Chrominance Table:

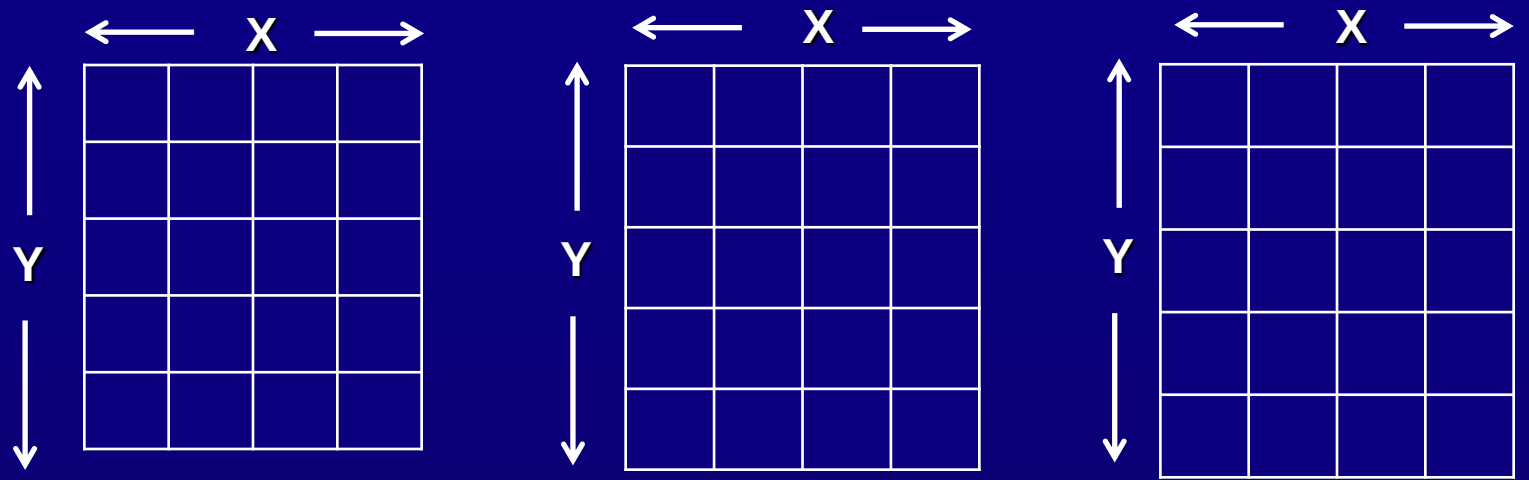
17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

Eye Sensitivity vs. Spatial Frequency

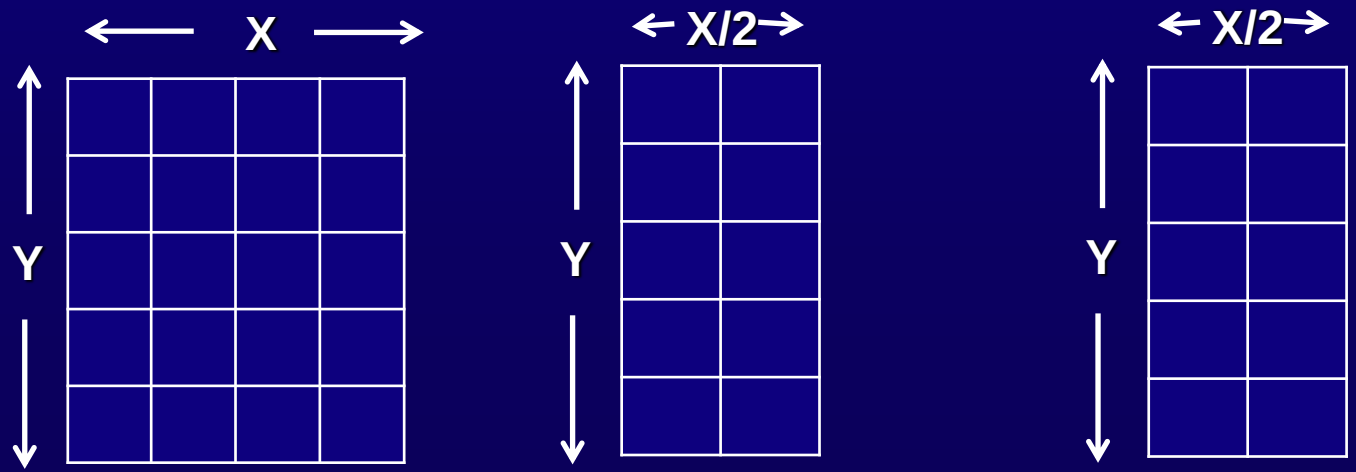


Y帶有訊息較多，CbCr相對少

Resolution of Color Images



Same resolution

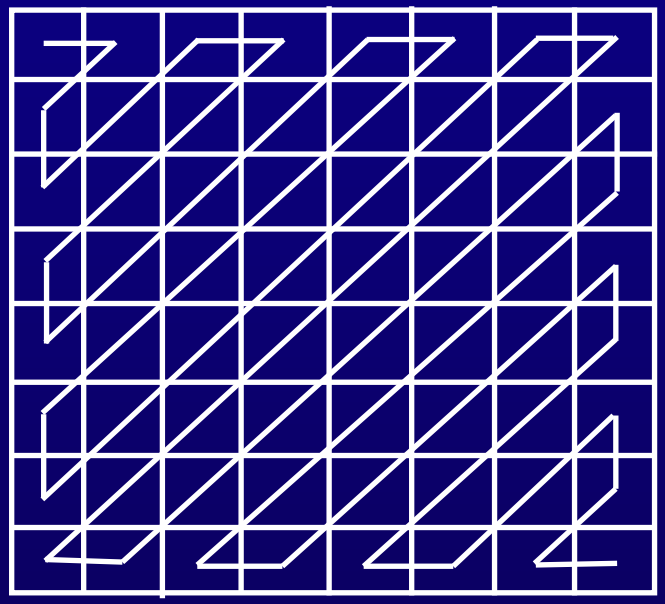


Different resolution

Quantization of DCT Coefficients

- (a),(c),(e) use the default quantization matrix
- (b),(d),(f) use 4 times the default quantization matrix

Zigzag scan



An Example of JPEG Encoding

- After zigzag scan:

[- 26 - 3 1 - 3 - 2 - 6 2 - 4 1 - 4 1 1 5
0 2 0 0 - 1 2 0 0 0 0 0 -1 -1 EOB]

- After DPCM of DC coefficient (assuming previous DC value of -17):

[- 9 - 3 1 - 3 - 2 - 6 2 - 4 1 - 4 1 1 5
0 2 0 0 - 1 2 0 0 0 0 0 -1 -1 EOB]

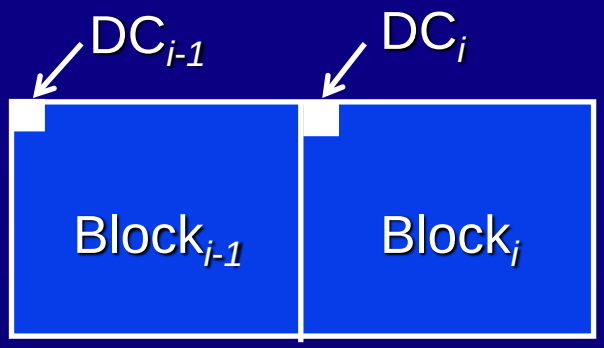
- After RLC:

-9 (0,-3) (0,1) (0,-3) (0,-2) (0,-6) (0,2) (0,-4) (0,1) (0,-4) (0,1)
(0,1) (0,5) (1,2) (2,-1) (0,2) (5,-1) (0,-1) EOB

- After VLC:

1010110 0100 001 0100 0101 100001 0110 100011
001 100011 001 001 100101 11100110 110110
0110 11110100 000 1010

Baseline DC Coefficient Coding



$DIFF=DC_i-DC_{i-1}$

- The difference DIFF is represented as a CAT.Value symbol

CAT: 0-11 (see the table on the next slide)
CAT=0: EOB (end of block)

- Value is a fixed-length code with length indicated by the Category:

If Value > 0: positive binary number representation,
if Value < 0: one's complement

Example:
DIFF=6: CAT=3 (100), Value=110 $\Rightarrow$ 100110
DIFF=-3: CAT=2 (011), Value=00 $\Rightarrow$ 01100

Baseline Coefficient Coding

DC Category (SSSS)	AC Category (SSSS)	Coefficient
0	N/A	0
1	1	-1,1
2	2	-3,-2,2,3
3	3	-7,…,-4,4,…,7
4	4	-15,…,-8,8,…,15
5	5	-31,…,-16,16,…,31
6	6	-63,…,-32,32,…,63
7	7	-127,…,-64,64,…,127
8	8	-255,…,-128,128,…,255
9	9	-511,…,-256,256,…,511
10	10	-1023,…,-512,512,…,1023
11	11	-2047,…,-1024,1024,…,2047
12	12	-4095,…,-2048,2048,…,4095
13	13	-8191,…,-4096,4096,…,8191
14	14	-16383,…,-8192,8192,…,16383
15	N/A	-32767,…,-16384,16384,…,32767

Default DC Codes (Luminance)

Category	Huffman Code
0	00
1	010
2	011
3	100
4	101
5	110
6	1110
7	11110
8	111110
9	1111110
10	11111110
11	111111110
…	...

Baseline AC Coefficient Coding

- The zigzag sequence of DCT coefficients is represented as a sequence of (RRRRSSSS).Value symbols

- RRRR: Zero-runlength
  SSSS: Category
  The maximum length of zero run is limited to 15
  11110000: zero-runlength=16 (15 zeros followed by a coefficient of zero amplitude)
  00000000:EOB (end of block)
  RRRRSSSS is Huffman coded

- Value is a fixed-length code with length indicated by the Category:
  If Value > 0: positive binary number representation,
  if Value < 0: one's complement

An Example of JPEG Baseline System

- Decoder input:

  1010110 0100 001 0100 0101 100001 0110 100011
  001 100011 001 001 100101 11100110 110110
  0110 11110100 000 1010

- After VLD:

  -9 (0,-3) (0,1) (0,-3) (0,-2) (0,-6) (0,2) (0,-4) (0,1) (0,-4) (0,1)
  (0,1) (0,5) (1,2) (2,-1) (0,2) (5,-1) (0,-1) EOB

- After RLD, Inverse DPCM, and Inverse zigzag scan:

- 26	- 3	- 6	2	2	0	0	0
1	- 2	- 4	0	0	0	0	0
- 3	1	5	- 1	- 1	0	0	0
- 4	1	2	- 1	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

An Example of JPEG Baseline System

- After inverse quantization:

- 416	- 33	- 60	32	48	0	0	0
12	- 24	- 56	0	0	0	0	0
- 42	13	80	- 24	- 40	0	0	0
- 56	17	44	- 29	0	0	0	0
18	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

- After IDCT:

- 70	- 64	- 61	- 64	- 69	- 66	- 58	- 50
- 72	- 73	- 61	- 39	- 30	- 40	- 54	- 59
- 68	- 78	- 58	- 9	13	- 12	- 48	- 64
- 59	- 77	- 57	0	22	- 13	- 51	- 60
- 54	- 75	- 64	- 23	- 13	- 44	- 63	- 56
- 52	- 71	- 72	- 54	- 54	- 71	- 71	- 54
- 45	- 59	- 70	- 68	- 67	- 67	- 61	- 50
- 35	- 47	- 61	- 66	- 60	- 48	- 44	- 44

An Example of JPEG Baseline System

- After level shift by +128:

58	64	67	64	59	62	70	78
56	55	67	89	98	88	74	69
60	50	70	119	141	116	80	64
69	51	71	128	149	115	77	68
74	53	64	105	115	84	65	72
76	57	56	74	75	57	57	74
83	69	59	60	61	61	67	78
93	81	67	62	69	80	84	84

- Error:

- 6	- 9	- 6	2	11	- 1	- 6	- 5
7	4	- 1	1	11	- 3	- 5	3
2	9	- 2	- 6	- 3	- 12	- 14	9
- 6	7	0	- 4	- 5	- 9	- 7	1
- 7	8	4	- 1	11	4	3	- 2
3	8	4	- 4	2	11	1	1
2	2	5	- 1	- 6	0	- 2	5
- 6	- 2	2	6	- 4	- 4	- 6	10

Reconstruction in Sequential Mode

(a) 25% received

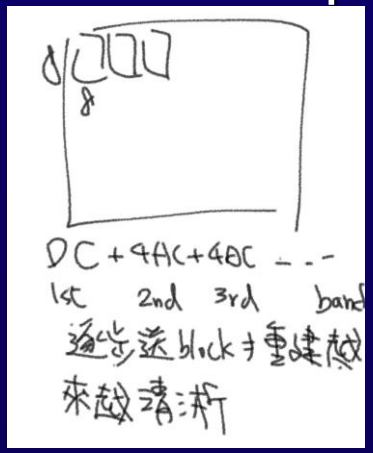
(b) 50% received

(c) 75% received

(d) 100% received

Progressive DCT Mode

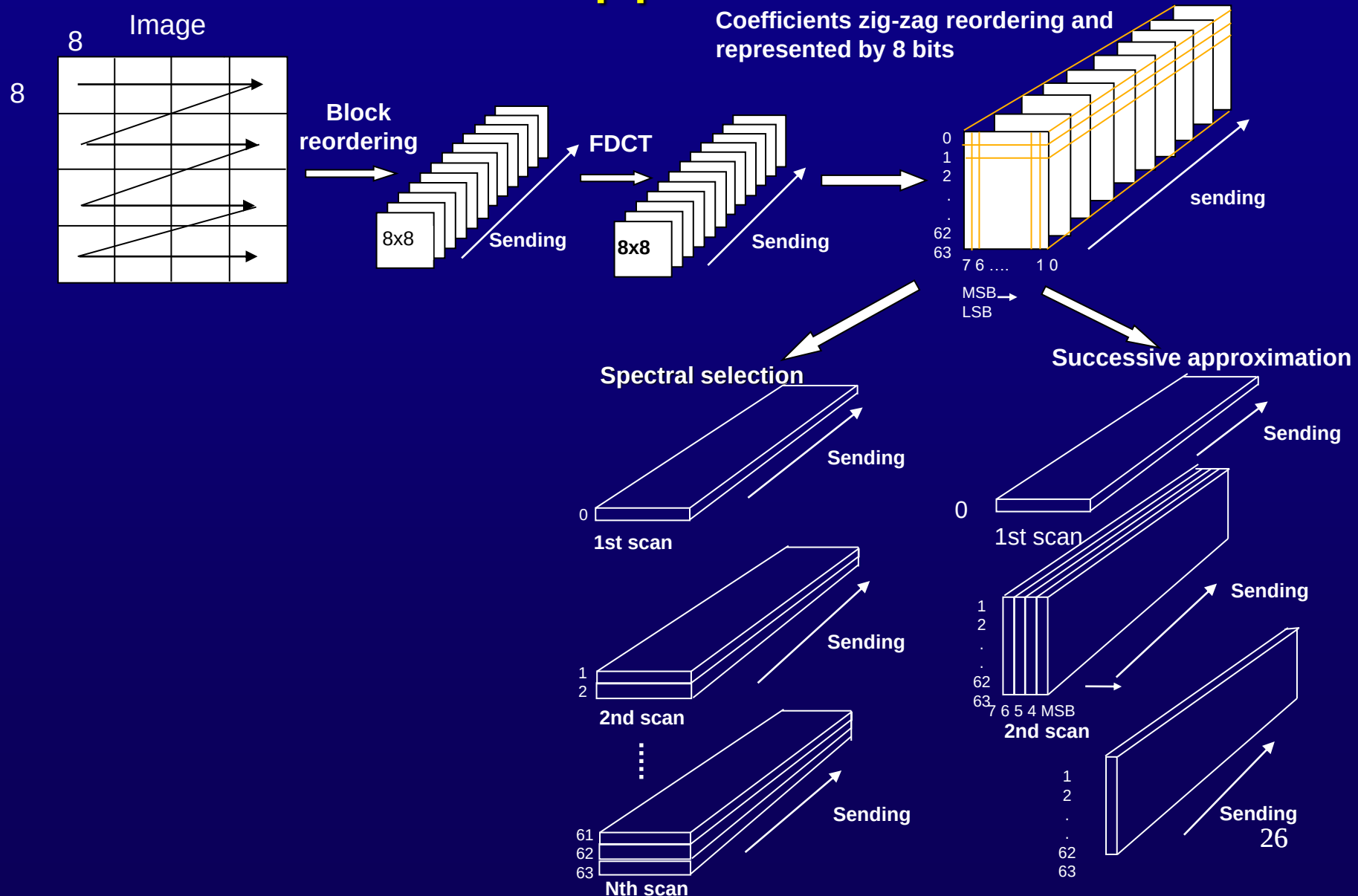
- Spectral selection: DCT coefficients are grouped into "spectral" bands of related spatial frequencies and the lower-frequency bands are (usually) sent first
- Successive approximation: the information is first sent with lower precision and then refined in later scans
- Require a coefficient memory buffer between quantizer and the entropy encoder
- Successive approximation gives better quality at low bit-rates



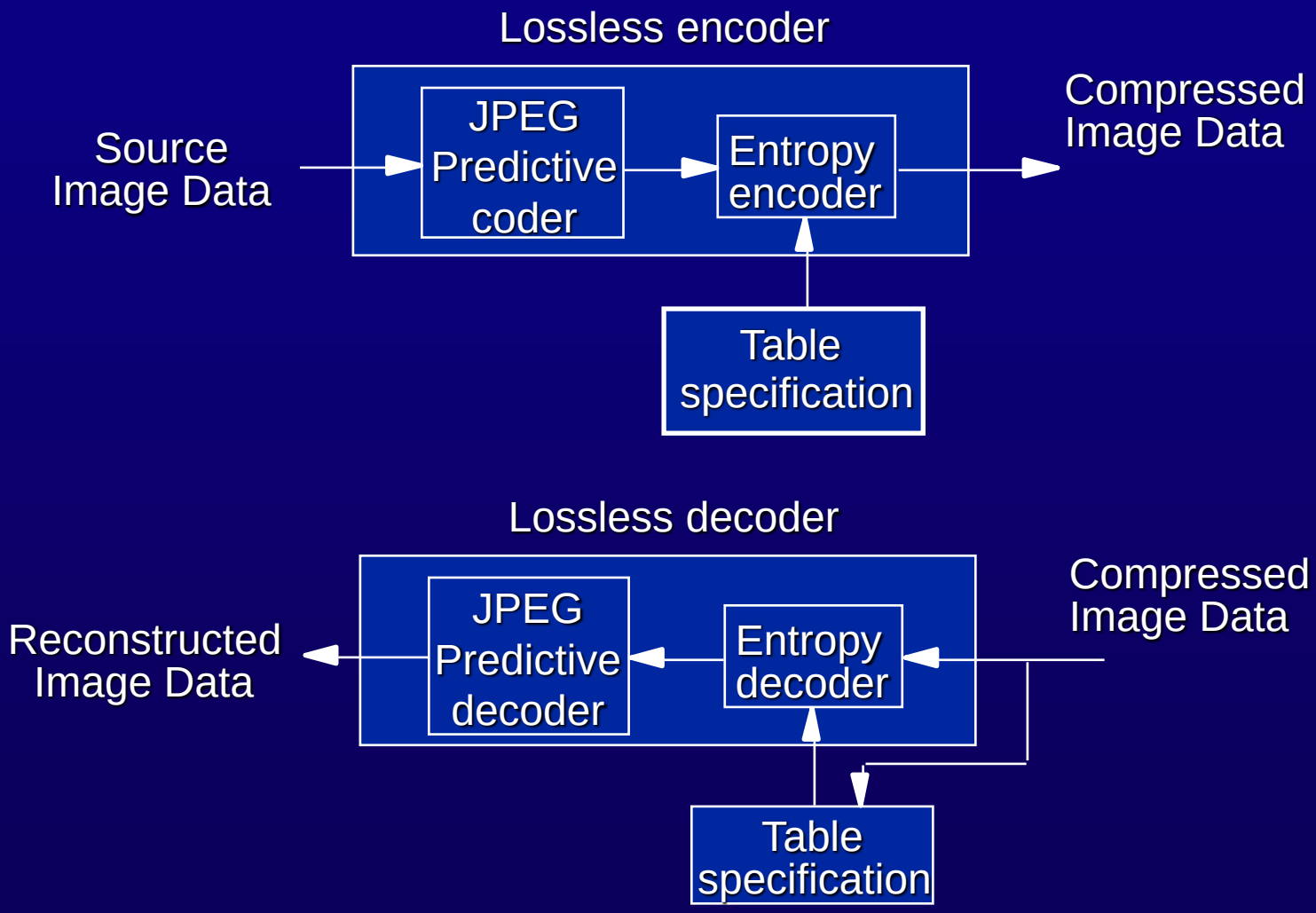
Reconstruction in Progressive Mode



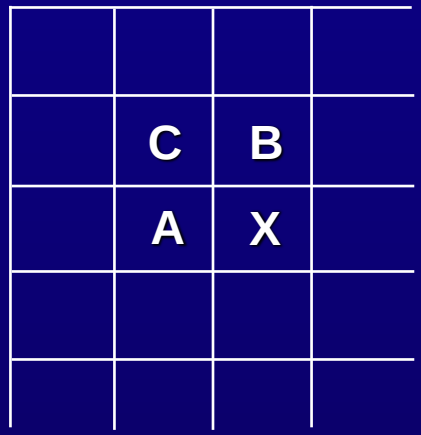

Progressive Coding with Spectral Selection and Successive Approximation



Sequential Lossless Systems



Predictors for Lossless Coding



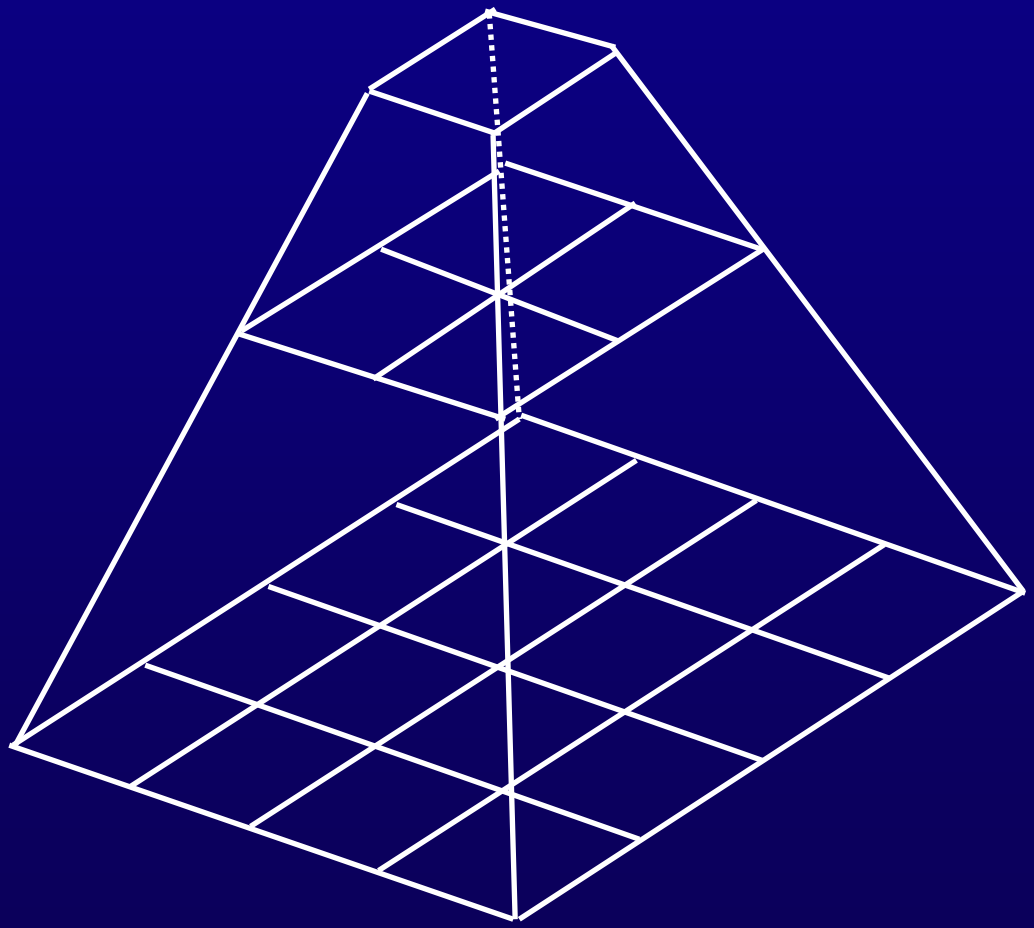
Selection-value	Prediction
0	no prediction
1	A
2	B
3	C
4	A + B - C
5	A + (B - C) / 2
6	B + (A - C) / 2
7	(A + B) / 2

*** JPEG lossless coding typically produce ~2:1 compression**

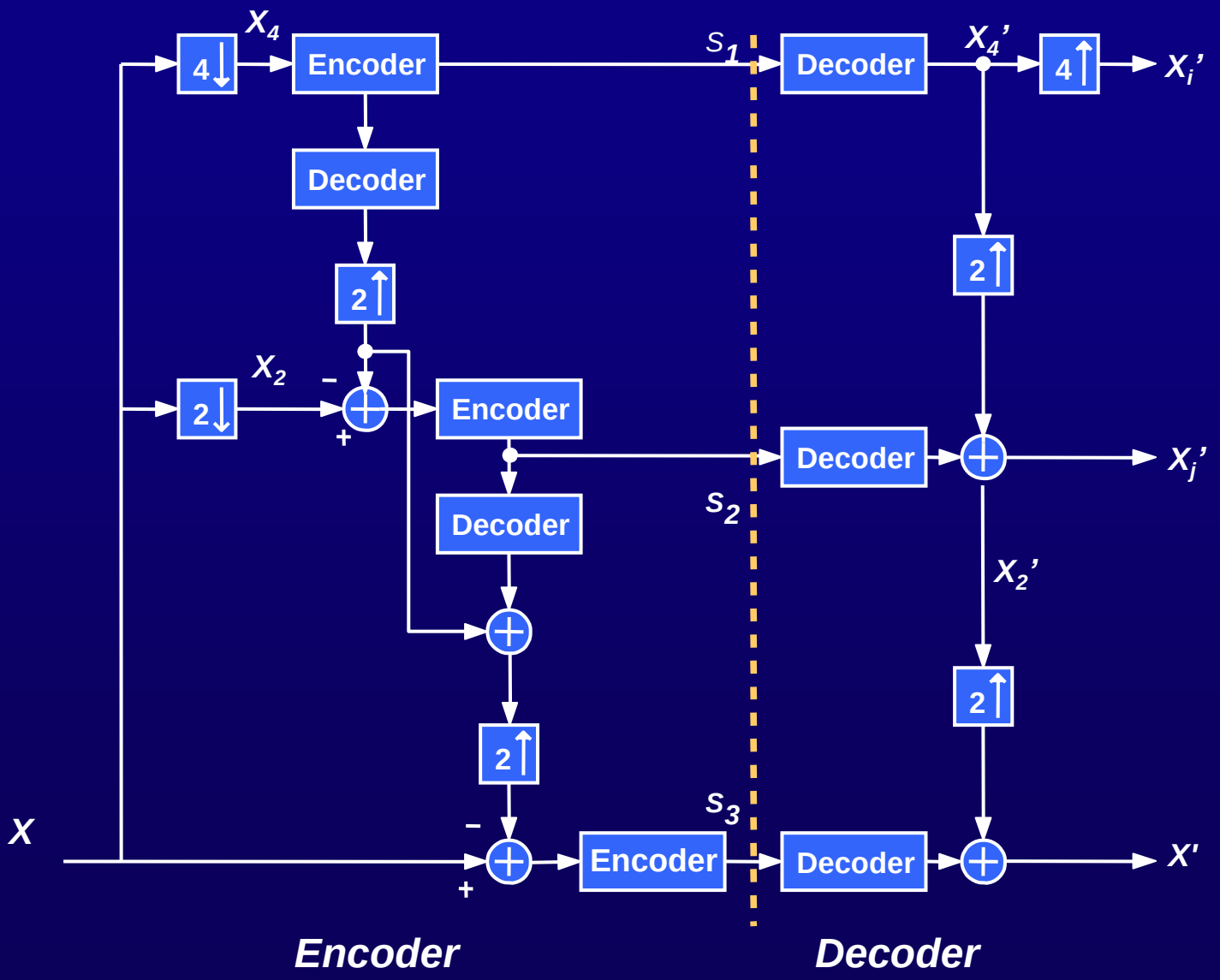
Hierarchical Mode

- The first stage (lowest resolution) is coded using one of the sequential or progressive JPEG modes
- The output of each hierarchical stage is then upsampled and becomes the prediction for the next stage

Hierarchical Multiresolution Encoding



A Three-level Hierarchical Coder



JPEG Performance

Bits/pixel	Quality	Compression Ratio
>= 2	Indistinguishable	8-to-1
1.5	Excellent	10.7-to-1
0.75	Very Good	21.4-to-1
0.5	Good	32-to-1
0.25	Fair	64-to-1

Example of Artifacts due to JPEG Coding

Original

JPEG Encoded

Subjective Quality of JPEG

Original Lena Image
File Size = 256K bytes

Compressed Lena Image
File Size = 13K bytes

H.261

H.261 Video Coding Standard

- Rec. H.261, "Video Codec for Audiovisual Services at *p*x64 kb/s," (*p*=1,2,...,30) completed and approved in December 1990
- Intended applications: videotelephony (*p* = 1 or 2) and videoconferencing ($p \geq 6$)
- Capable of real-time operation with minimum delay
- Offer a factor of about three in compression compared to JPEG by using interframe coding to remove the temporal redundancy

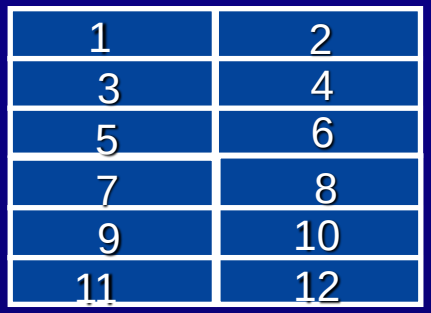
H.261 Video Source Formats

Format	CIF		QCIF	
Signal Component	Lines/ Frame	Pixels/ Line	Lines/ Frame	Pixels/ Line
Luminance (*Y*)	288	360 (352)	144	180 (176)
Chrominance (C_b)	144	180 (176)	72	90 (88)
Chrominance (C_r)	144	180 (176)	72	90 (88)

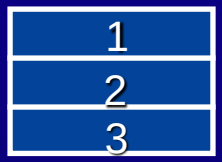
* 4:2:0 *Y* C_b C_r
* The numbers in the parentheses represent active pixels
- CIF (Common Intermediate Format) is optional and QCIF (Quarter CIF) is mandatory
- Uncompressed bit-rate for transmitting CIF and QCIF at 30 fps are 36.5 Mb/s and 9.12 Mb/s, respectively

Arrangement of Blocks in A Picture

GOB (Group Of Blocks) in a picture



CIF

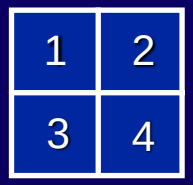


QCIF

Macroblocks in a GOB

1	2	3	4	5	6	7	8	9	10	11
12	13	14	15	16	17	18	19	20	21	22
23	24	25	26	27	28	29	30	31	32	33

Blocks in a macroblock

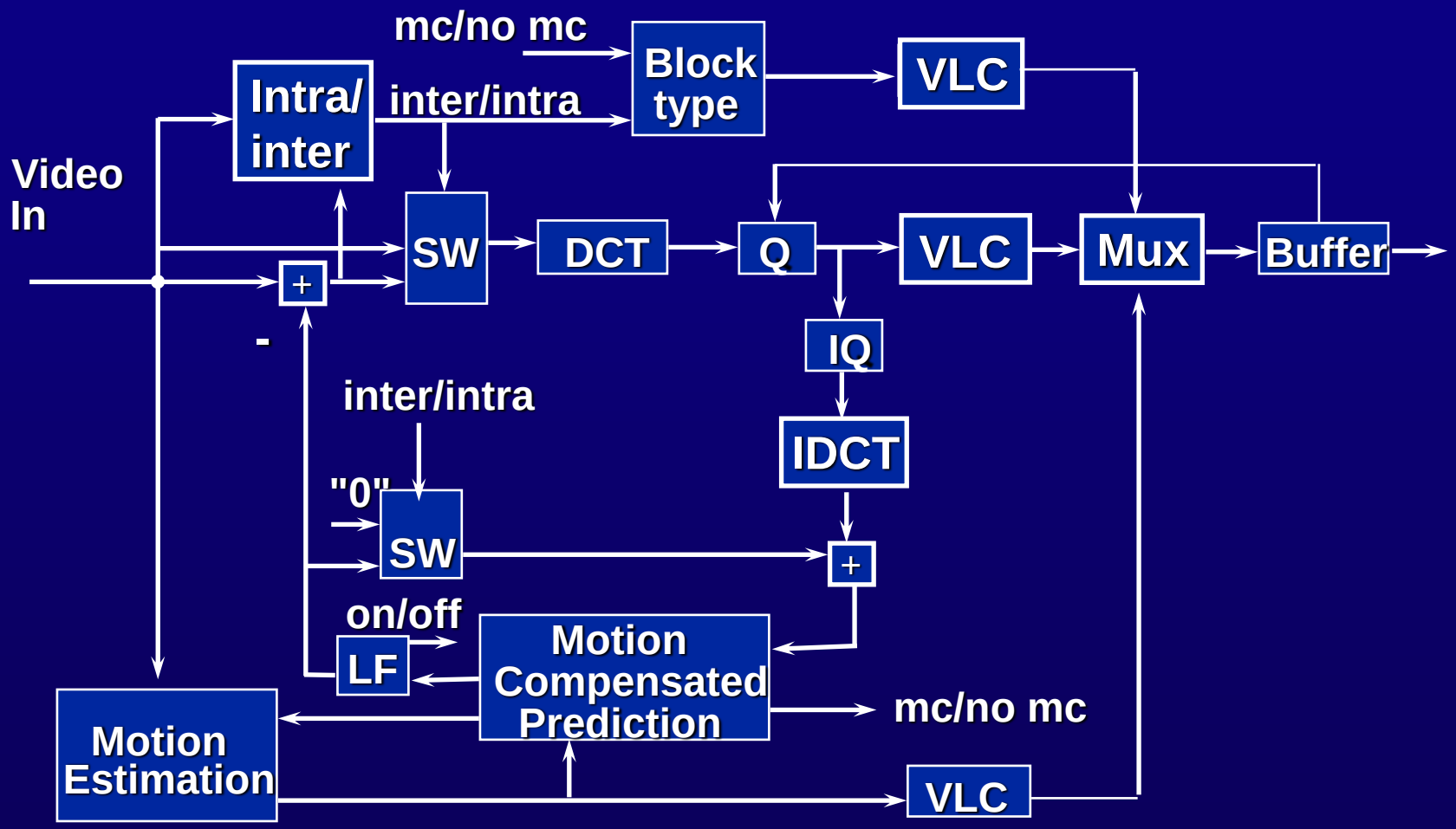


Y

C_b

C_r

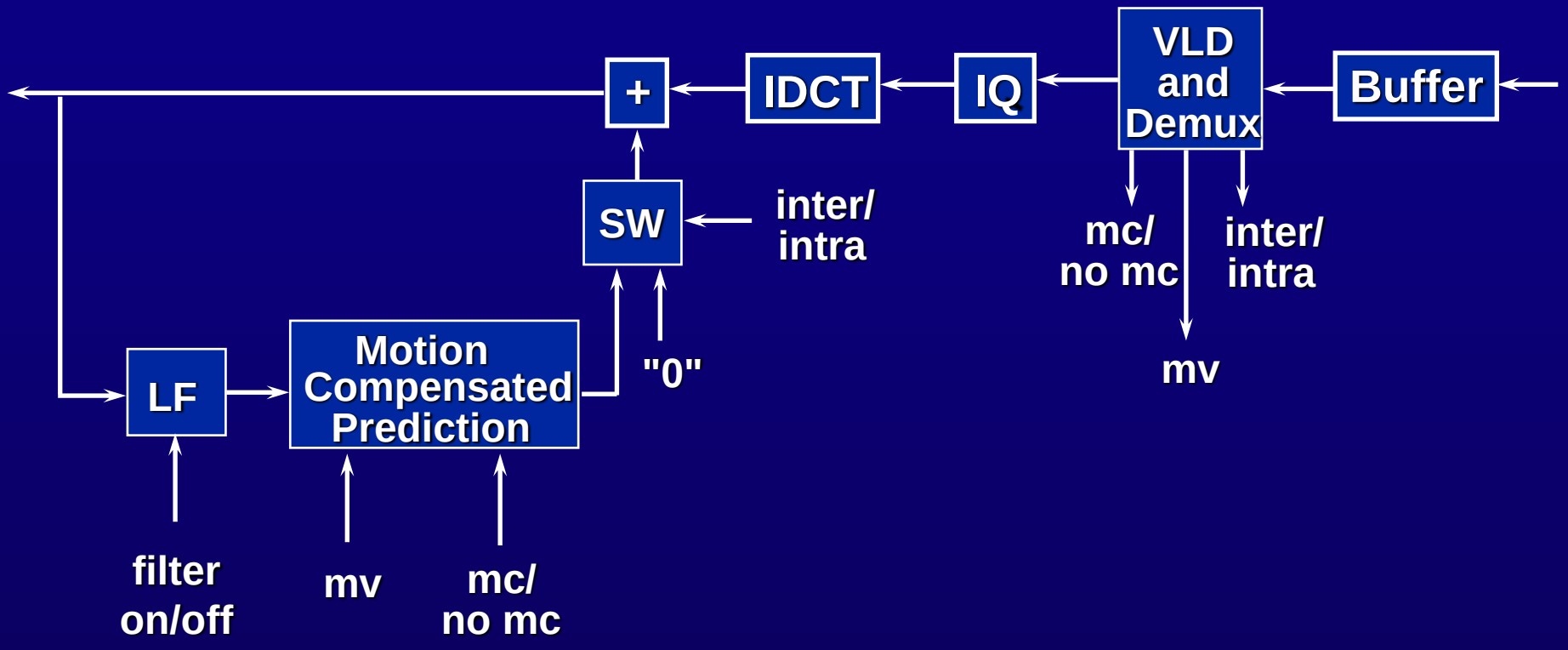
Block Diagram of An H.261 Encoder



mc: motion compensated prediction
LF: loop filter

SW: switch, Mux: multiplexer
VLC: variable-length coding

Block Diagram of An H.261 Decoder



Motion Estimation and Compensation

- 16 x 16 Block Matching; the search algorithm is not specified (can be full search, 3-step search, etc. using mean-absolute-errors, mean-squared-errors, ...)
- Motion estimation and compensation is optional in the encoder
- Maximum search range: +/- 15 pixels
- Motion vectors are differentially coded with VLC
- Motion estimation uses luminance signal only

Coding of Motion Vectors

DMV	Code
...	...
-7 & 25	0000 0111
-6 & 26	0000 1001
-5 & 27	0000 1011
-4 & 28	0000 111
-3 & 29	0001 1
-2 & 30	0011
-1	011
0	1
1	010
2 & -30	0010
3 & -29	0001 0
4 & -28	0000 110
5 & -27	0000 1010
6 & -26	0000 1000
7 & -25	0000 0110
...	...

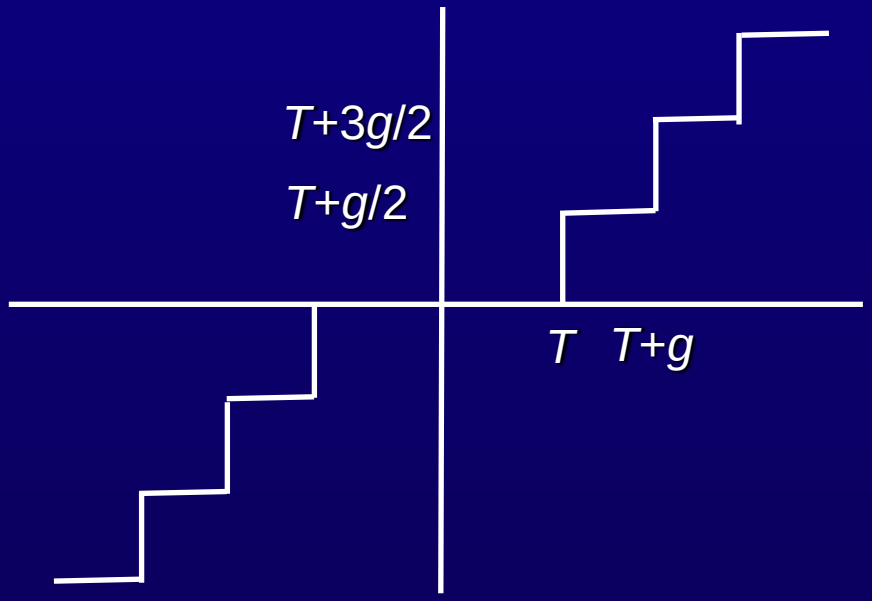
- Differential coding
- VLC for MV differences

Quantizer

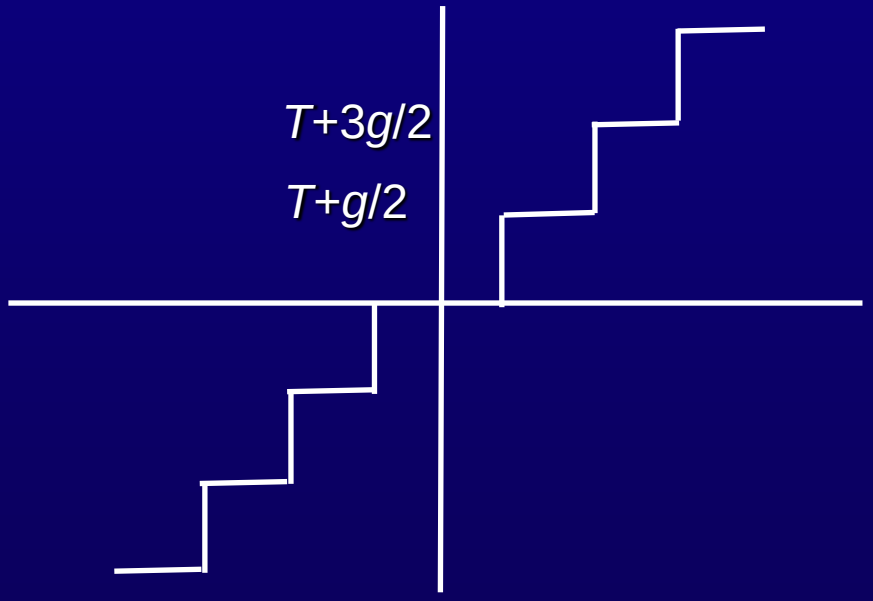
- The number of quantizer is 1 for the INTRA DC coef. and 31 for all other coefficients.
- Within a macroblock the same quantizer is used for all coefficients except the INTRA dc one.
- The reconstruction levels are defined, but the decision levels are not defined.
- The INTRA dc is linearly quantized with a stepsize of 8 and no dead-zone.
- Each of the other 31 quantizers is linearly quantized with a central dead-zone and with a step-size g of an even value in the range 2 to 62.

Quantizer

- Step-size *g*
- Dead-zone of *T* for noise removal



- Without dead-zone



2-D VLC

run	Absolute level: 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	…	128
0	3	5	6	8	9	9	11	13	13	13	13	14	14	14	14			
1	4	7	9	11	13	14	14											
2	5	8	11	13														
3	6	9	13	14														
4	6	11	13															
5	7	11	13															
6	7	13																
7	7	13																
8	8	13																
9	8	14																
10	9	14																
11	9																	
12	9																	
13	9																	
14	11																	
…	...																	
26	14																	
27																		
…																		
63																		

Others are 20-bit fixed length codes:
Escape (6 bits) + Run (6 bits) + level (8 bits)

- EOB: 2 bits (10)

Example

Run	Level	Code
EOB		10
0	1	1s If first coefficient in block
0	1	11s Not first coefficient in block
0	2	0100 s
0	3	0010 1s
0	4	0000 110s
0	5	0010 0110 s
0	6	0010 0001 s
0	7	0000 0010 10s
0	8	0000 0001 1101 s
0	9	0000 0001 1000 s
0	10	0000 0001 0011 s
0	11	0000 0001 0000 s
0	12	0000 0000 1101 0s
0	13	0000 0000 1100 1s
0	14	0000 0000 1100 0s
0	15	0000 0000 1011 1s
1	1	011s
1	2	0001 10s
1	3	0010 0101 s
1	4	0000 0011 00s
1	5	0000 0001 1011 s
1	6	0000 0000 1011 0s
1	7	0000 0000 1010 1s
2	1	0101 s
2	2	0000 100s
2	3	0000 0010 11s
2	4	0000 0001 0100 s
2	5	0000 0000 1010 0s
3	1	0011 1s
3	2	0010 0100 s
3	3	0000 0001 1100 s
3	4	0000 0000 1001 1s

S (sign):
1 negative, 0 positive

Macroblock Types

Prediction	MQUANT	MVD	CBP	TCOEFF	VLC
Intra				x	0001
Intra	x			x	0000001
Inter			x	x	1
Inter	x		x	x	00001
Inter + MC		x			000000001
Inter + MC		x	x	x	00000001
Inter + MC	x	x	x	x	0000000001
Inter + MC + FIL		x			001
Inter + MC + FIL		x	x	x	01
Inter + MC + FIL	x	x	x	x	000001

H.261

Does not specify:

- preprocessing and post processing
- the criteria for choosing the mode for coding a macroblock
- the use of BCH (511,493) in the decoder
- motion estimation in the encoder
- the quantizer decision levels
- rate-control algorithm
- frame-rate

Specifies:

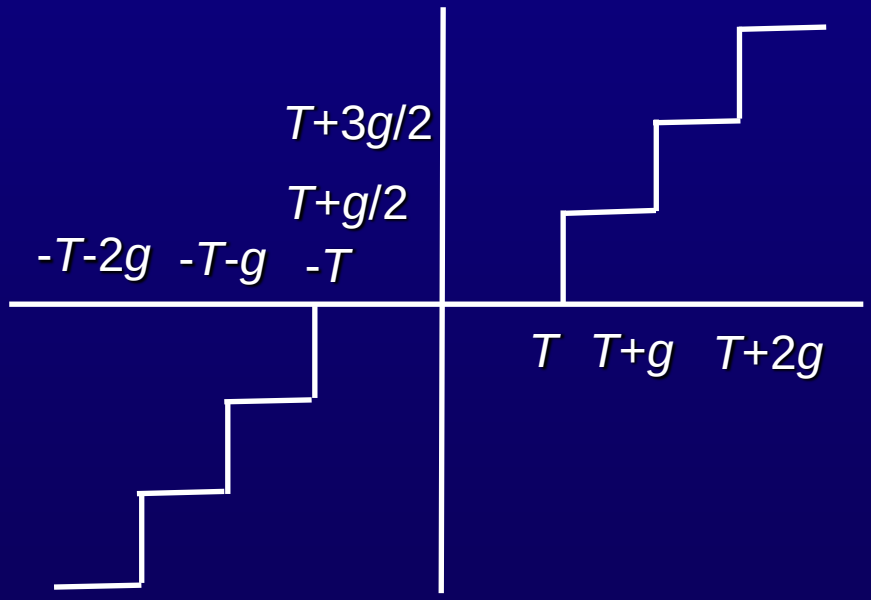
- a macroblock should be forcibly updated at least once per every 132 times it is transmitted
- for CIF, the number of bits created by coding any single picture must not exceed 256 kb; for QCIF, 64 kb
- Hypothetical reference decoder (HRD)

Reference Model 8 (RM8)

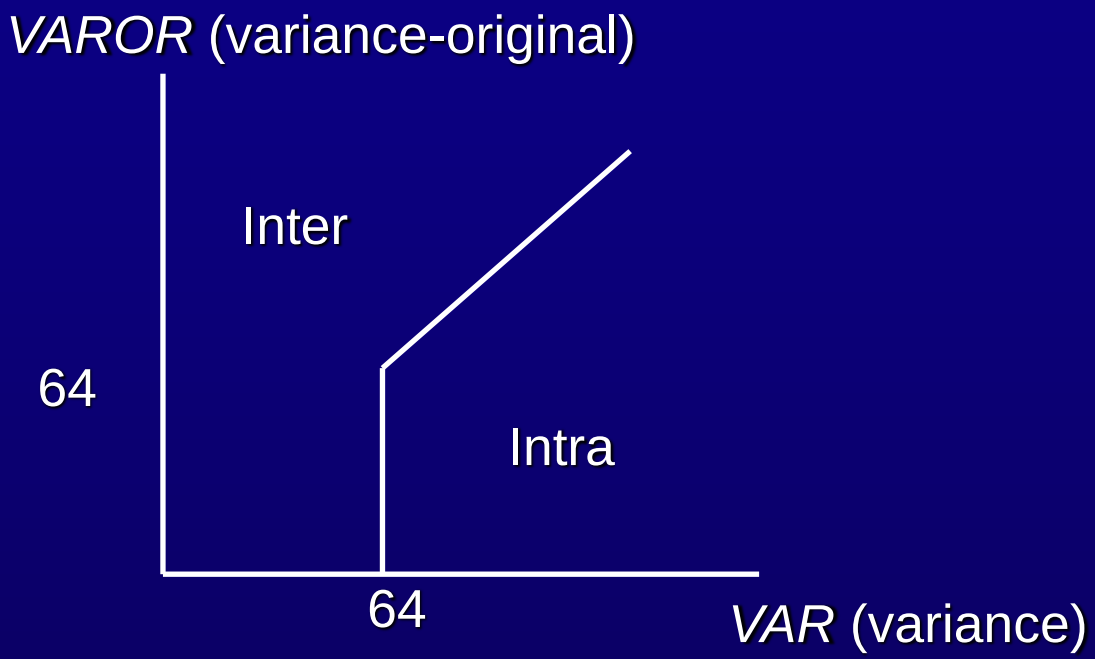
- A specific reference implementation of H.261 encoder including details, which were not specified in the standard
- Motion estimation: +/- 7 pixels, 3-step search
- Quantizer stepsize varies from 4 to 64
  with increment of 2 depending on the buffer fullness
- Methods for MC / No MC and Intra / Inter mode decision
- Force Intra-Update (every 132 frames)
- Rate control

Quantizer

- Stepsize *g*
- Dead-zone of *T* for noise removal



Decision of Inter/intra



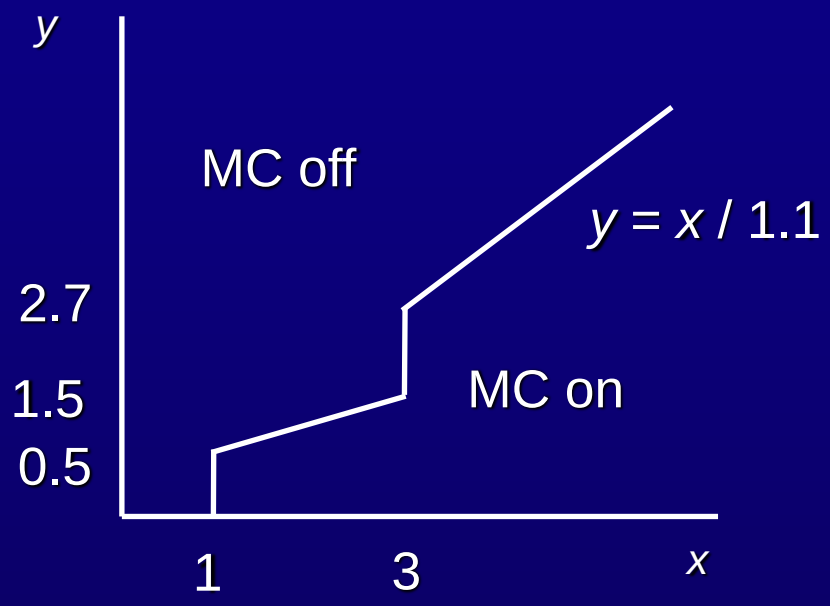
$VAR = (\Sigma(pel - mc\_pel)^2) / 256$

$VAROR = (\Sigma(pel - blk\_ave)^2) / 256$

Inter mode includes the solid line

Question: reasons for choosing Intra mode?

Decision of MC/no MC



$$x = \frac{|\,bd\,|}{256} \qquad y = \frac{|\,dbd\,|}{256}$$

MC off includes the solid line

bd: block difference, *dbd*: displaced block difference

Loop Filter

Impulse response: $h(x,y) = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$

9	3							
3	1							
			1	2	1			
			2	**4**	2			
			1	2	1			
			1	2	1			
			3	**6**	3			

Loop Filter (Cont.)

without loop filter 128 kbps

64 kbps

loop filter 128 kbps

64 kbps

Rate Control

- Buffer size = *q* * 6.4 kb (100ms)
- Quantization step size varies from 4 to 64 with step 2
- When the buffer fullness exceeds *q**6.4 kb, the coefficients and the motion vectors are set to zero in the next MB

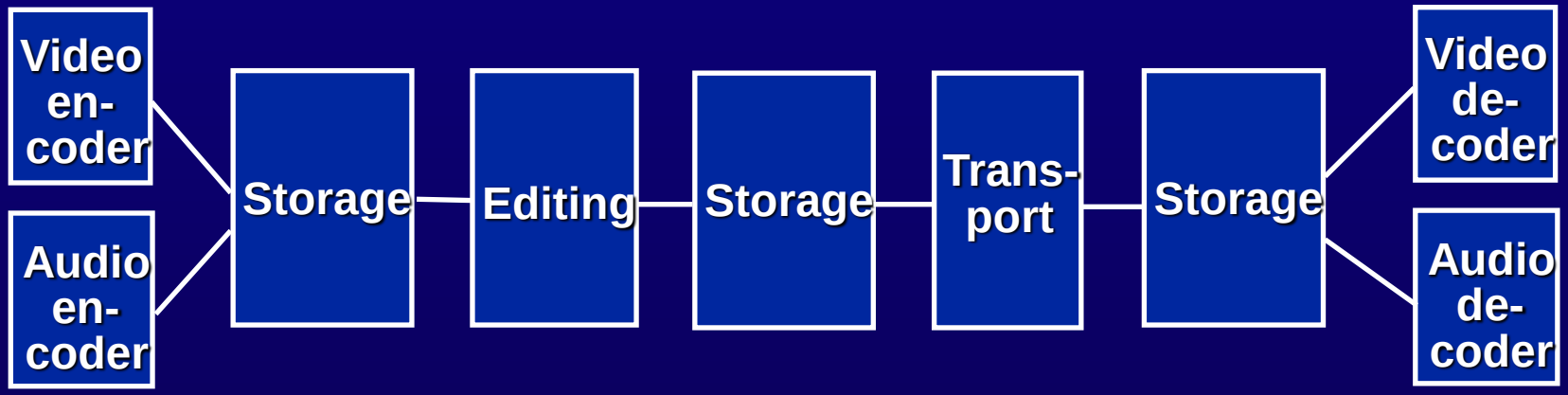
Buffer Fullness	Quantizer Step Size
< 400**q*	4
< 600**q*	6
< 800**q*	8
...	...
< 6200**q*	62
>= 6200**q*	64

MPEG-1

History of MPEG-1 Standard

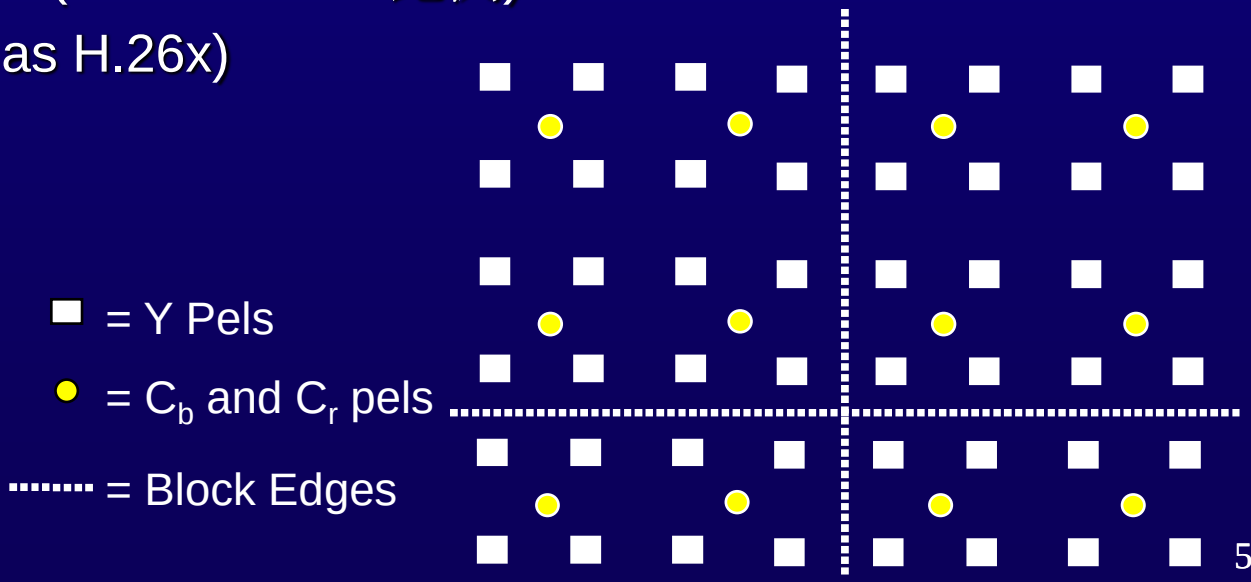
- Applications
  - Video on digital storage media
  - Computer and telecommunication networks

For basic digital storage media application



Parameters of MPEG-1

- Picture size up to 4096 × 4096 supported
  - Normally at 360 × 240
- Pel aspect ratio: 14 choices
- Picture rates: 23.976, 24 (movies), 25 (PAL), 29.97, 30, 50, 59.94, 60
- 4:2:0 format(macro block比例)
  - (same as H.26x)



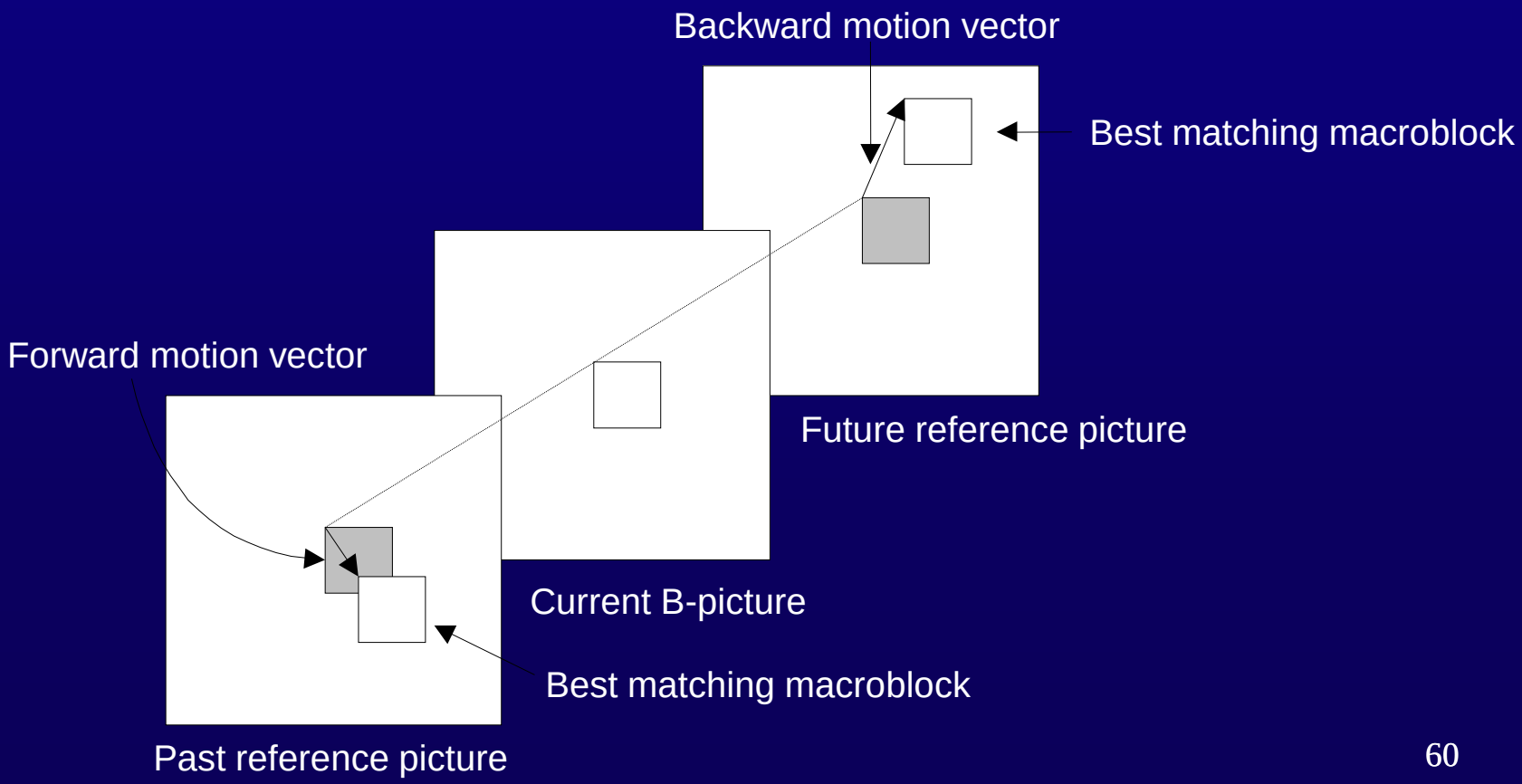
Typical MPEG-1 Video Source Format

Format	SIF (525/625)	
Signal component	Lines/Frame	Pixels/Line
Luminance (Y)	240/288	352
Chrominance (Cb)	120/144	176
Chrominance (Cr)	120/144	176

- Uncompressed bit-rate for transmitting SIF at 30 fps is 30.4 Mb/s

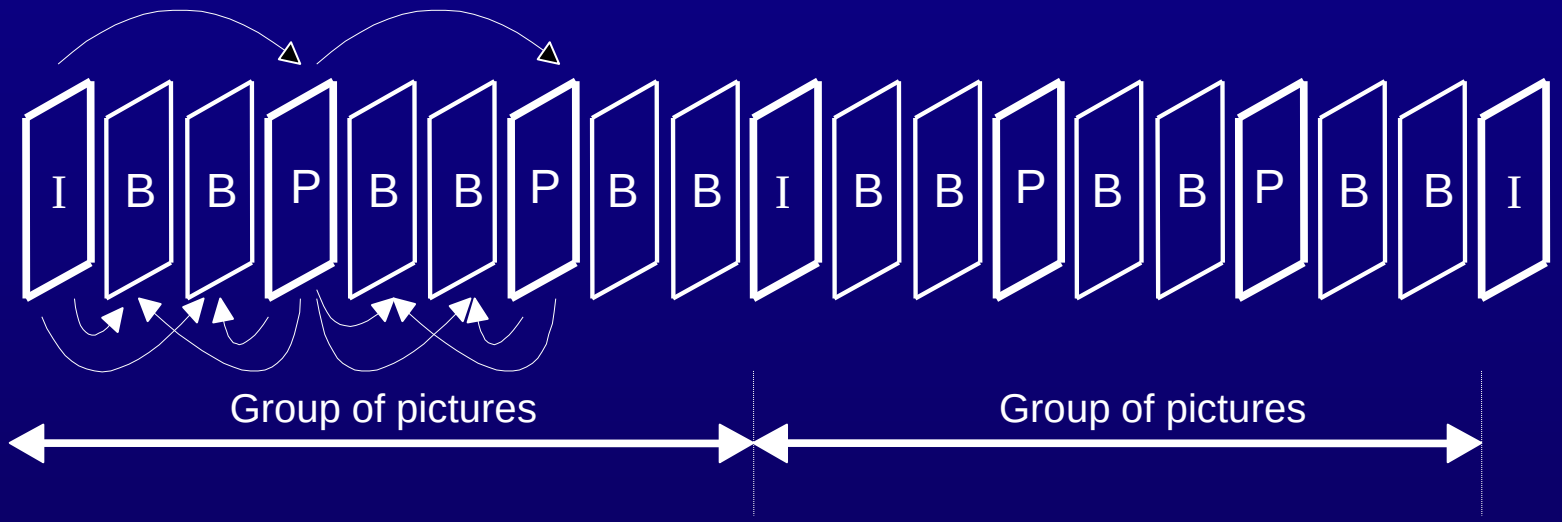
Bi-directional Motion Estimation

- Forward, backward, or average prediction: one or two motion vectors per 16x16 block with half-pixel accuracy



Coding Structure of GOP

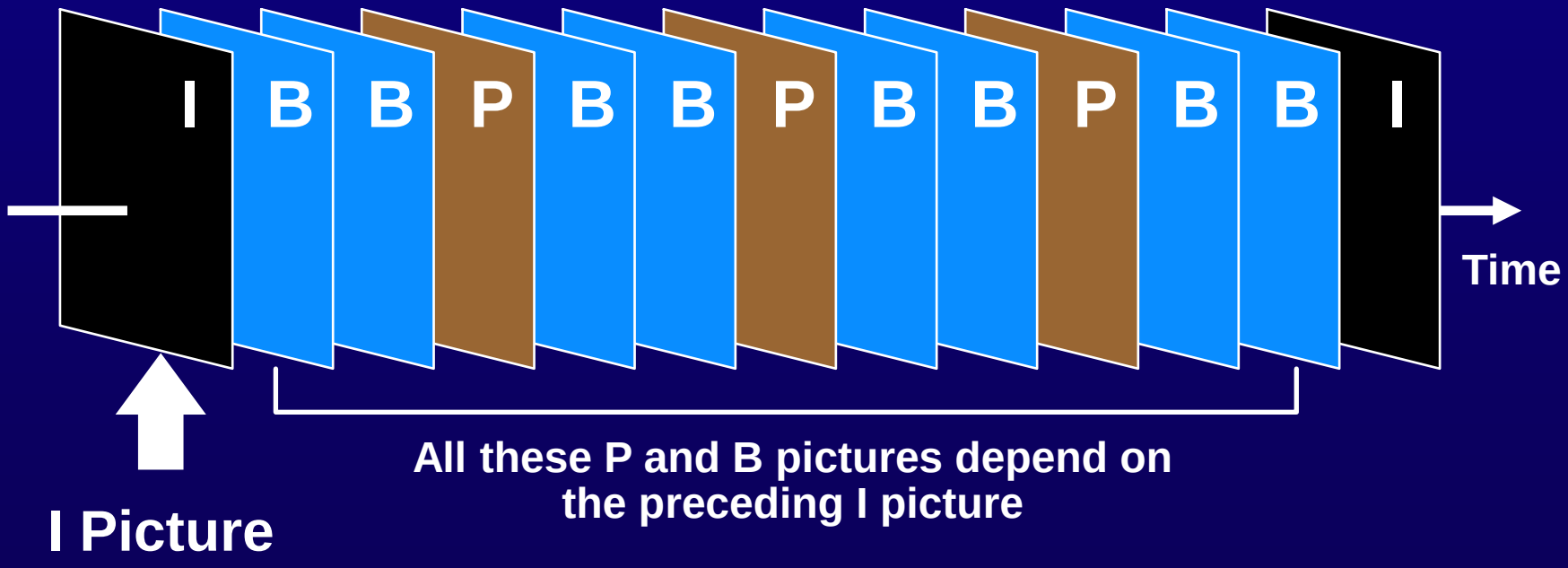
Random access可以從任意地方播，就會從此處找尋最近的I開始



I = **Intra-Picture** Coding, allow random access, for reference
P = **Predictive** coding, causal prediction only, can be referenced
B = **Bi-directional** coding, noncausal prediction, never referenced

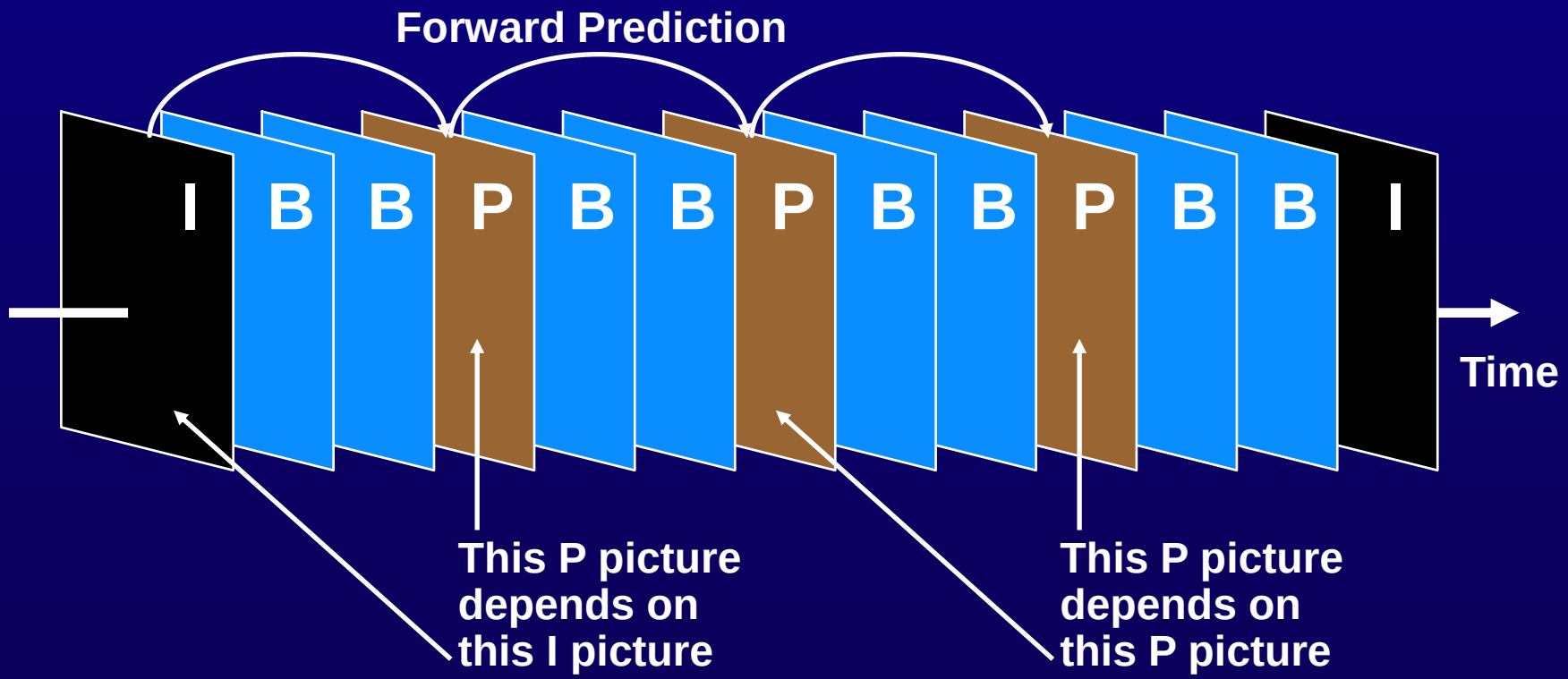
I-Pictures

- DCT coded without reference to any other pictures
- Stored in a frame buffer in encoder and decoder
- Used as basis of prediction for entire GOP



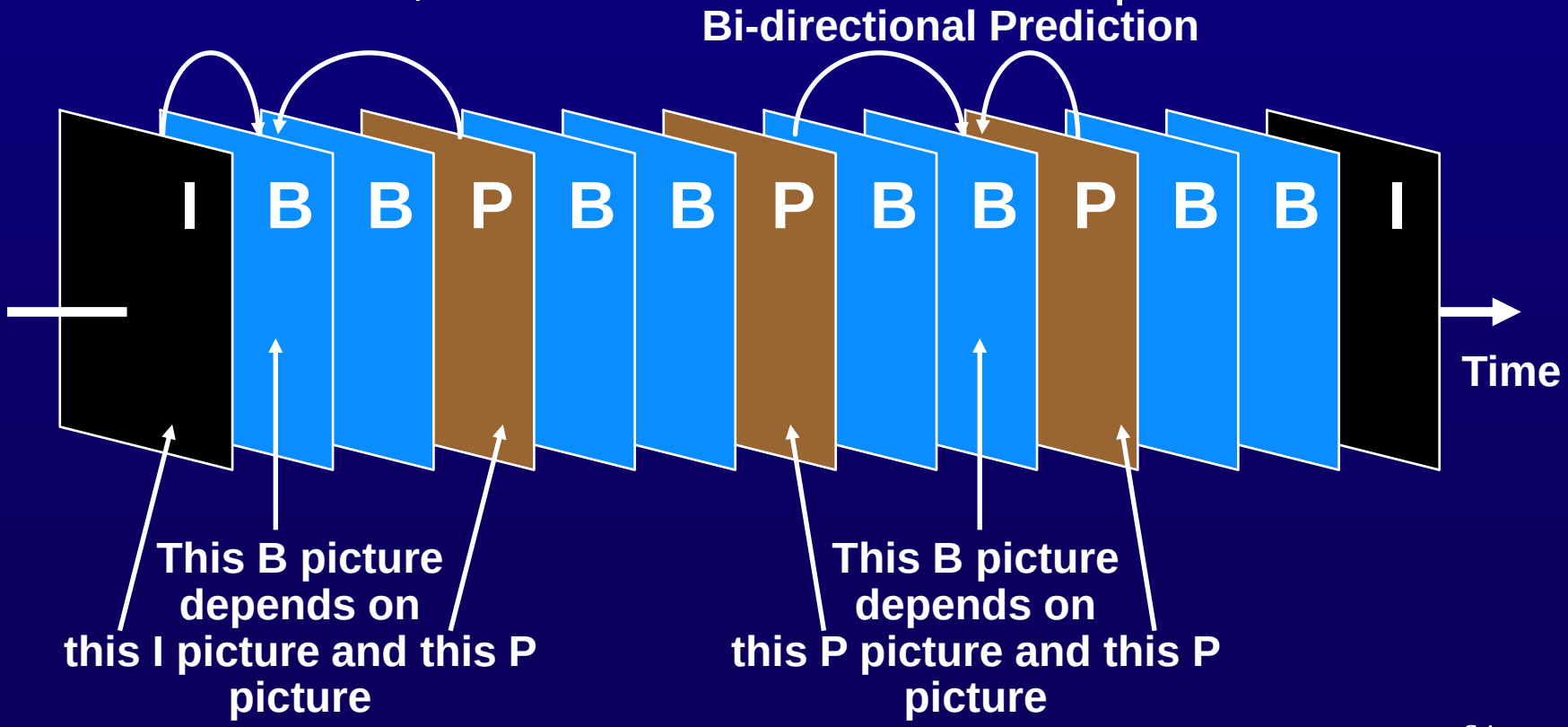
P-Pictures

- DCT coded with reference to the preceding anchor picture
- Stored in a frame buffer in encoder and decoder
- Use forward prediction only



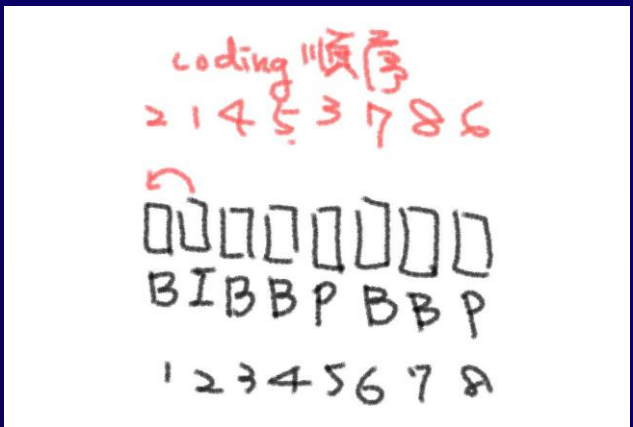
B-Pictures

- DCT coded with reference to either the preceding anchor picture, the following anchor picture(ref. pic), or both
- Use forward, backward or bi-directional prediction



GOP Rules

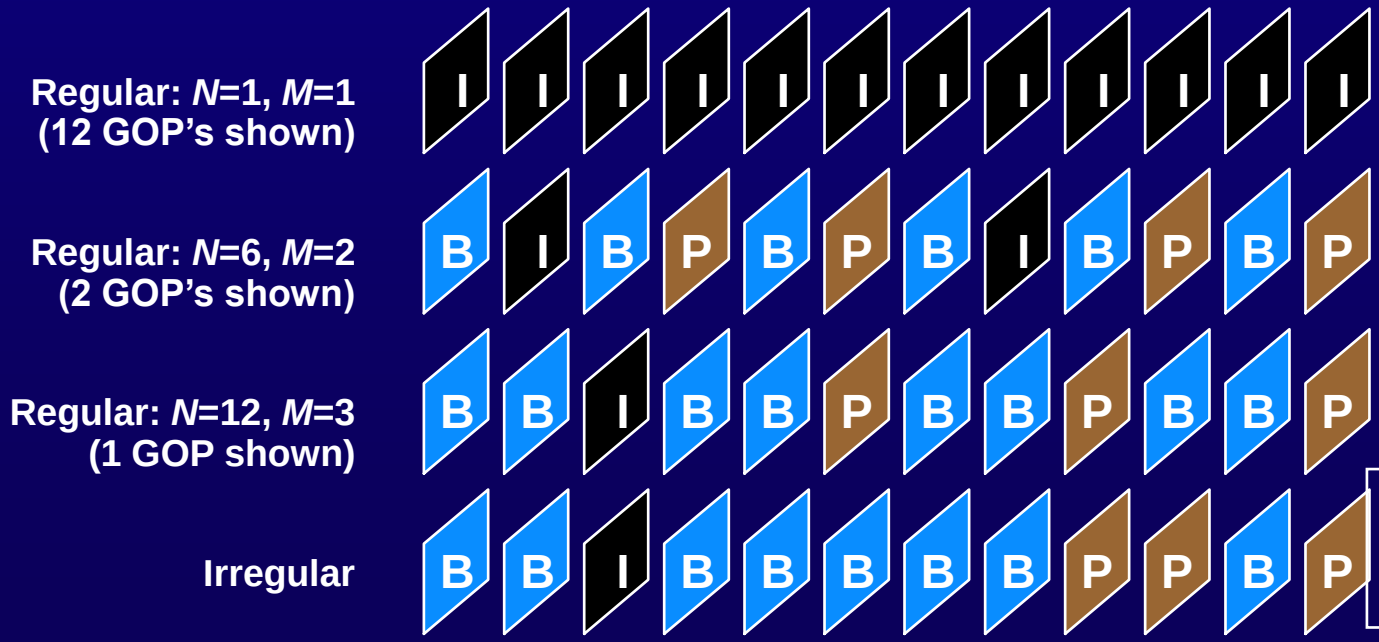
- A GOP must contain at least one I picture
- This I picture may be followed by any number of I and P pictures
- Any number of B pictures may occur between anchor pictures, and B pictures may precede the first I picture
- A GOP, in coding order, must start with an I picture
- A GOP, in display order, must start with an I or B picture and must end with an I or P picture



=>delay

Regular and Irregular GOPs

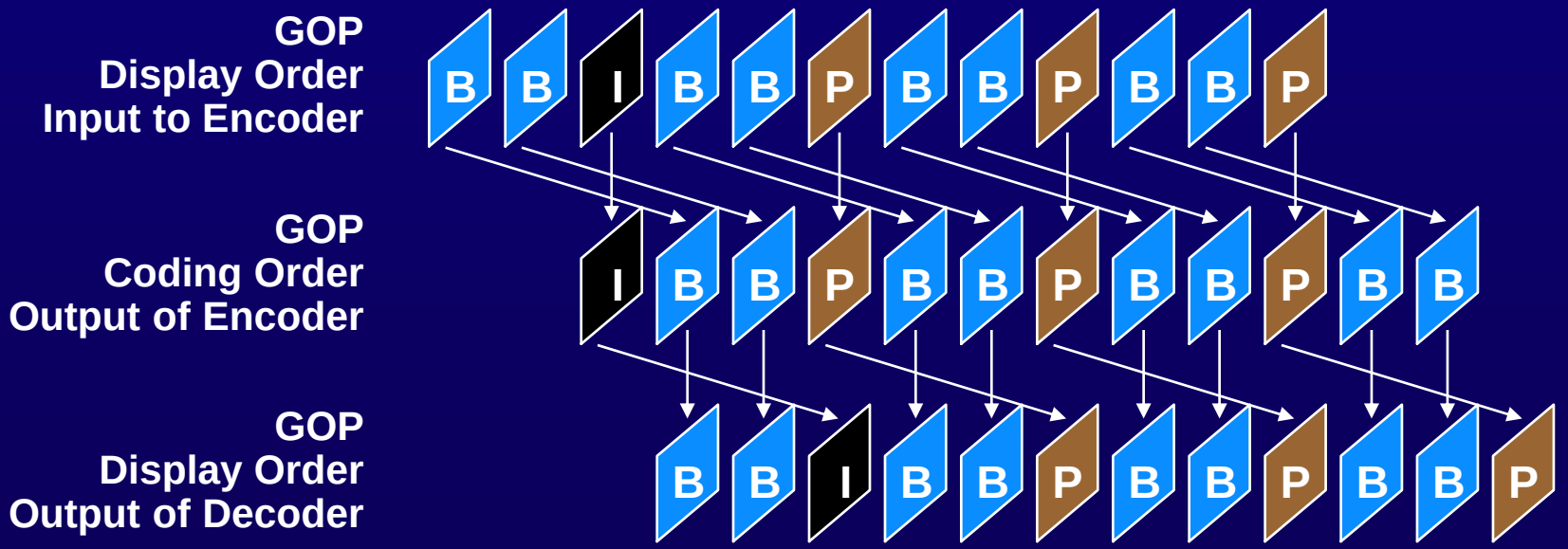
- Regular GOP's are defined by *N* and *M**:
  - *N* is the I picture interval
  - *M* is the anchor picture interval. There are *M*-1 B pictures between anchor pictures
- Irregular GOP's are not defined by *N* and *M*, but are still allowed as long as they follow the GOP Rules.



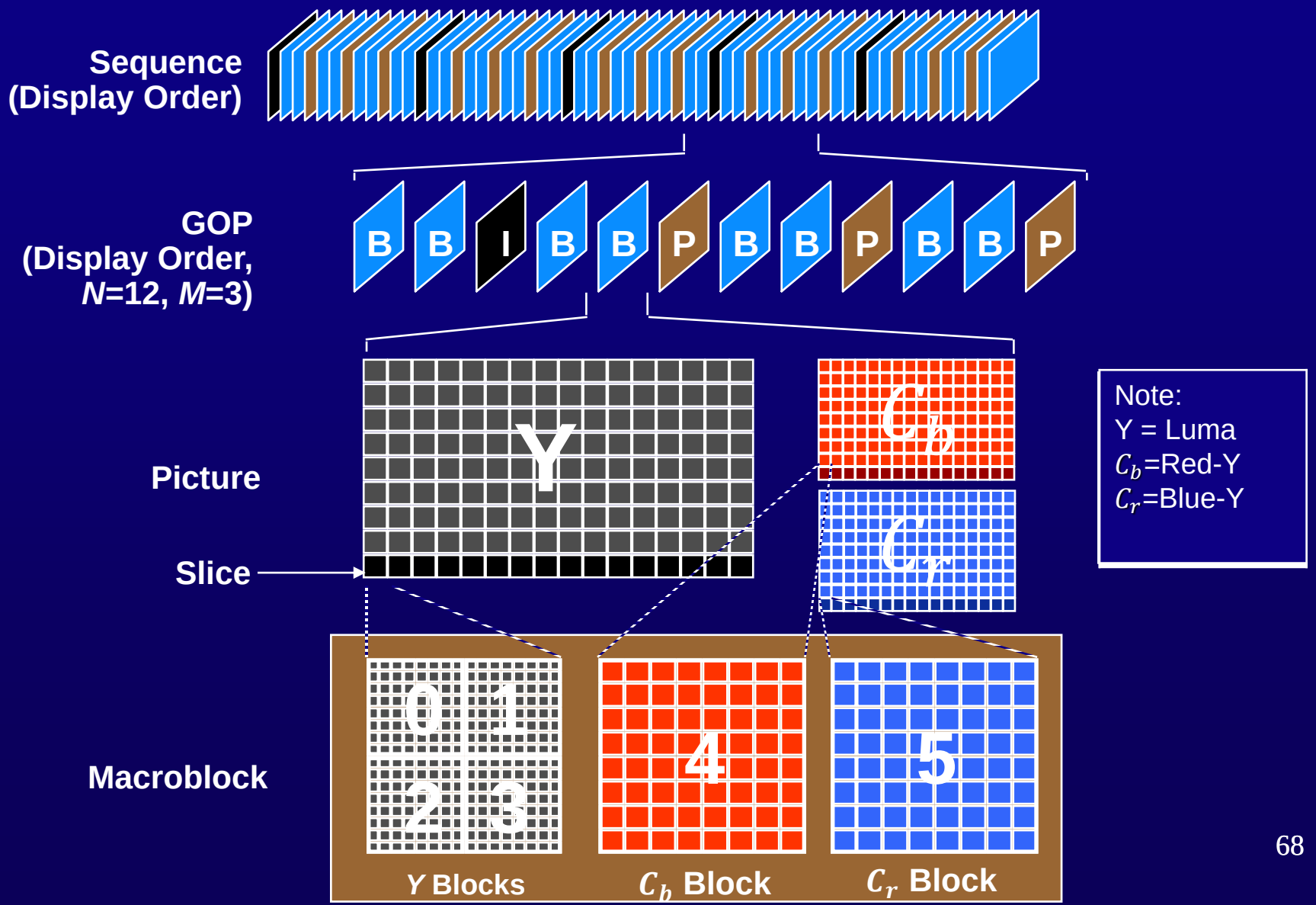
*N and M are not MPEG syntax elements and are not used in any way by the specification.

GOP Picture Orderings

- **Two Distinct Picture Orderings**
  - **Display Order (input to encoder, output of decoder)**
  - **Coding Order (output of encoder, input to decoder)**
  - **These are different if B frames are present**
  - **B frames must be reordered so that "future" anchor pictures are available for prediction.**



Six Hierarchical Layers of MPEG

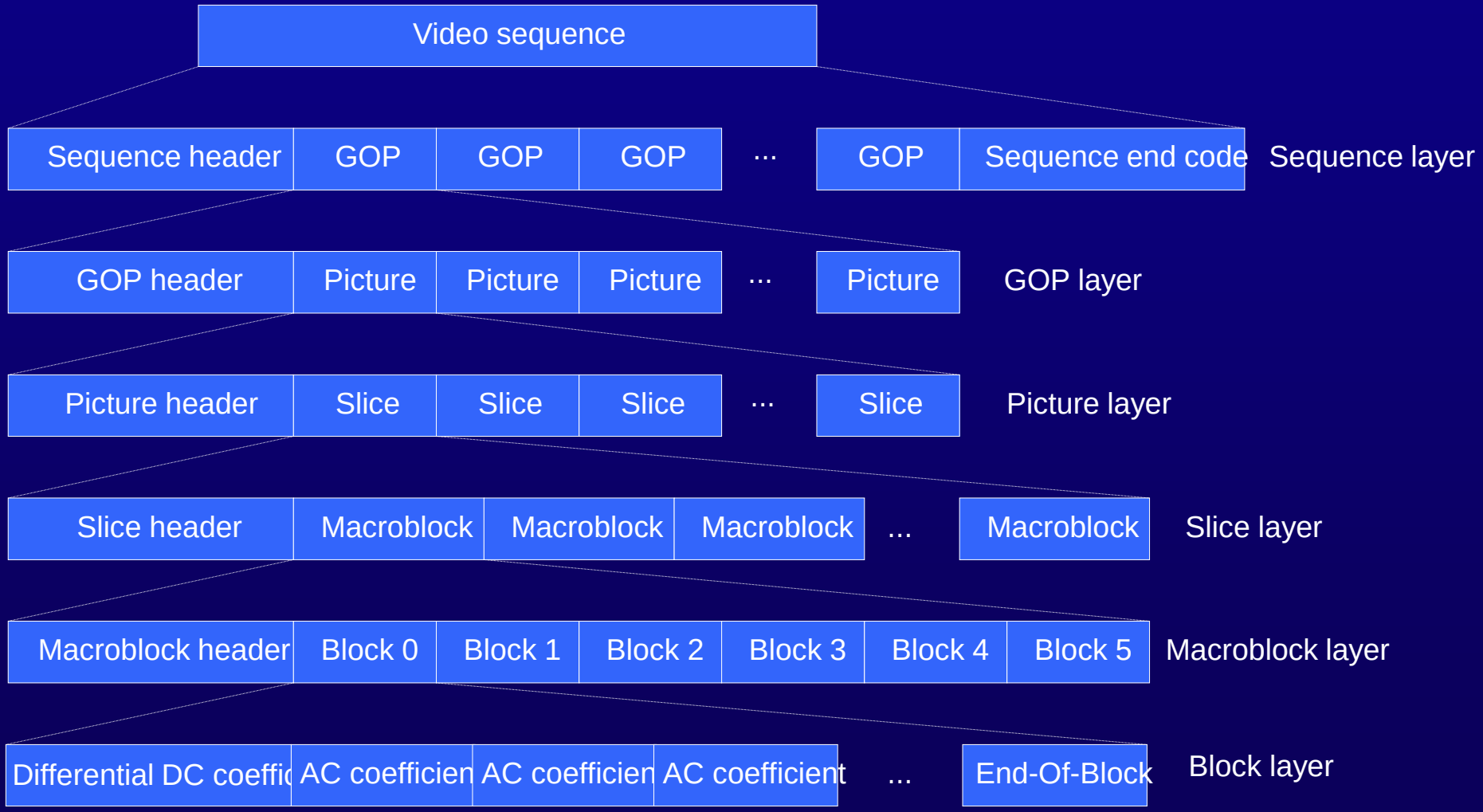


Six Hierarchical Layers of MPEG (Cont.)

- Important syntax elements in each layer:

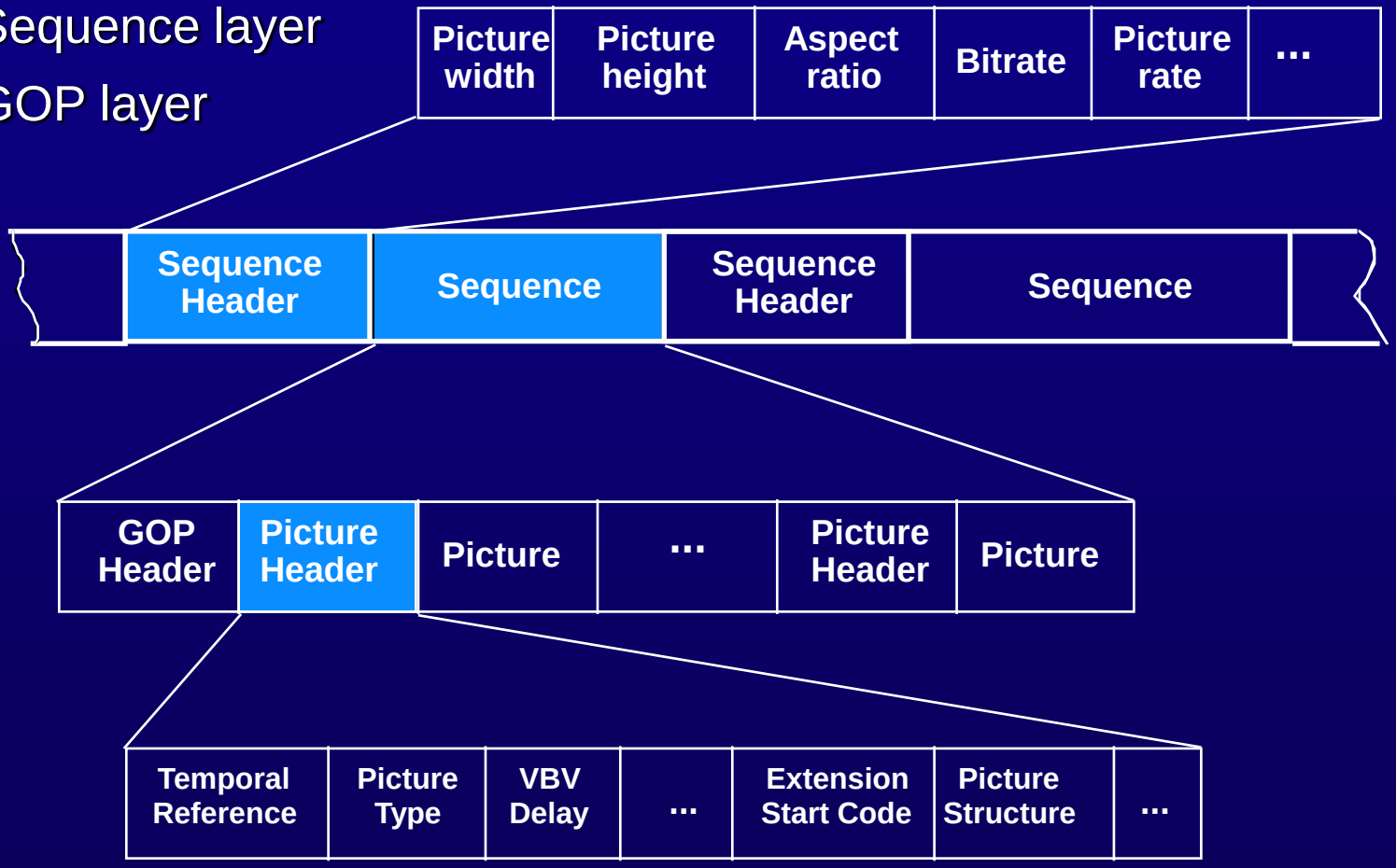
Sequence	Picture Size; Frame Rate Bit Rate; Buffering Requirements Programmable Coding Parameters
GOP	Random Access Unit SMPTE Time-Code
Picture	Timing information (buffer fullness, temporal reference), Coding type (I, P, or B)
Slice	Intra-frame addressing information Coding re-initialization (error resilience)
Macroblock	Basic coding structure, Coding method, Motion Vectors, Quantization
Block	DCT coefficients

MPEG-1 Video Bit-stream



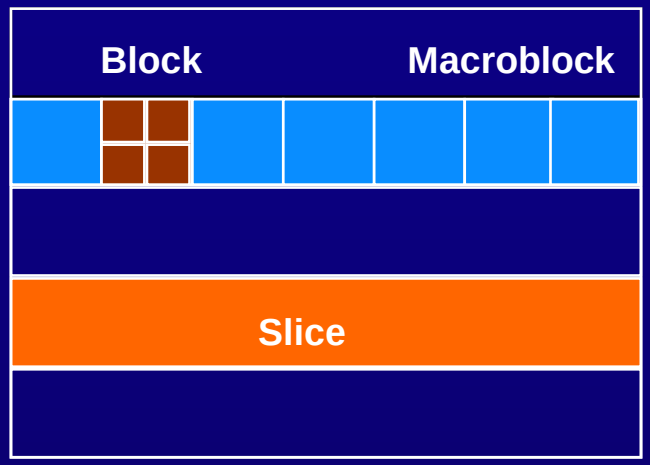
MPEG Bit-Stream Structure

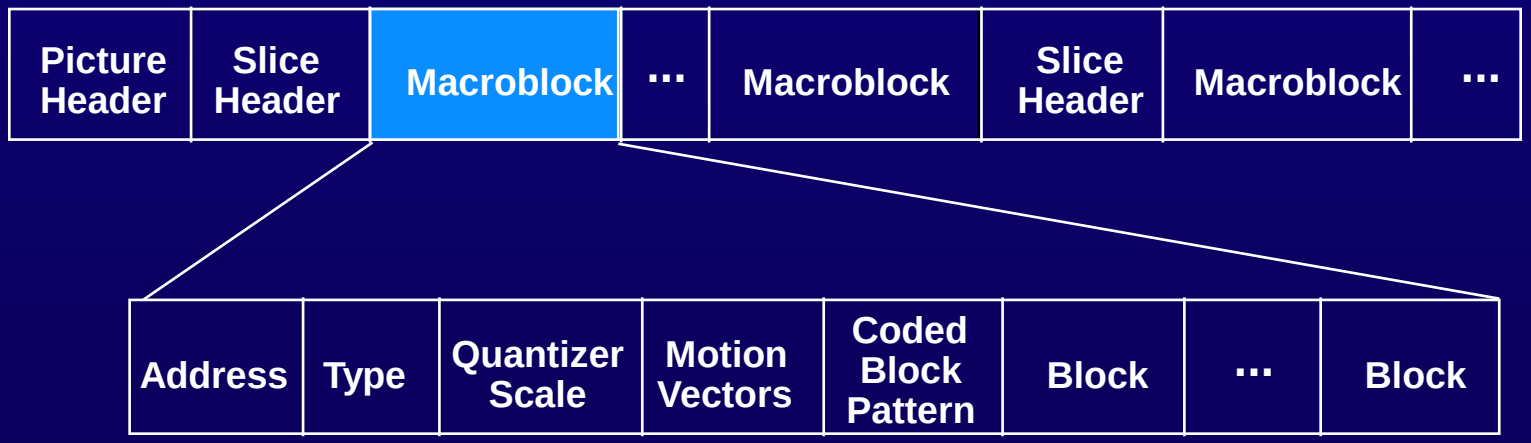
- Sequence layer
- GOP layer



MPEG Bit-Stream Structure (Cont.)

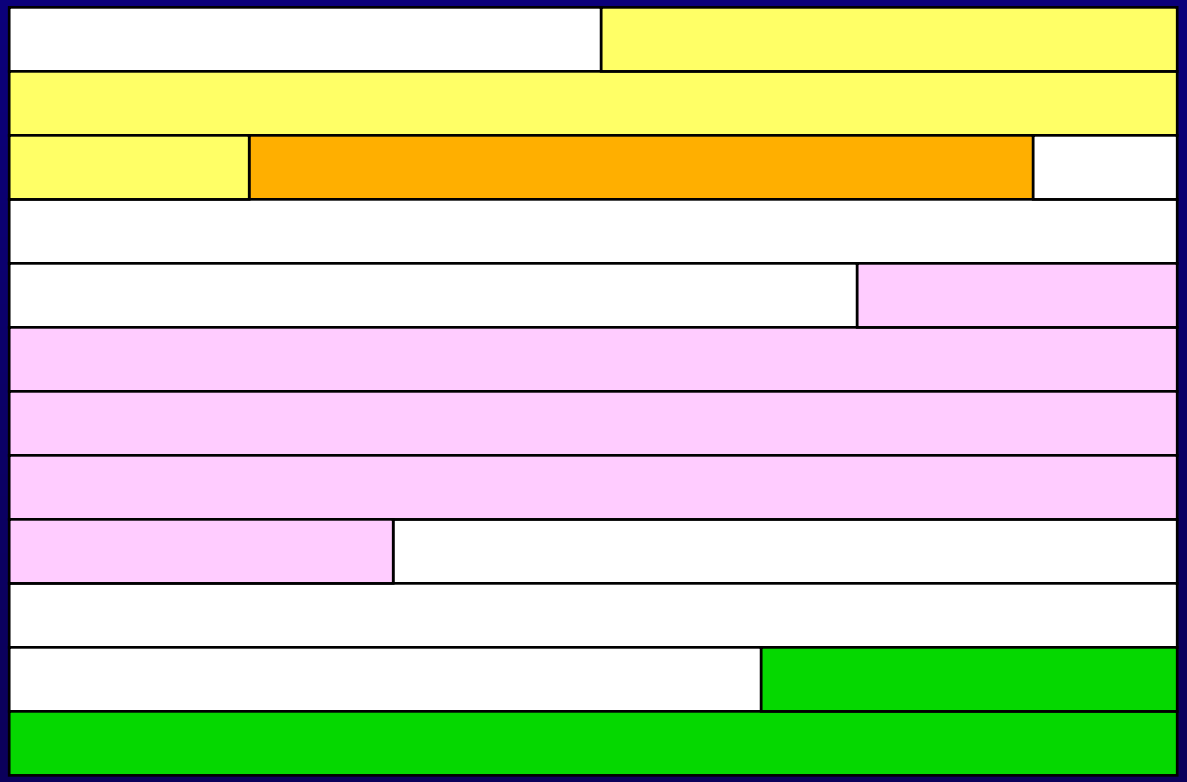
- Picture layer
- Slice layer
- Macroblock layer
- Block Layer





Slice Structure

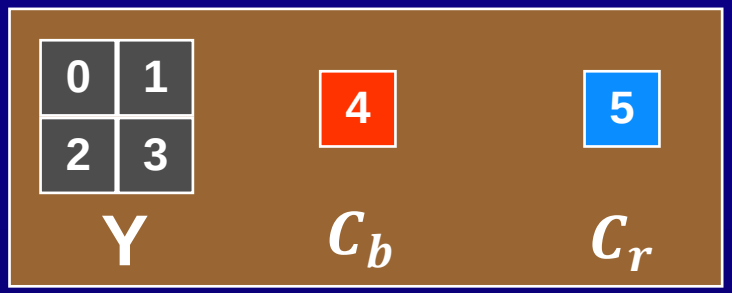
- A slice is a collection of macroblocks in raster scan order.
- A slice in MPEG-1 video frame can be single MB or entire picture.

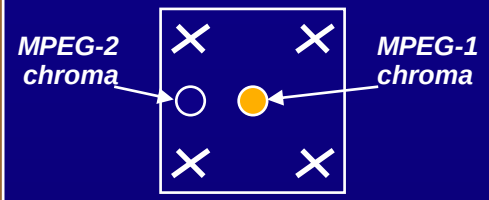


Macroblock Structure

4:2:0

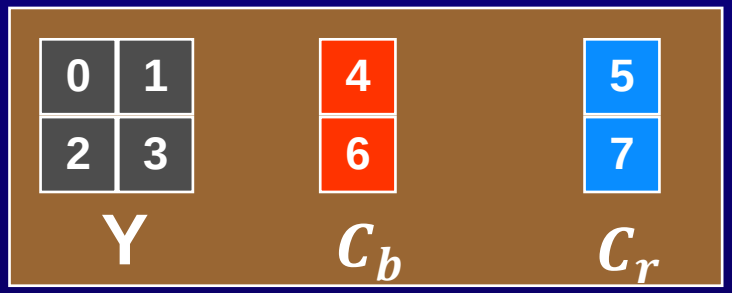
6 Blocks

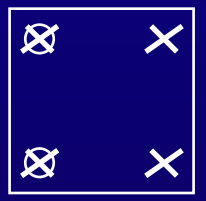




4:2:2

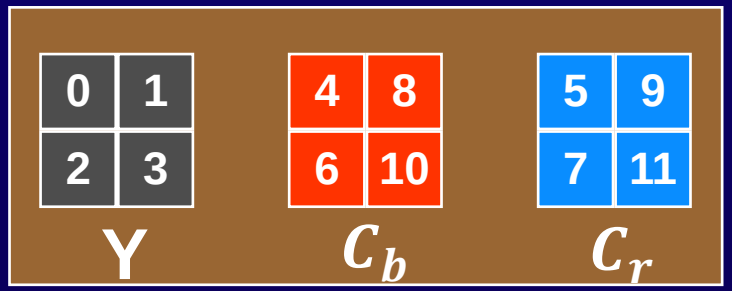
8 Blocks





4:4:4

12 Blocks

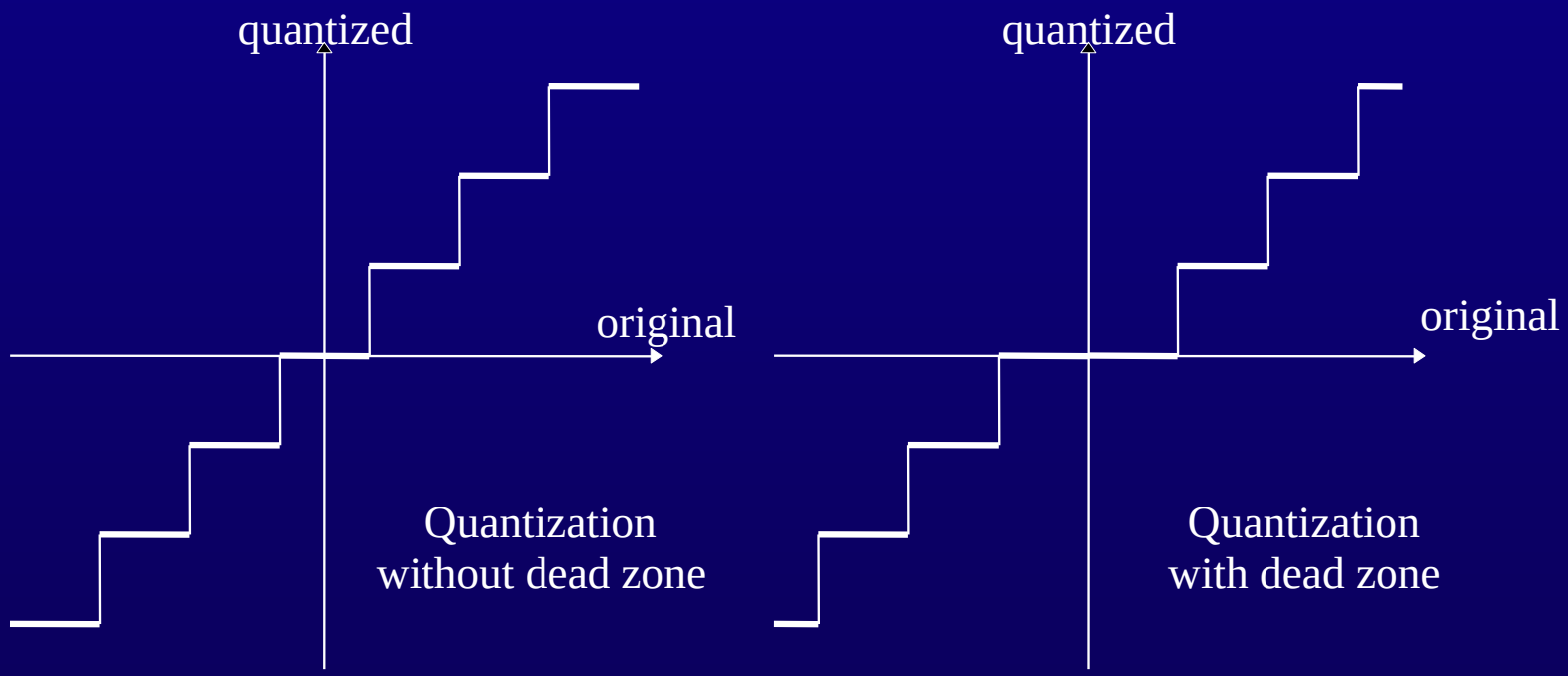


Spatial Sampling Relationship

✕ *luma* ○ *chroma*

Quantization

- Same as H.263



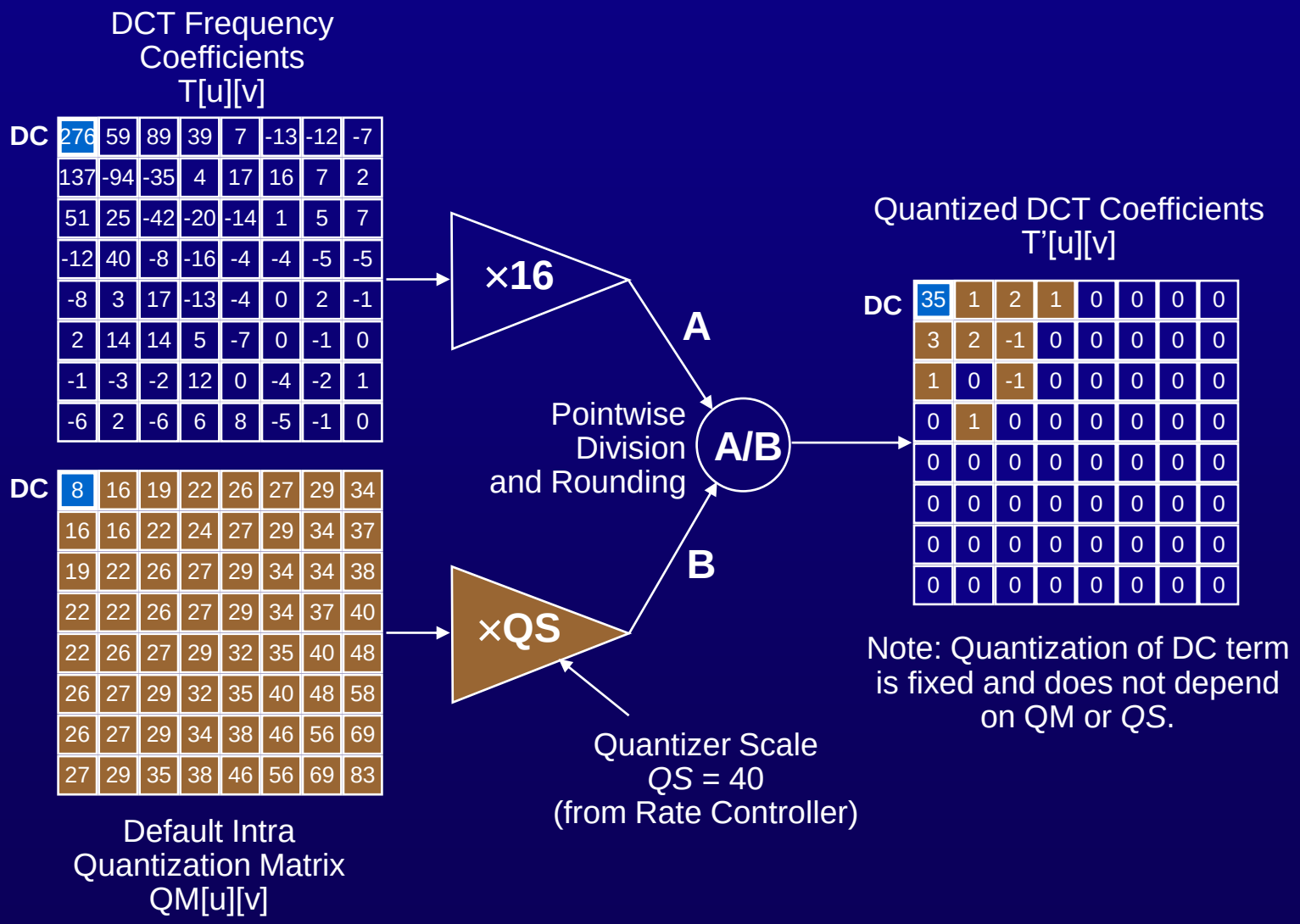
Intra(I pic)

(nearest integer)

Nonintra(P/B pic)

(rounded down)

Quantization Example



Default Quantization Matrices

DC

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

Intra Matrix: $QM_I[u][v]$

Note: AC coefficients (all coefficients except DC) are first multiplied by 16, then divided by $QS*QM_I[u][v]$.

DC term is treated specially.

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

Non-Intra Matrix: $QM_N[u][v]$

Note: All coefficients are first multiplied by 16, then divided by $QS*QM_N[u][v]$.

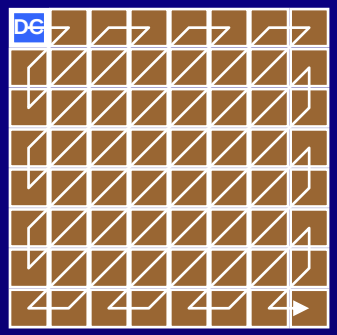
Downloadable Quantization Matrices

- **For improved quality in certain coding situations, quantization matrices for Intra and Non-Intra macroblocks can be downloaded.**
- **The decoder uses these instead of the defaults (which are not sent in the bitstream)**
- **The example at right shows an improved Non-Intra Quant Matrix used by the MPEG-2 Test Model 5 (TM5)**

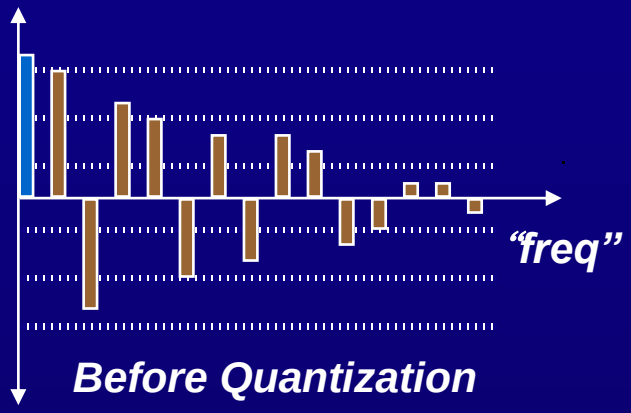
16	17	18	19	20	21	22	23
17	18	19	20	21	22	23	24
18	19	20	21	22	23	24	25
19	20	21	22	23	24	26	27
20	21	22	23	25	26	27	28
21	22	23	24	26	27	28	30
22	23	24	26	27	28	30	31
23	24	25	27	28	30	31	33

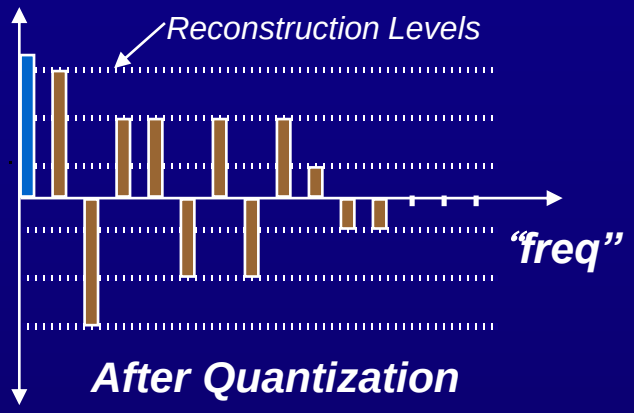
Example of Downloadable Matrix (TM5 Non-Intra Matrix)

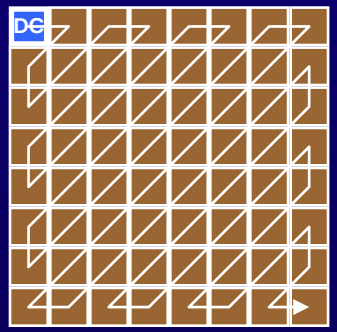
Quant Matrix Effect

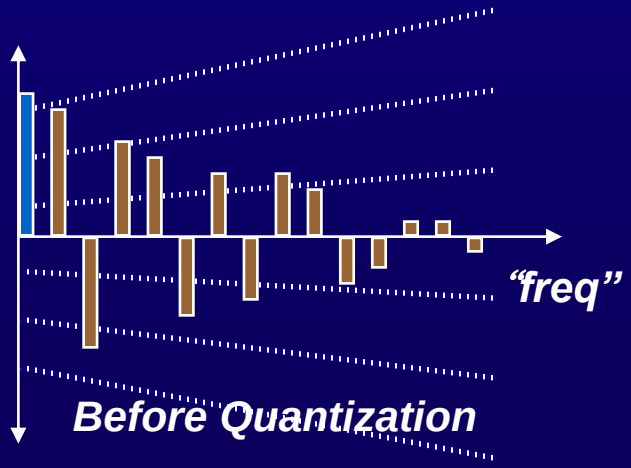


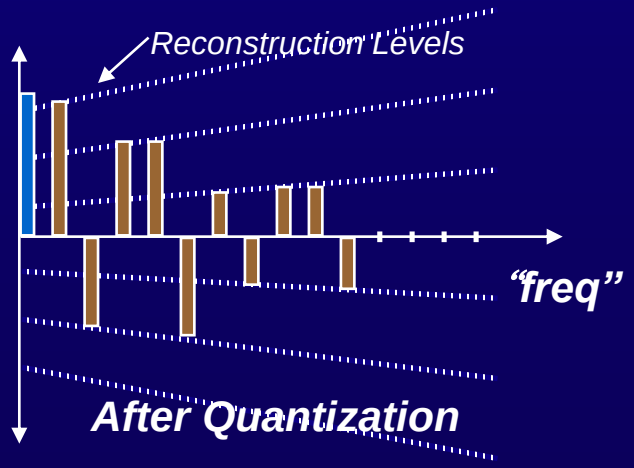
Flat Matrix







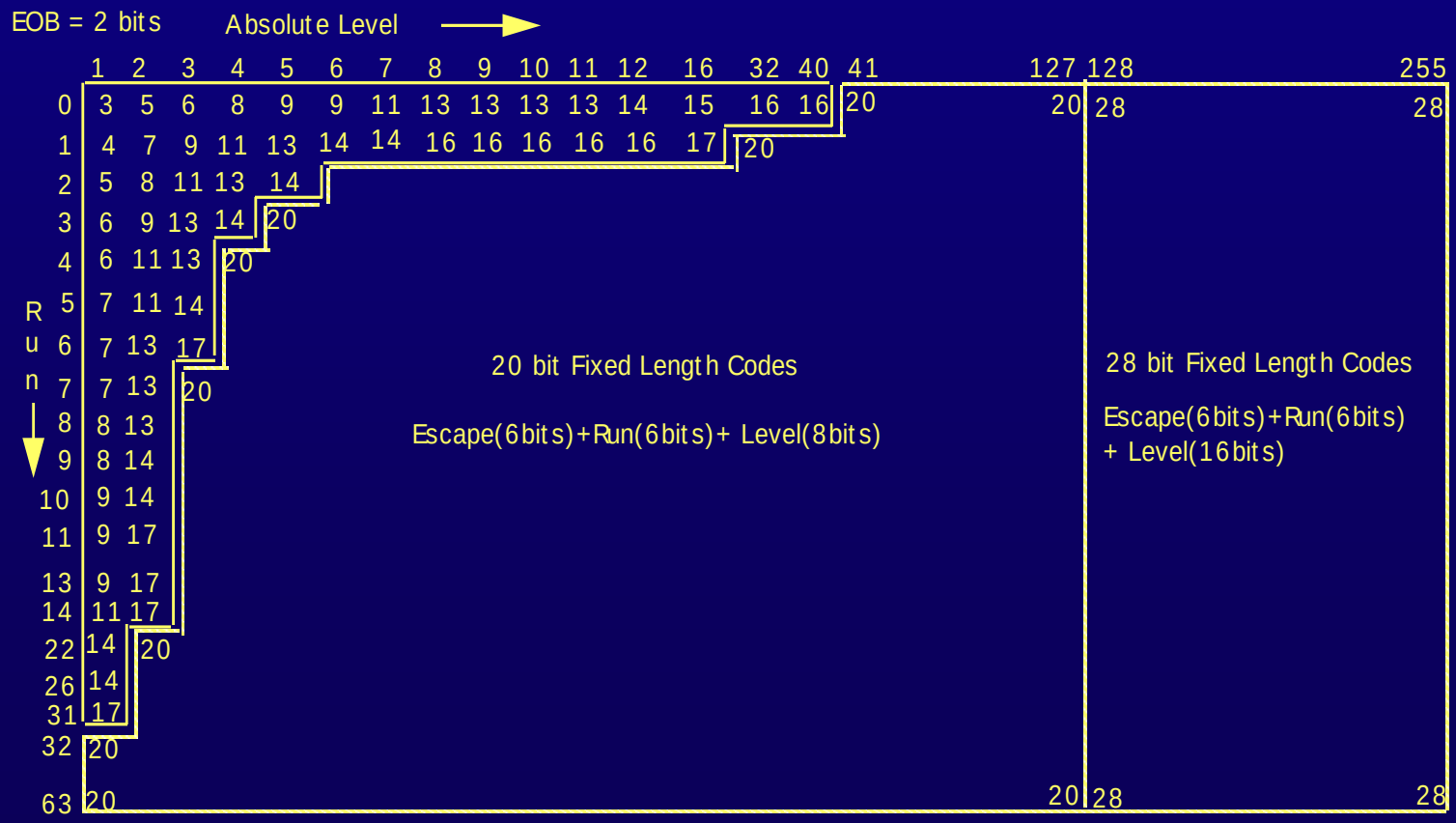




Tilted Matrix

2-D VLC

- Similar to H.261
  - Most codes same, some codes shorter
  - Levels up to 255



MPEG-1 Video Coding Standard

- Similar to H.261, MPEG-1 specifies
  - Bit stream syntax and semantics
  - Decoding to raster representation
  - VBV (video buffer verifier)

  doesn't specify
  - Pre-processing to raster representation
  - Encoding
  - Post-processing from raster representation
- Left flexible (parameters in the bit stream)
  - Coded bit-rate (constant or variable)
  - Lines per picture (< 4096)
  - Pels per line (< 4096)
  - Picture rate (24, 25, or 30 per second)
  - Pel aspect ratio (14 choices)

Simulation Model 3 (SM3)

- A specific reference implementation of MPEG-1 encoder including details which were not specified in the standard
- Motion estimation: one forward and/or one backward vector per MB with half-pixel resolution; 2-step search: (1) full search in the range of +/- 7 pixels, (2) search 8 neighboring half-pel positions
- Methods for MC / No MC and Intra/Inter decision
- Quantizer, rate control

一個方法決定是否使用motion compensate / intra等等，通常使用於此block可能是新的，對於計算(mvx,mvy)及residue會比直接intra該block還差，就可能使用該block進行intra計算即可（減少傳輸量)，rate control則是看buffer狀況調整quantization step size以控制輸出量

MPEG-2

MPEG-2 Video Coding Standard

- Primarily for coding interlaced video at 4 - 15 Mb/s for digital broadcast TV and high quality digital storage Media; also for HDTV, cable/satellite TV, video services over networks (e.g., ATM), and 2-way communications
- MPEG only specifies bitstream syntax and decoding process
- Encoding algorithms (e.g., motion estimation, rate Control and mode decisions) are open to invention and proprietary techniques
- MPEG is asymmetric in that much less computational power is required in the decoder.

Encoding的computation花費>>decoding的
相反：distributed system(ex. Sensor)就會變成decoding那邊計算量大

Features

Scalability縮小原圖，並再次放大，這個過程會有誤差，因此可以與原圖相減得到residue，網路好就basic和residue一起送，反之就送basic就好

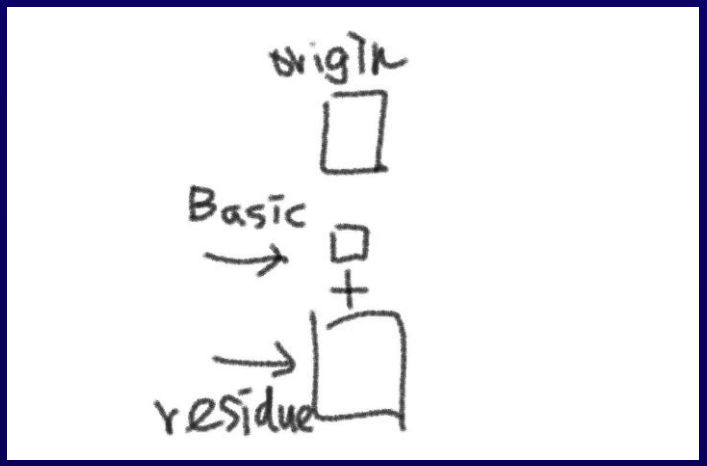
- Picture quality - good quality NTSC (4-6 Mb/s)
  excellent quality NTSC (8-10 Mb/s)
- Random access/channel switching in limit time - intra-pictures
- Trick modes - basic VCR functions
- Delay - low delay mode using Simple Profile for visual communications
- Error resilience - intra-mv, data-partitioning, priority assignment to video layers
- Allow higher chroma resolution - e.g., 4:2:2 and 4:4:4
- Scalability - data partition, SNR scalability, spatial scalability, temporal scalability, hybrid scalability (up to 3 layers)
- Compatibility - decodes MPEG-1 bit-stream, base layer may be decoded by MPEG-1 decoder　若支援mpeg2，就也要支援mpeg1
- Support multiple video formats and frame rates
- Subset of the standard permit real-time encoders of reasonable complexity

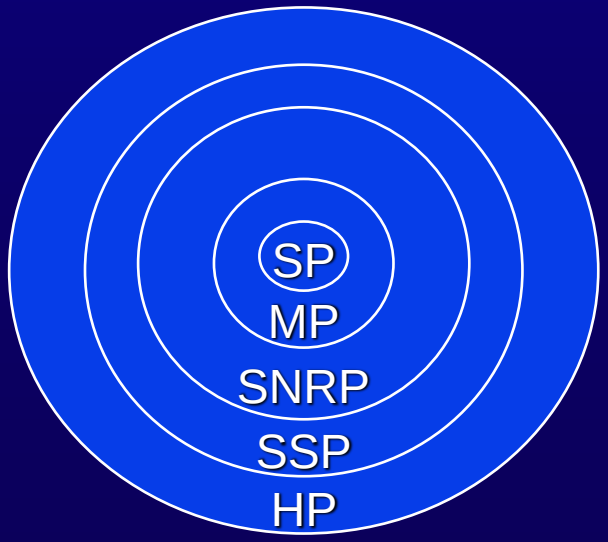
Profiles and Applications

- Each profile supports groups of features for an application area
- Simple Profile: low-delay videoconferencing
- Main Profile: most important, for general applications
- SNR Profile: multiple grades of quality
- Spatially Scalable Profile: multiple grades of quality and resolution
- High Profile: multiple grades of quality, resolution, and chroma format

New profiles:

- 4:2:2 profile
- Multiview profile





Profiles and Levels (Cont.)

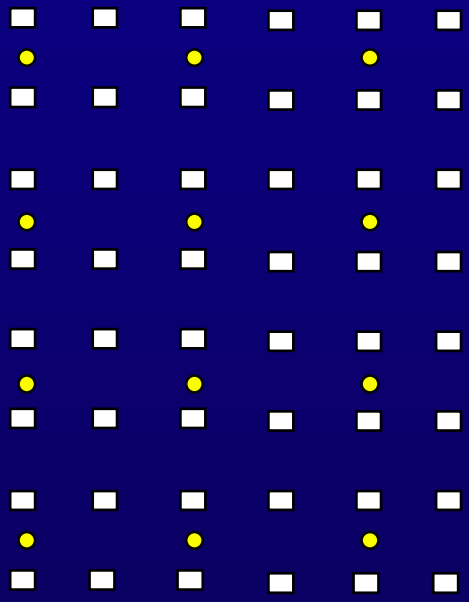
Level	Profile				
	Simple 4:2:0	Main 4:2:0	SNR Scalable 4:2:0	Spatially Scalable 4:2:0	High 4:2:0 or 4:2:2
High 1920x1152 (60 frames/s)		62.7 Ms/s 80 Mbit/s			100 Mbit/s for 3 layers
High-1440 1440x1152 (60 frames/s)		47 Ms/s 60 Mbit/s		47 Ms/s 60 Mbit/s for 3 layers	80 Mbit/s for 3 layers
Main 720x576 (30 frames/s)	10.4 Ms/s 15 Mbit/s	10.4 Ms/s 15 Mbit/s	10.4 Ms/s 15 Mbit/s for 2 layers		20 Mbit/s for 3 layers
Low 352x288 (30 frames/s)		3.04 Ms/s 4 Mbit/s	3.04 Ms/s 4 Mbit/s for 2 layers		

* numbers in the table are maximum allowed

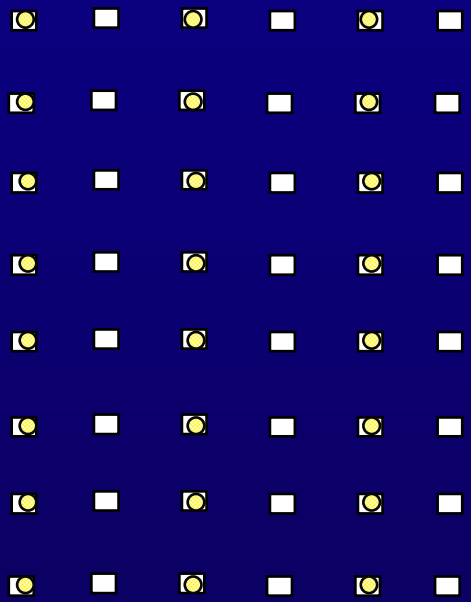
MPEG-2: Resolutions and Formats

- Picture sizes extension up to 16kx16k; 720x480 ~ TV resolution
- Support picture rates: 23.98, 24, 25, 29.97, 30, 50, 59.94, 60
- Support both progressive and interlaced formats
- Support 4:2:0, 4:2:2, and 4:4:4 sampling formats

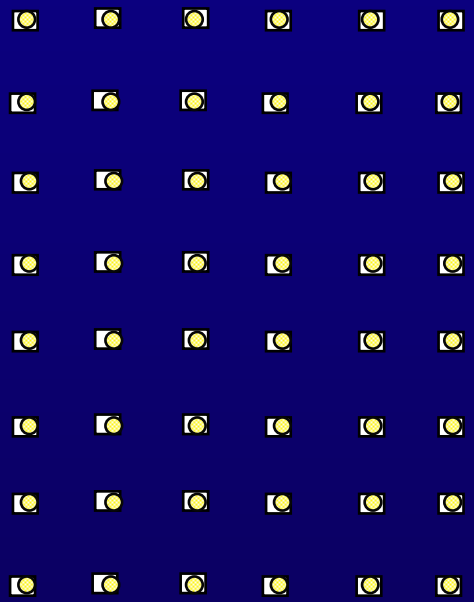
Chrominance Sampling



4:2:0

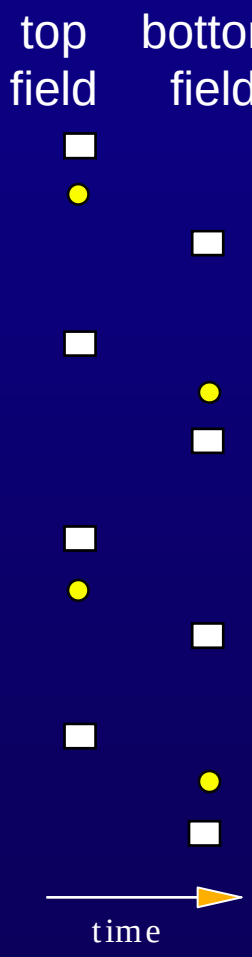


4:2:2

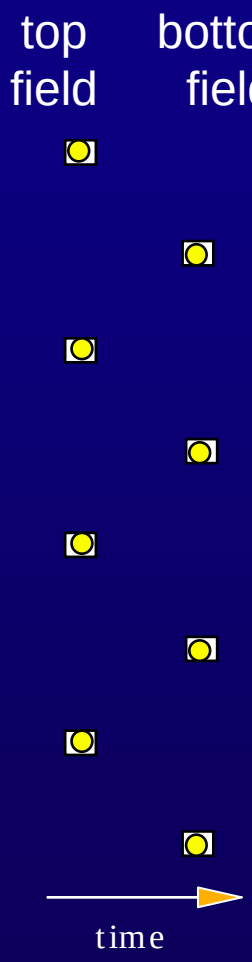


4:4:4

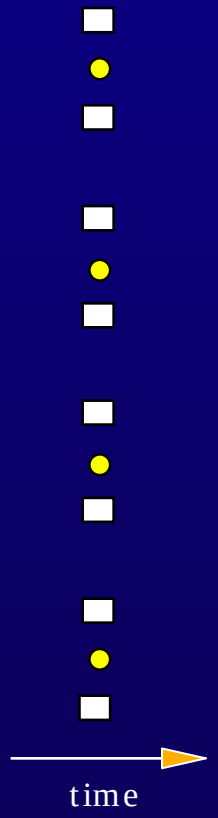
Chrominance Sampling (Cont.)



interlaced 4:2:0



interlaced 4:2:2/4:4:4

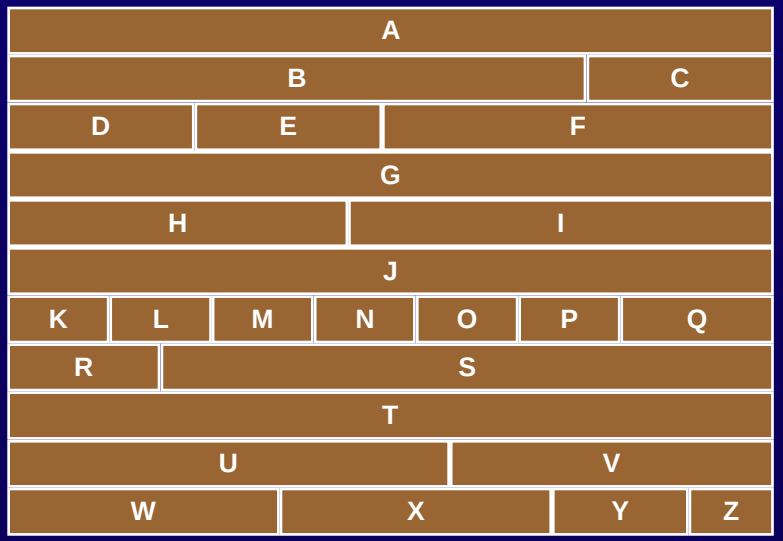


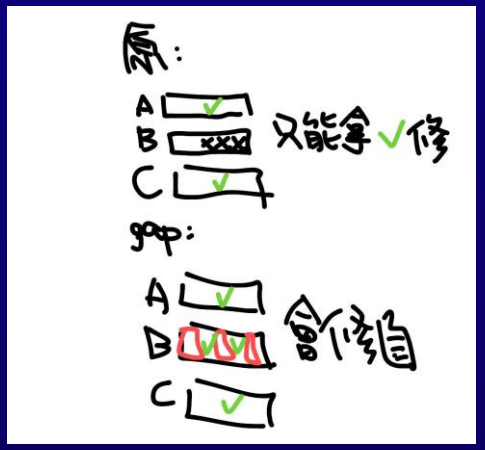
progressive

MPEG-2 Slice Structure

- A slice is a collection of macroblocks in raster scan order.
- Restriction on slice sizes:
  - MPEG-2 restricts a slice to be contained within a row of macroblocks
- MPEG-2 allows gaps between slices in “general slice structure”
- MPEG-2 defines “restricted slice structure,” in which no gaps are allowed. This is used in most Profiles and Levels.

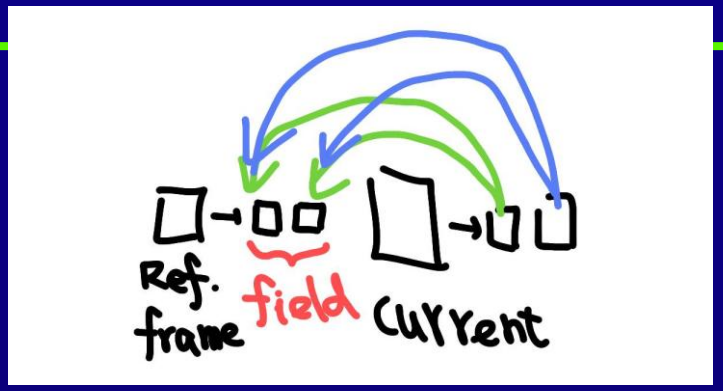
Mpeg1 slice任意，mpeg2則最多只能一整條





Example of restricted slice structure

Interlaced Video Coding



- Frame-pictures or field-pictures
- Motion compensation
  - Frame prediction for frame-pictures (same as MPEG-1)
  - Field prediction for field-pictures
  - Field prediction for frame-pictures
  - Dual-prime
    - Field-pictures or frame-pictures
    - Only for P-pictures
  - 16×8 MC for field pictures

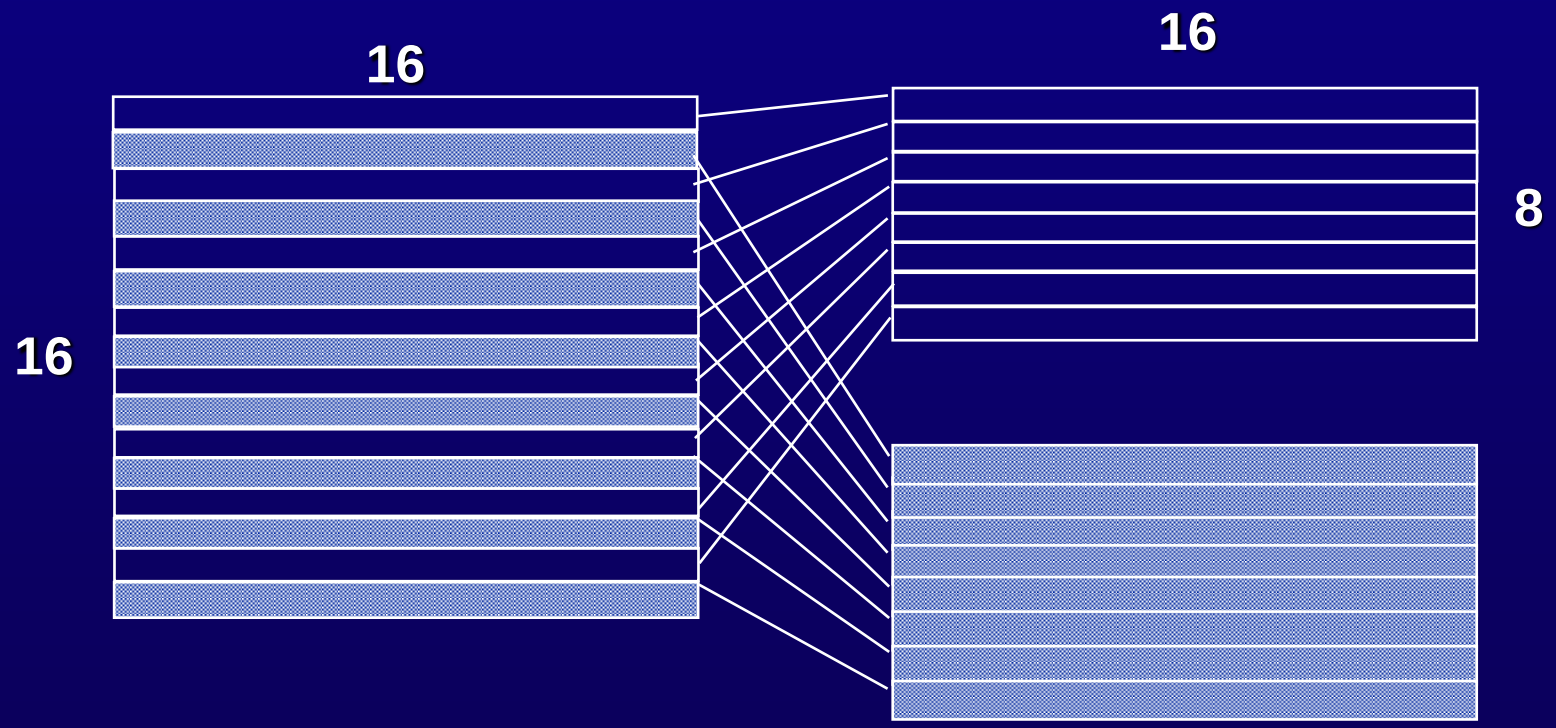
Same except bitstream

Allowable Prediction Modes

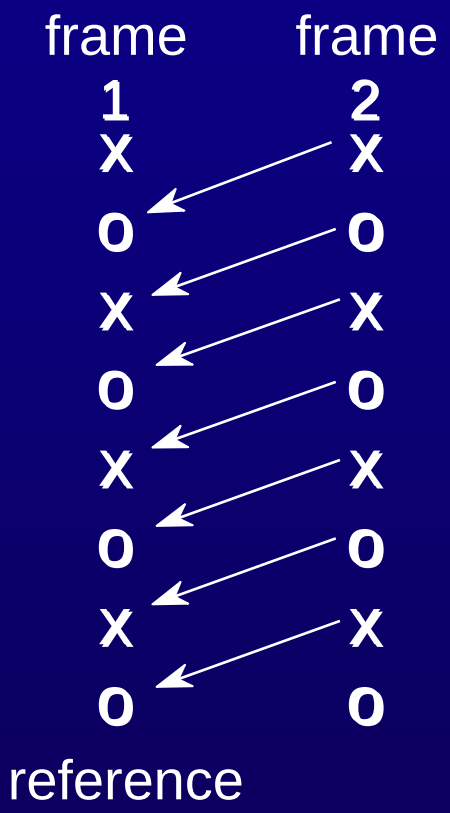
Frame Pictures
Frame Prediction
Field Prediction
Dual Prime

Field Pictures
16x8 Motion Compensation
Field Prediction
Dual Prime

Frame/Field Motion Compensated Predictive Coding

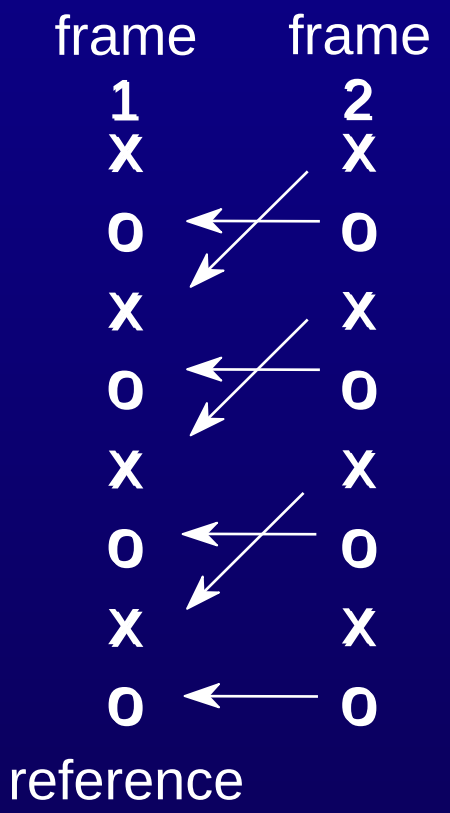


Prediction Modes



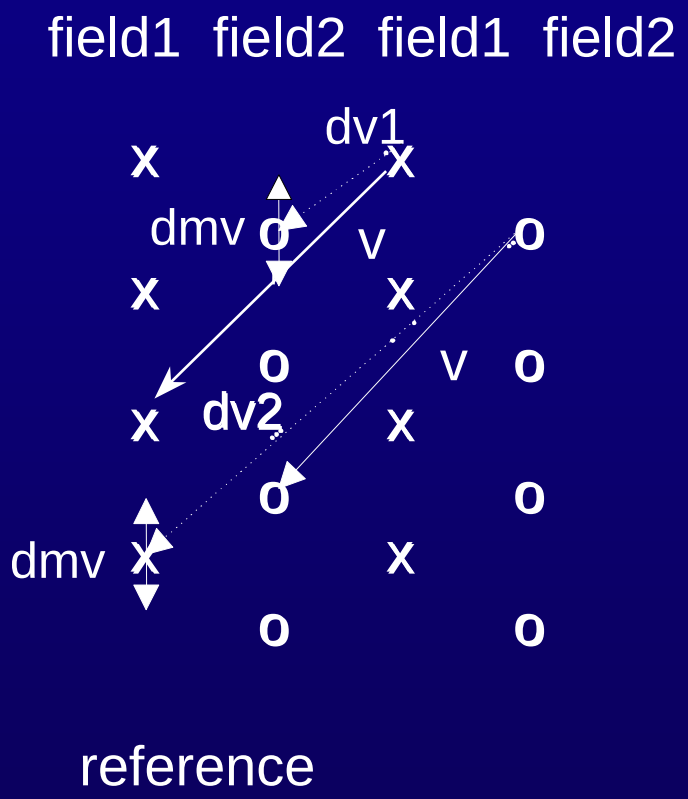
Frame Prediction

(a)



Field Prediction

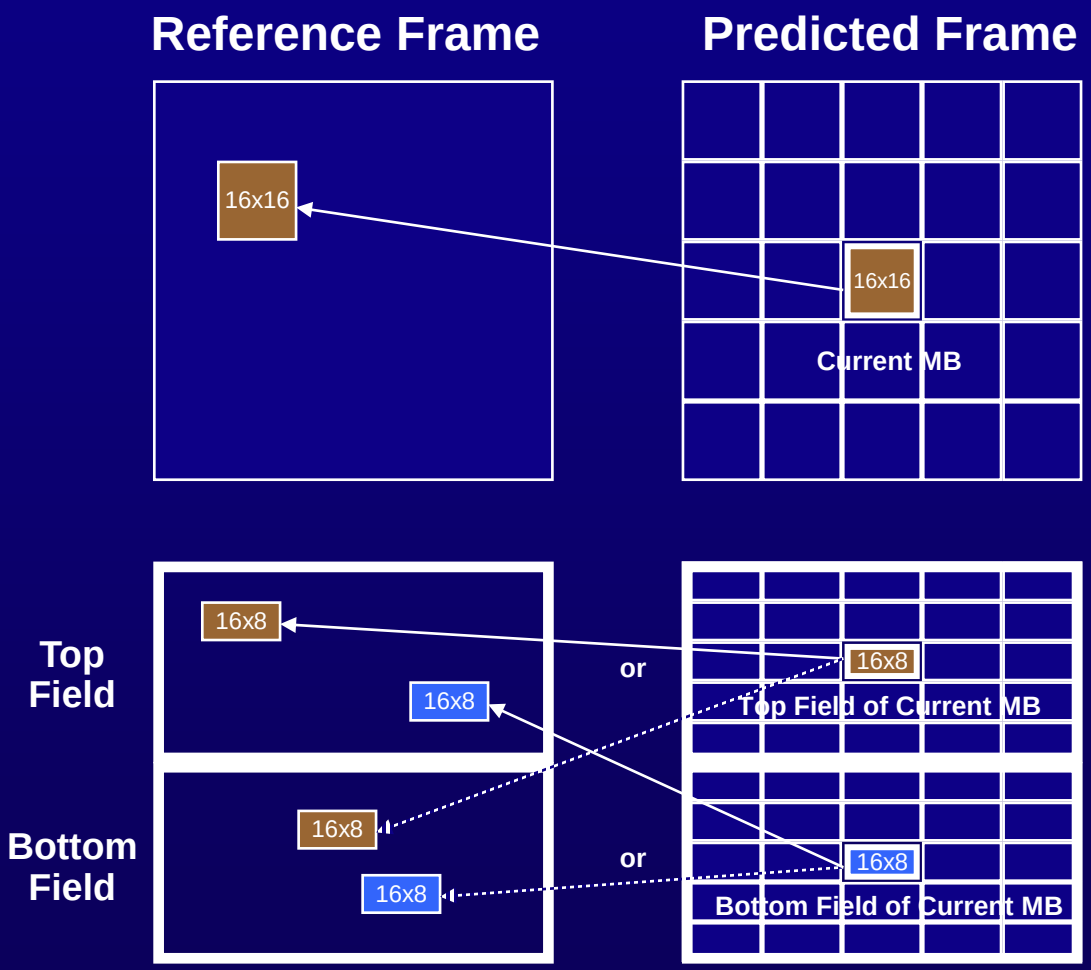
(b)



Dual Prime Prediction

(c)

Frame & Field predictions



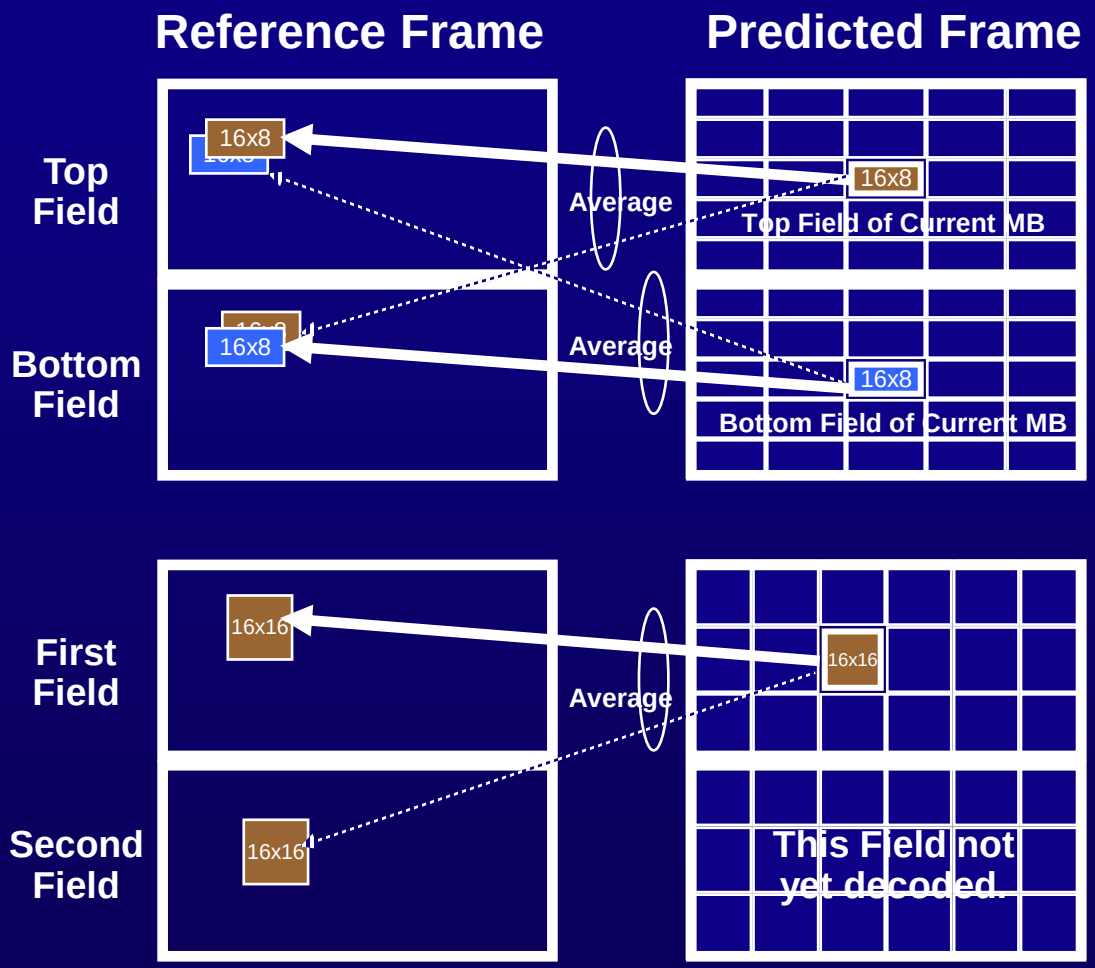
Frame Prediction

Best 16x16 region in reference picture determines frame MV for 16x16 MB. Only mode allowed in MPEG-1.

Field Prediction

Best 16x8 region in top or bottom field in reference picture determines field MV's for top and bottom portions of 16x16 MB.

Dual-Prime Prediction



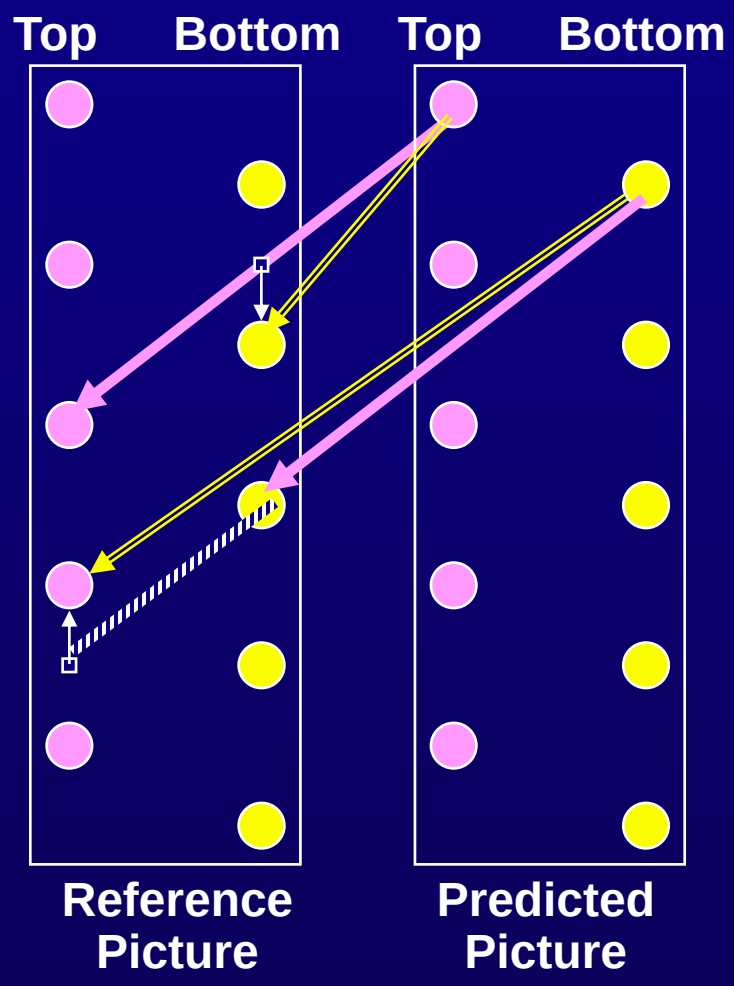
In Frame Pictures

Single MV (heavy arrow) sent in bitstream; this represents predictions from fields of same parity. Small differential MV's are also sent; these represent offset predictions from fields of opposite parity. Same and opposite field predictions are averaged to form final prediction for each 16x8 region of current MB.

In Field Pictures

Single MV (heavy arrow) sent in bitstream; this represents prediction from field of same parity. A small differential MV is also sent; this represents an offset prediction from field of opposite parity. Same and opposite field predictions are averaged to form final prediction for current 16x16 MB.

Dual-Prime Prediction (Cont.)



Vector Transmitted in Bitstream for Same Parity Fields

Differential Vector Transmitted in Bitstream (limited to values -1, 0, +1)

Vector Derived at Decoder for Opposite Parity Fields

Prediction Modes

- PSNR at 4 Mbits/s
  - Frame-pictures, M=1

Sequence	Frame MC	Field MC	Frame/Field MC	Dualp MC	Frame/Field/Dualp MC
Flowergarden	27.72	28.06 (+0.34)	28.22 (+0.50)	28.39 (+0.67)	29.38 (+1.66)
Mobile & Cal	25.69	25.86 (+0.17)	26.04 (+0.35)	25.51 (-0.18)	26.63 (+0.94)
Football	34.20	35.60 (+1.40)	35.69 (+1.49)	35.69 (+1.49)	36.04 (+1.84)
Bus	28.99	30.26 (+1.27)	30.43 (+1.44)	30.70 (+1.71)	31.31 (+2.32)
Carousel	28.67	29.97 (+1.30)	30.07 (+1.40)	29.99 (+1.32)	30.53 (+1.86)

  - Frame-pictures, M=3

Sequence	Frame MC	Field MC	Frame/Field MC
Flowergarden	29.07	29.20 (+0.13)	29.63 (+0.56)
Mobile & Cal	28.11	27.86 (-0.25)	28.27 (+0.16)
Football	34.54	35.01 (+0.47)	35.12 (+0.58)
Bus	30.79	31.32 (+0.53)	31.60 (+0.81)
Carousel	29.22	29.54 (+0.32)	29.73 (+0.51)

  - Field-pictures, M=1

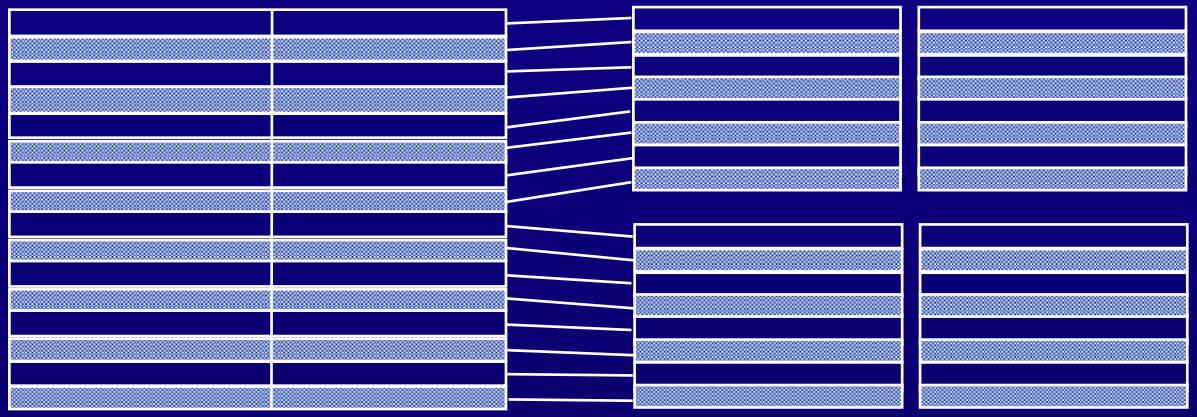
Sequence	Field MC	16x8 MC	Field/16x8 MC
Flowergarden	26.99	25.94 (-1.05)	27.18 (+0.19)
Mobile & Cal	25.02	23.61 (-1.41)	25.21 (+0.19)
Football	36.07	35.07 (-1.00)	35.89 (-0.18)
Bus	29.63	28.76 (-0.87)	29.83 (+0.20)
Carousel	30.31	29.30 (-1.01)	30.29 (+0.12)

Low Delay Coding

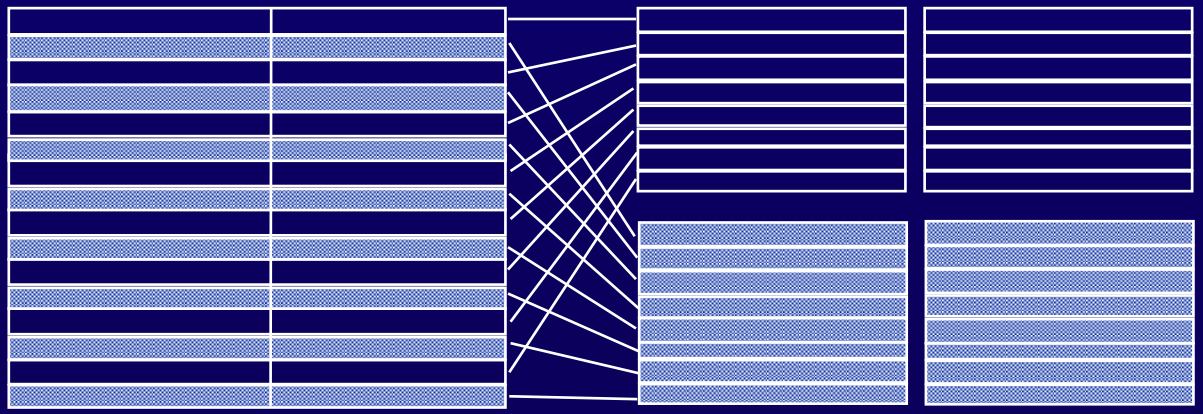
- For face-to-face applications
- Total encoding and decoding delay of less than 150 ms can be achieved
- Low delay coding by not using B-pictures, using dual-prime prediction for P-frames, intra slices, skip frames

Frame/Field DCT

Frame format

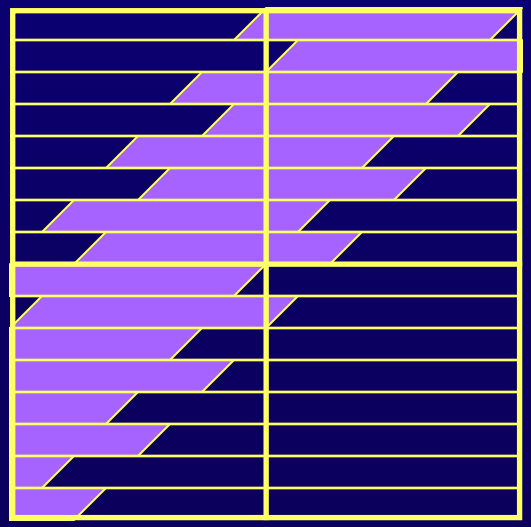


Field format

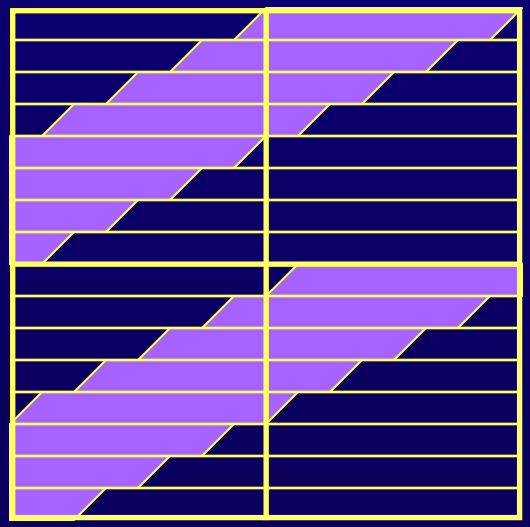


Frame/Field Adaptive DCT

- Organize 16×16 block as frame blocks or filed blocks
- Compute correlation in vertical direction in each case
- Choose the case that has higher correlation



Frame blocks



Field blocks

Frame/Field Adaptive DCT (Cont.)

- PSNR (dB) at 4 Mbits/s
  - M=1

Sequence	Frame DCT	Field DCT	Frame/Field DCT
Flowergarden	29.36	29.04 (-0.32)	29.38 (+0.02)
Mobile & Cal	26.66	25.87 (-0.79)	26.63 (-0.03)
Football	35.54	35.95 (+0.41)	36.04 (+0.50)
Bus	31.05	31.00 (-0.05)	31.31 (+0.26)
Carousel	29.68	30.36 (+0.68)	30.53 (+0.85)

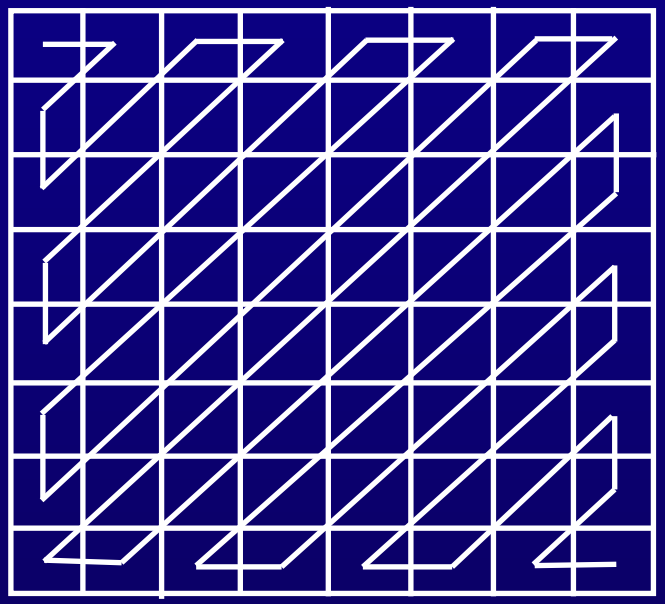
  - M=3

Sequence	Frame DCT	Field DCT	Frame/Field DCT
Flowergarden	29.61	29.46 (-0.15)	29.63 (+0.02)
Mobile & Cal	28.34	27.74 (-0.60)	28.27 (-0.07)
Football	34.67	35.04 (+0.37)	35.12 (+0.45)
Bus	31.34	31.41 (+0.07)	31.60 (+0.26)
Carousel	29.04	29.59 (+0.55)	29.73 (+0.69)

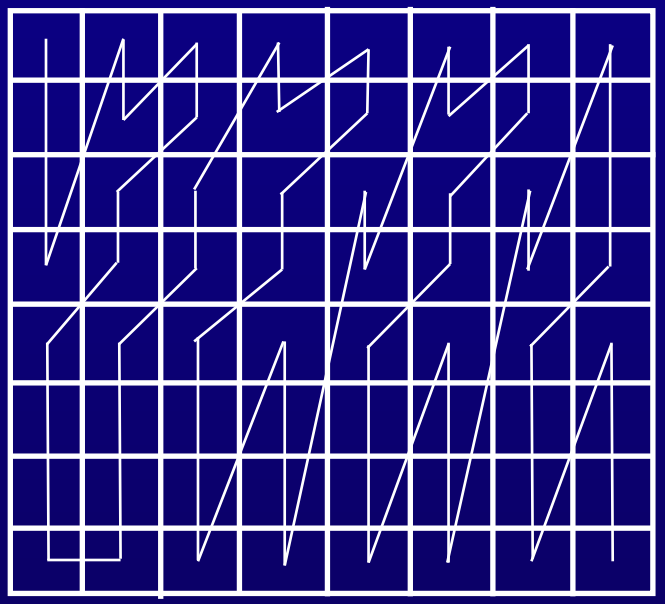
Progressive/Interlaced Scan

因為field DCT這樣做導致關係跑掉了，因此不能再使用zigzag scan

Zigzag (progressive)



Alternate (interlaced)



Sequence	Zigzag Scan (dB)	Alternate Scan (dB)
Flowergarden	29.36	29.61 (+0.25)
Mobile & Cal	28.20	28.24 (+0.04)
Football	34.77	35.07 (+0.30)
Bus	31.35	31.57 (+0.22)
Carousel	29.57	29.68 (+0.11)

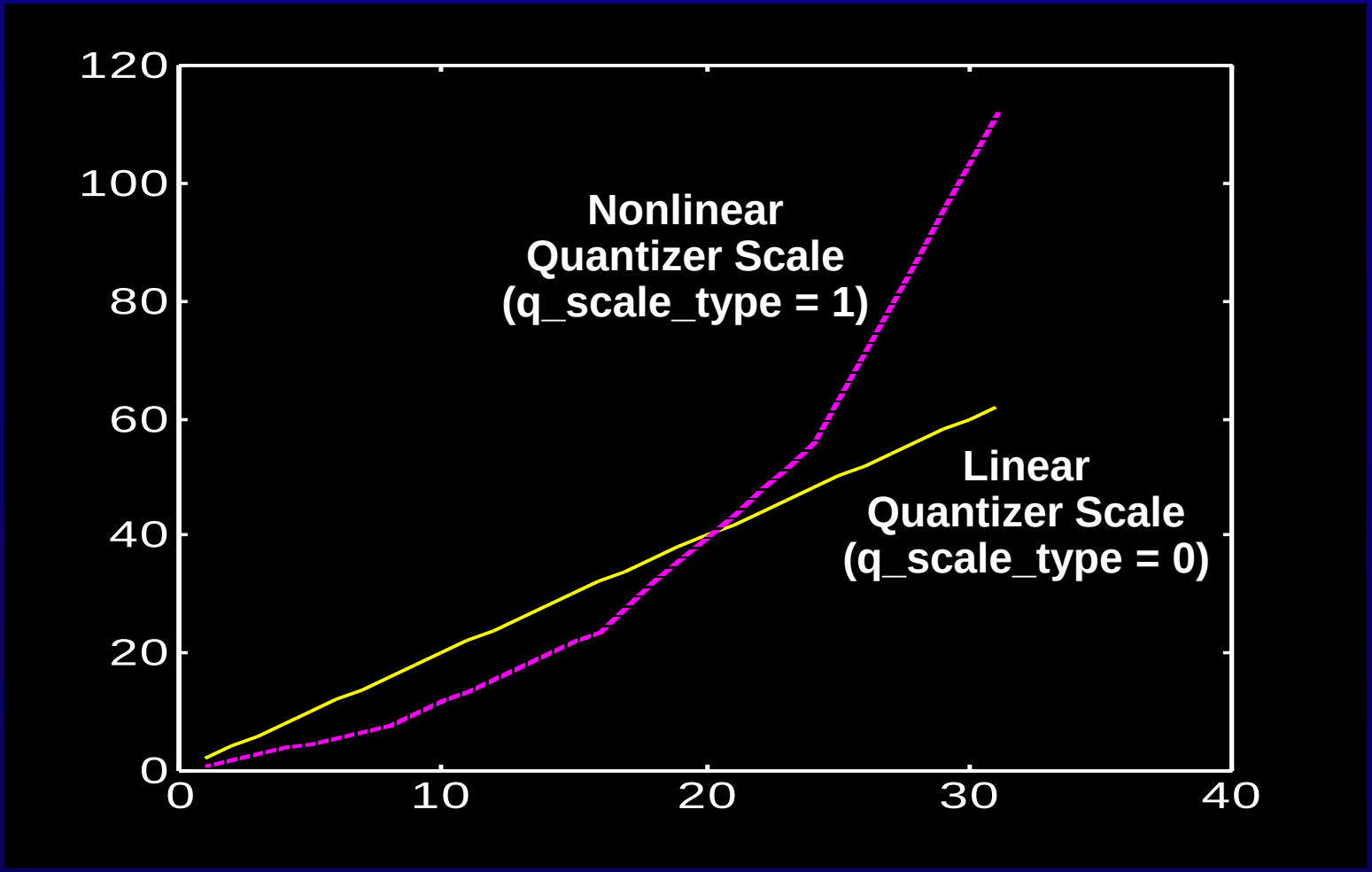
Quantization Tools

- Quantization Matrix (QM)
  - 8x8 matrix can be shaped so that coarser quantization of high spatial frequencies occurs
  - coarser quantization of high spatial frequencies saves bits but causes little or no subjective degradation
  - In MPEG-2, up to four QM's (luma intra/non-intra and chroma intra/non-intra) can be changed at the picture rate
  - Default matrices are specified and need not be sent, but different ones can be downloaded

- Quantizer Scale (QS)
  - QS can change on a macroblock basis
  - rate control's job is to modify QS in a way that keeps picture quality high for a given bit rate

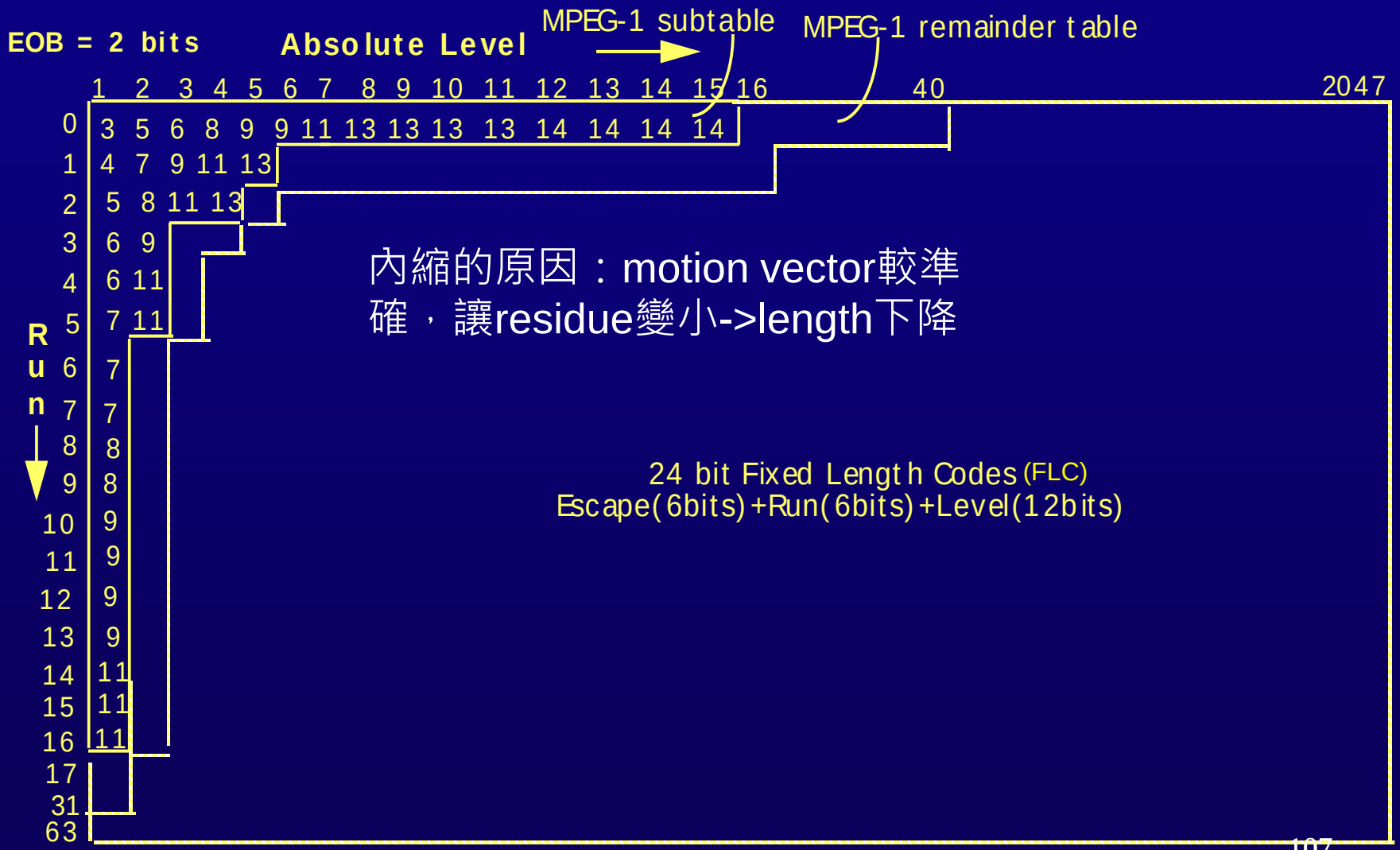
mpeg1 QS都一樣，mpeg2則可以各自不同

MPEG-2 Quantizer Scale Types

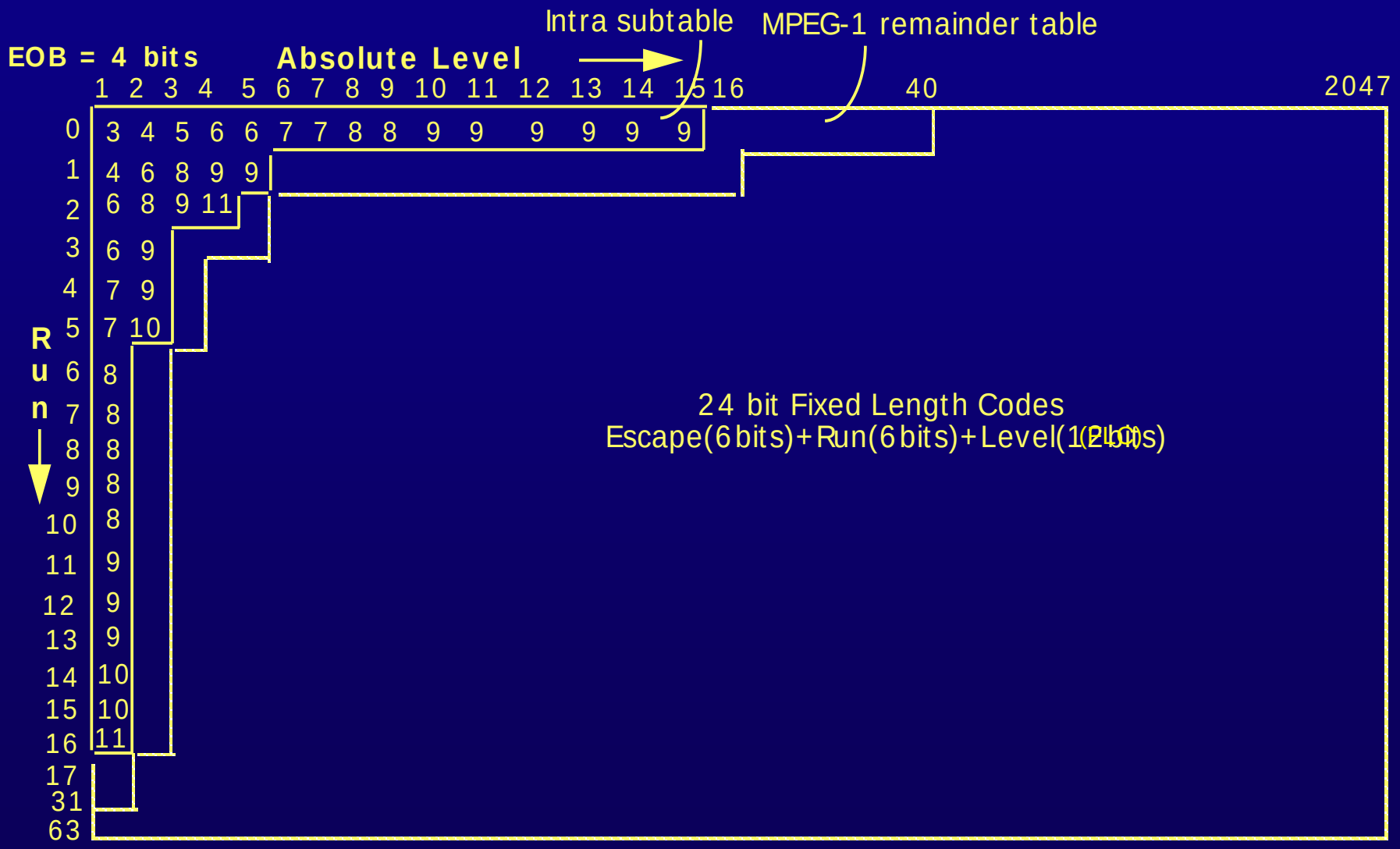


quantizer_scale_code [1, 31]
(sent in bitstream)

2-D VLC (Inter/Intra)



2-D VLC (Intra)


Intra subtable
MPEG-1 remainder table
EOB = 4 bits
Absolute Level
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 40 2047
Run
0 3 4 5 6 6 7 7 8 8 9 9 9 9 9 9
1 4 6 8 9 9
2 6 8 9 11
3 6 9
4 7 9
5 7 10
6 8
7 8
8 8
9 8
10 8
11 9
12 9
13 9
14 10
15 10
16 11
17
31
63
24 bit Fixed Length Codes
Escape(6bits)+Run(6bits)+Level(12bits)

Coding for DC

Range of Differential DC (DIFFs)	SIZE	SIZE VLC Luminance	SIZE VLC Chrominance	VLIs
-2047 to -1024	11	9*1	9*1 1	9*0 00 to 0 9*1 1
-1023 to -512	10	8*1 0	9*1 0	9*0 0 to 0 9*1
-511 to -256	9	7*1 0	8*1 0	9*0 to 0 8*1
-255 to -128	8	6*1 0	7*1 0	8*0 to 0 7*1
-127 to -64	7	5*1 0	6*1 0	7*0 to 0 6*1
-63 to -32	6	4*1 0	5*1 0	6*0to 0 5*1
-31 to -16	5	1110	4*1 0	5*0 to 0 4*1
-15 to -8	4	110	1110	4*0 to 0111
-7 to -4	3	101	110	000 to 011
-3 to -2	2	01	10	00 to 01
-1	1	00	01	0
0	0	100	00	
1	1	00	01	1
2 to 3	2	01	10	10 to 11
4 to 7	3	101	110	100 to 111
8 to 15	4	110	1110	1000 to 4*1
16 to 31	5	1110	4*1 0	1 4*0 to 5*1
32 to 63	6	4*1 0	5*1 0	1 5*0 to 6*1
64 to 127	7	5*1 0	6*1 0	1 6*0 to 7*1
128 to 255	8	6*1 0	7*1 0	1 7*0 to 8*1
256 to 511	9	7*1 0	8*1 0	1 8*0 to 9*1
512 to 1023	10	8*1 0	9*1 0	1 9*0 to 9*1 1
1024 to 2048	11	9*1	9*1 1	1 9*0 0 to 9*1 11

Range larger then MPEG1

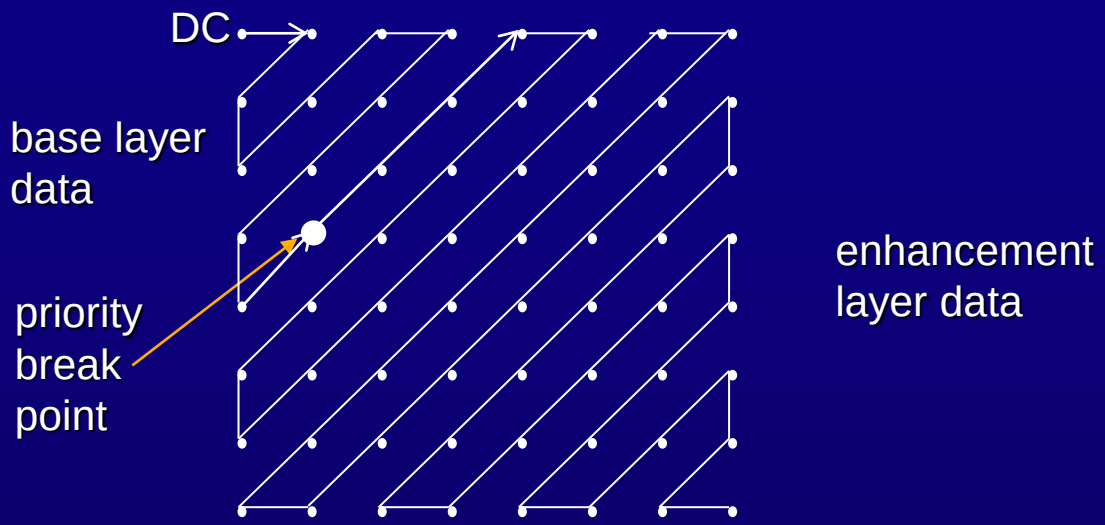
Test Model 5 (TM5)

- Frame/field/dual prime and forward/backward ME
- Integer pel full search followed by half-pel search
- MPEG-1 mode decision: MC/no MC, inter/intra
- MPEG-1 and nonlinear quantizer tables
- Zigzag scan for inter; alternate scan for intra coding
- Quantizer and rate control

Scalability Types

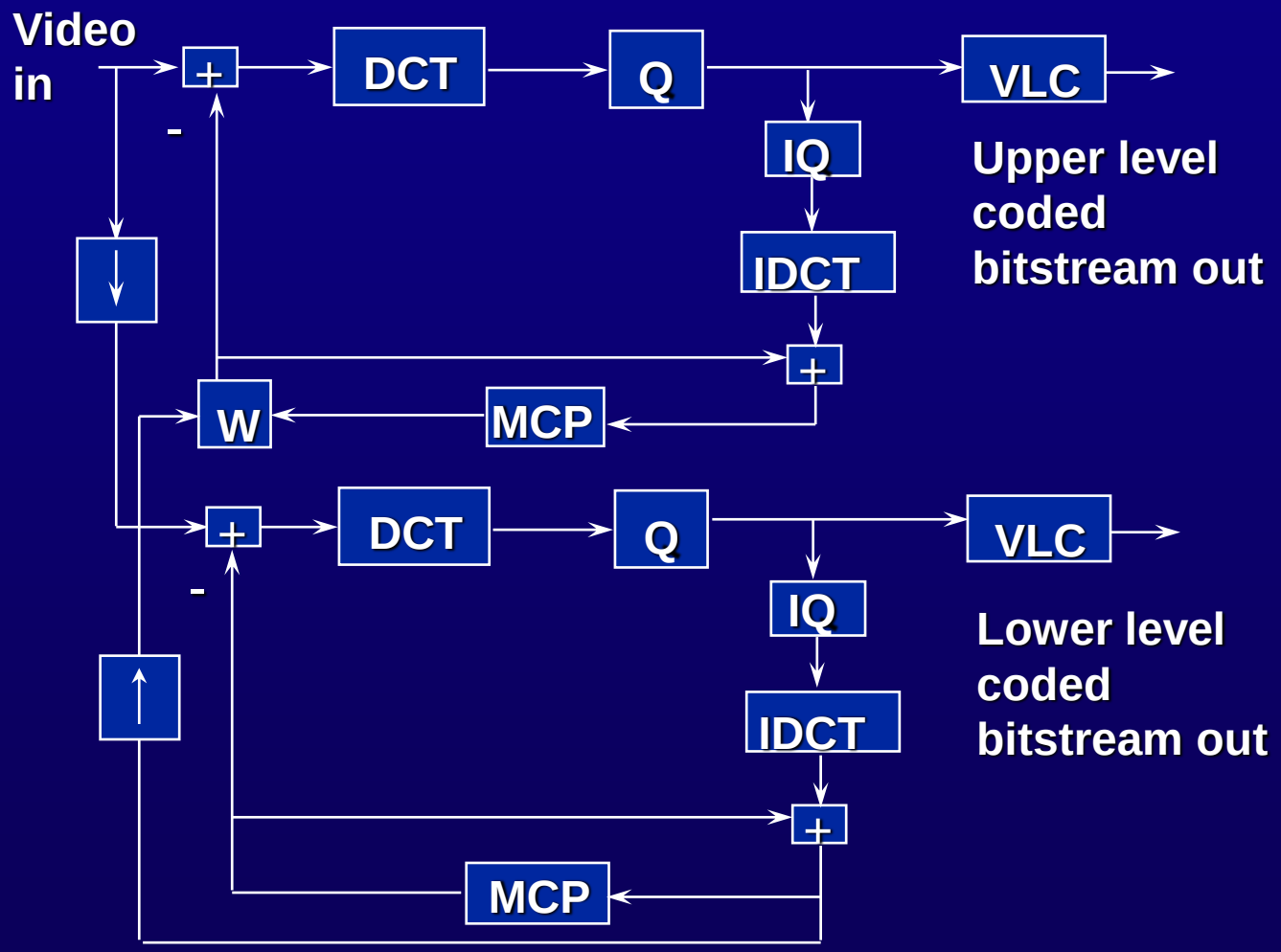
- Data partitioning
- SNR scalability
- Spatial scalability
- Temporal scalability
- Hybrid scalability

Data Partitioning

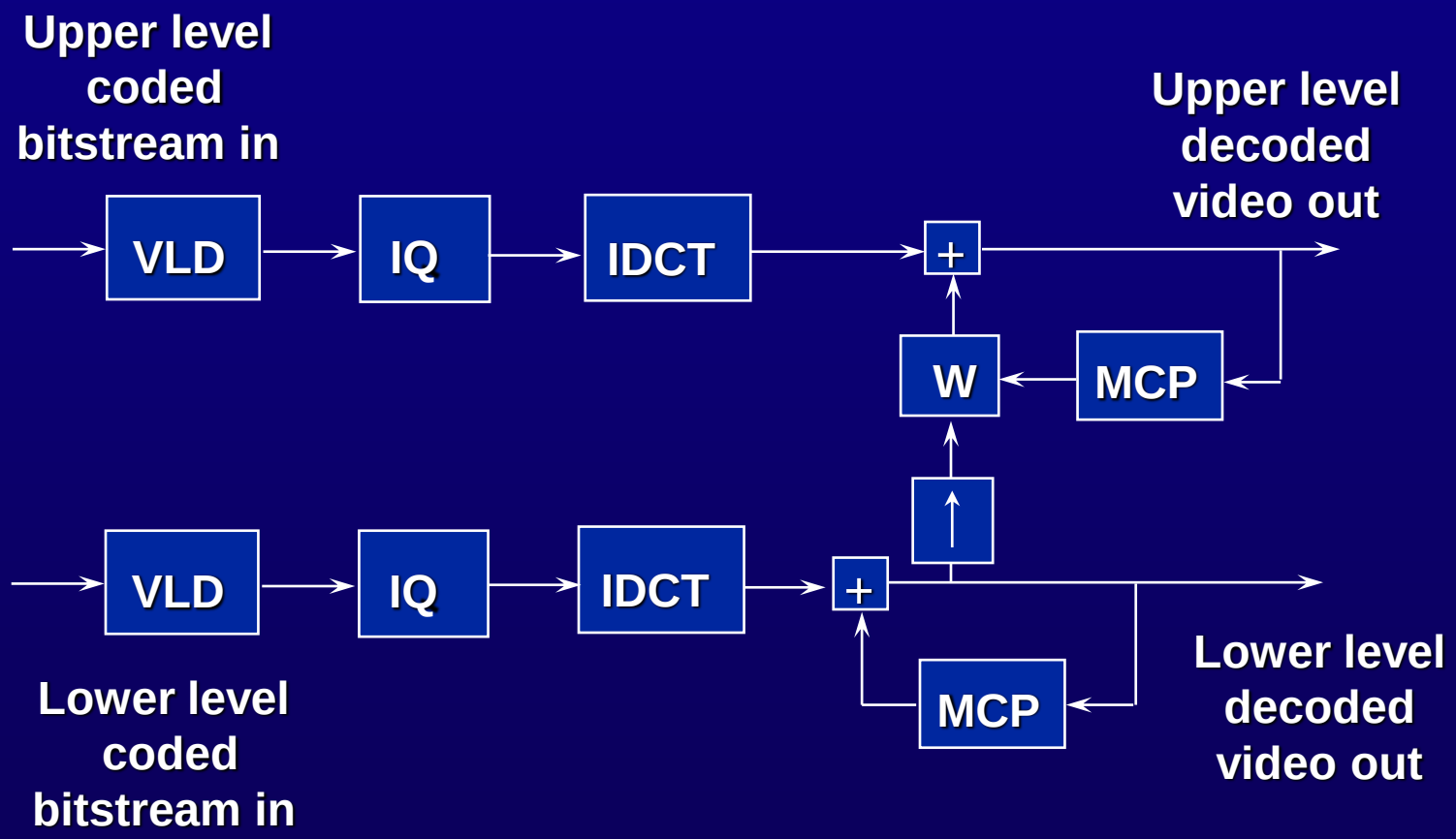


- May cause *picture drift*
- I-pictures can clean up the drift, but cause higher bit rates
- One of the limitations in data partitioning is the need for a high allocated bit rate to the base layer to avoid “blockiness”
- The simplest kind of scalability, has no extra complexity over the nonscalable encoder

Spatial Scalable Encoder

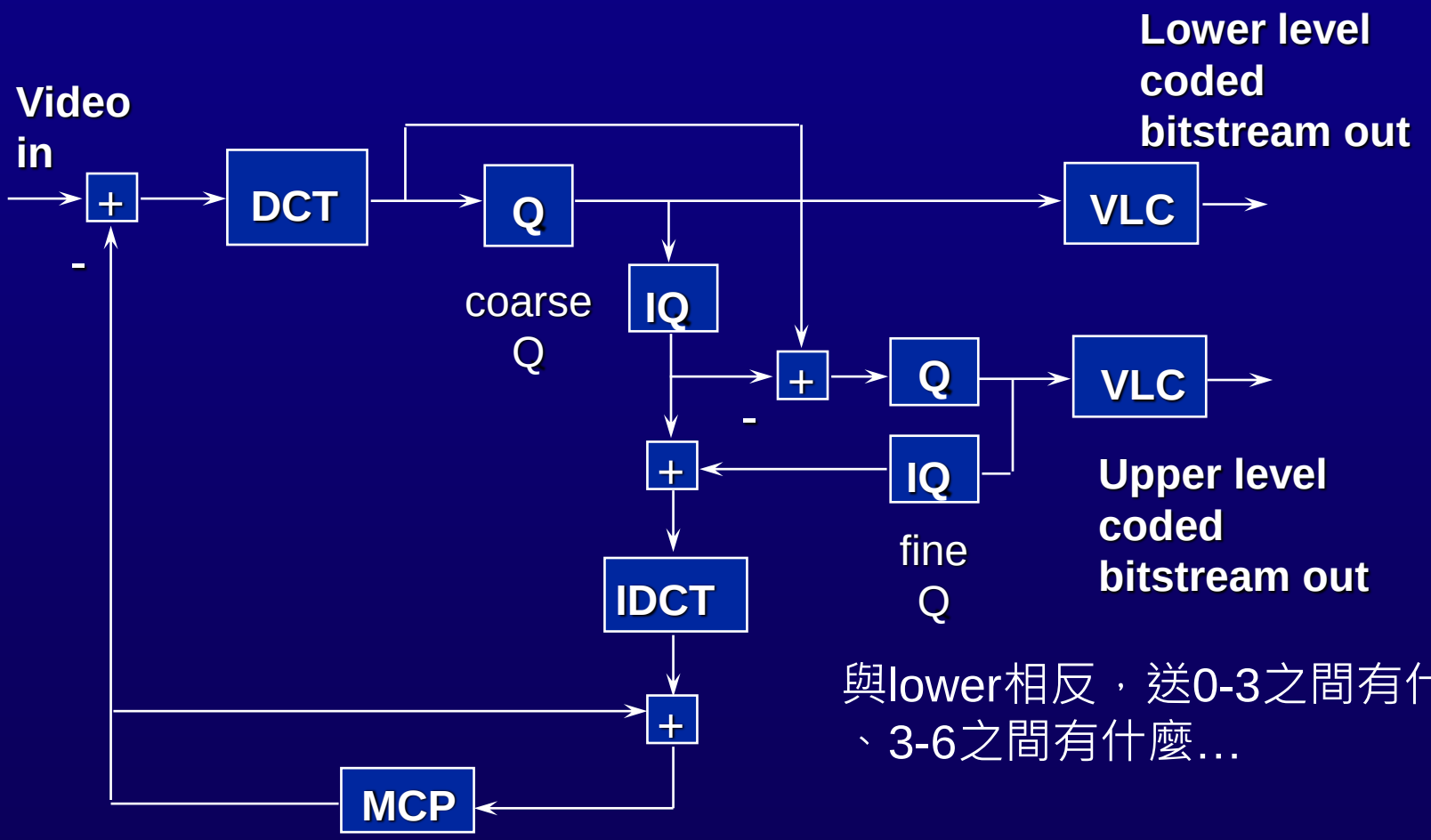


Spatial Scalable Decoder



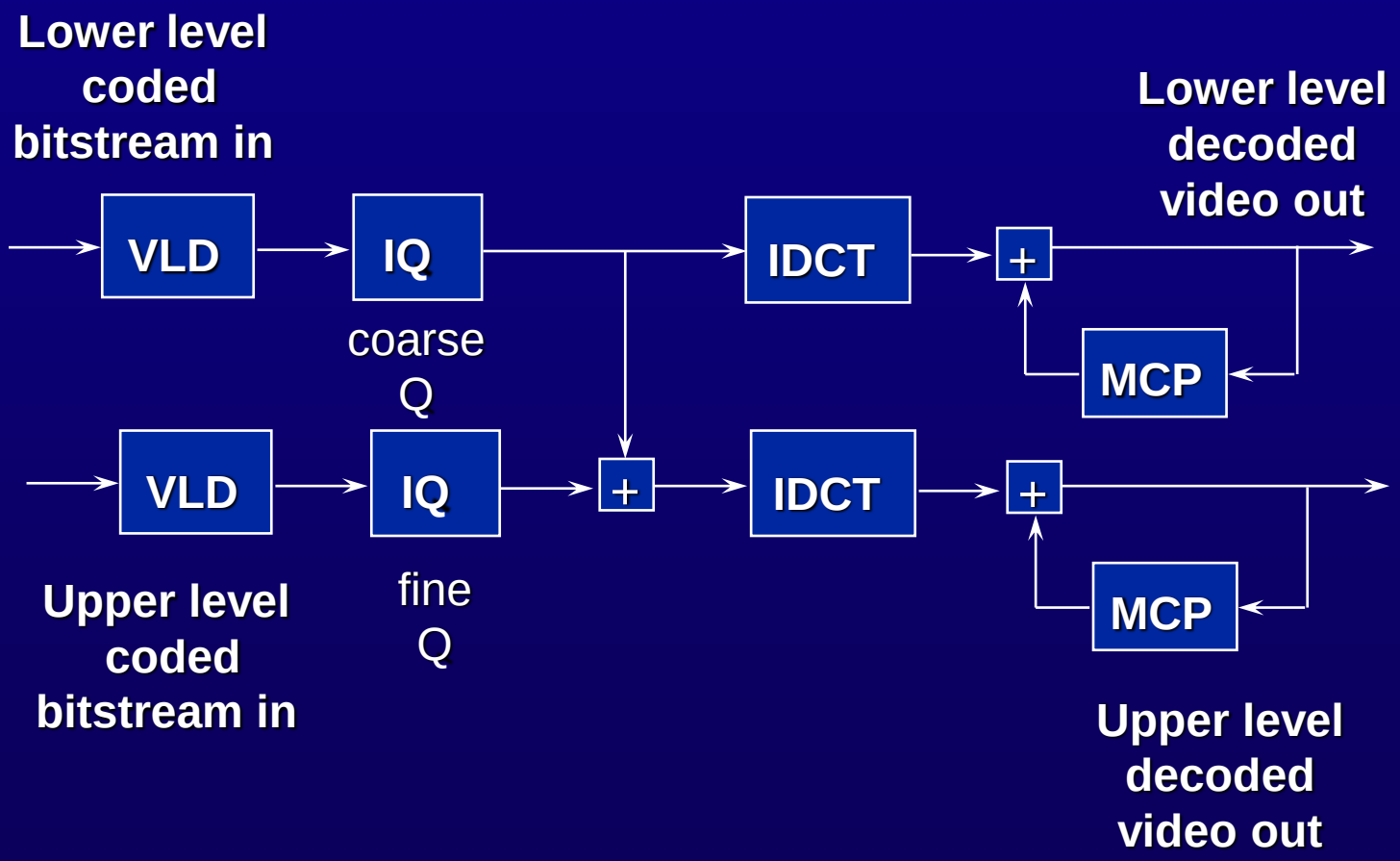
SNR Scalable Encoder

QP大->粗糙(例如先送0 3 6 9 12 15…)



與lower相反，送0-3之間有什麼、3-6之間有什麼…

SNR Scalable Decoder



Applications of Scalability

- Data partitioning (simplest):
  - Video over low-loss networks (e.g., ATM with congestion control)
- SNR scalability:
  - Transmission of video at different qualities
    - multiquality video, video on demand, broadcasting of TV and enhanced TV
  - Video over networks with high error or packet loss rates
    - the Internet,
    - heavily congested ATM networks

Applications of Scalability

- Spatial scalability (most complex):
  - Interworking between two different standard video codecs or heterogeneous data networks
  - Simulcasting of drift-free, good-quality video at two spatial resolutions, such as standard TV and HDTV
  - Distribution of video over computer networks
  - Video browsing
  - Reception of good quality low spatial resolution pictures over mobile networks
  - Similar to other scalable coders, transmission of error resilient video over packet networks.

Error Resilience

- Slice structure
- Concealment motion vectors
- Data partition
- SNR scalability
- Spatial scalability
- Temporal scalability
- Intra pictures
- Intra slices

送1 3 5 7 9...，中間用平均重建
當然有誤差->enhancement時是用
residue送(例如送2 4 6 8 10等等)

MPEG Average Quality

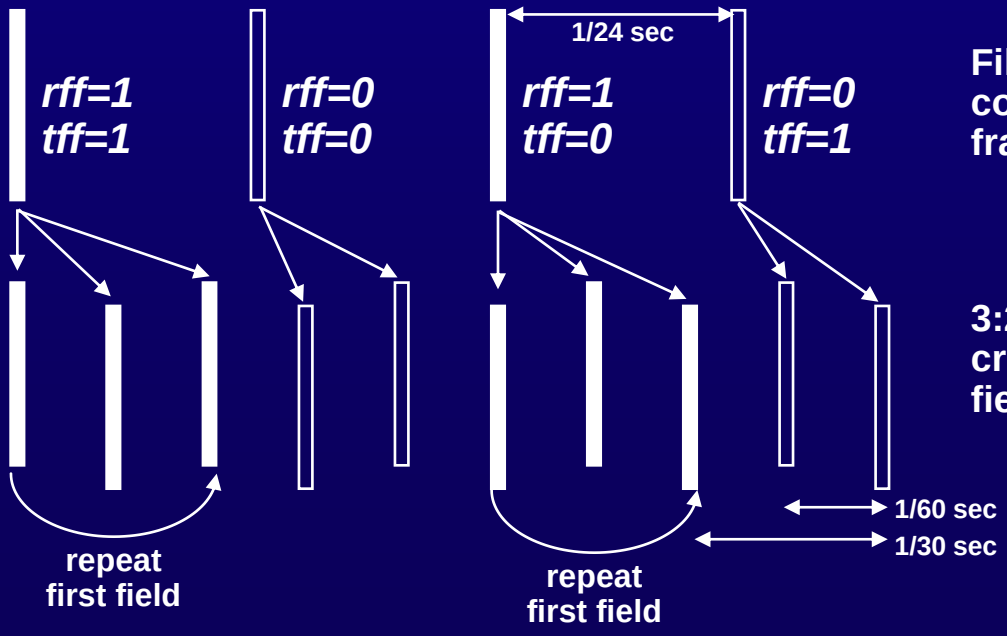
Bit Rate (Mbits/sec)	SIF-30 ~CVGA	CCIR 601 29.97 FPS ~VGA	HDTV 29.97 FPS	HDTV 60 FPS ~SVGA
1.1 Mbs	good	poor		
4.0 Mbs	excellent	good		
9.0 Mbs	excellent++	excellent		
18.0 Mbs		excellent++	good	good
28.0 Mbs			excellent	excellent

	SIF-30 ~CGA	CCIR 601 29.97 FPS ~VGA	HDTV 29.97 FPS	HDTV 60 FPS ~SVGA
Pels	352	704	1920	1280
Lines	240	480	1080	720
Uncompressed Bit Rates (Mbps)	30.4	121.5	745.7	663.6

3:2 Pull Down

解決電影(24)到電視(30)的frame rate問題

- MPEG-2 provides a mechanism for film-originated content to be coded at 24 frame/sec but displayed at 30 frames/sec
- The lower frame rate of film means it can be coded at the same quality as 30 frame/sec video, but at a lower bit rate.
- The repeat_first_field (rff) and top_field_first (tff) flags allow decoders to recreate the 3:2 pulldown sequence for display.

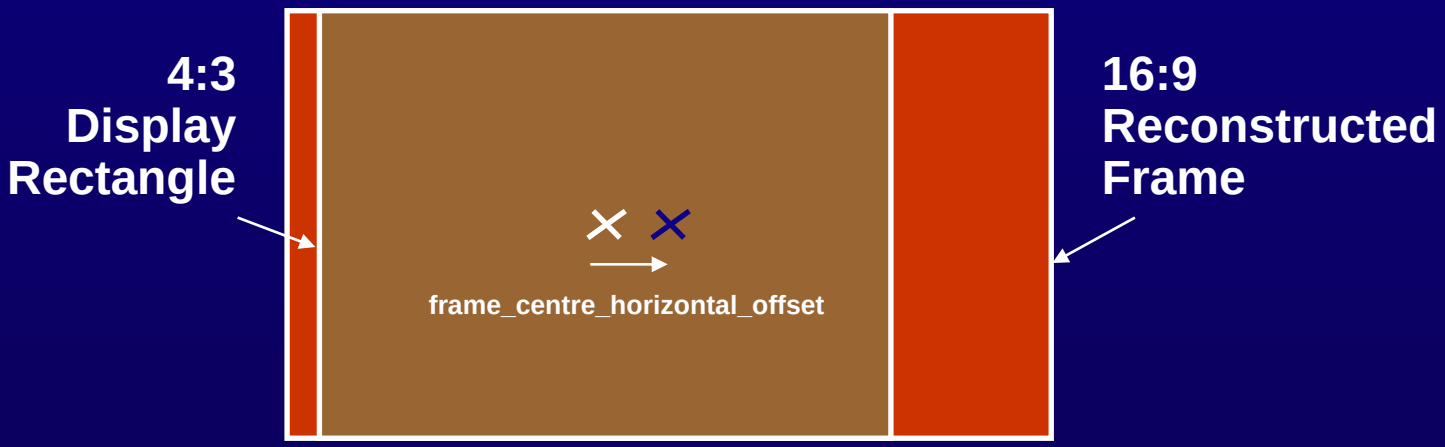


Film Frames coded as progressive frames at 24 frames/sec

3:2 pulldown alternately creates 3 and 2 displayed fields for each input frame

Pan-and-Scan

- MPEG-2 provides a mechanism for panning a display rectangle around a reconstructed frame
- Horizontal and vertical offsets are specified to 1/16 pixel resolution and can be sent for every displayed field.
- This allows widescreen material to be viewed on 4:3 displays.



In this example the horizontal frame center offset is a positive number.

H.263 & H.263+

H.263: ITU-T Very Low Bit Rate Video Coding

- ITU-T SG16/LBC Near Term
- Optimized at bitrate < 22 kb/s (overall < 28.8 kb/s)
- TMN5:
  - 3-4 dB higher PSNR than H.261 at < 64kb/s for all ITU test sequences
  - 30% saving compared with MPEG1 SM3 at 512 kb/s for “football” at CIF resolution

Picture Formats Supported

Picture format	Luminance pixels	Luminance lines	H.261 support	H.263 support	Uncompressed bitrate(Mbps)			
					10 frames/sec		30 frames/sec	
					Mono	Color	Mono	Color
SQCIF	128	96	No	Yes	1.0	1.5	3.0	4.4
QCIF	176	144	Yes	Yes	2.0	3.0	6.1	9.1
CIF	352	288	Optional	Optional	8.1	12.2	24.3	36.5
4CIF	704	576	No	Optional	32.4	48.7	97.3	146.0
16CIF	1408	1152	No	Optional	129.8	194.6	389.3	583.9

H.263 Input Formats

- Sub-QCIF (128x96), QCIF, CIF, 4CIF (704x576), 16 CIF (1408x1152)
- Encoder shall be able to support either sub-QCIF or QCIF
  Decoder shall be able to support both sub-QCIF and QCIF

H.263: Main Features wrt H.261

- **Half-pixel motion estimation (range -16 to 15.5) with motion vector prediction**
- **Four negotiable options:**

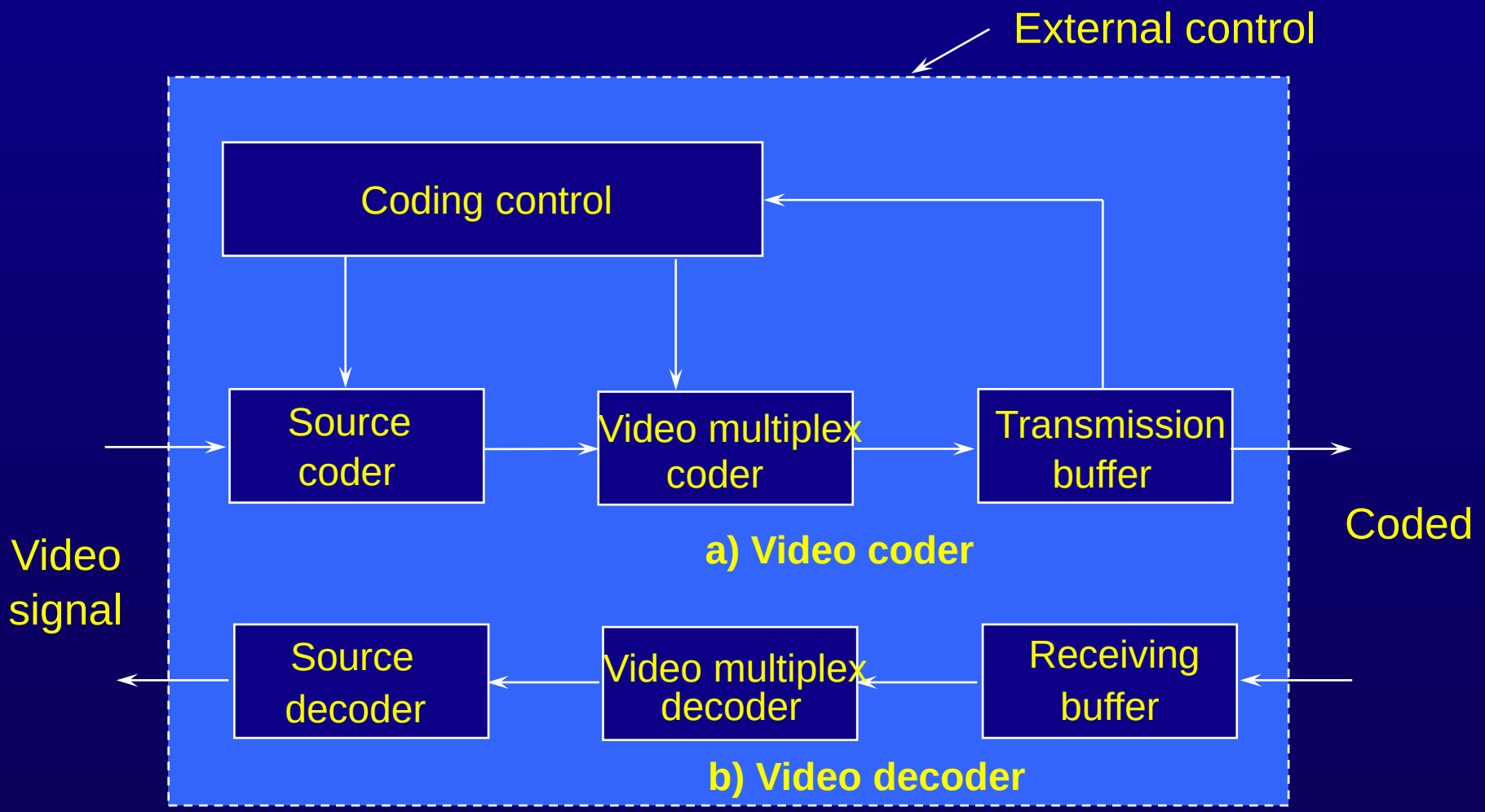
  - Unrestricted motion vector: motion vectors can point outside the picture, maximum range for motion vectors: -31,5 to 31.5 instead of -16 to 15.5

  - Advanced prediction mode: 8x8 motion vectors, overlapped block motion compensation (OBMC), motion vector can point outside the picture

  - Syntax-based arithmetic coding (about 5% decrease in bit-rate)

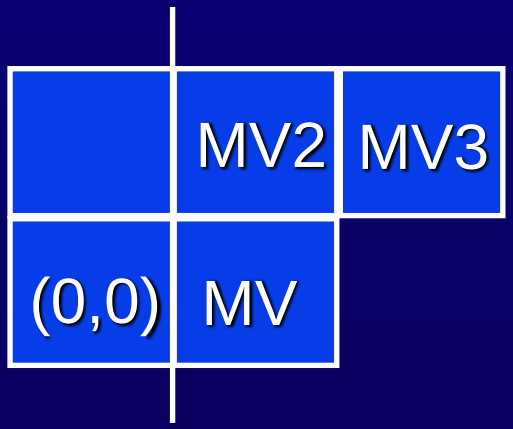
  - PB-frame mode (similar to dual-prime motion estimation)
- **Reduced overhead: optional GOB header and BCH, 3-D VLC**
- **Different VLC tables at macroblock and block levels**

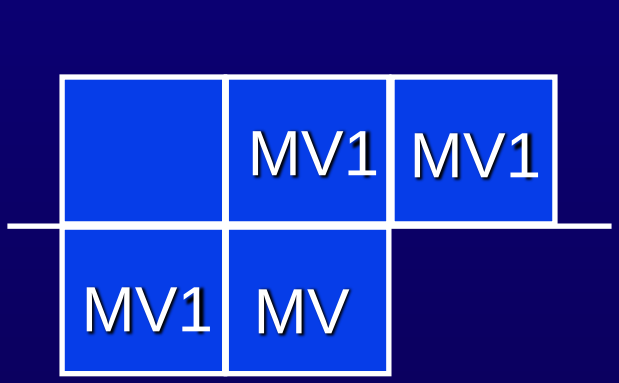
Block Diagram of Video Codec

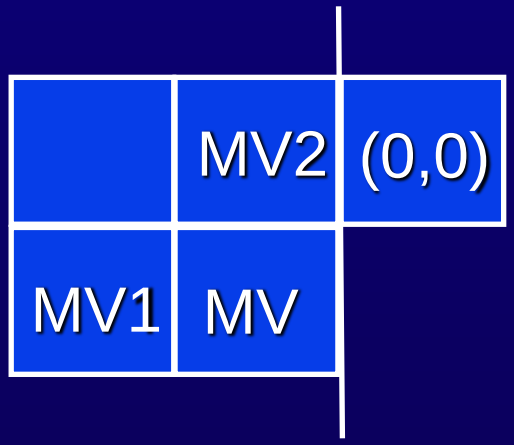


Differential Motion Vector

- $MVD_x = MV_X - P_X$
- $MVD_Y = MV_Y - P_Y$
- P_X=Median(MV_{1X},MV_{2X},MV_{3X})
- P_Y=Median(MV_{1Y},MV_{2Y},MV_{3Y})
- $P_X = P_Y = 0$ if MB is INTRA coding



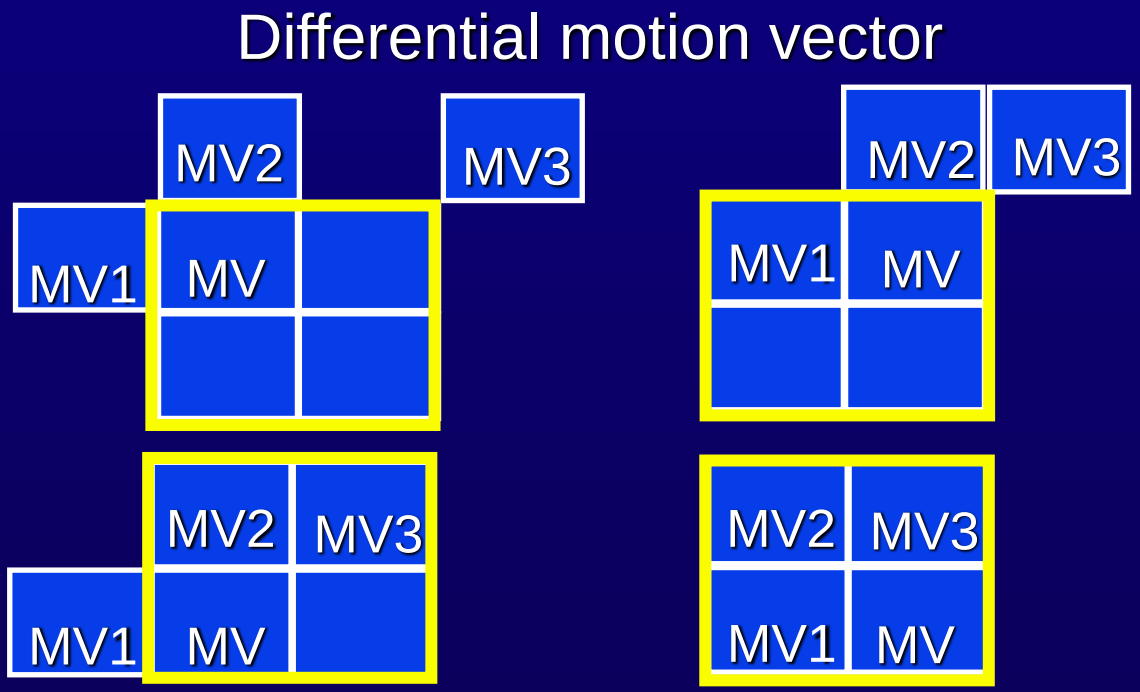




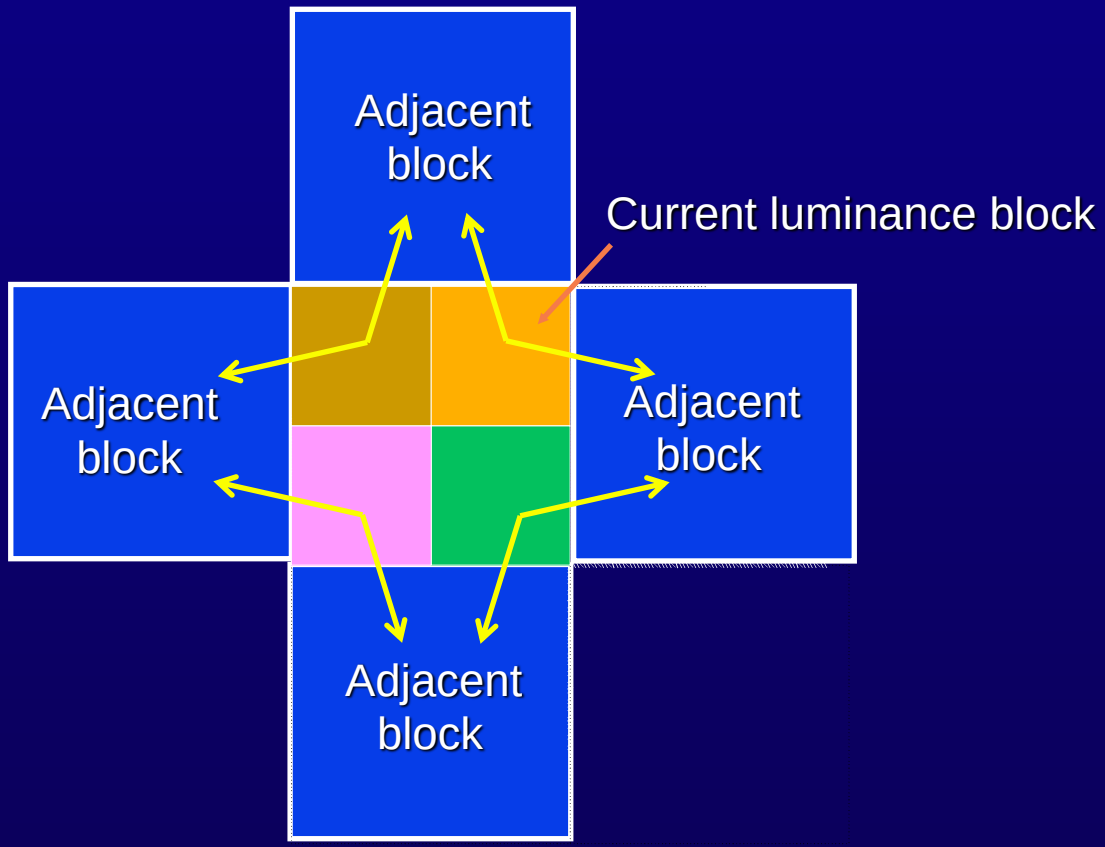
: Picture or GOB border

Advanced Prediction Mode

- Four 8x8 vectors instead of one 16x16 vector



Advanced Prediction Mode (Cont.)



Advanced Prediction Mode (Cont.)

- Overlapped block motion compensation for P-frames: each pixel in a prediction block is a weighted sum of three prediction values (blocks pointed by motion vectors of current, top/bottom, left/right)
- Motion vectors can point outside the picture

weighting

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

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

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

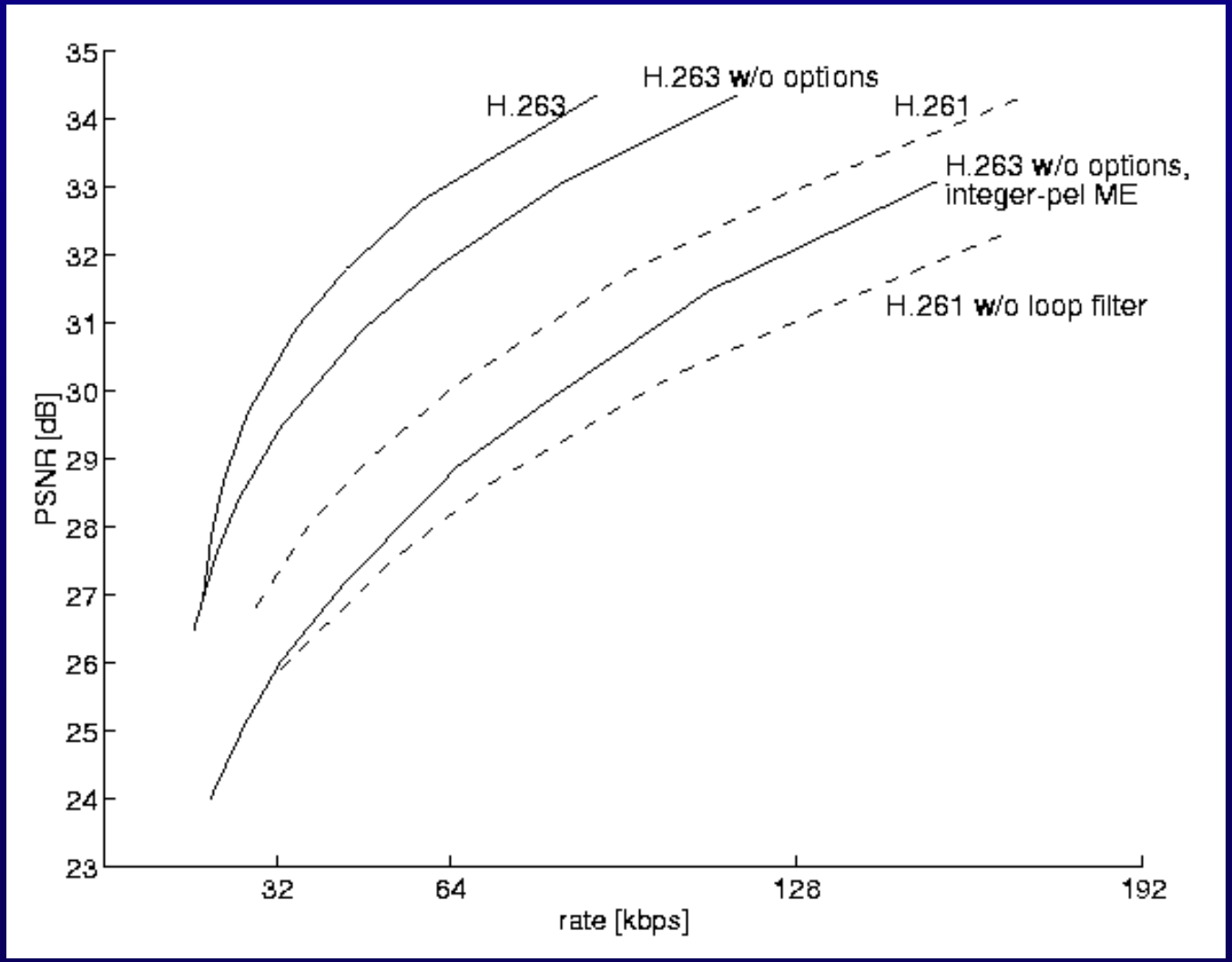
Syntax-Based Arithmetic Coding (SAC) Mode

- Arithmetic coding is used instead of VLC
- Cumulative frequencies are provided
- About 5% less bits at the same SNR

TMN6

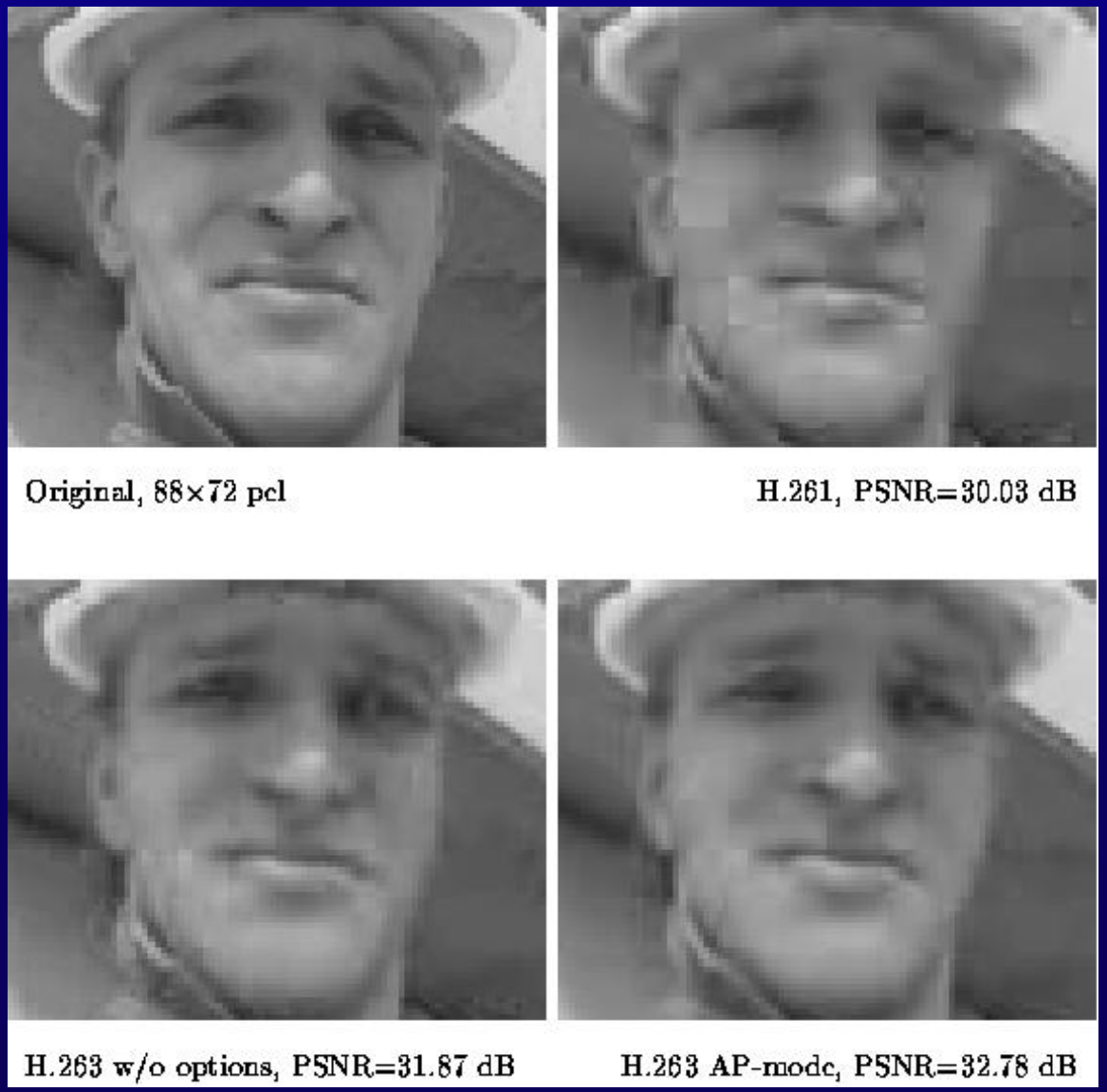
- Test model near-term 6
  - Advanced motion prediction (AP)
  - Half-pel motion accuracy
  - Overlapped block motion compensation
  - Adaptive motion vectors for 8x8 or 16x16
  - Syntax-adaptive arithmetic coding (SAC)
  - PB frames

Performance Comparison of H.263 & H.261



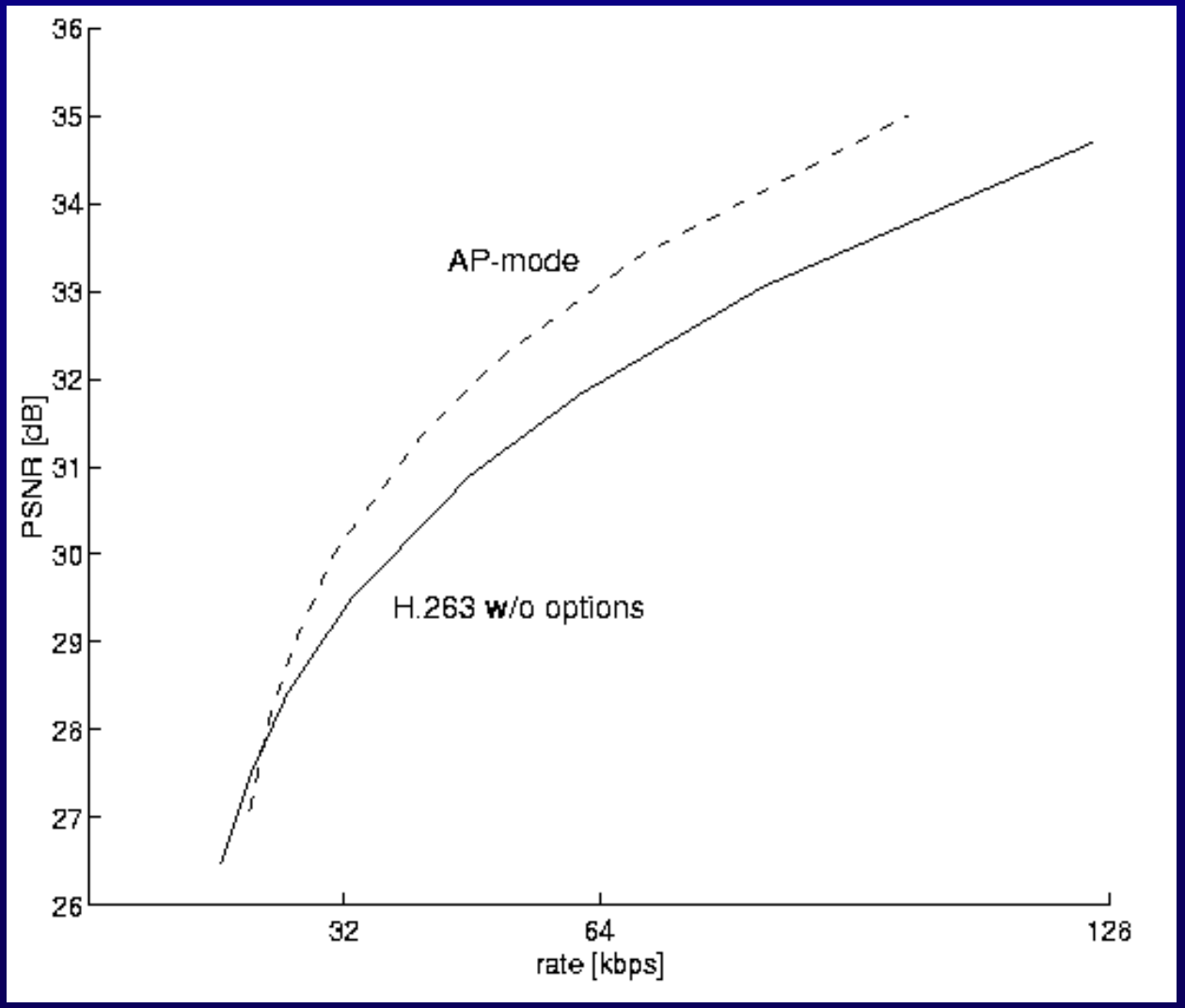
Foreman at 12.5 fps

Performance Comparison of H.263 & H.261 (Cont.)



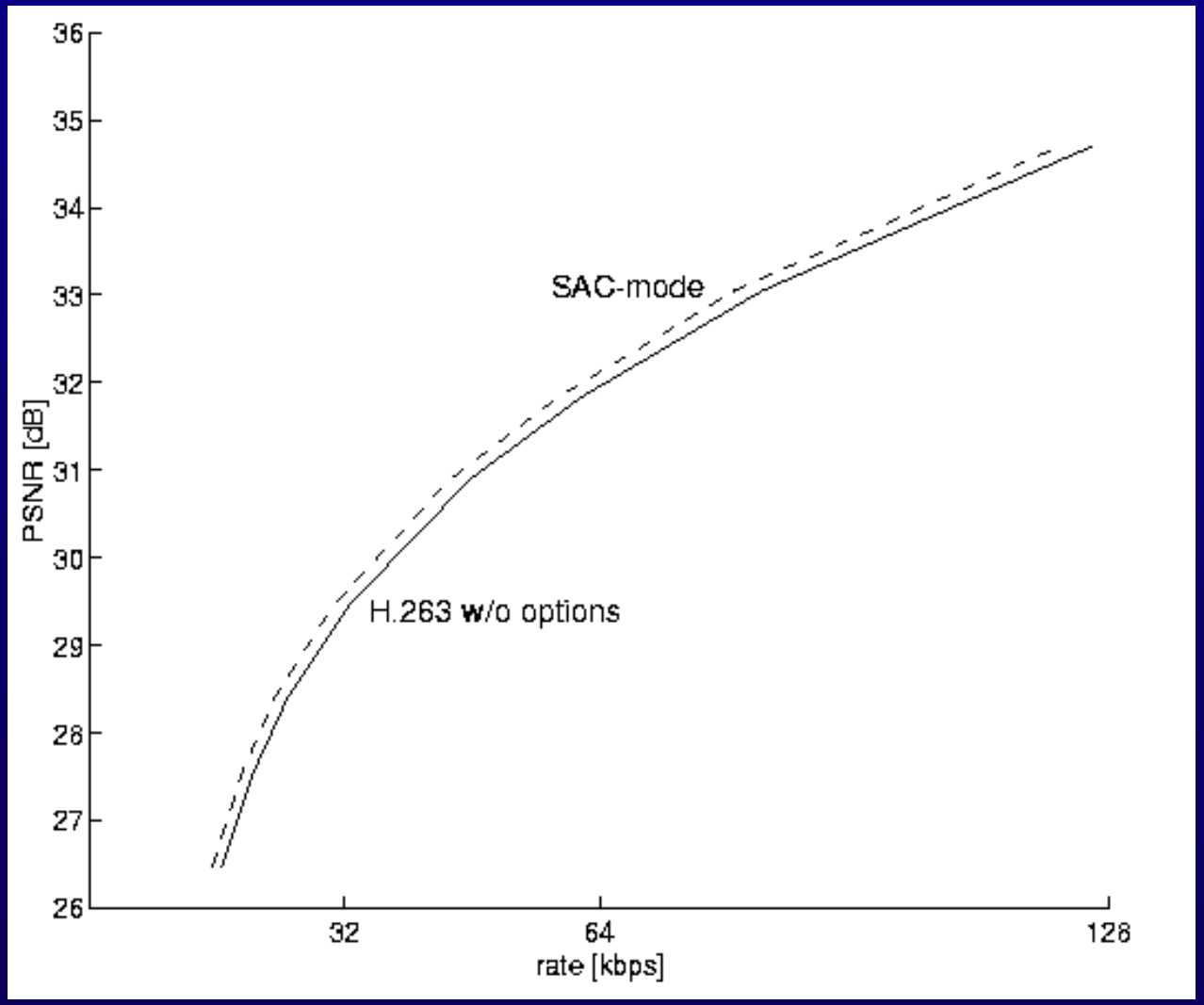
Foreman
at 12.5 fps
64 kbps

Performance Comparison of H.263 w/wo AP Mode



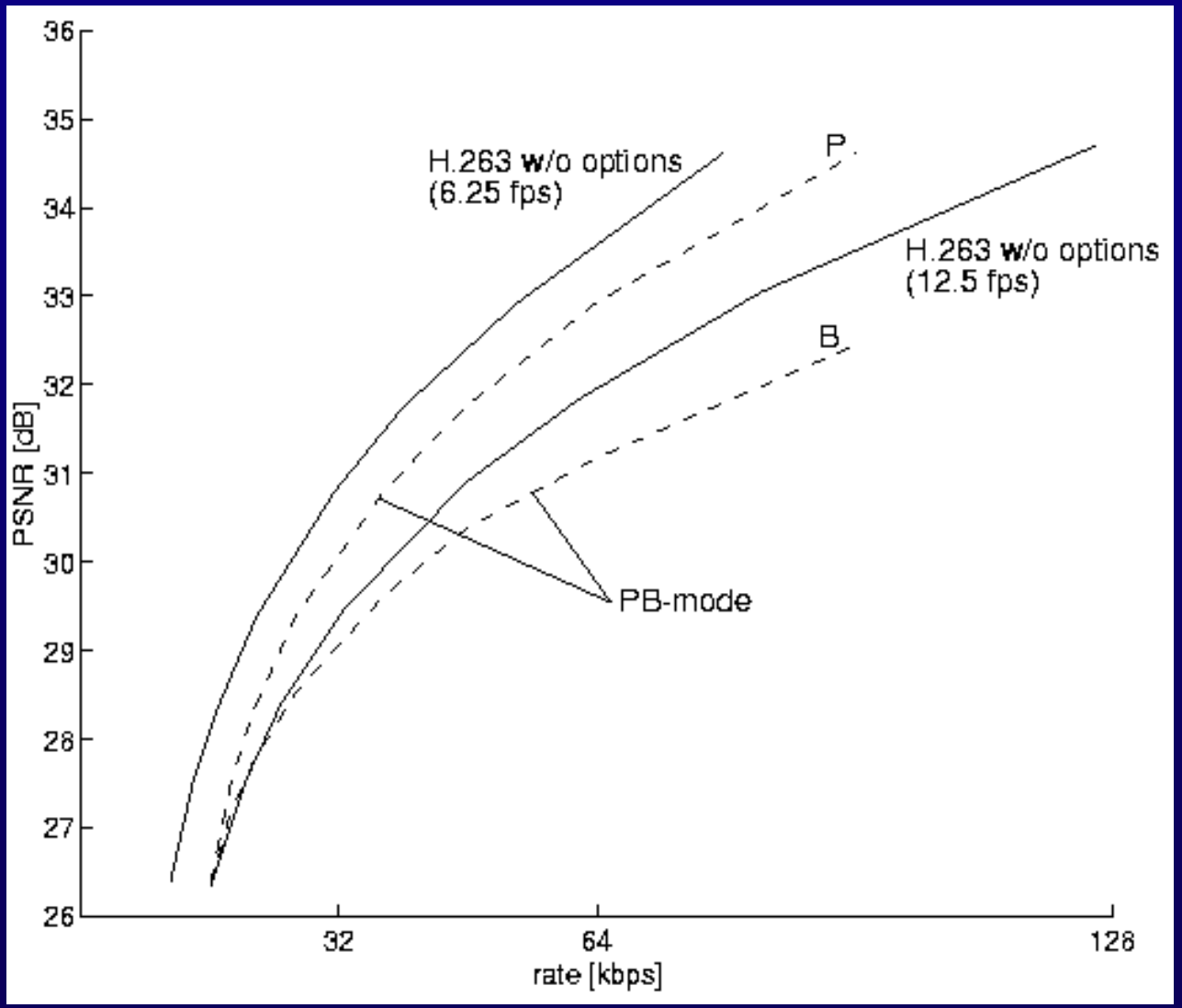
Foreman at 12.5 fps

Performance Comparison of H.263 w/wo SAC Mode



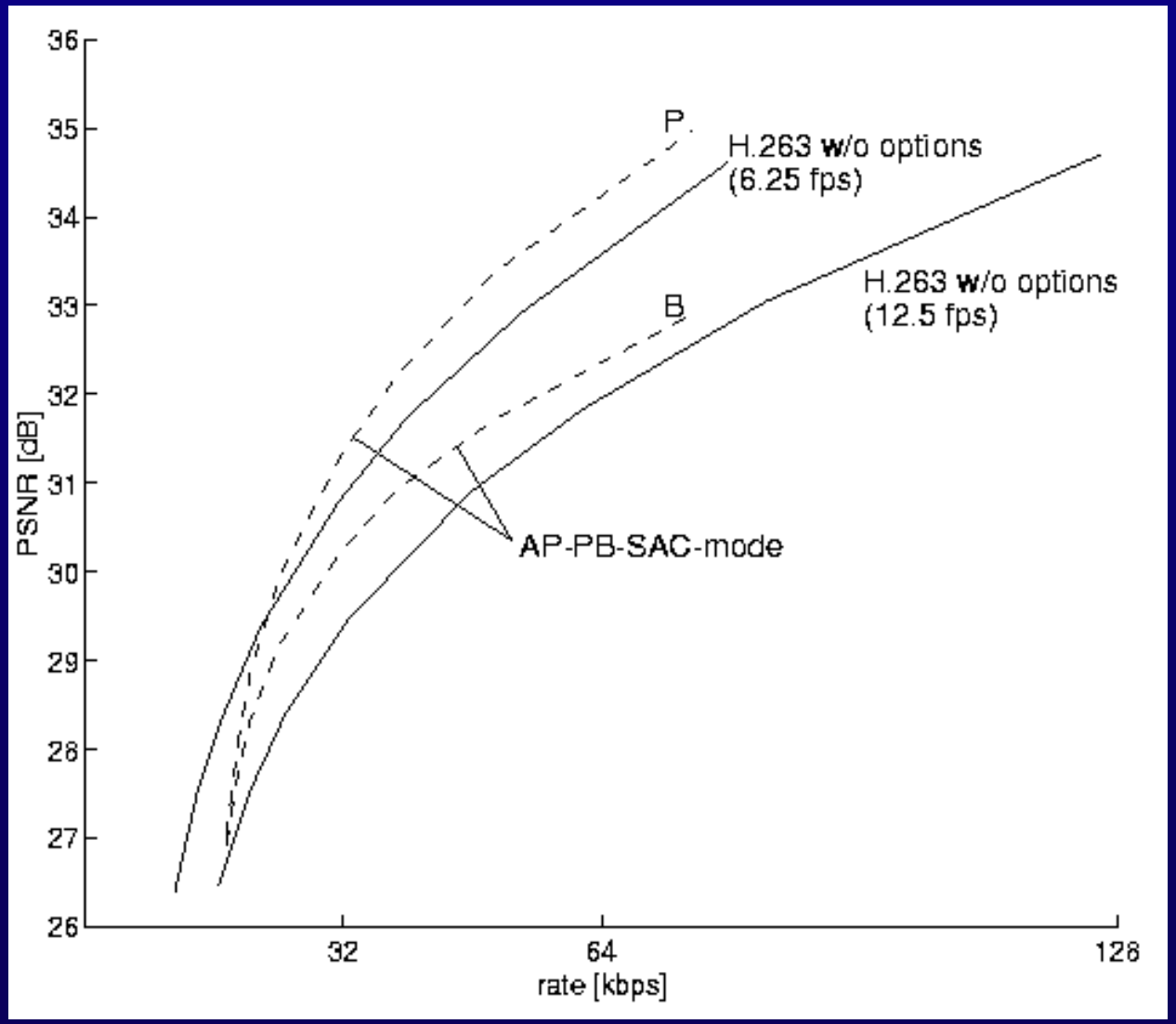
Foreman at 12.5 fps

Performance Comparison of H.263 w/wo PB Mode



Foreman at 12.5 fps

Performance Comparison of H.263 w/wo All Options



Foreman at 12.5 fps

H.263+

- Formally H.263 Version 2
- Enhancements of H.263
- 12 new negotiable modes
- Backward compatible with H.263
- Allows a wide range of custom source formats

Custom Source Formats

- Higher picture clock frequency (PCF)
- Custom picture formats
- Custom pixel aspect ratios (PAR)

Pixel aspect ratio	**pixel width : pixel height**
Square	1:1
CIF	12:11
525-type for 4:3 picture	10:11
CIF for 16:9 picture	16:11
525-type for 16:9 picture	40:33
Extended PAR	*m*:*n*, *m* and *n* are relatively prime

H.263+ Optional Modes

- Annex D: New Unrestricted Motion Vector Mode (UMV)
- Annex I: Advanced Intra Coding Mode
- Annex J: Deblocking Filter Mode
- Annex M: Improved PB-Frame Mode
- Annex O: Temporal, Spatial, and SNR Scalability Mode
- Annex P: Reference Picture Resampling Mode
- Annex Q: Reduced Resolution Update Mode
- Annex S: Alternative Inter VLC Mode
- Annex T: Modified Quantization Mode

Error resilience:

- Annex K: Slice Structured Mode
- Annex R: Independent Segment Decoding Mode
- Annex N: Reference Picture Selection Mode

New Optional Modes for Coding Efficiency Improvement

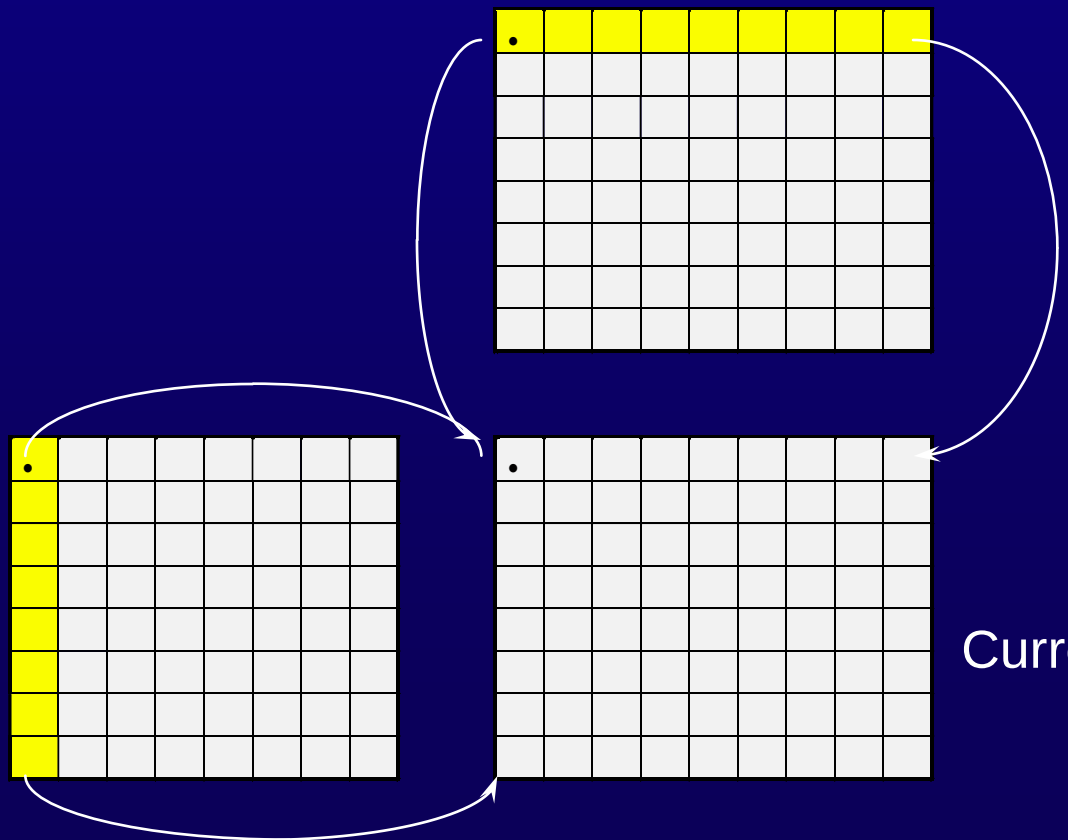
- Coding efficiency
  - New UMV mode
  - Advanced Intra Coding Mode
  - Alternate Inter VLC Mode
    - Use intra table for inter DCT
  - Deblocking Filter Mode
    - Depending on quantization step size
  - Modified Quantization Mode
    - More flexible changes of quantization step sizes
    - Finer quantization for chrominance
    - Extended DCT range
  - Improved PB-Frame Mode
    - Forward, backward, or bi-directional

Modifications to UMV Modes

- Single value VLC
  - Easy implementation
- Reversible VLC table
  - Better error resilience
- Larger motion vector range
  - Depending on the picture size
  - Up to [–256, 255.5]

Advanced Intra Coding

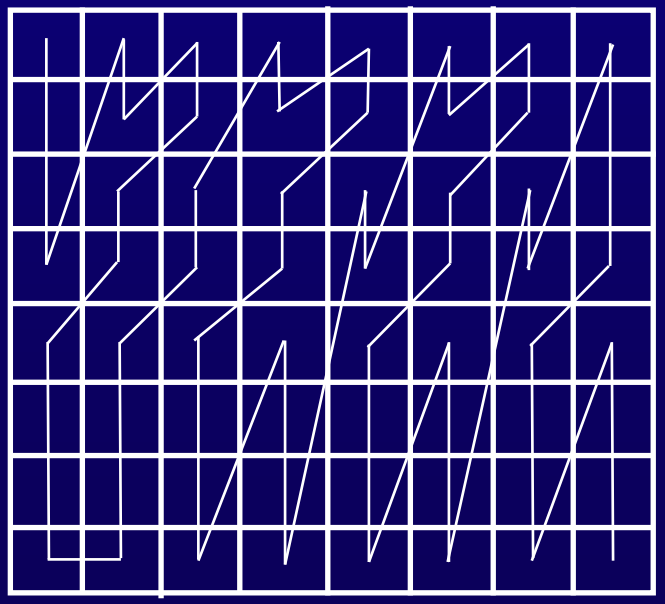
- A separate VLC table for intra DCT
- Modified inverse quantization
- Spatial prediction of DCT coefficients
  - DC only, vertical DC & AC, horizontal DC & AC



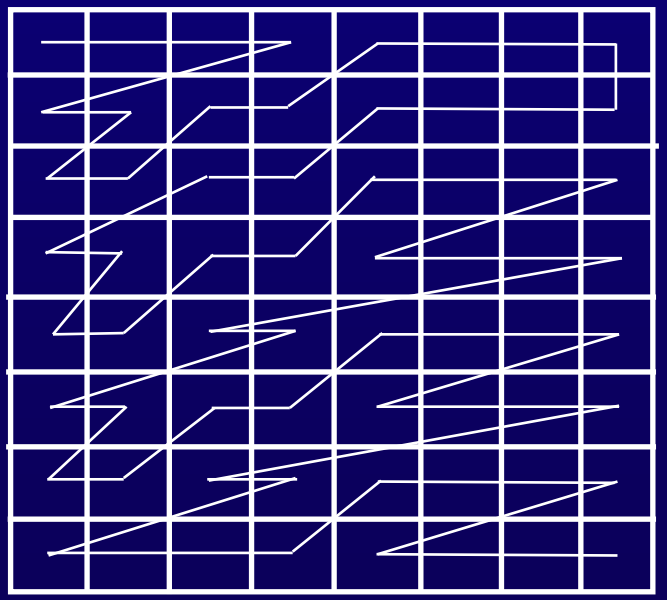
Advanced Intra Coding (Cont.)

- If the prediction refers to the horizontally adjacent block
  => alternate-vertical scan
- If the prediction refers to the vertically adjacent block
  => alternate-horizontal scan

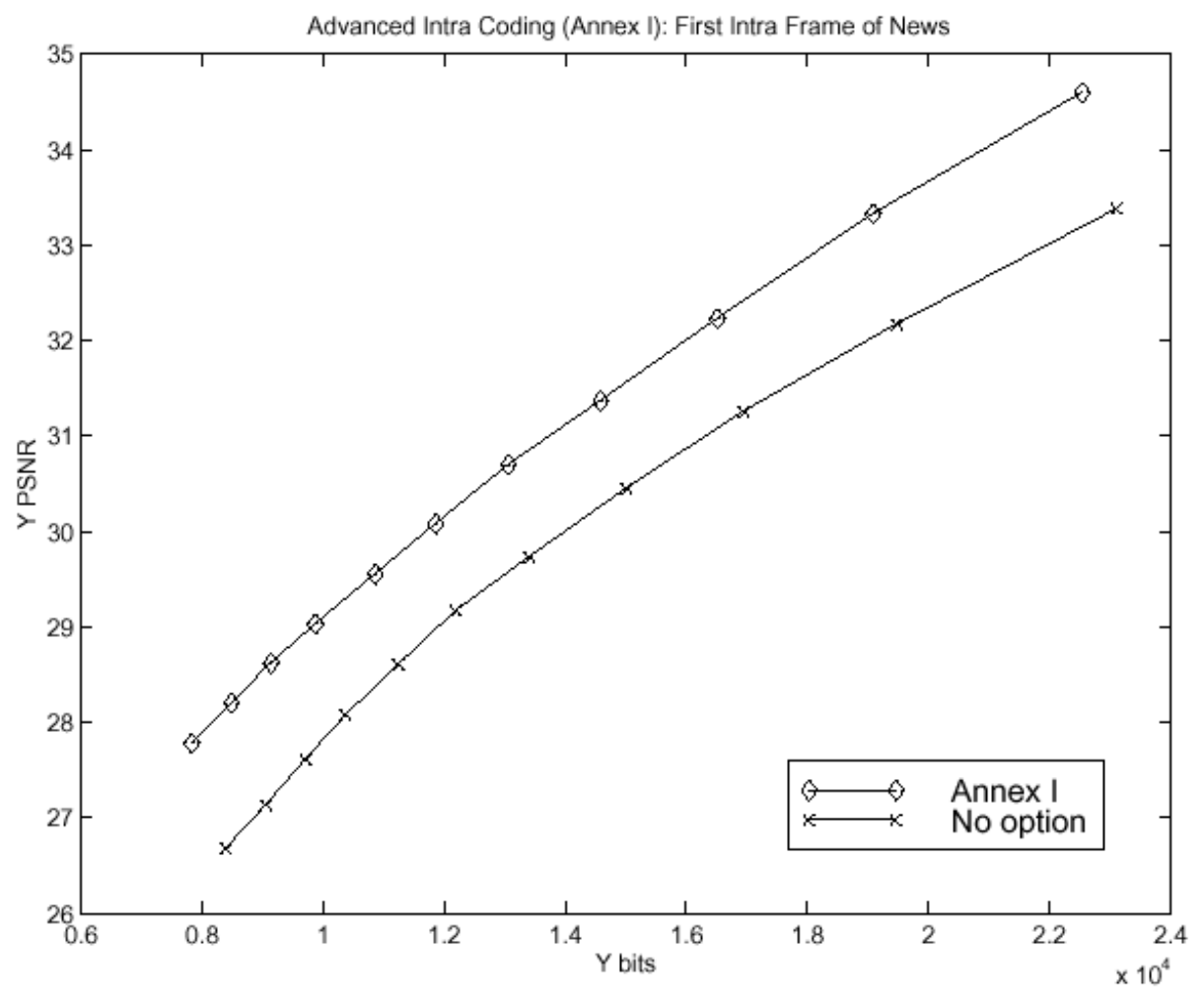
Alternate (vertical)



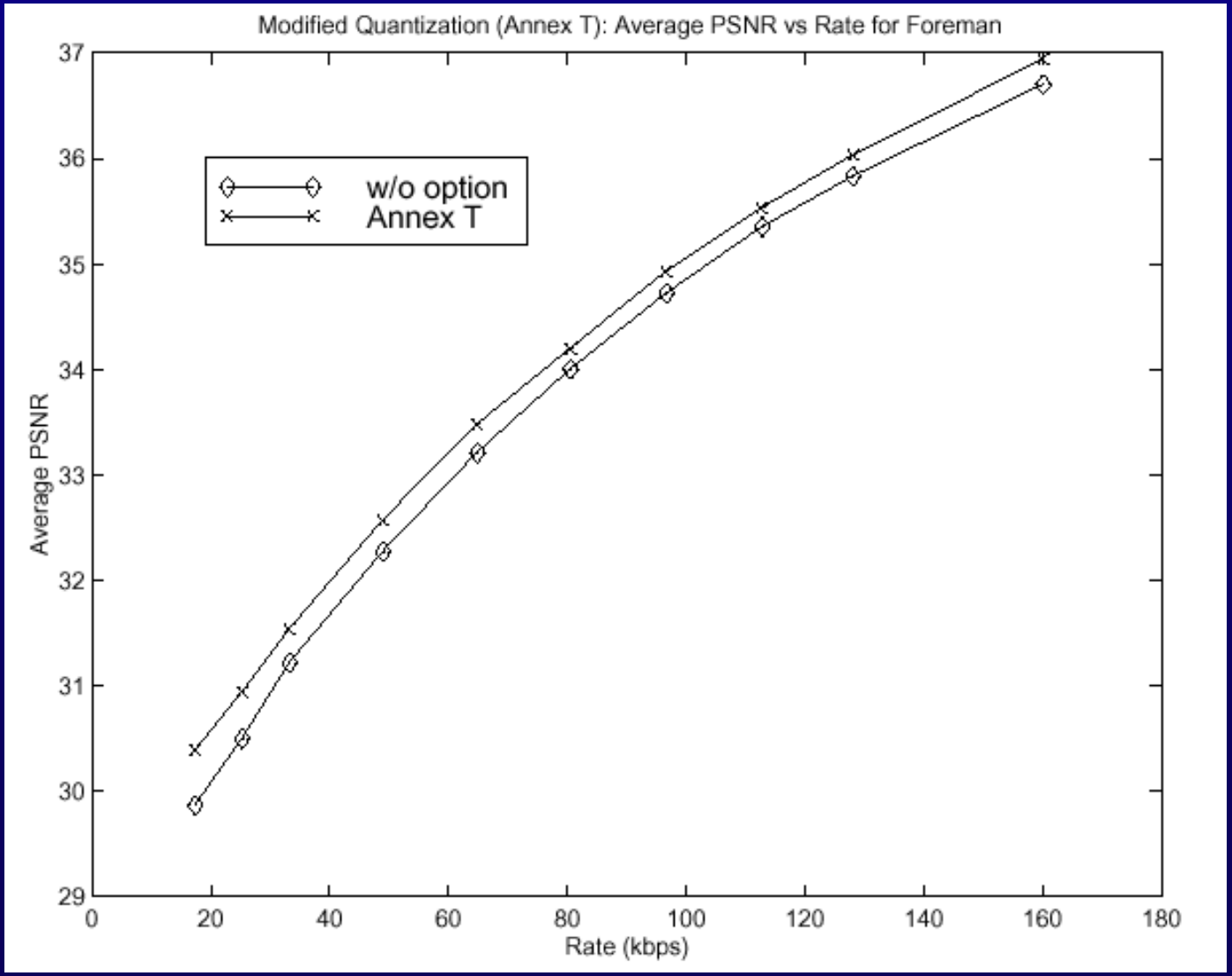
Alternate (horizontal)



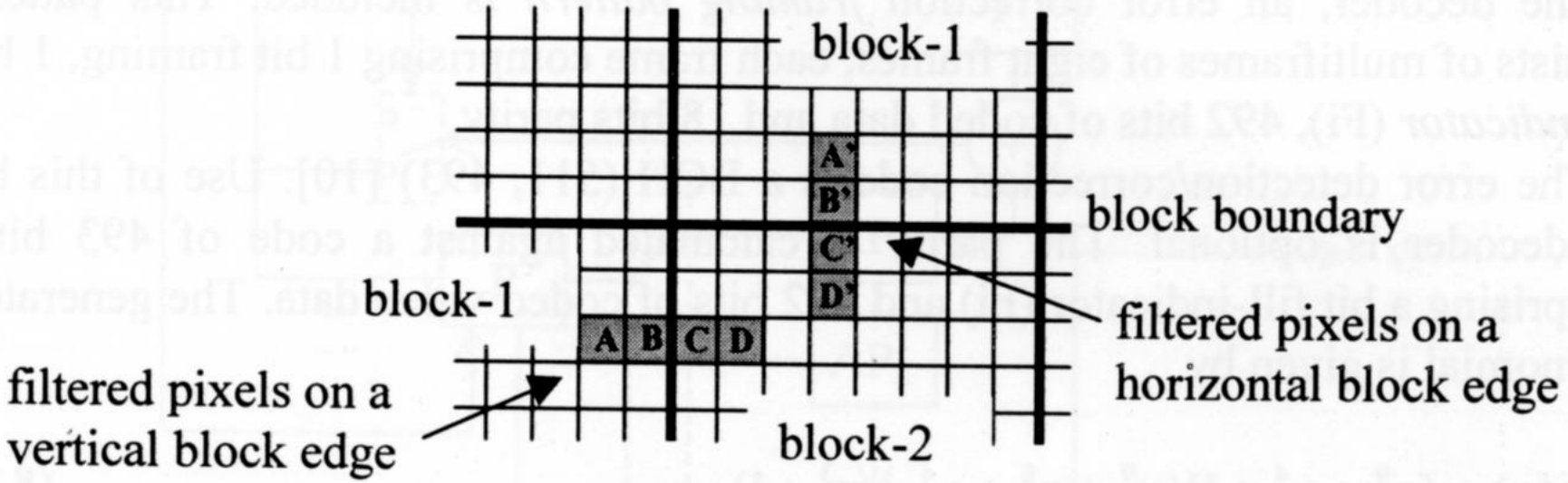
Advanced Intra Coding (Cont.)



Modified Quantization Mode



Deblocking Filter Mode



$$B_1 = B + d_1$$

$$C_1 = C - d_1$$

$$d_1 = sign(d) \times (Max(0,|d| - Max(0,2\times|d| - QP))$$

$$d = (3A - 8B + 8C - 3D)/16$$

$$QP = \text{quantisation parameter of block - 2}$$

Deblocking Filter Mode (Cont.)

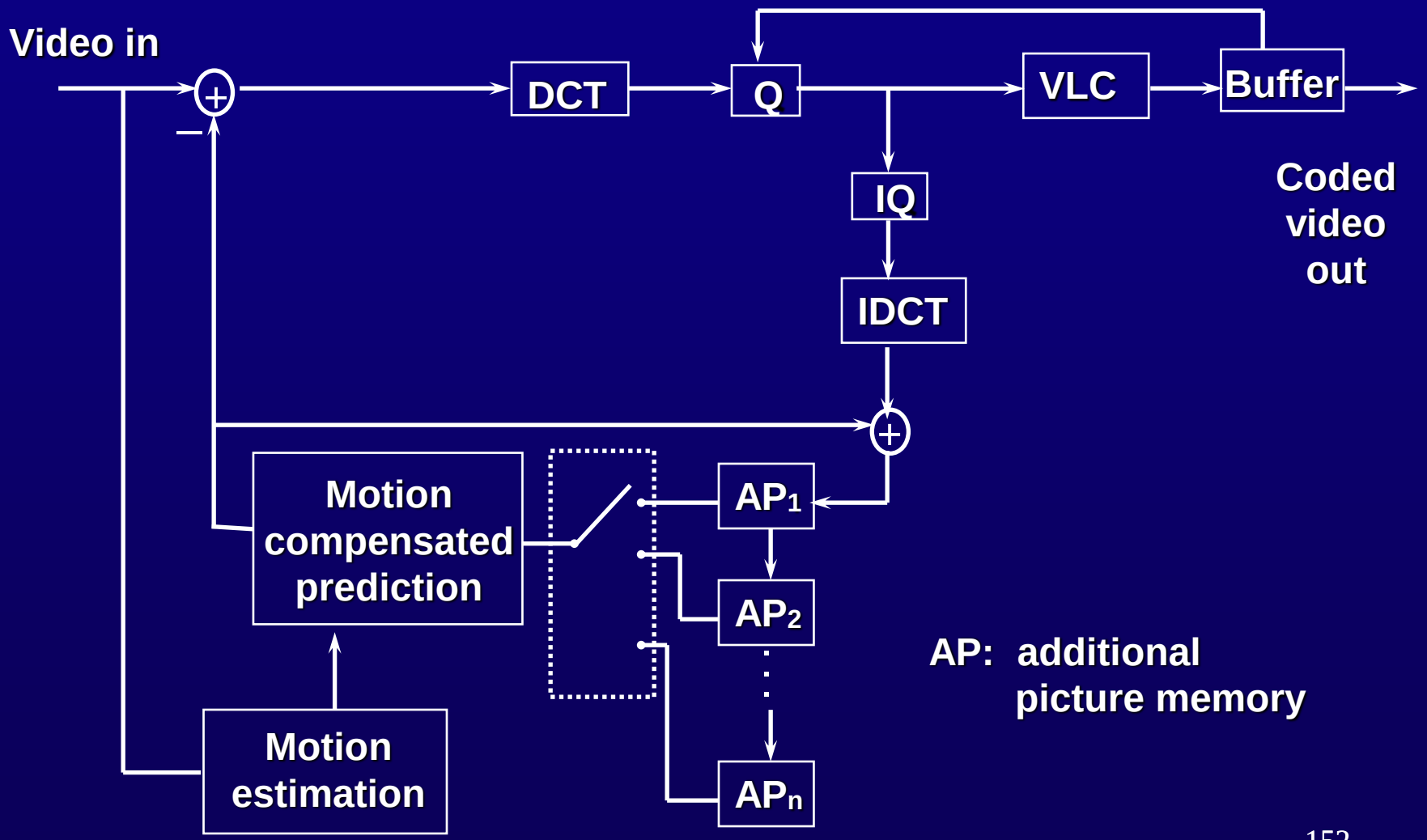
(a) No Option

(b) Post Filter Only

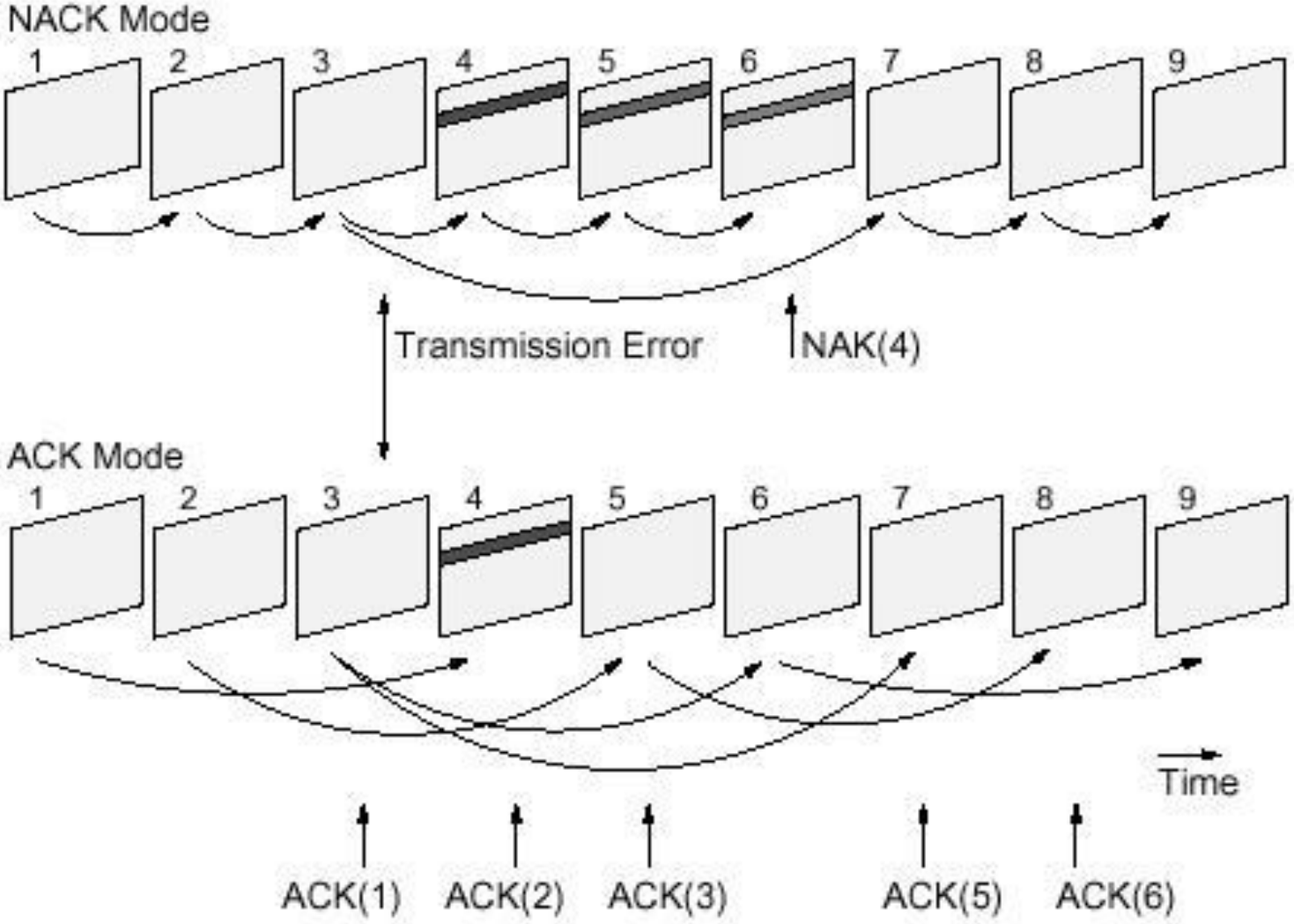
(b) Deblocking Filter Only

(c) Deblocking Filter Combined with Post Filter

Reference Picture Selection Mode



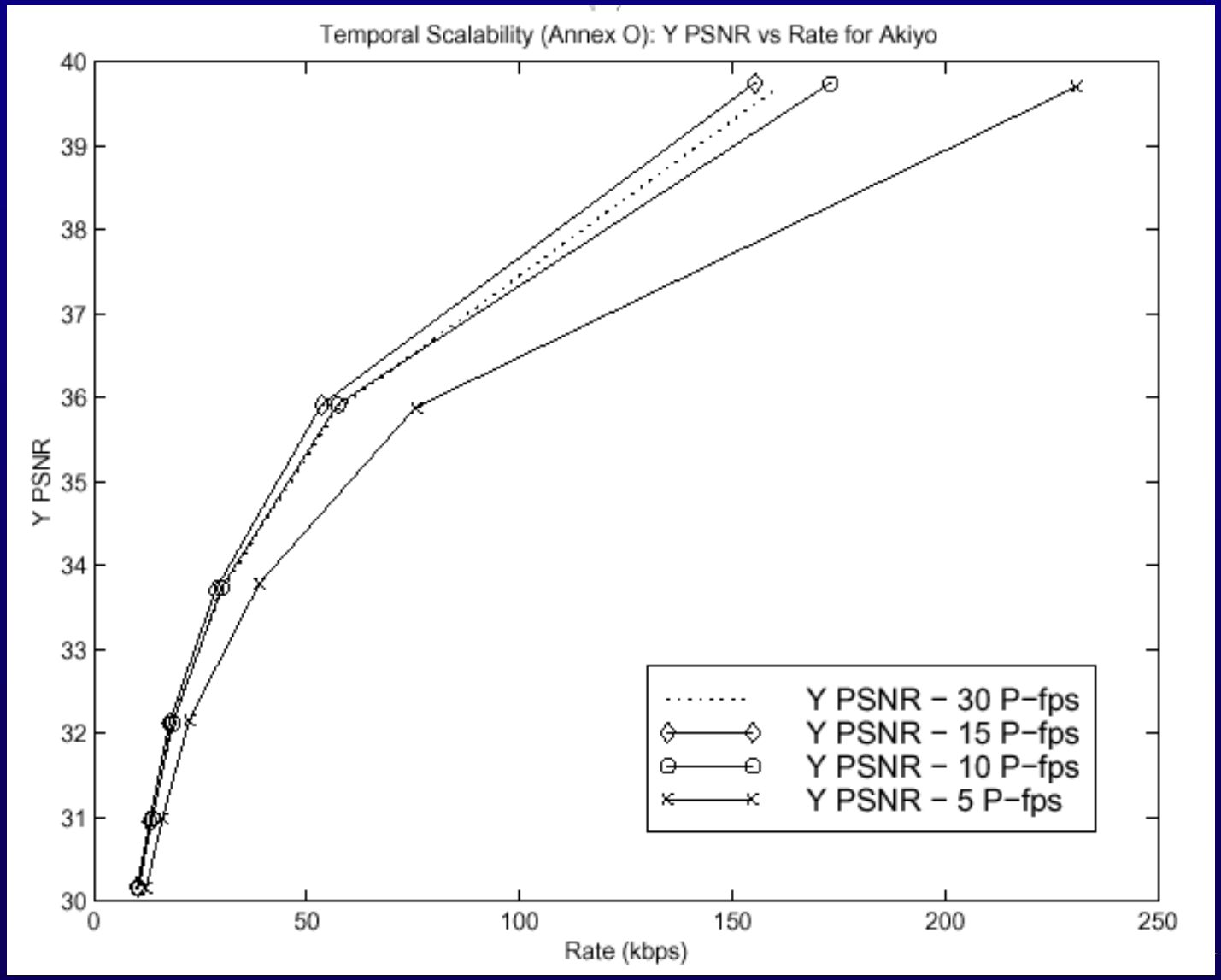
Spatio-Temporal Error Propagation with RPS



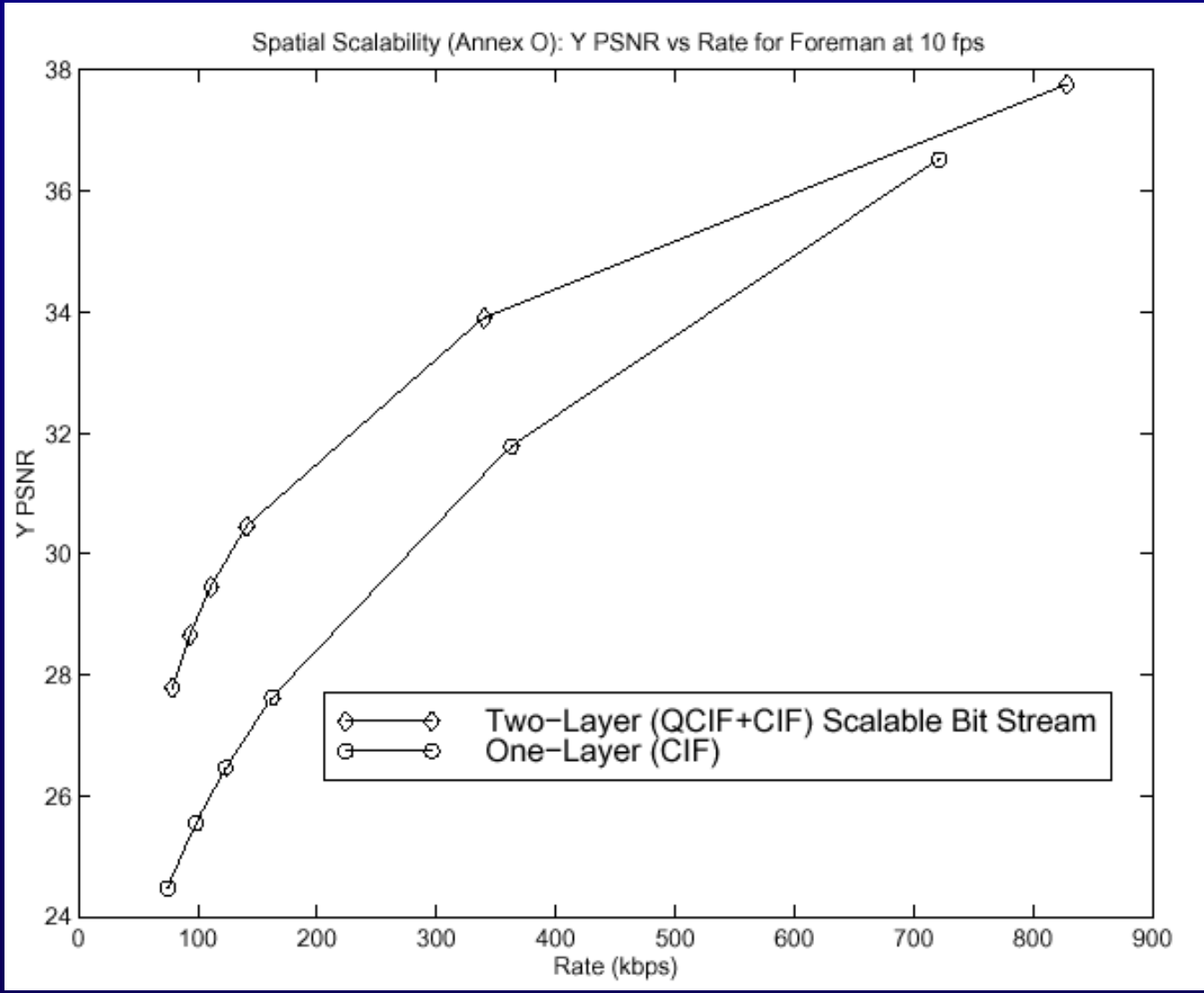
Scalability

- Scalability
  - Decode partial information from partial bitstream
  - To fit various bandwidth requirements
  - To fit terminals with different capabilities
- Scalability Modes
  - Temporal scalability: bi-directional prediction
  - Spatial scalability
  - SNR scalability

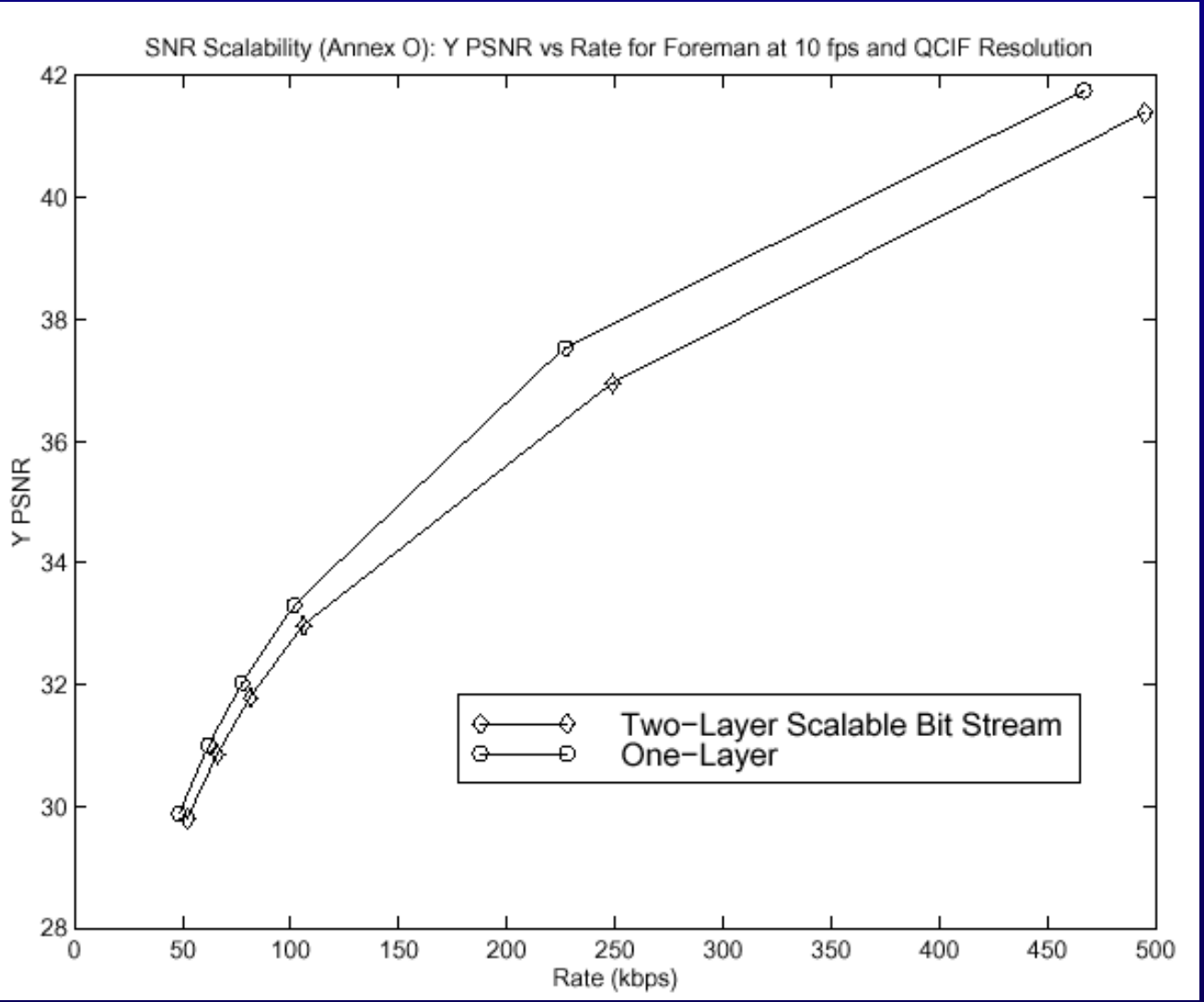
Temporal Scalability (Cont.)



Spatial Scalability (Cont.)

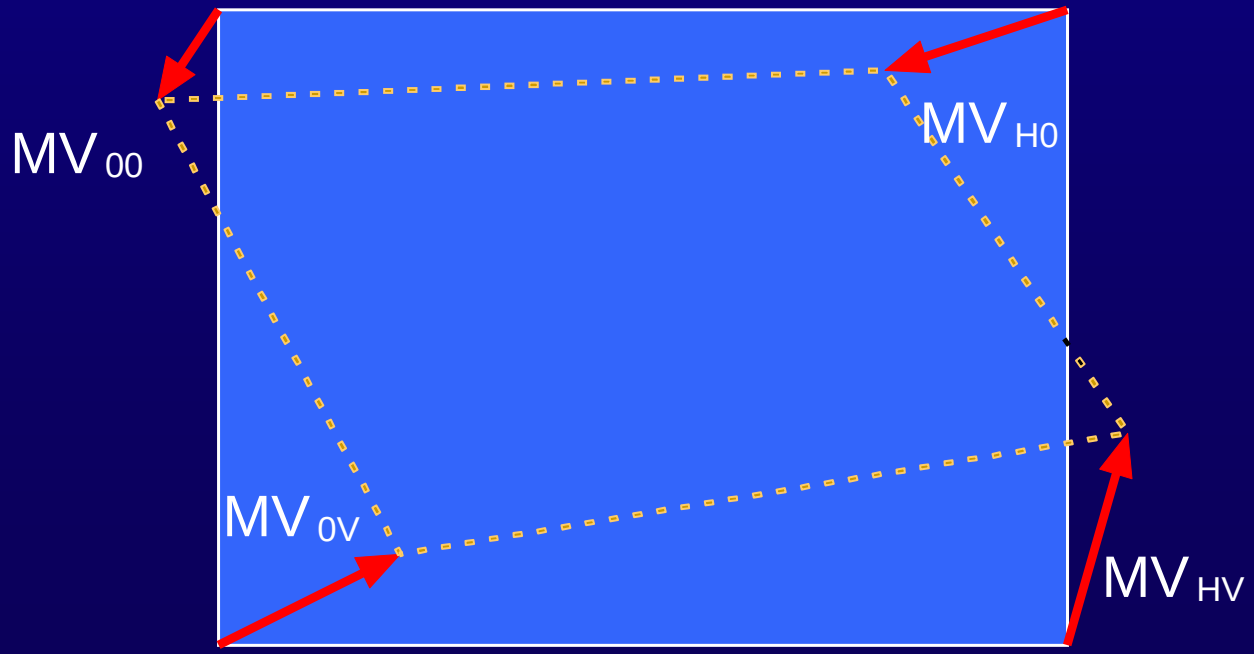


SNR Scalability (Cont.)



Reference Picture Resampling

- Global motion compensation
- Rotating motion
- Special-effect warping



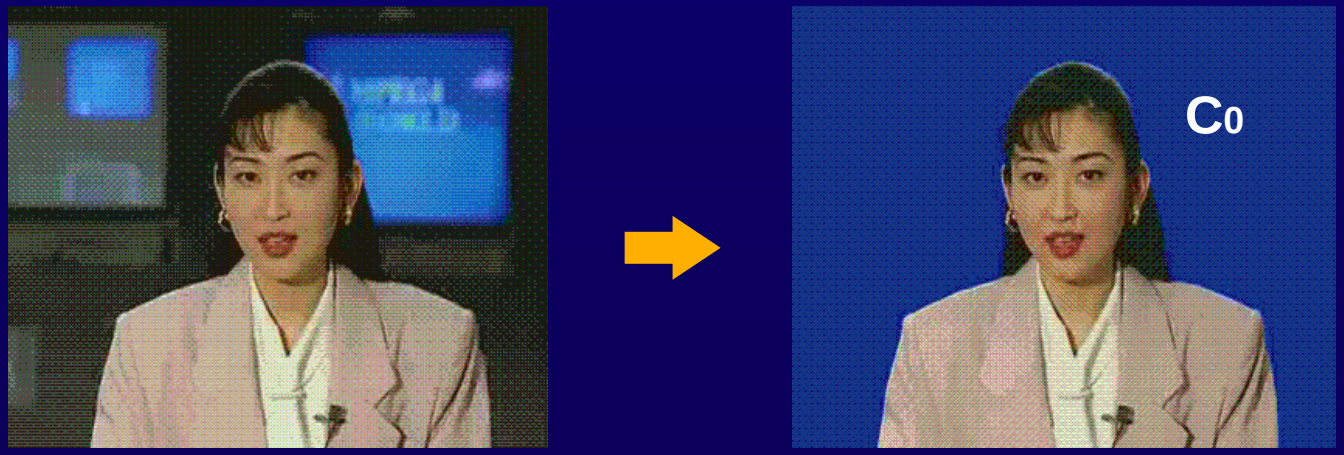
Supplemental Enhancement Information

- Enhanced features
  - Picture freeze and release
  - Tagging information
    - Snapshot
    - Video segment start/end
    - Progressive refinement start/end
  - Chroma key
- Can be discarded by decoders that do not understand

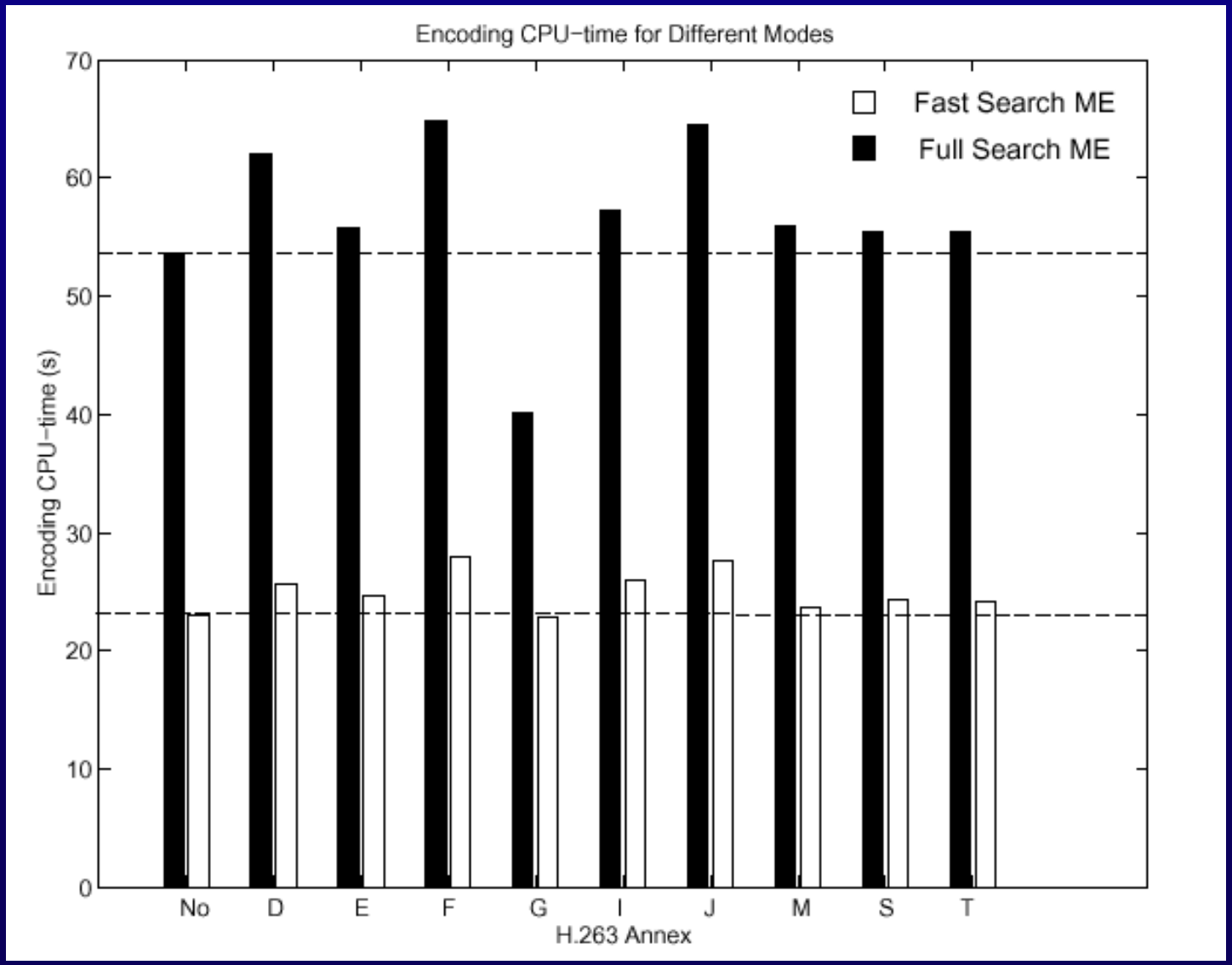
Chroma Key

- A video sequence $f(x,n)$ with regions R_i
- For each region R_i, replace the non-object area with a special color C_0

$$g(\mathbf{x},n)=\begin{cases} f(\mathbf{x},n) & \text{if } \mathbf{x}\in\Re_i \\ C_0 & \text{if } \mathbf{x}\in\overline{\Re_i} \end{cases}$$



Computational Complexity Comparison of Optional Modes



TMN8

- Motion estimation and mode selection
- Rate control algorithm
  - Frame skipping
    - If the buffer fullness is larger than a threshold *M*, the encoder skips encoding frames until the buffer fullness is below *M*.
  - Frame-layer rate-control
    - Frame target bits selection: the target bits is selected depending on the buffer fullness
  - Macroblock-layer rate control
    - Initialization: calculating the standard deviations σ_i for all MBs
    - Compute *Q* for *i*-th MB according to σ_i and encode the macroblock
    - Update counters and model parameters

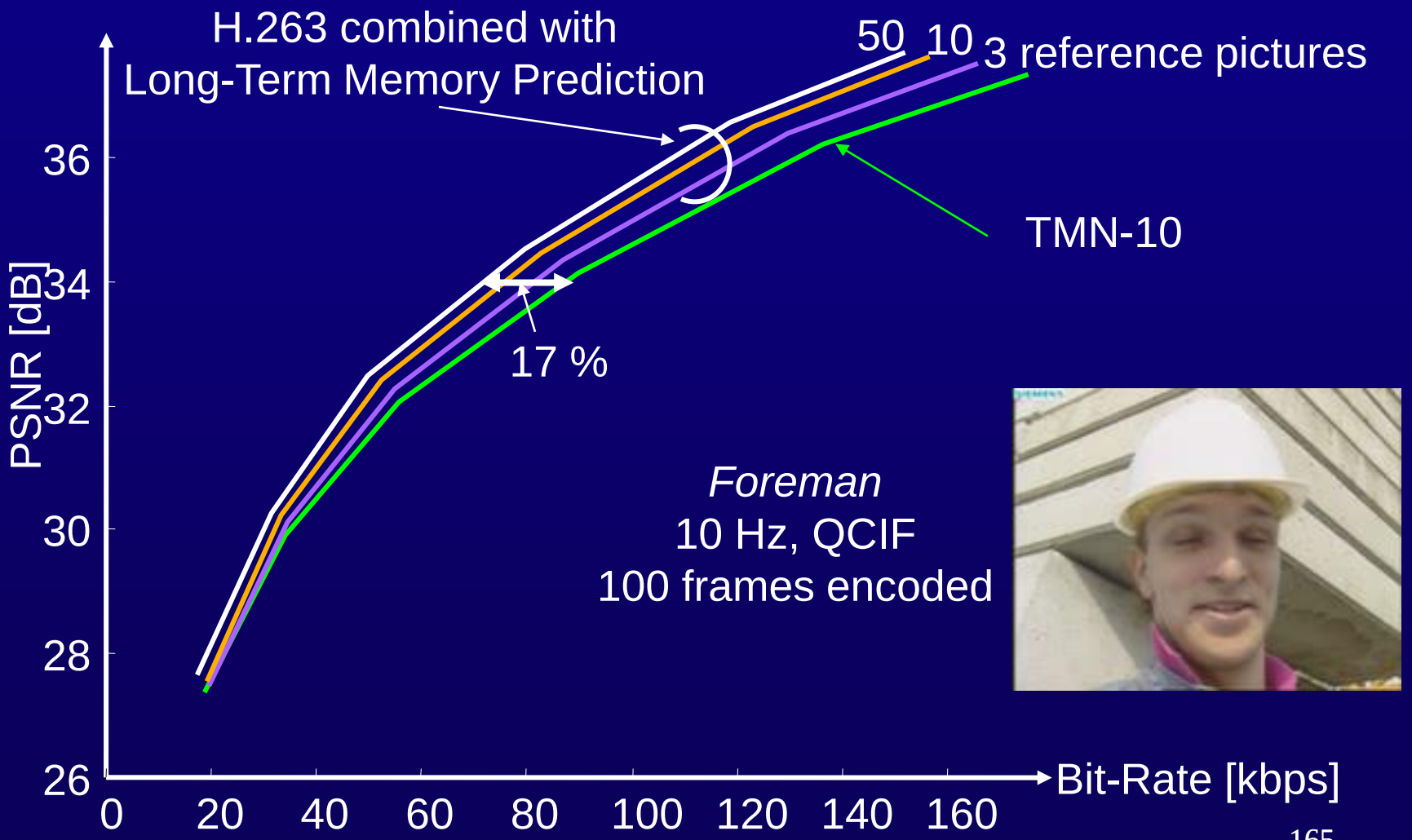
H.263++ New Features

- **Annex U**: Fidelity enhancement by macroblock and block-level reference picture selection – a significant improvement in compression quality
- **Annex V**: Packet loss & error resilience using data partitioning with reversible VLCs

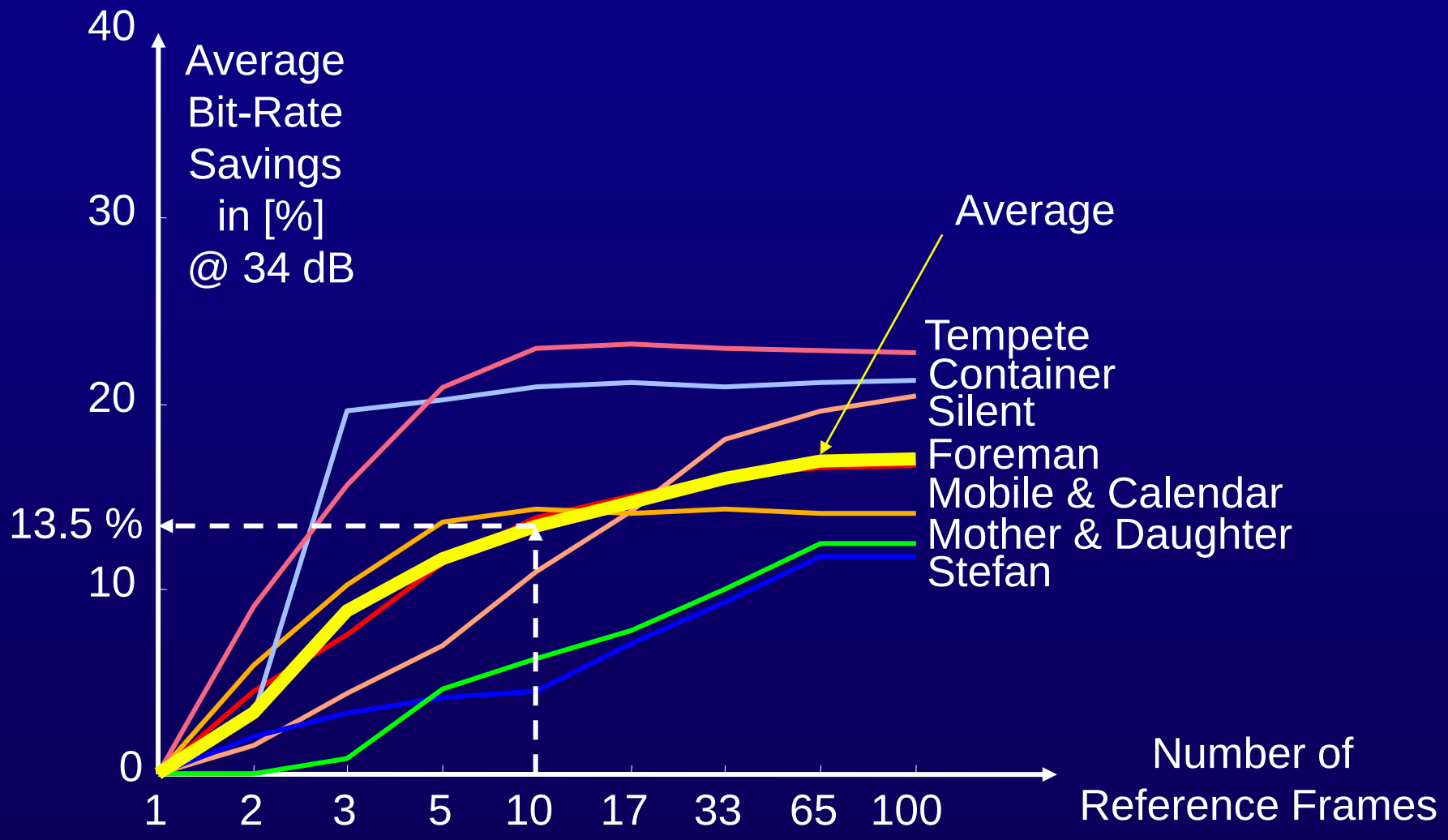
H.263++ New Features (Cont.)

- **Annex W**:additional supplemental enhancement information
  - IDCT mismatch elimination (specific fixed-point fast IDCT)
  - Arbitrary binary user data
  - Text messages (arbitrary, copyright, caption, video description, and URI)
  - Error resilience:
    - Picture header repetition (current, previous, next+TR, next-TR)
    - Spare reference pictures for error concealment
  - Interlaced field indications (top & bottom)

H.263++ Performance: Annex U



H.263++ Average Bit-rate Saving

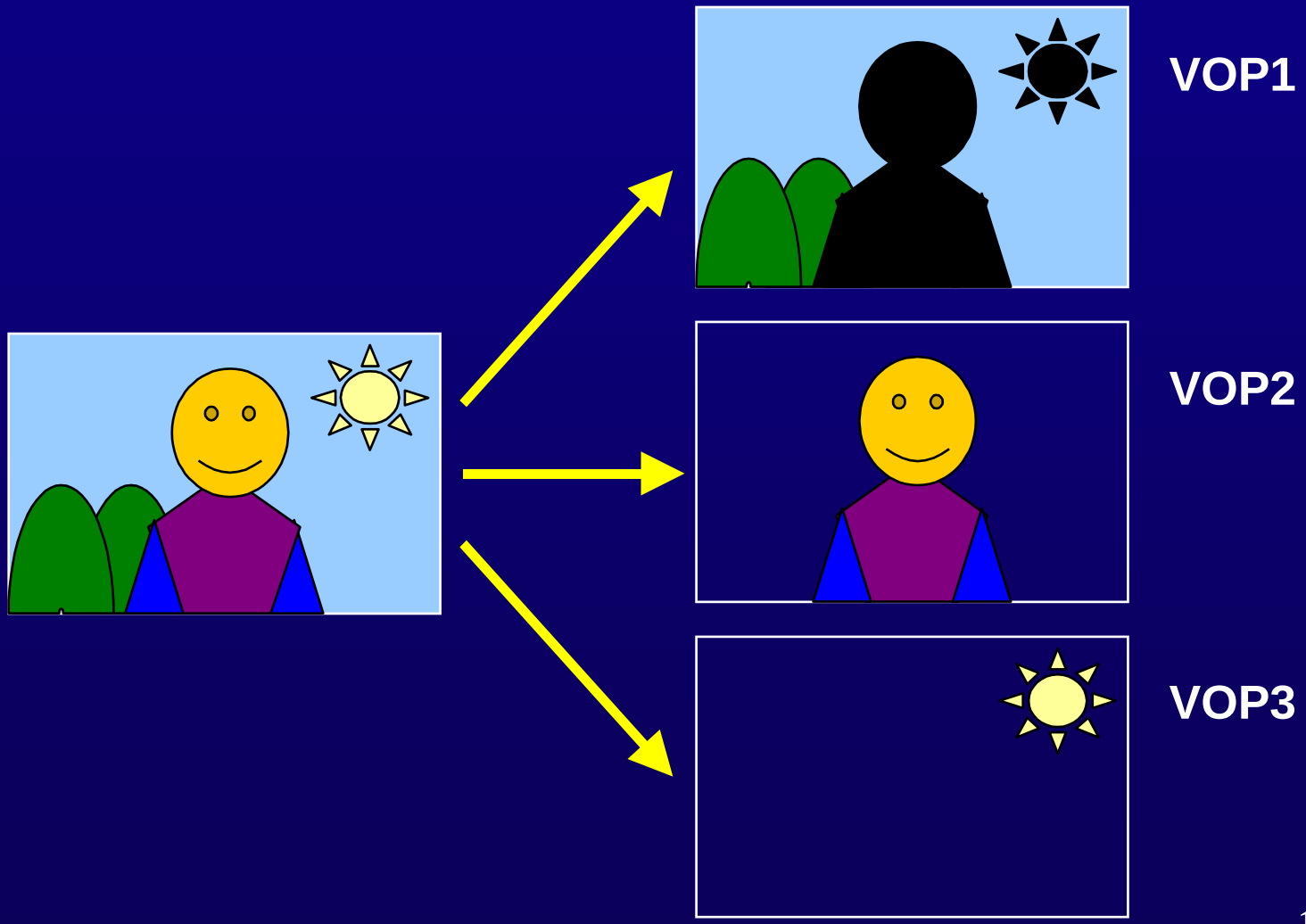


MPEG-4

MPEG-4

- Initial driving application:
  1. Video coding for very low bit-rate video
  2. Object-based coding
- **MPEG-4 attempts to become *THE* standard for streaming AV media on the Internet and via wireless networks**
- Error-resilience features, optimized for low-bit-rate applications, fine granular scalability (FGS) coding
- MPEG-4 file format defines first standard for streaming AV media on the Internet

MPEG-4 Video Object Plane (VOP)



Integration of Natural and Synthetic Content

MPEG-4 Video: Core & Generic Coder

MPEG-4 VLBV Core Coder

Video Object Plane → Motion (MV) | Texture (DCT) → bitstream

(Similar to H.263/MPEG-1)

Generic MPEG-4 Coder

Video Object Plane → Shape | Motion (MV) | Texture (DCT) → bitstream

Coding of AV Objects

- AV scenes consist of “objects”
- Objects can be natural or synthetic (A&V, text & graphics, animated faces, arbitrarily shaped or rectangular)
- A “compositor” composes objects in a scene (A&V, 2 & 3D)
- Binary format for scenes: BIFS

VOP-Based Encoding

video objects
segmentation
& object mask

VOP1

VOP2

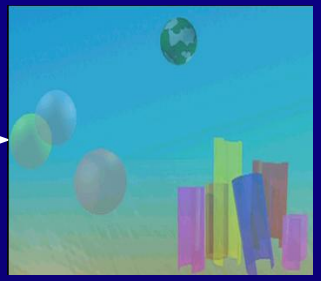
VOP3

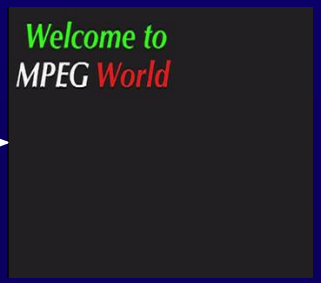
layered encoding

bitstream
(VOP1)

bitstream
(VOP2)

bitstream
(VOP3)



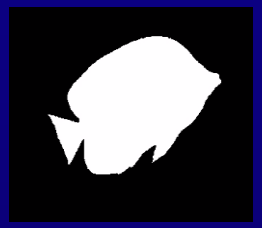


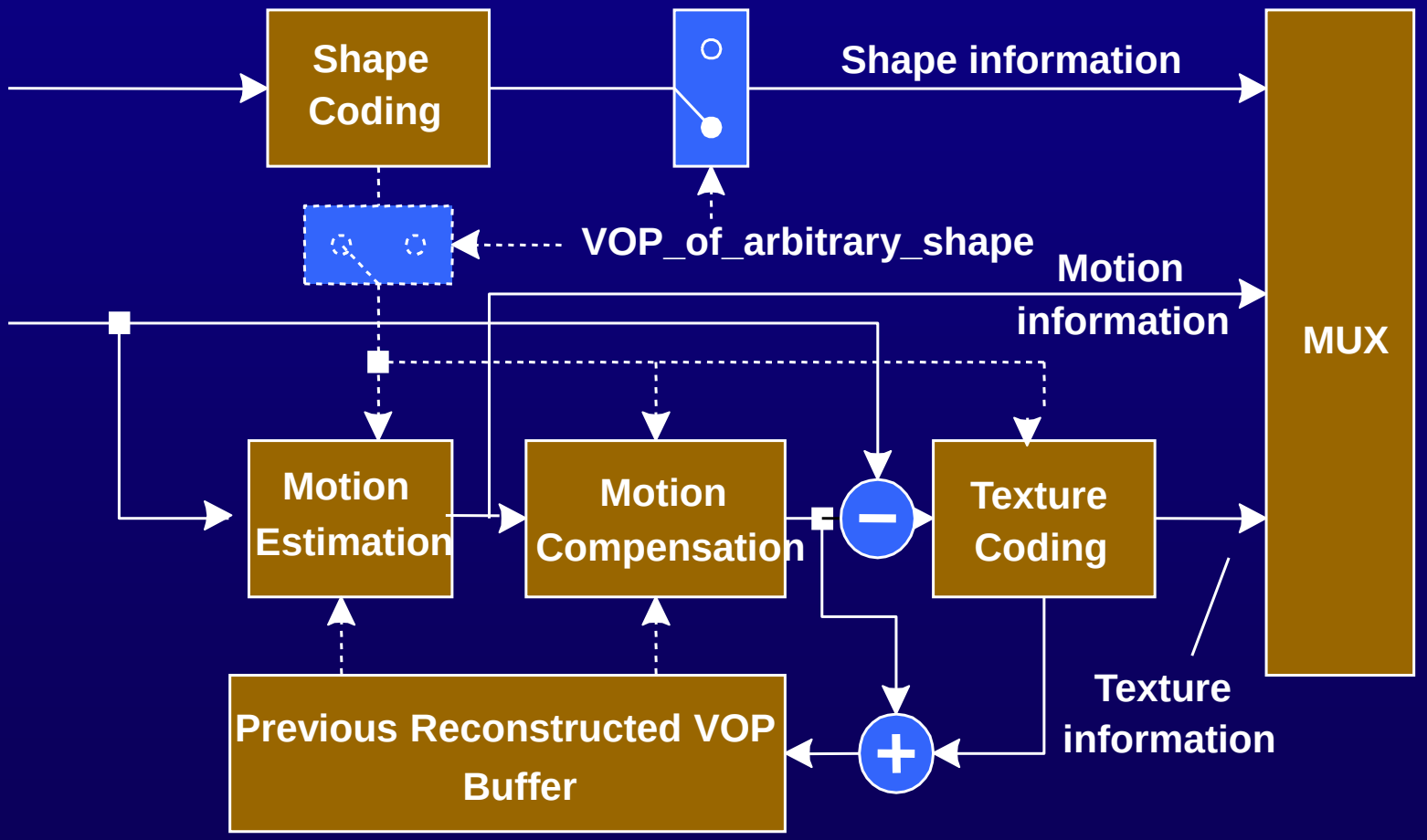
separate decoding

content-based
scalability

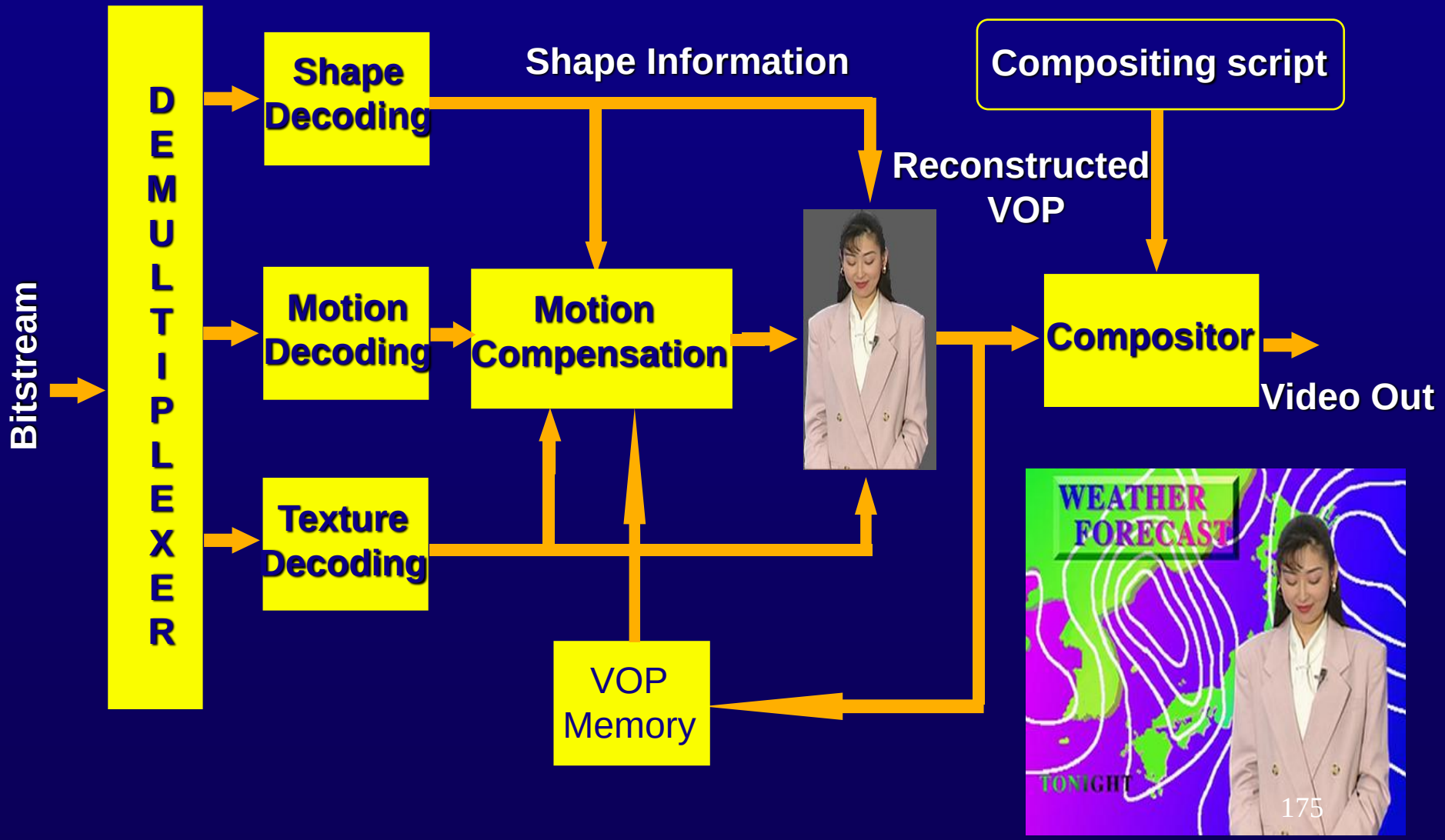
content-based
bitstream access
& manipulation

MPEG-4 Video Encoder



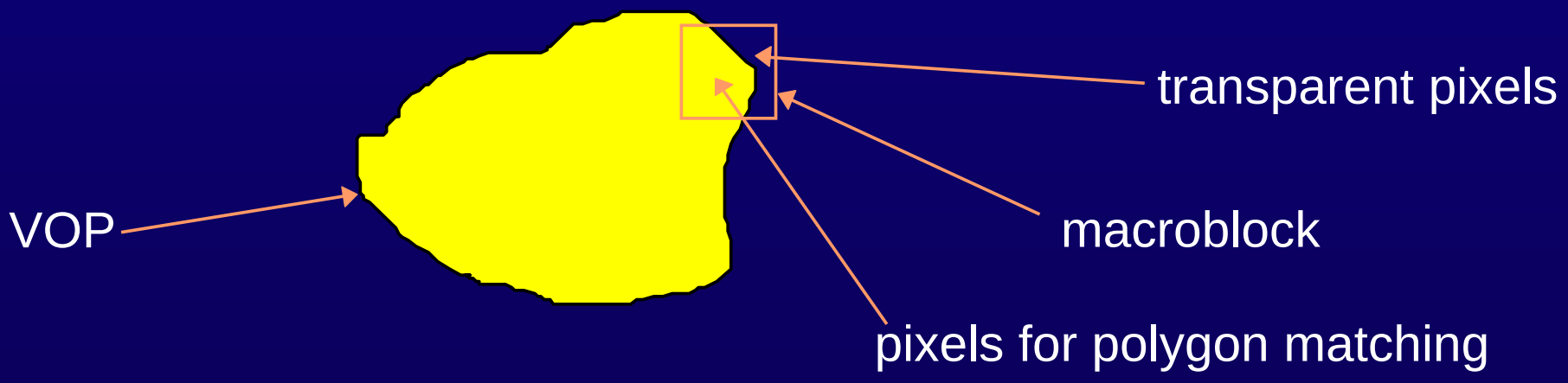


VOP Decoder

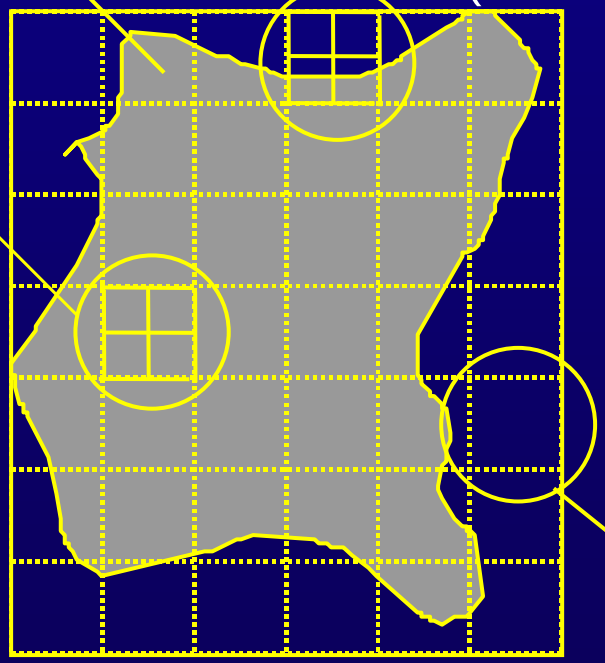

Bitstream
DEMULTIPLEXER
Shape Decoding
Motion Decoding
Texture Decoding
Shape Information
Motion Compensation
VOP Memory
Reconstructed VOP
Compositing script
Compositor
Video Out
WEATHER FORECAST
TONIGHT

Texture Coding (1/3)

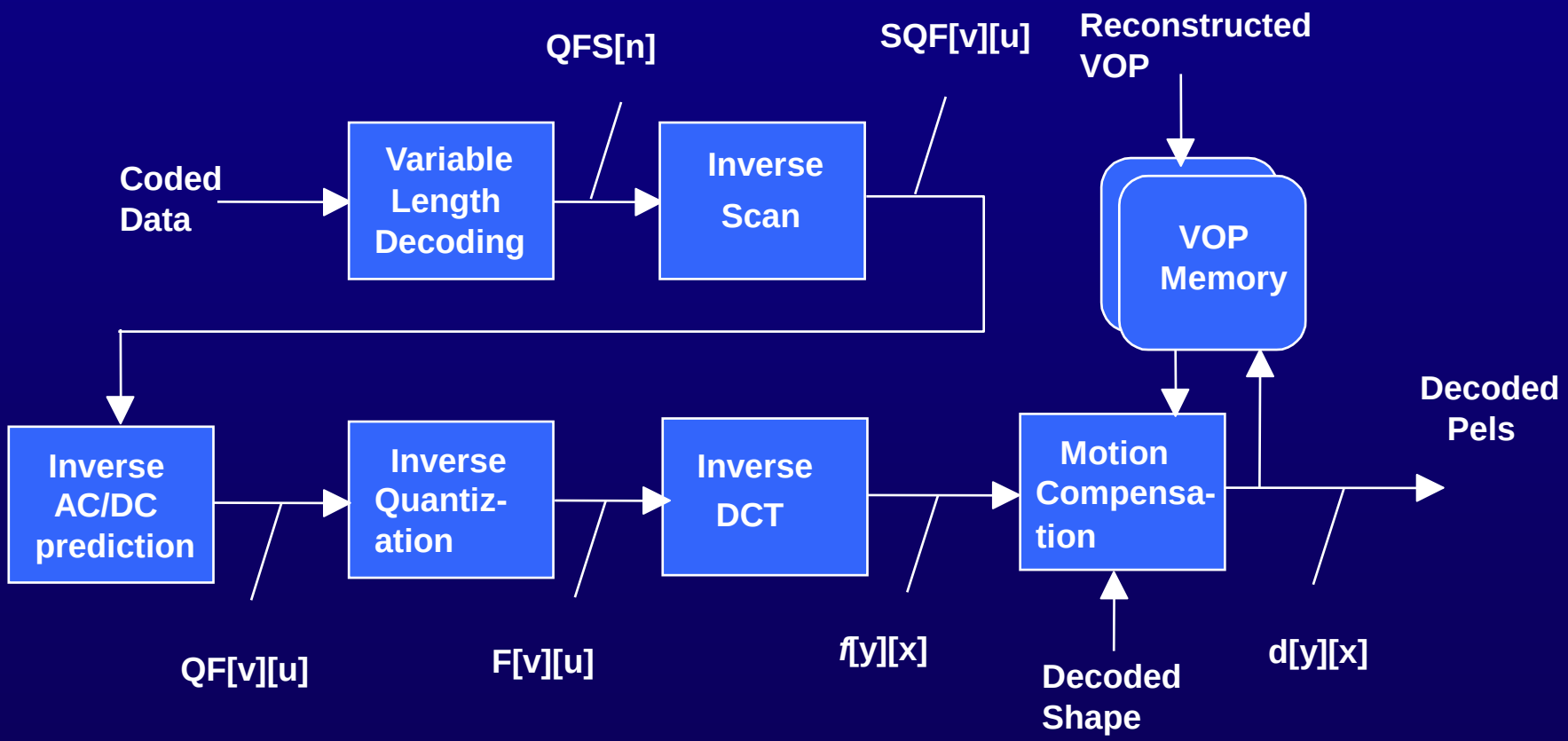
- Motion compensated DCT
  - Very similar to H.263
- Polygon matching



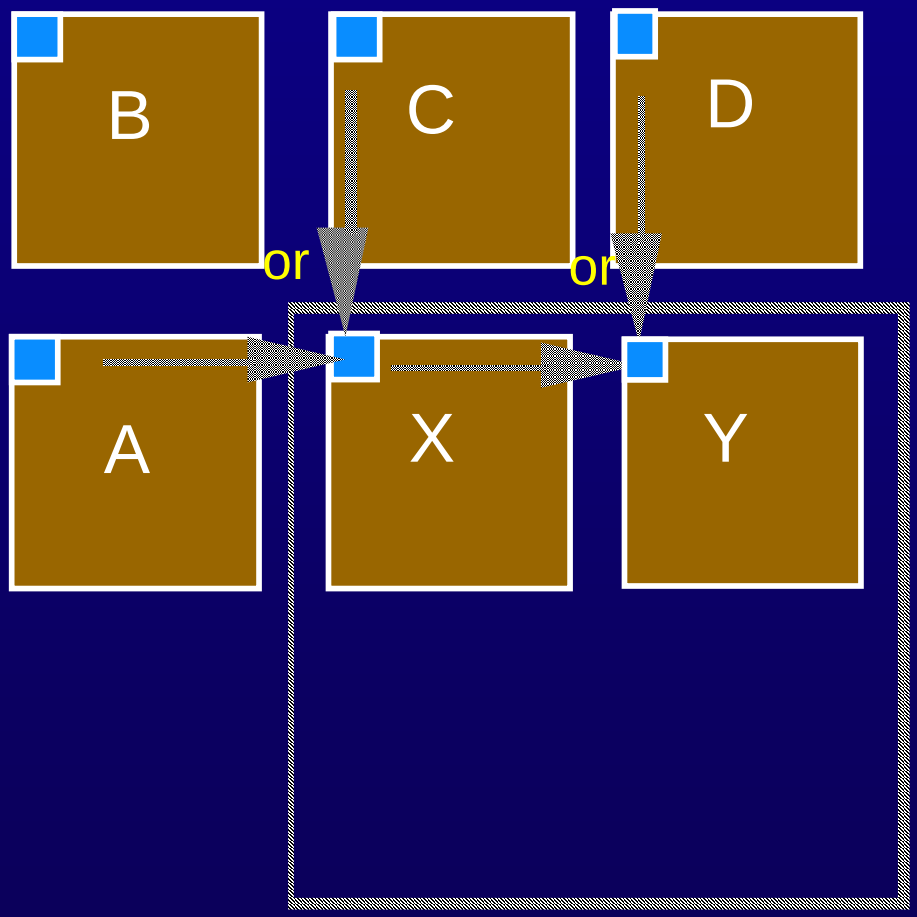
Texture Coding (2/3)



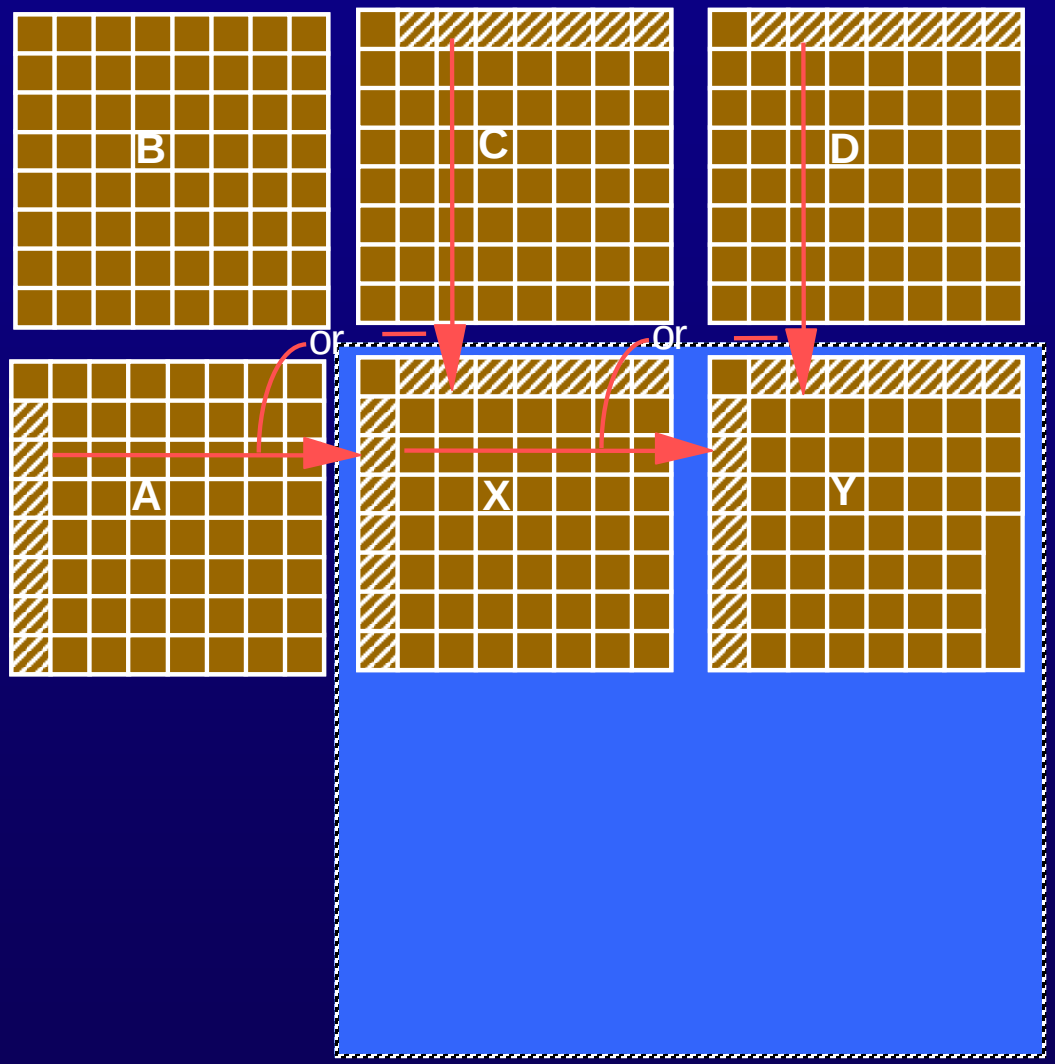
Texture Coding (3/3)



Adaptive DC Prediction

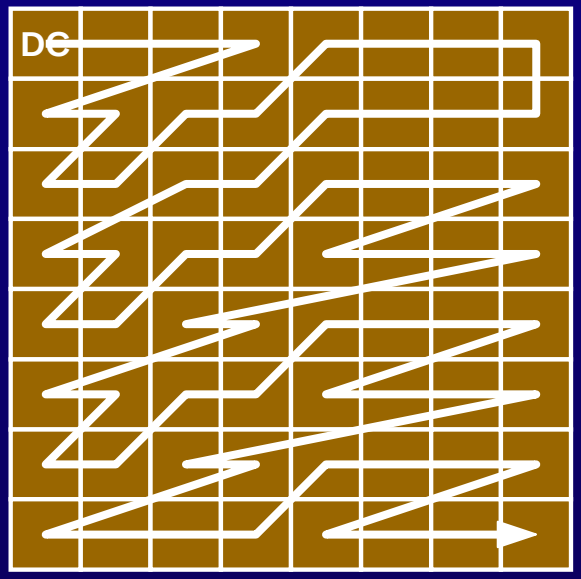


Adaptive AC Prediction

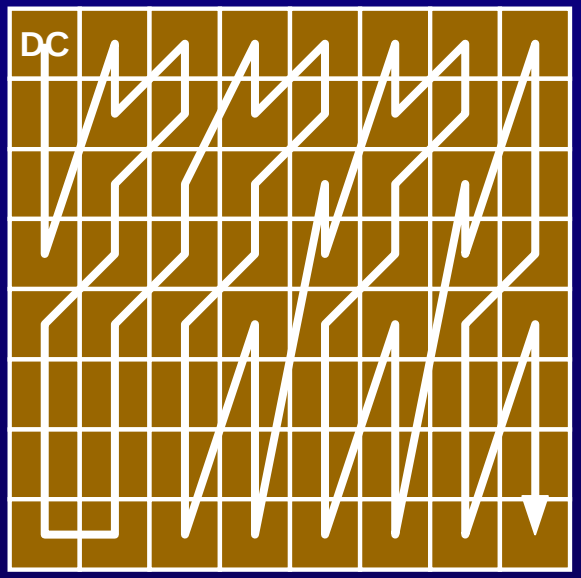


macroblock

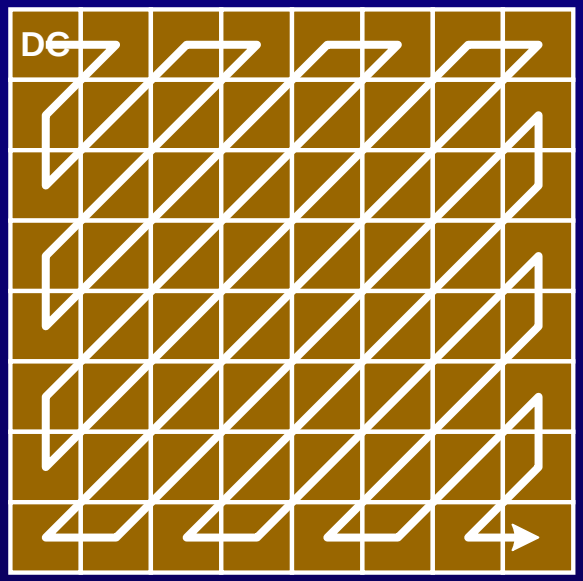
Scanning Pattern



Alternate-horizontal scan

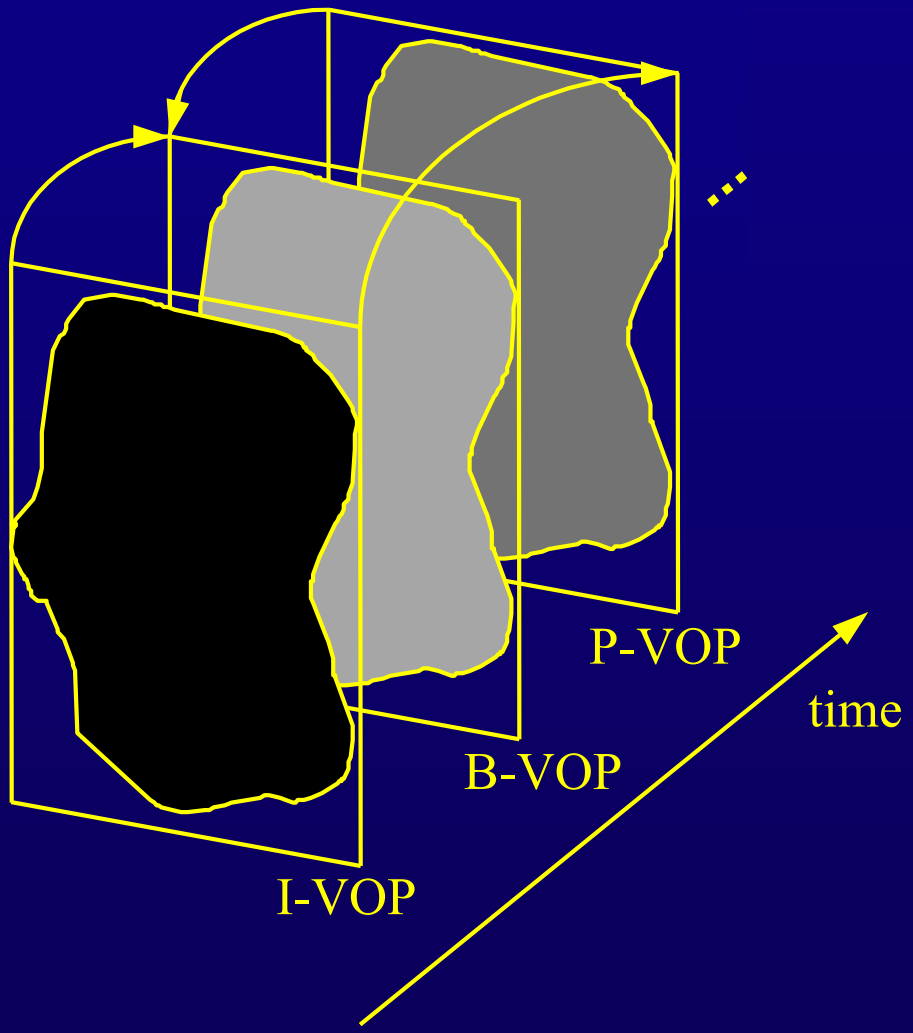


Alternate-Vertical scan

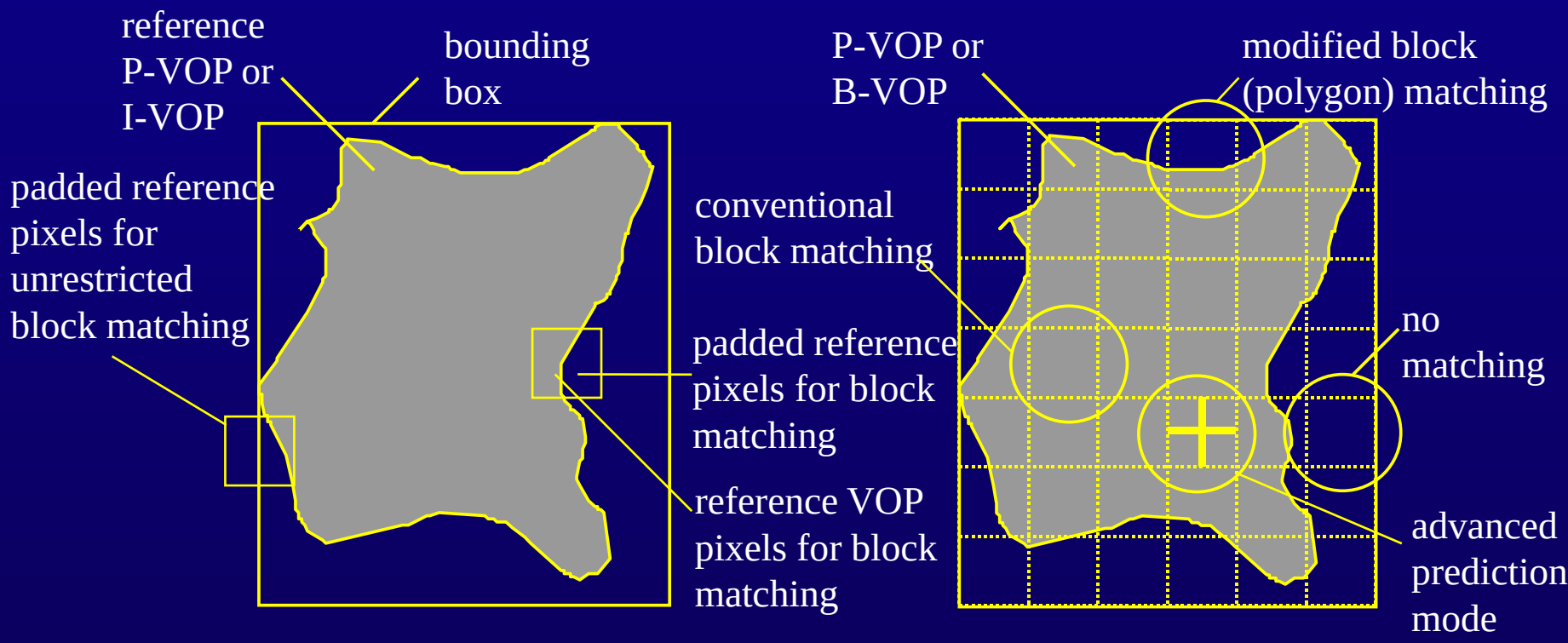


Zig-Zag scan

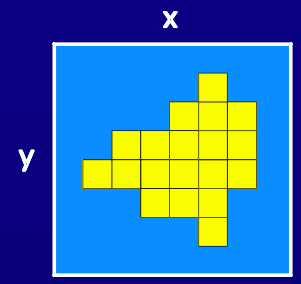
Motion Estimation/Compensation (1/3)



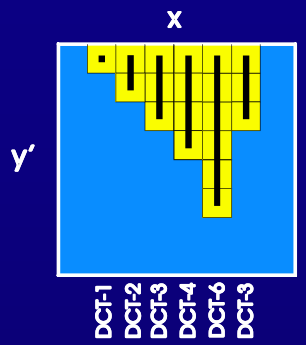
Motion Estimation/Compensation (2/3)



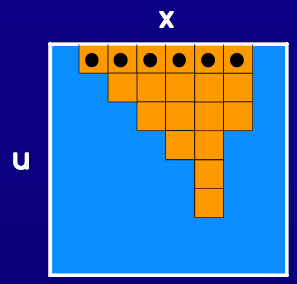
Shape Adaptive DCT for Texture Coding



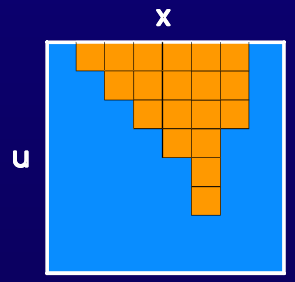
(A) Original Segment



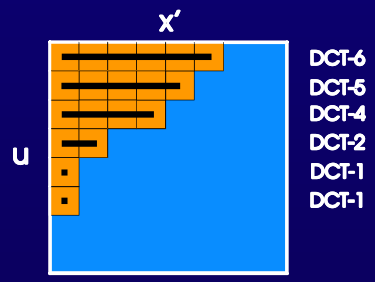
(B) Ordering of Pels and Vertical SA-DCT Used



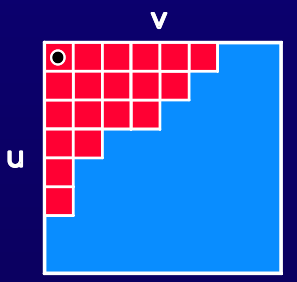
(C) Location of Pels after Vertical SA-DCT



(D) Location of Pels Prior to Horizontal SA-DCT



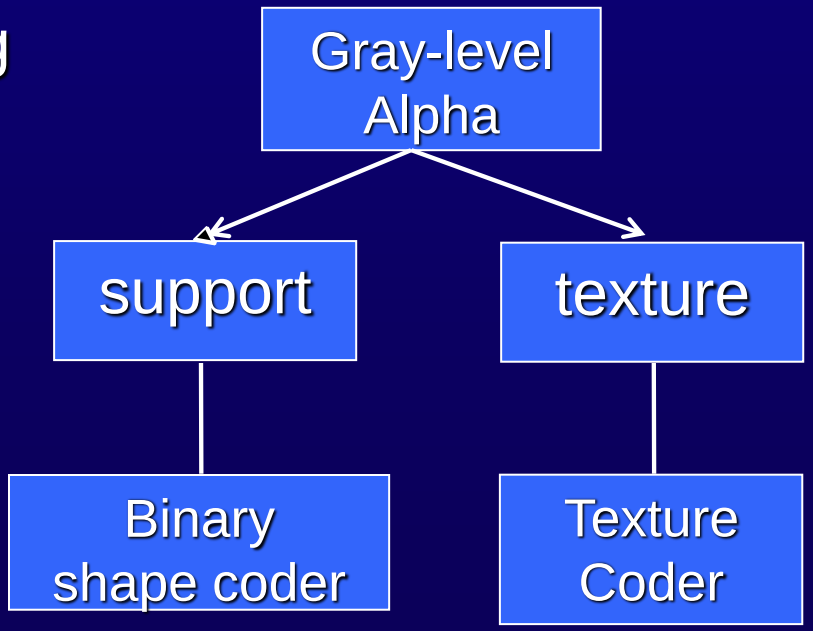
(E) Ordering of Pels and Horizontal SA-DCT Used



(F) Location of 2-D SA-DCT Coefficients

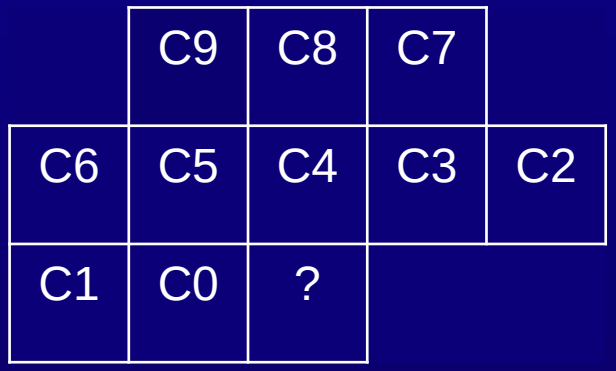
Shape Coding

- Binary shape
  - Context-based arithmetic encoding (CAE)
- Gray scale alpha plane
  - Motion compensated DCT
    - Similar to texture coding

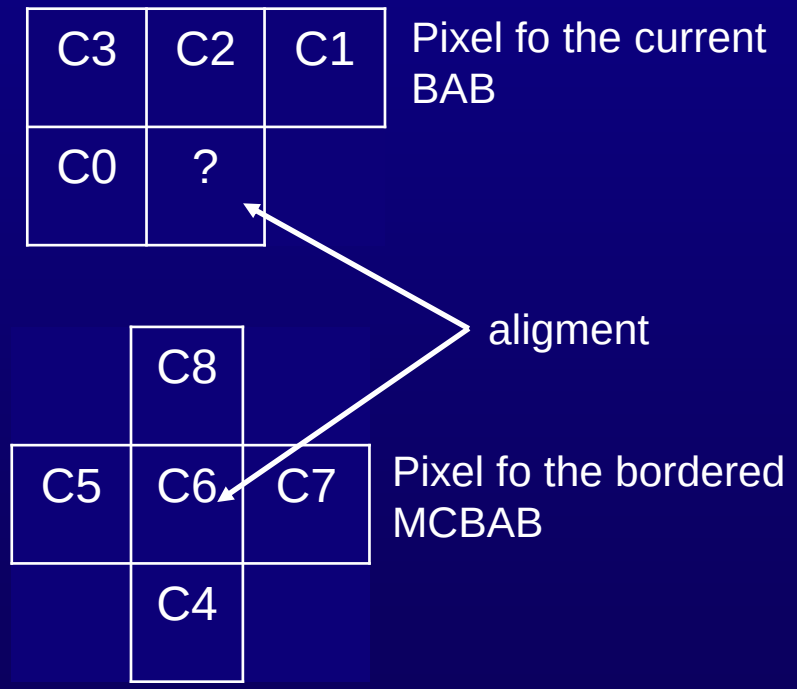


Block-based CAE

- The context



Intra



BAB: binary alpha block

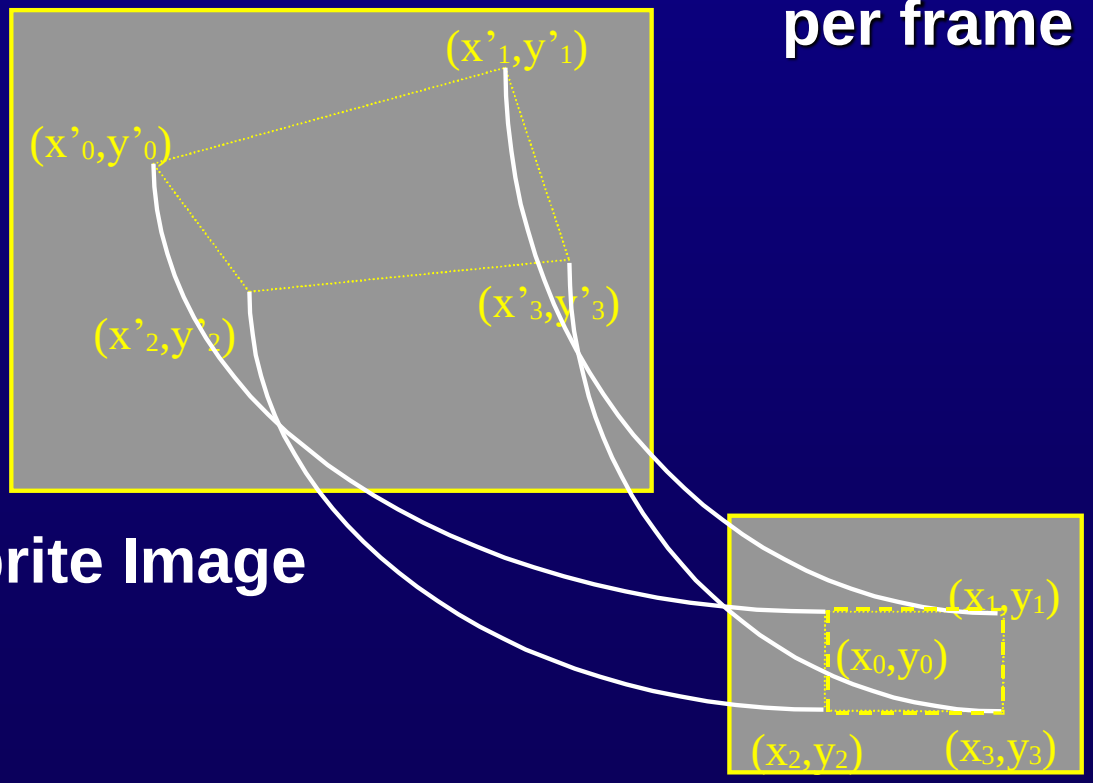
Inter

Sprite Coding

Warping of Reference Sprite

Sprite and sprite points

Only 2-8 global motion parameters are transmitted per frame



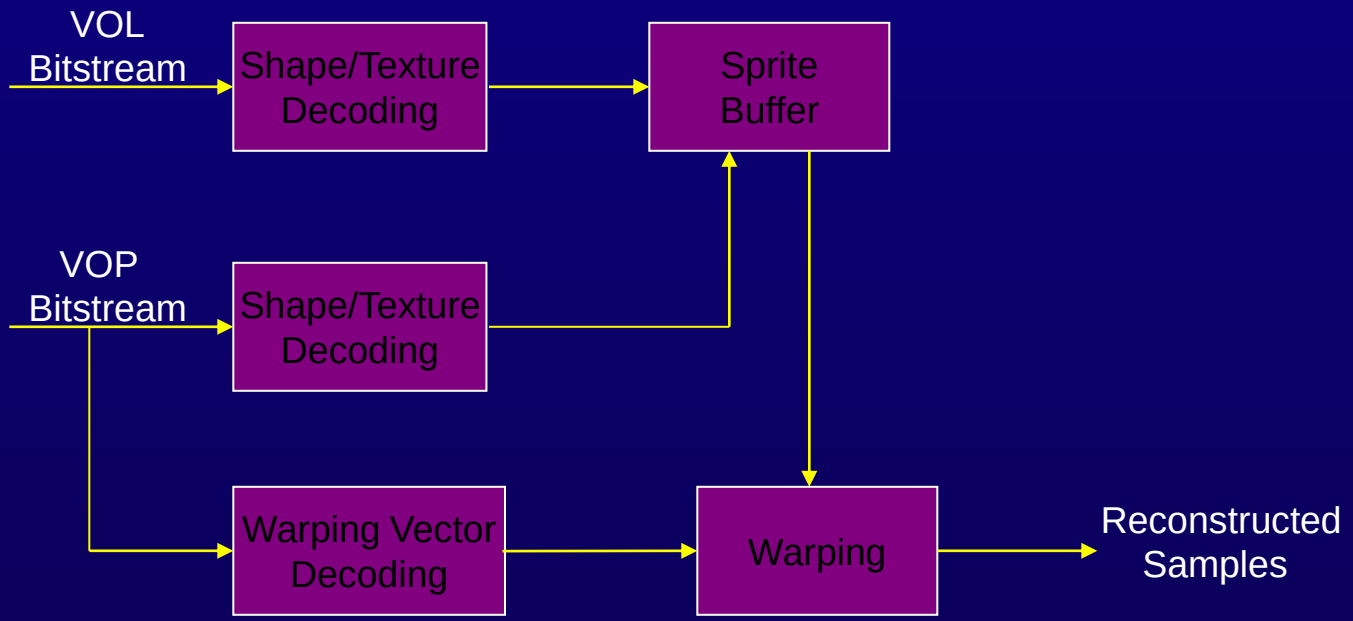
Sprite Image

Actual Frame

VOP and reference points

Static Sprite Coding Tools

- Based sprite coding
- Low latency sprite coding
- Scalable sprite coding



MPEG-4 Video verification model

- Verification model (VM8)
  - Video object plane (VOP) structure
  - Polygonal matching for motion estimation
  - Padding
  - Motion/texture coding derived from H.263
  - Binary and gray-scale shape coding
  - B-VOPs (video object plane) derived from H.263 B-pictures and MPEG-1/2 B-pictures

SNHC

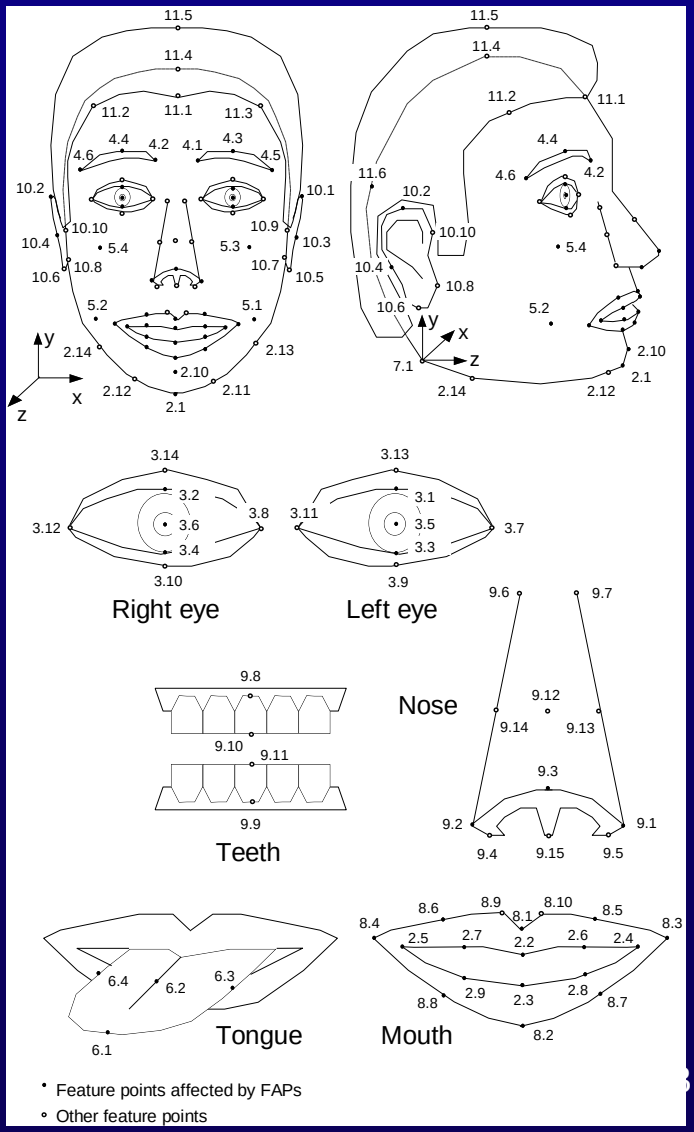
- Synthetic & Natural Hybrid Coding (SNHC)
  - To provide efficient representation and composition of synthetically and naturally generated audiovisual data
  - Compression of geometry, integration of mixed media, synthetic and spatial audio, facial and body animation
  - Applications

    Virtual environment, conferencing, education/entertainment, media production, and real-time interactive and broadcast media experiences

MPEG-4 Visual Face Animation

- Application:
  - very low rate video (as low as 100 b/s)
  - user interface to intelligent agents
- Model-based coding of faces
- Models are not normative
- Type of data:
  - Facial definition parameter
  - Facial animation parameter
  - Rendering

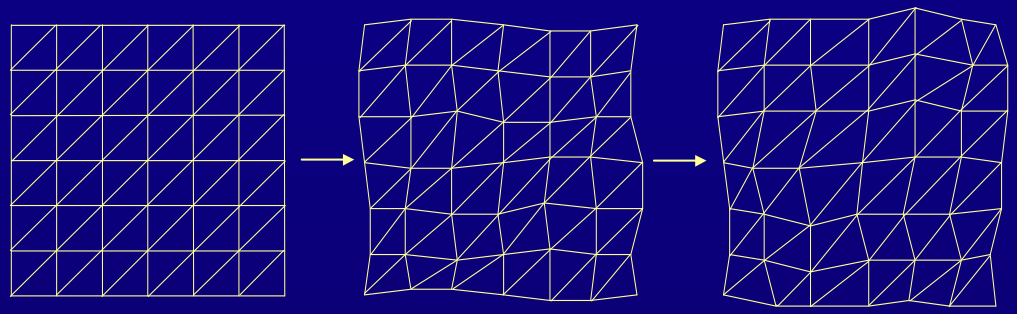
MPEG-4 Visual Face Animation



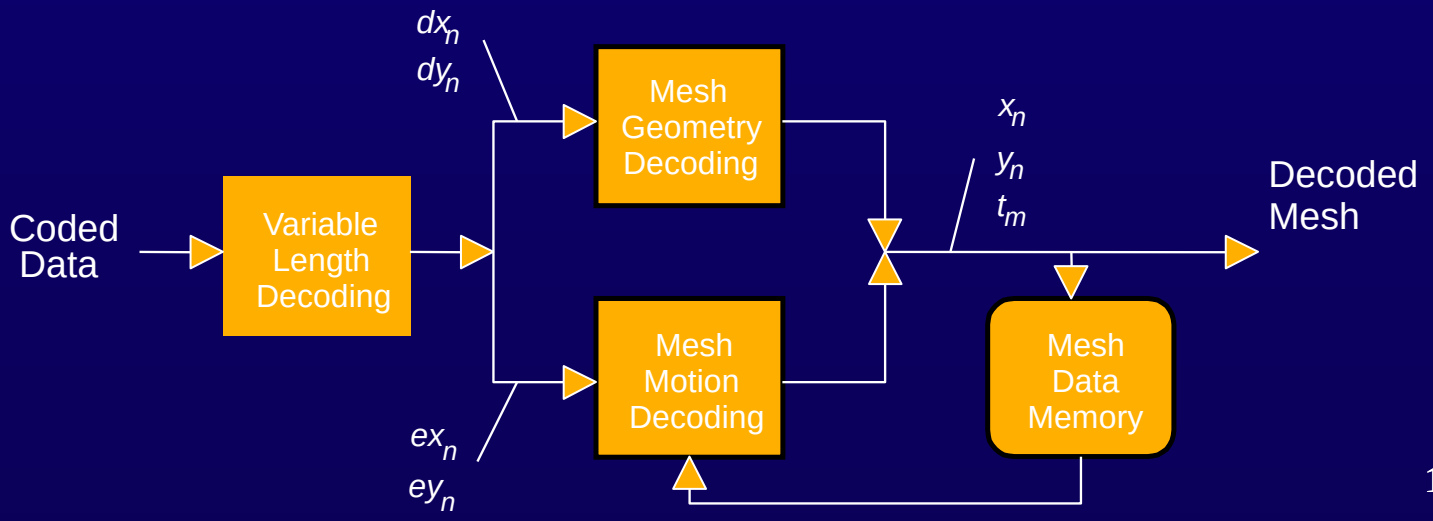
MPEG-4 Visual Mesh Coding

- Used to code and animate 2D/3D objects
- Mesh is a set of connected polygons
- Image and video can be rendered onto mesh
- Error robustness
- Mesh + still texture -> synthetic video
- Mesh is modified by motion vectors transmitted for each node

2-D Mesh Coding

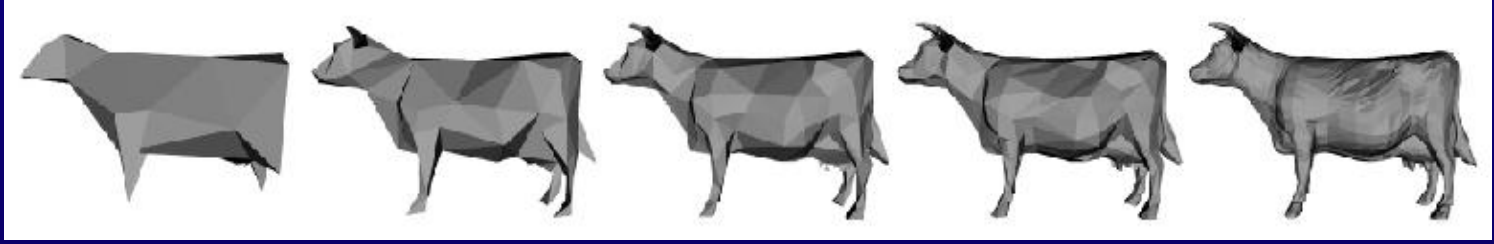


- Decoding process



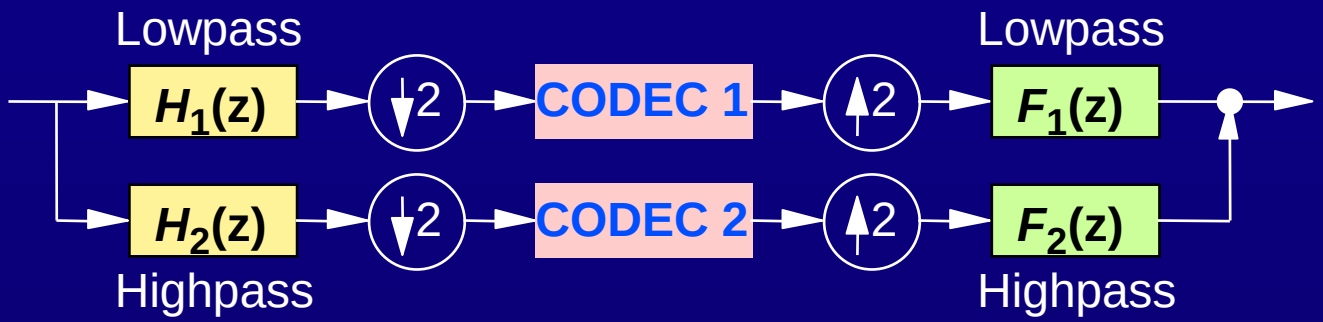
3-D Mesh Coding

- Progressive representation
  - Streaming of 3D objects
  - Both spatially and temporally



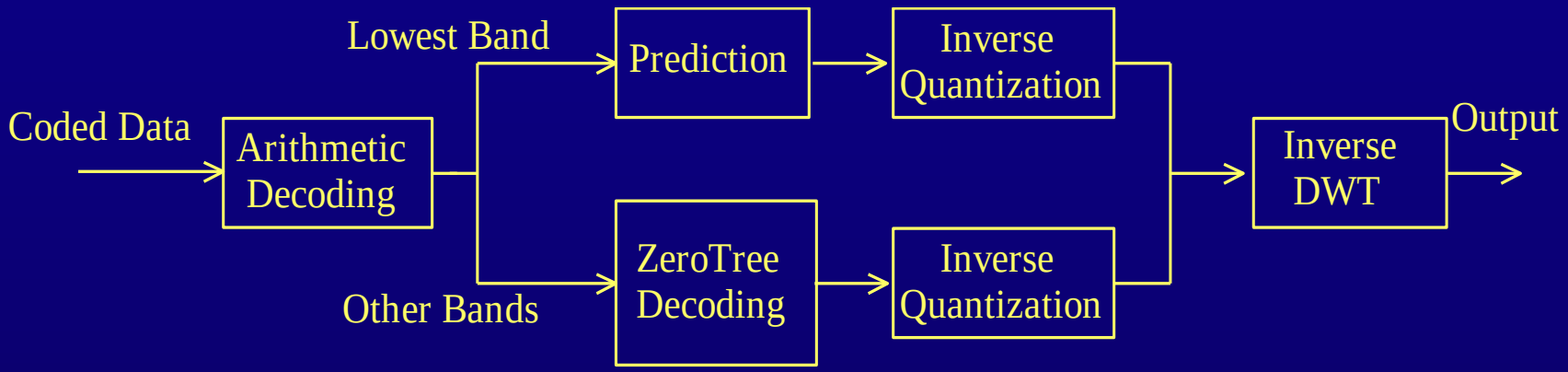
- Indexing and retrieval of 3D meshes
  - Multiresolution databases
  - Related to MPEG-7

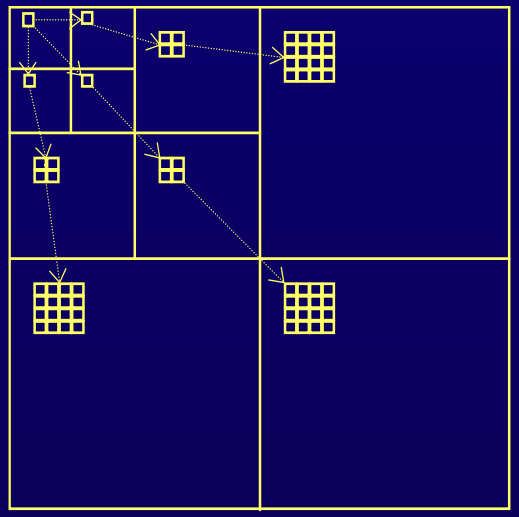
Wavelet for Scalable Texture Coding



- Decompose the signal in the frequency domain
- Critical downsampling maintains the number of samples in the subbands
- For 2D case, separable filters are often used. Decompose into four bands: LL, LH, HL, HH
- Decompose the LL band iteratively

Wavelet for Scalable Texture Coding (Cont.)





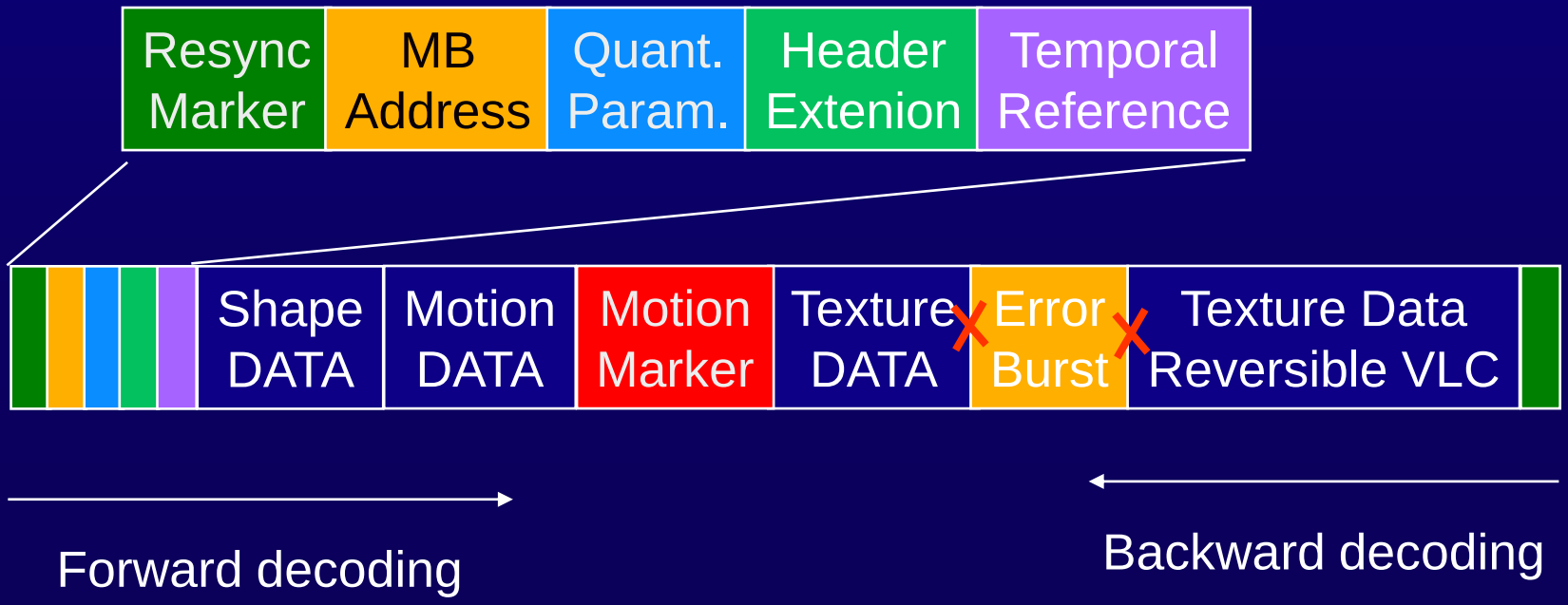
"Zero Tree"

MPEG-4 Visual Still Texture Coding

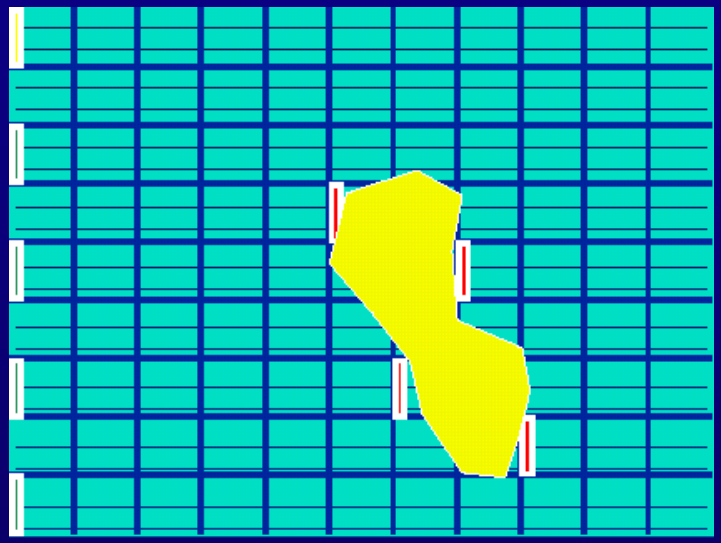
- Wavelet image coder
- Intended for rendering of mesh and face objects
- Fine granularity scalability (FGS) supported
- Error robustness

Error Resilience Tools

- Resync marker can be placed at any MB boundary (resync with MB address) (e.g. resync evenly spaced in bit-stream)
- Separating motion vectors and the texture data with resync
- RVLC (reversible variable length coding)
- HEC (header extension code)



Error Resilience Tools- Resync Marker



Picture Start Code

MPEG-4 Resync Marker

H.263 Resync Marker

H.263 Birstream

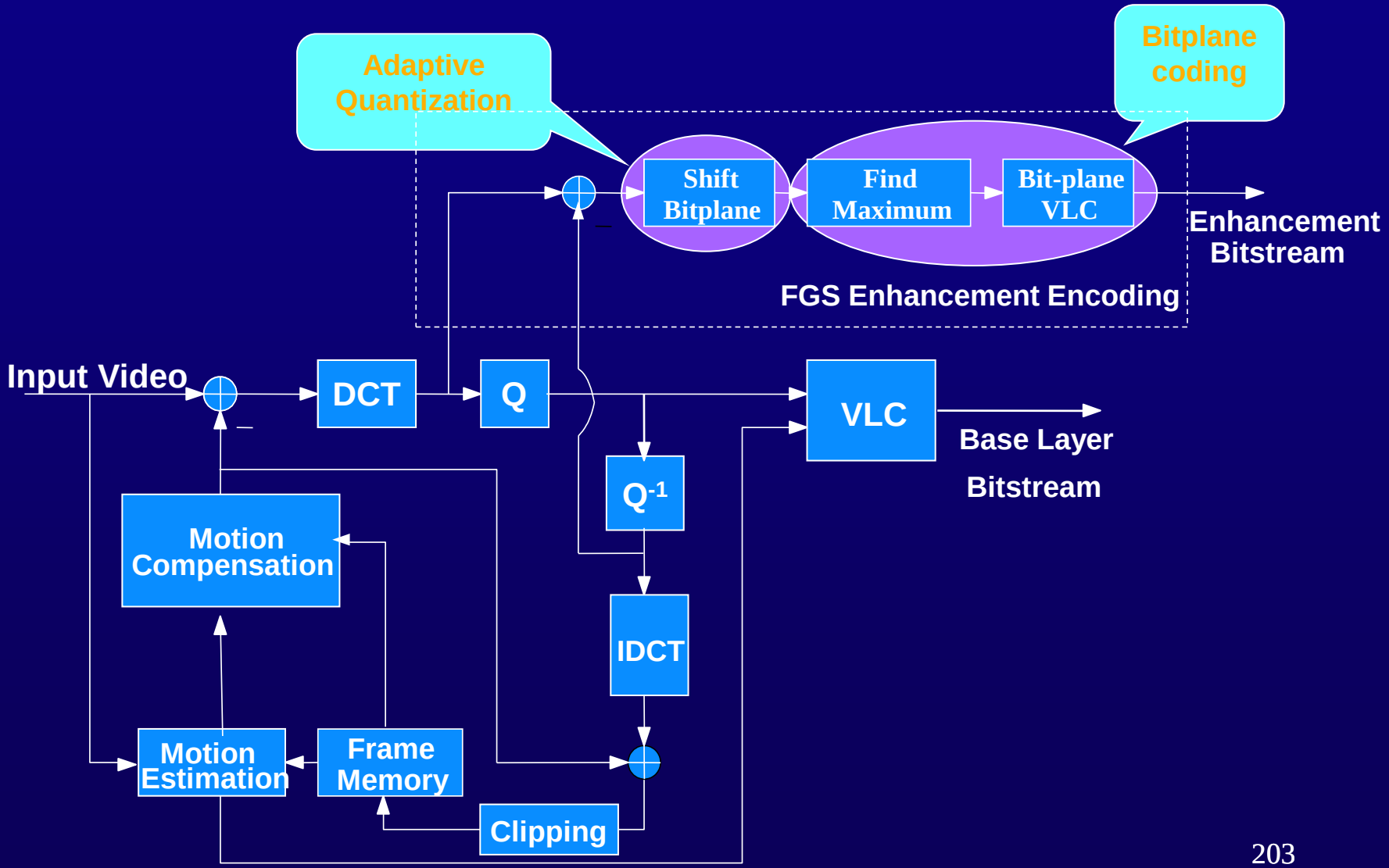
MPEG-4 Bitstream

MPEG-4 Fine Granularity Scalability (FGS)

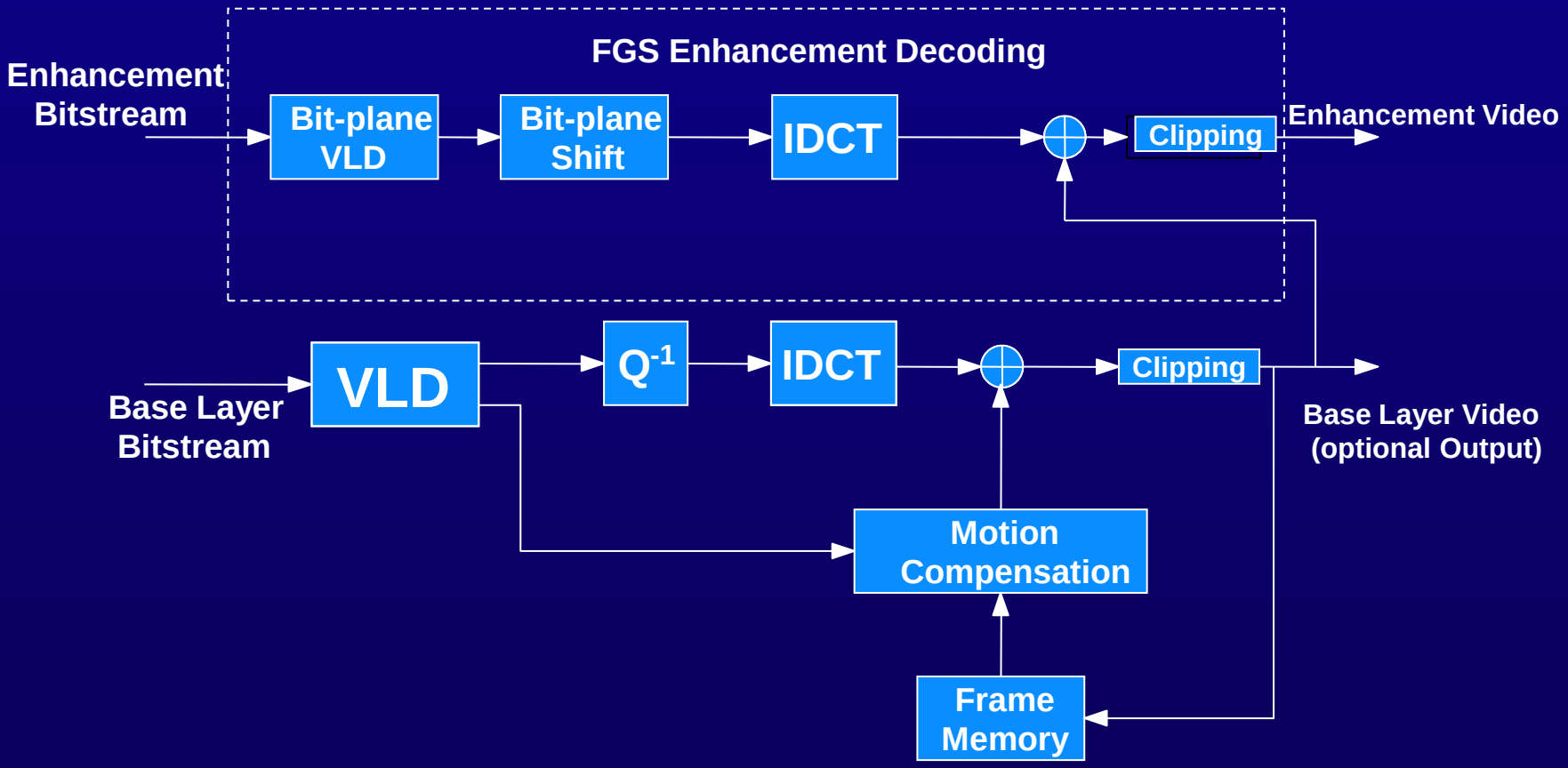
- ***Internet applications***
- ***broadcast applications over packet networks***

- Low complexity
- Supports both unicast & multicasting capabilities
- Supports various layers of SNR enhancements
- Covers a “range” of bitrates instead of a few discrete bitrates
- Base-layer compatible to MPEG-4
- Error robustness

FGS Encoder


Adaptive Quantization
Bitplane coding
Shift Bitplane
Find Maximum
Bit-plane VLC
Enhancement Bitstream
FGS Enhancement Encoding
Input Video
DCT
Q
VLC
Base Layer Bitstream
Q-1
Motion Compensation
IDCT
Motion Estimation
Frame Memory
Clipping

FGS Decoder


FGS Enhancement Decoding
Enhancement Bitstream
Bit-plane VLD
Bit-plane Shift
IDCT
Clipping
Enhancement Video
Base Layer Bitstream
VLD
Q-1
IDCT
Clipping
Base Layer Video (optional Output)
Motion Compensation
Frame Memory

Profiles: / VISUAL TOOLS:	Simple	Core	Main	Simple Scalable	N-Bit	Hybrid	Basic Animated Texture	Still Scalable Texture	Simple Face
Basic (I-VOP, P-VOP, AC/DC Prediction, 4-MV, Unrestricted MV)	X	X	X	X	X	X			
Error Resilience	X	X	X	X	X	X			
Short Header	X	X	X		X	X			
B-VOP		X	X	X	X	X			
P-VOP with OBMC (Texture)									
Method 1/2 Quantization		X	X		X	X			
P-VOP based temporal scalability (rectangular, arbitrary shaped)		X	X		X	X			
Binary Shape		X	X		X	X	X		
Gray Shape			X						
Interlace			X						
Sprite			X						
Temporal Scalability (rectangular)				X					
Spatial Scalability (rectangular)				X					
N-Bit					X				
Scalable Still Texture						X	X	X	
2D Dynamic Mesh with uniform topology						X	X		
2D Dynamic Mesh with Delaunay topology						X			
Facial Animation Parameters						X	X		X

Profiles

Profiles:

Simple: H.263-like
Simple Scalable: Simple + rectangular scalability
Core: Simple + binary shape + scalability
Main: Core + gray shape + interlace + sprite

...

Performance Baseline

- ***Coding Efficiency (MAIN Profile)***

  (optimized headers, VLC tables)

  - 30 % reduction in bitrate compared to MPEG-1

    (tested between 40-768 kbit/s)

- ***Coding Efficiency (Advanced Coding Eff. Profile (ACE))***

  (with global motion estimation)

  - 30-50% reduction in bitrate compared to MAIN Profile

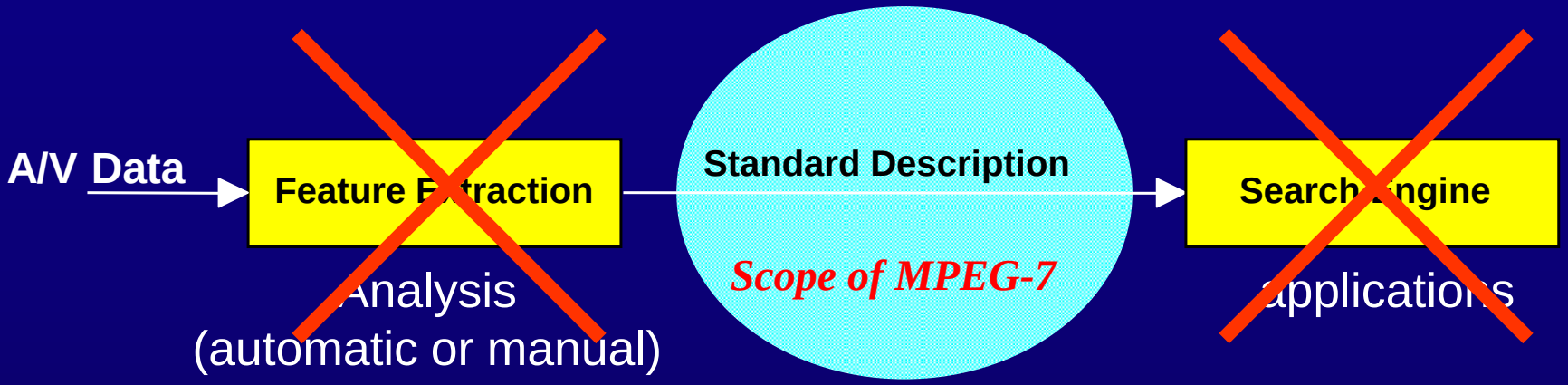
    (218 - 1000 kb/s) (critical sequences)

Universal Multimedia Access (UMA)

- Any Network:
  - wireline or wireless
  - circuit-switched or packet-switched
  - symmetrical or asymmetrical
  - broadband or bandwidth-limited
  - LAN, MAN or WAN
- Any Terminal:
  - various processing capability
  - various storage capability
  - various display capability
- Any Where:
  - seamless ubiquitous access

MPEG-7

Scope of MPEG-7

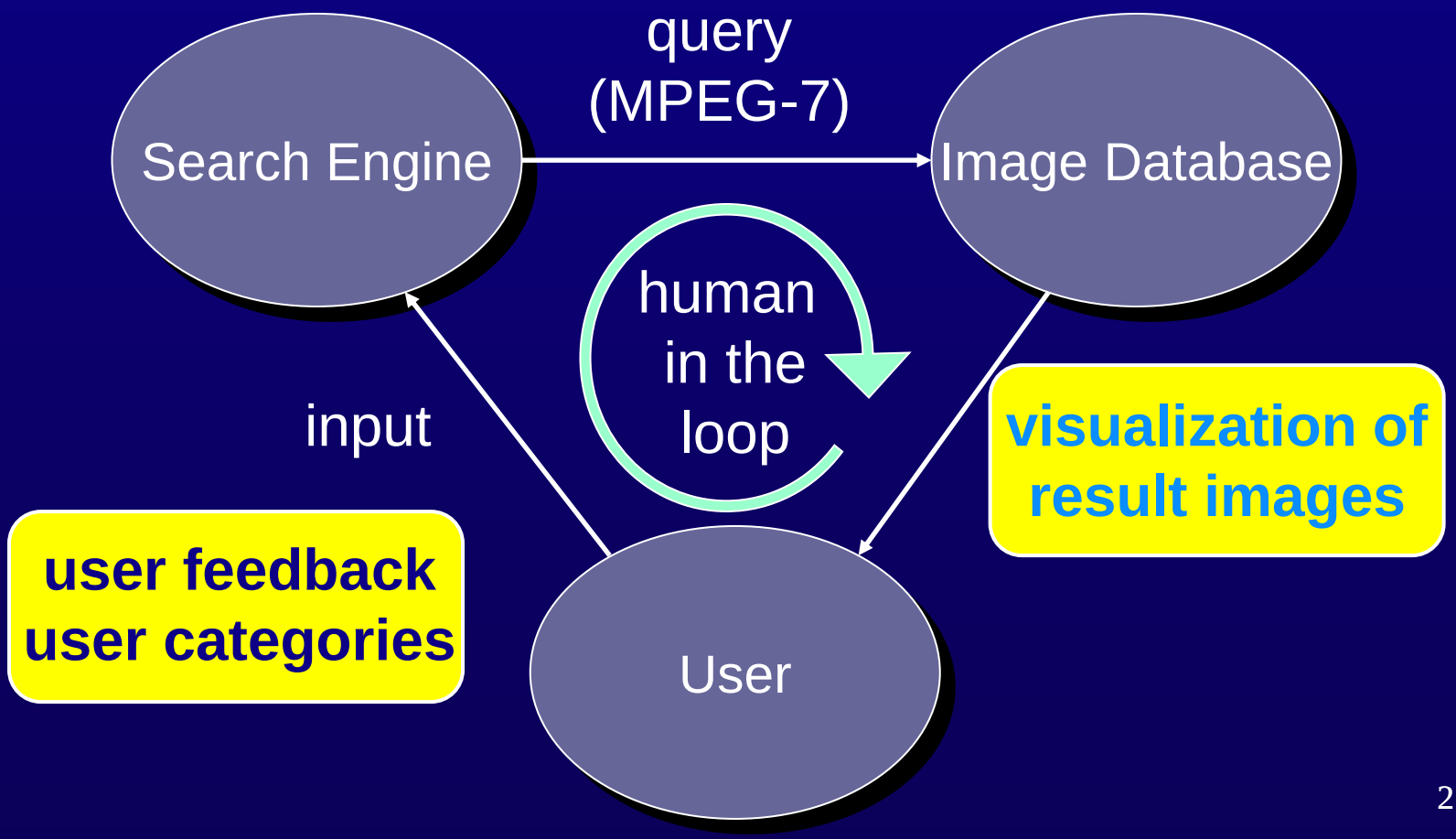


Feature extraction is outside MPEG-7

Search and query are outside MPEG-7

MPEG-7: Multimedia Content Description Interface

To define standardized solutions that allow users or agents to ***search, browse, filter*** audiovisual content.



MPEG-7: Multimedia Content Description Interface (Cont.)

How can "Non-Text-Information" be described in a useful way?

Text

Speech- and Audio-Descriptors

- *Pitch*
- *Melody*
- *Frequency*
- *Sex of speaker*

Visual Clues

- ***Texture***
- ***Color***
- ***Form and Shape (2D/3D)***
- ***Motion Activity***
- ***Higher Order Object-Models***
- ***Persons, Gestures***

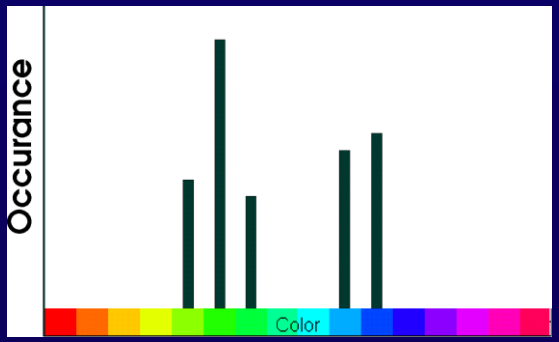
Concepts in MPEG-7

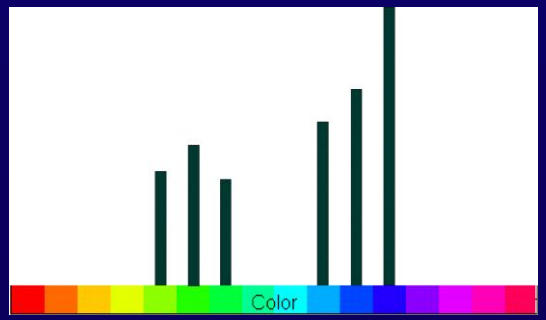
- Data
- Feature: e.g., color, motion
- Descriptor (D): e.g., histogram, motion vectors
  - Mapping between representation values and the feature
  - Basic unit of a description scheme
- Description Scheme (DS)
  - A framework that defines the descriptors and their relationships
- Description
  - An instantiation of a DS
  - Combination of descriptors and DS's

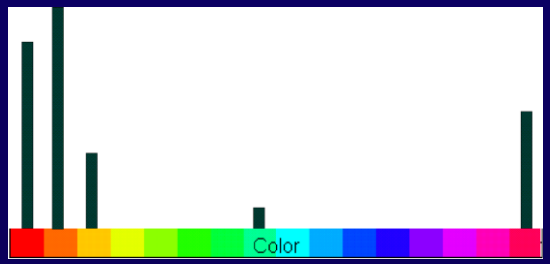
Descriptors (D's)

Color histograms
Dominant colors
DCT coefficient color description
Color composition (local)
Gabor wavelet (global)
Edge descriptor (Local)
Shape
Contour
Camera motion (global)
Object motion (Local)
MB motion activity (local and global)
...

Color Histogram

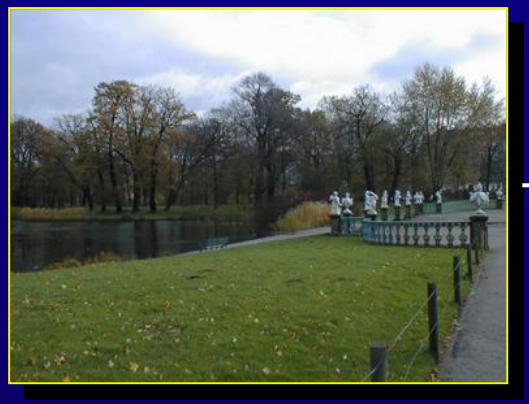


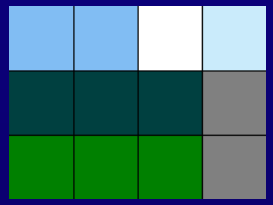


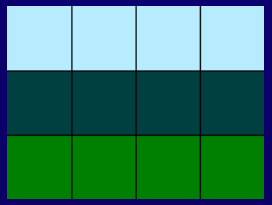


Color Composition

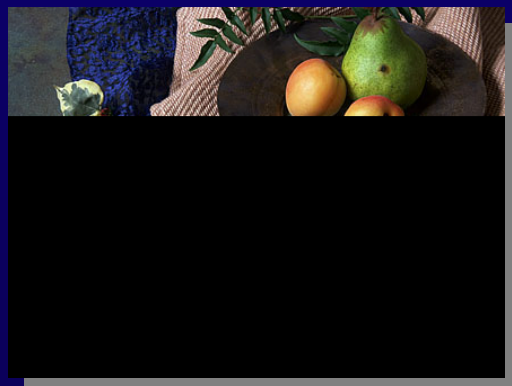
Image Data

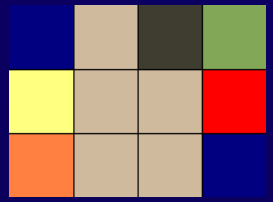




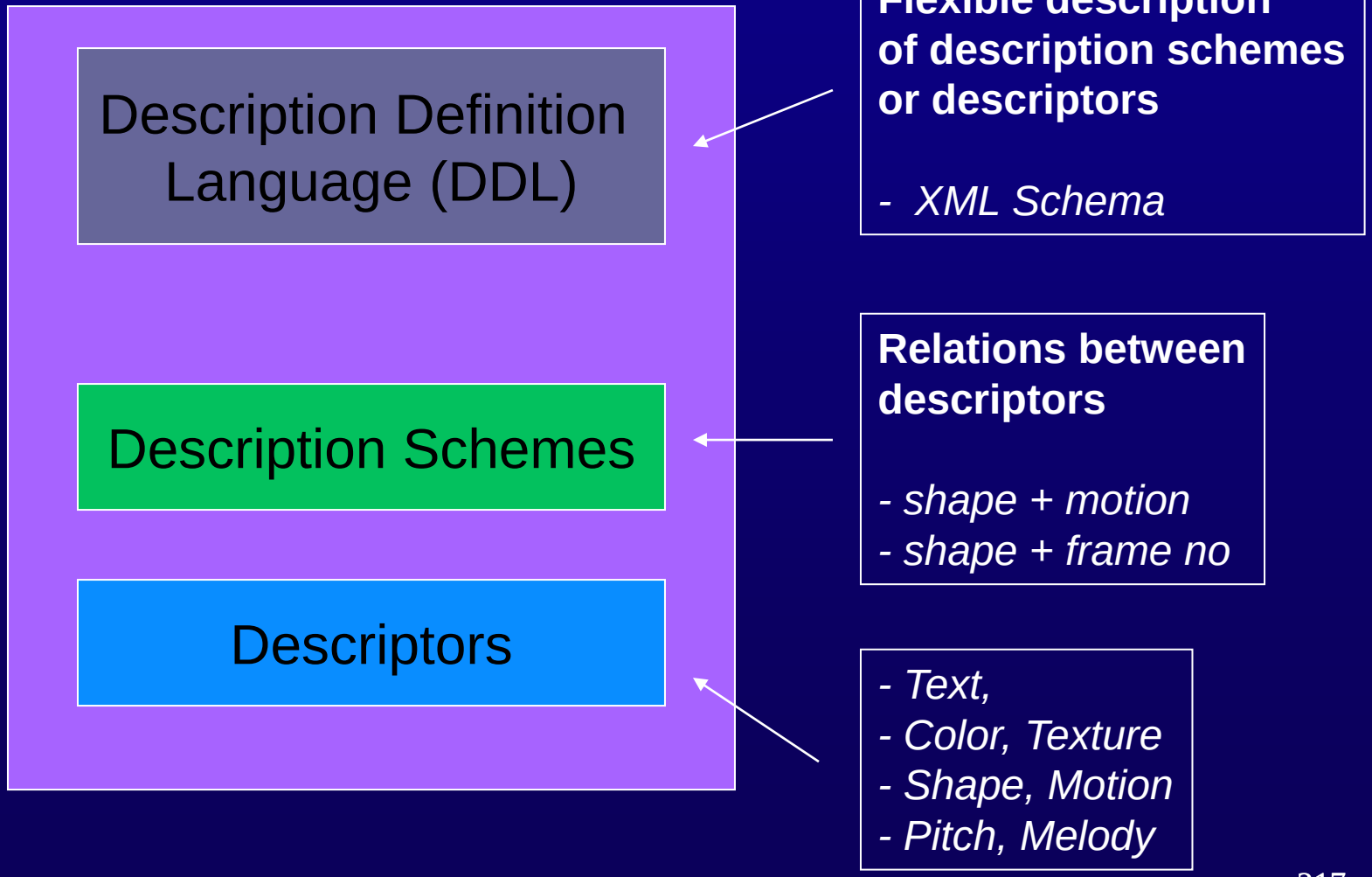


Query





What MPEG7 Standardize?



Example Uses

- Music
  - Play a few notes on a keyboard or whistle a melody
  - Retrieve a list of musical pieces that are similar
- Graphics
  - Draw a few lines on a screen
  - Retrieve a set of images containing similar graphics
- Images
  - Define objects, including color patches or textures
  - Retrieve examples among which you select the interesting objects to compose your image

Example Uses (Cont.)

- Movement
  - On a given set of objects, describe movements and relations between objects
  - Retrieve a list of animations fulfilling the described temporal and spatial relations
- Scenario
  - On a given content, describe actions and get a list of scenarios where similar actions happen
- Voice
  - Using an excerpt of Pavarotti's voice to retrieve a list of Pavarotti's records or video clips

MPEG-4 AVC/H.264

Design Goals of MPEG-4 AVC

- High compression efficiency
- Flexible application to delay constraints appropriate to a variety of services
- Error resilience capability
- Complexity scalability
- Full specification of decoding (no mismatch)
- High quality application
- Network friendliness

Video Coding Hierarchy

- Sequence, consisting of
- Pictures, consisting of
- Slices, consisting of
- Macroblocks, consisting of
- Blocks, consisting of
- Pixels / pels

NAL

VCL

Note: for interlaced video, a picture consists of either one frame or two fields

The Features of VCL (1/3)

- Transformation
  - Integer 4x4 block transform for residual coding
  - Hardamard
    - A 4x4 transform on the DC coefficients of the 4x4 blocks in a 16x16 macroblock
    - A 2x2 transform for the DC coefficients of the 4x4 chroma blocks in a 8x8 macroblock
- Motion Estimation
  - Variable block-size motion prediction (7 block sizes: 16x16, 8x8, 4x4, 16x8, 8x16, 8x4, 4x8)
  - Integer, 1/2-, and 1/4-pixel motion vector accuracy
  - Multiple reference frames (max. 15) may be used for prediction

The Features of VCL (2/3)

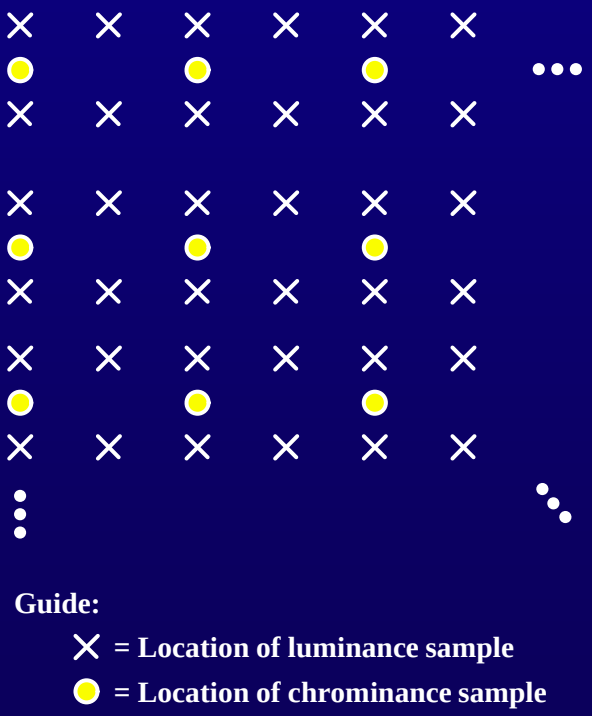
- Entropy coding:
  - Context-based adaptive variable length coding (CAVLC)
  - Context-based adaptive binary arithmetic coding (CABAC)
- Others:
  - Space-domain Intra prediction (10 prediction modes)
  - De-blocking loop filter
  - Motion vector prediction
  - Slice structure
  - Interlace coding tools

Frame Types

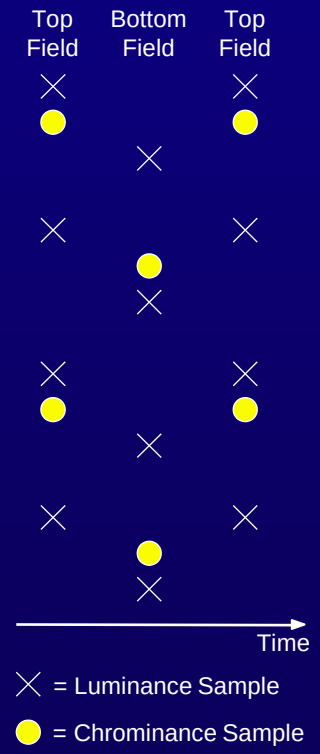
- I-frame
- P-frame
- B-frame
- SP- and SI-frames
  - SP and SI frames provide functionalities for bit-stream switching, splicing, random access, VCR functionalities, and error resilience/recovery

Picture Formats

- Color sequences using 4:2:0 chroma sub-sampling



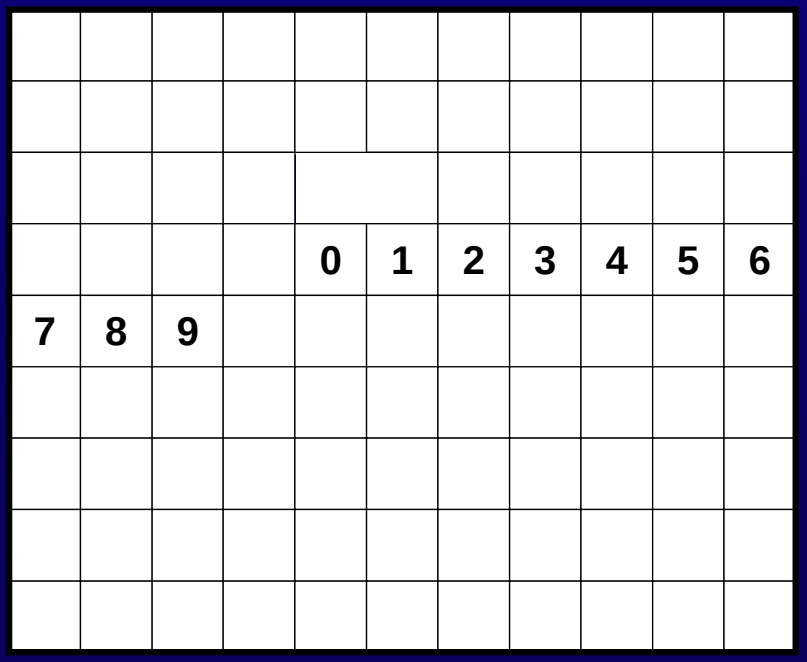
progressive frames

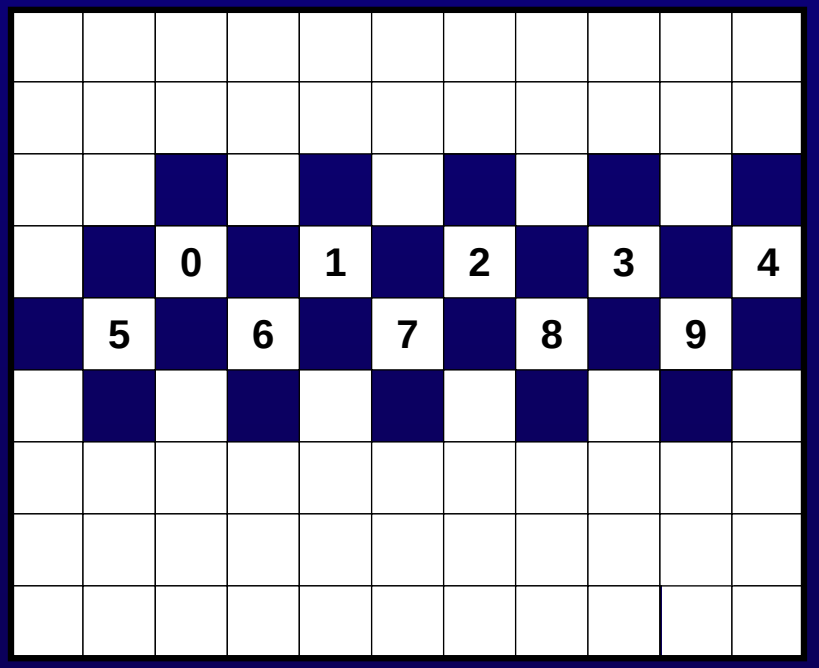


interlaced frames

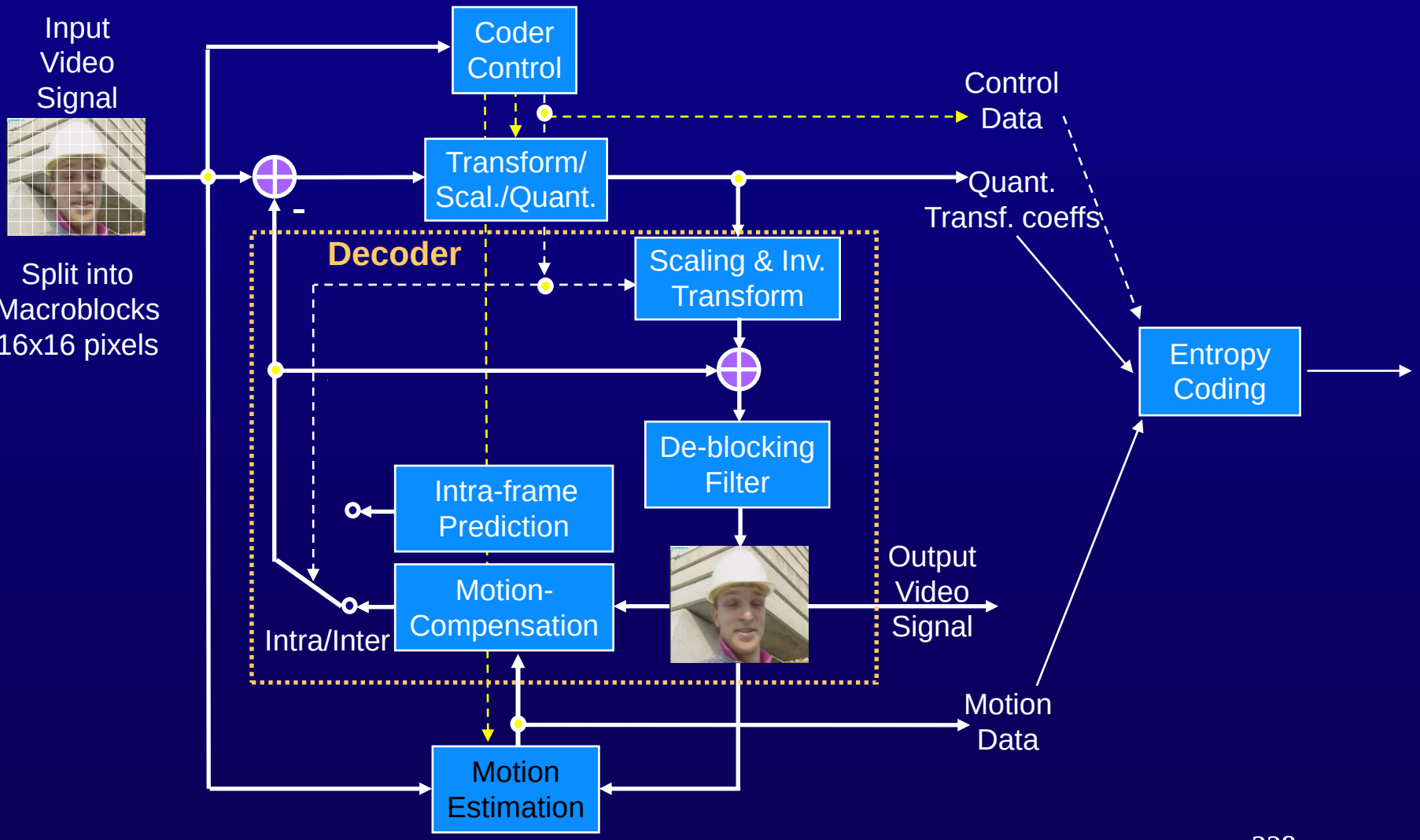
Macroblock Subdivision

- Each picture is divided into 16x16 macroblocks.
- The order of the macroblocks in the bitstream depends on the macroblock allocation map (MAM) and is not necessarily raster scan order

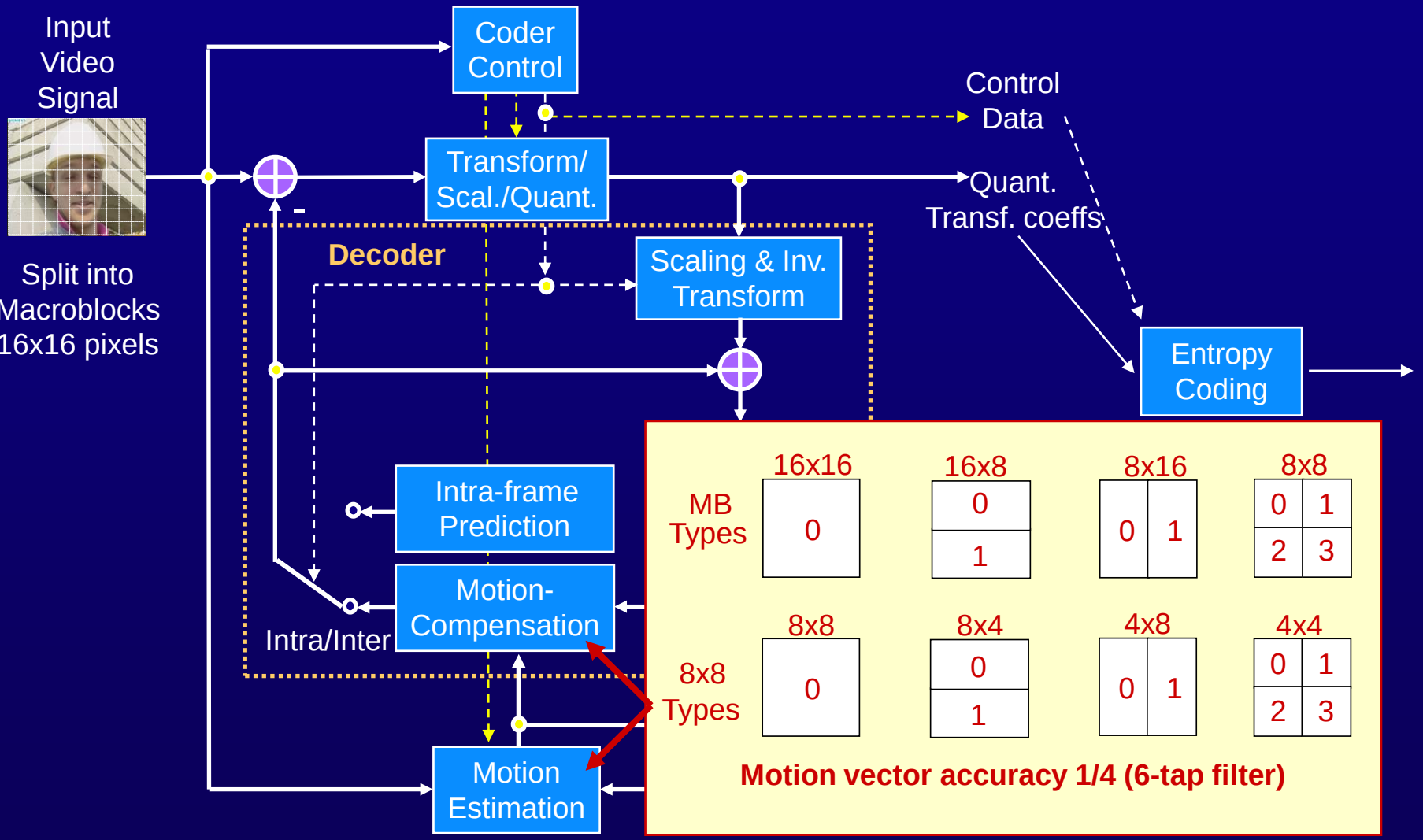




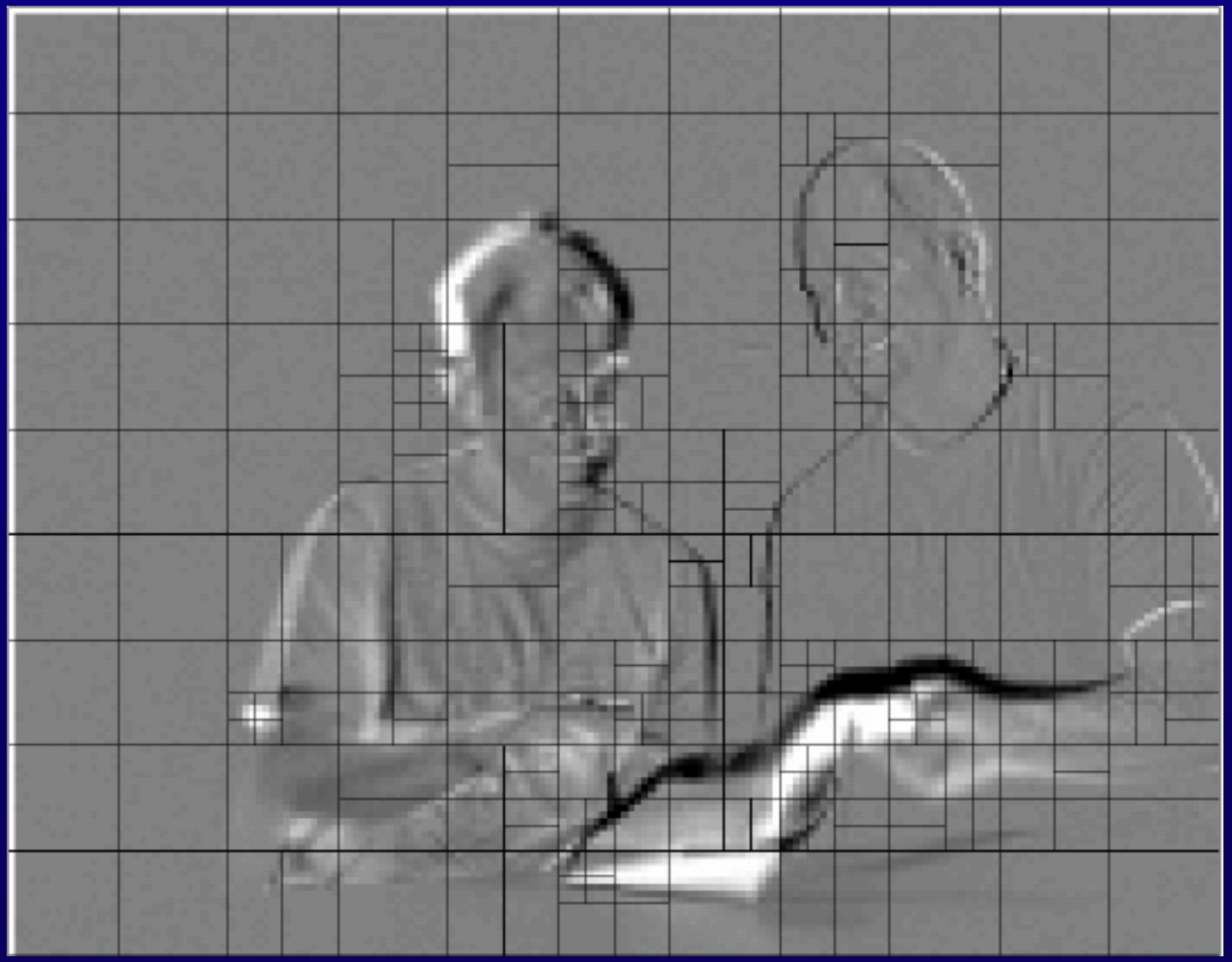
MPEG-4 AVC/H.264: Encoder Architecture



MPEG-4 AVC/H.264: Motion Compensation



Motivation of Variable Block-Size Coding



Motion Compensation

- Various block sizes and shapes for motion compensation
- 1/4 sample accuracy
  - 6 tap filtering to 1/2 sample accuracy
  - simplified filtering to 1/4 sample accuracy
  - special position with heavier filtering
- Multiple reference pictures
- Temporally-reversed motion and generalized B-frames
- B-frame prediction weighting

Block Modes of P Pictures

- Macroblock: 16x16
- 7 motion prediction modes
  - 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4
  - Motion vectors accuracy: integer, ½-, and ¼-pixel

Mode 1

0

Mode 2

0	1

Mode 3

0
1

Mode 4

0	1
2	3

Mode 5

0	1	2	3
4	5	6	7

Mode 6

0	1
2	3
4	5
6	7

Mode 7

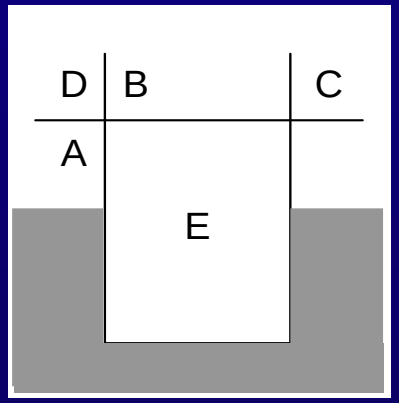
0	1	2	3
4	5	6	7
8	9	10	11
12	13	14	15

Motion Vector Search

- Motion Mstimation
  - Integer pixel search
  - Fractional pixel search (1/2- and 1/4-pixel)
  - Reference frames selection from multiple reference frames (max. 15 frames)
  - Search range:
    - horizontal [-2048, 2047.75] (max)
    - vertical [-512, 511.75] (max)

Motion Estimation

- Motion vector prediction
  - In same slice
  - Median prediction (except 16x8 and 8x16 blocks)



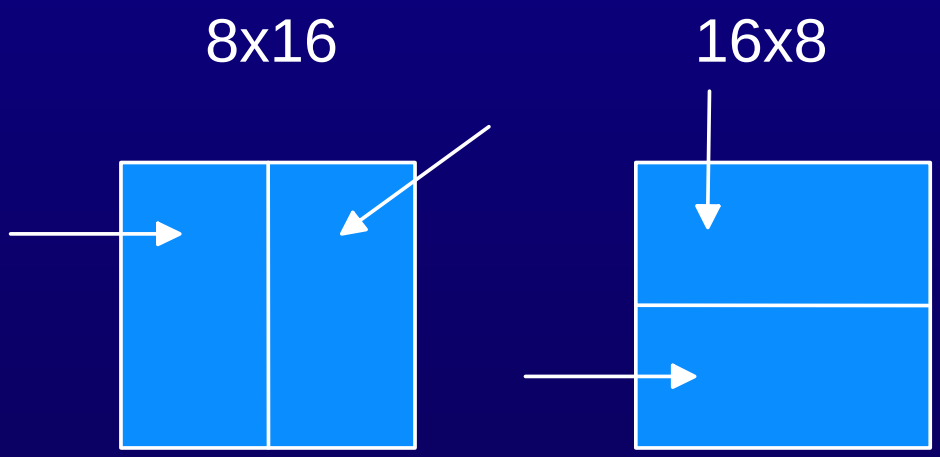
A, B, C, D and E may come from different reference picture

$$V_1 = median\{V_A, V_B, V_C, V_D\}$$

1. C is not available, $V_C = V_D$
2. B,C, and D are not available, $V_B = V_C = V_D = V_A$
3. Any predictor is neither of above two rules, its MV is 0.

Motion Estimation

- Motion vector prediction for 16x8 and 8x16 blocks
  - Directional segmentation prediction



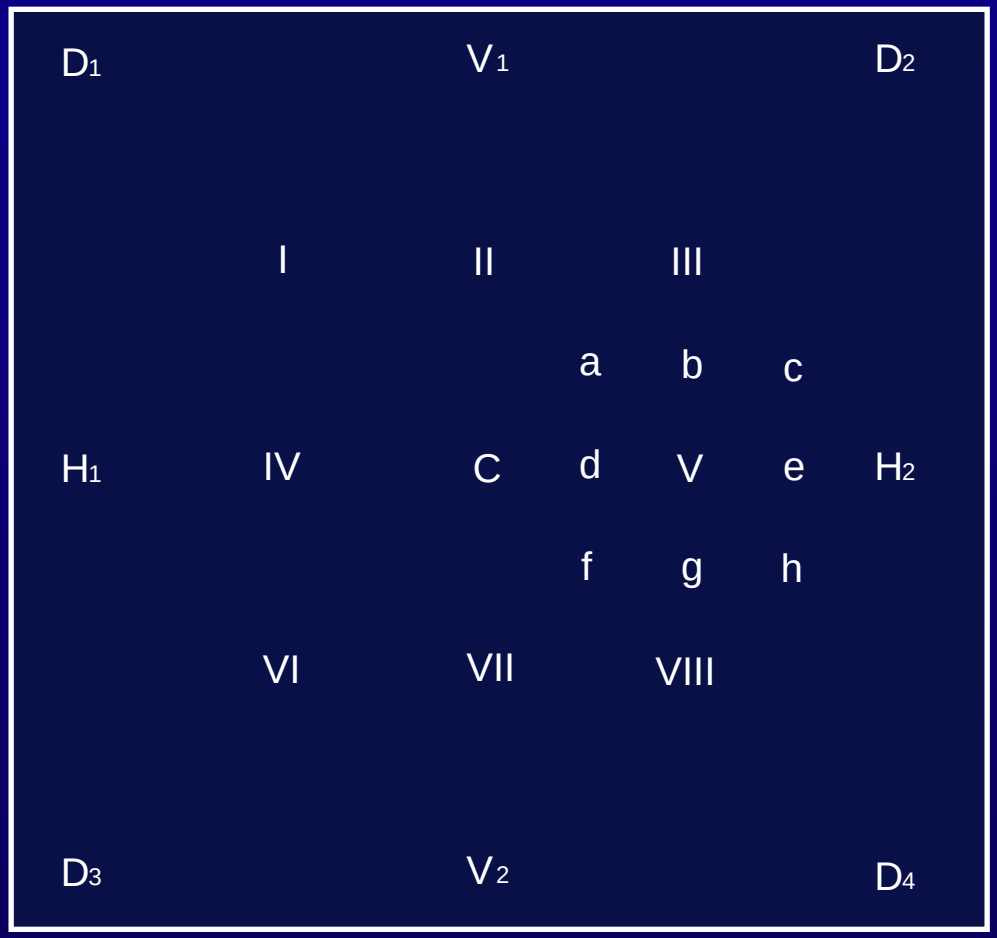
Motion Estimation

- Integer pixel search
  - search positions are organised in a "***spiral***" structure around the predicted vector

```
. . . . . .
. 15  9 11 13 16
. 17  3  1  4 18
. 19  5  0  6 20
. 21  7  2  8 22
. 23 10 12 14 24
```

Motion Estimation

- full fractional-pixel search (1/2- and ¼-pixel)



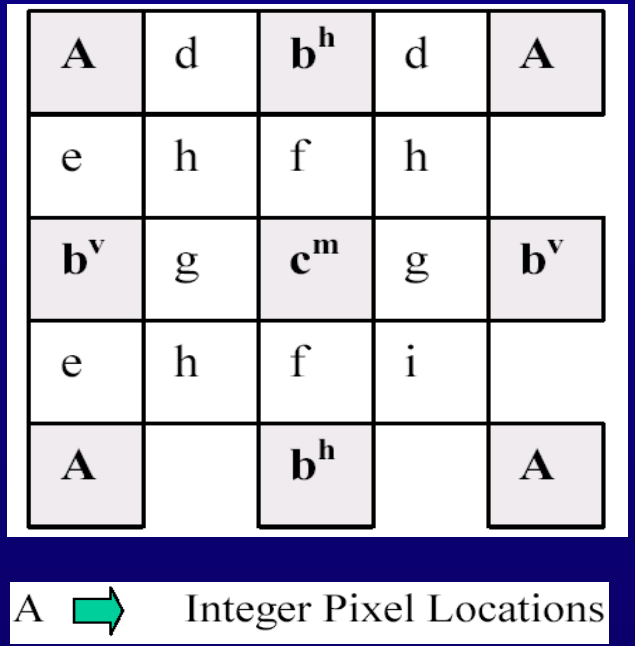
Capital letters ($C, H_1, H_2 \ldots$) : integer pixel positions
Roma numbers (I,II,III...): 1/2-pel positions
Lower case letters(a,b,c...):1/4-pel positions

Motion Estimation

- Fractional pixel search
  - Check the eight 1/2-pel candidates, I ~ VIII around the best integer-pel C; decide the best 1/2-pel V subject to the minimal cost among the 1/2-pel candidates
  - Check the eight 1/4-pel candidates, a ~ h around the best 1/2-pel V, decide the best 1/4-pel h subject to the minimal cost among the 1/4-pel candidates
  - Select the motion vector and block-size pattern, which produces the lowest cost

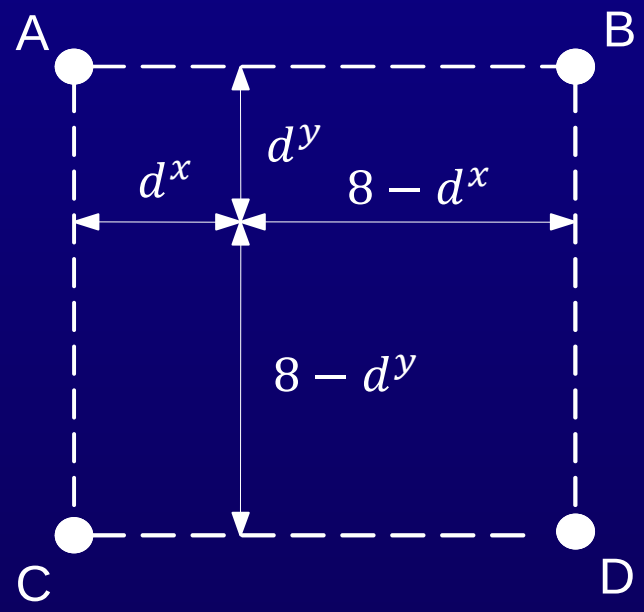
Fractional Pel Value Interpolation: Luma

- Calculate half-pel values
  - use 6-tap filter {1, -5, 20, 20, -5, 1} to get b
  - b^h= clip(((b+16)>>5))
  - c from b values using the 6 tap filter
  - c^m = clip(((c+512)>>10))
- Average of integer and half-pel values to find d, e, f, g
  - e.g., d = (A+ b^h)>>1
- h = (b^h + b^v)>>1 (diagonal direction averaging)
- i = (A1+A2+A3+A4+2)>>2



Fractional Pel Value Interpolation: Chroma

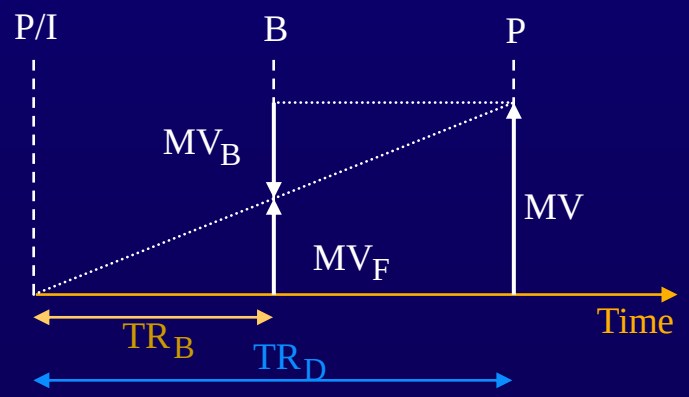
- d^y are the fractional position in units of one eighth samples
- A, B, C, and D are integer pixels



$$v = ((8 - d^x)(8 - d^y)A + d^x(8 - d^y)B + (8 - d^x)d^yC + d^xd^yD + 8^2/2)/8^2$$

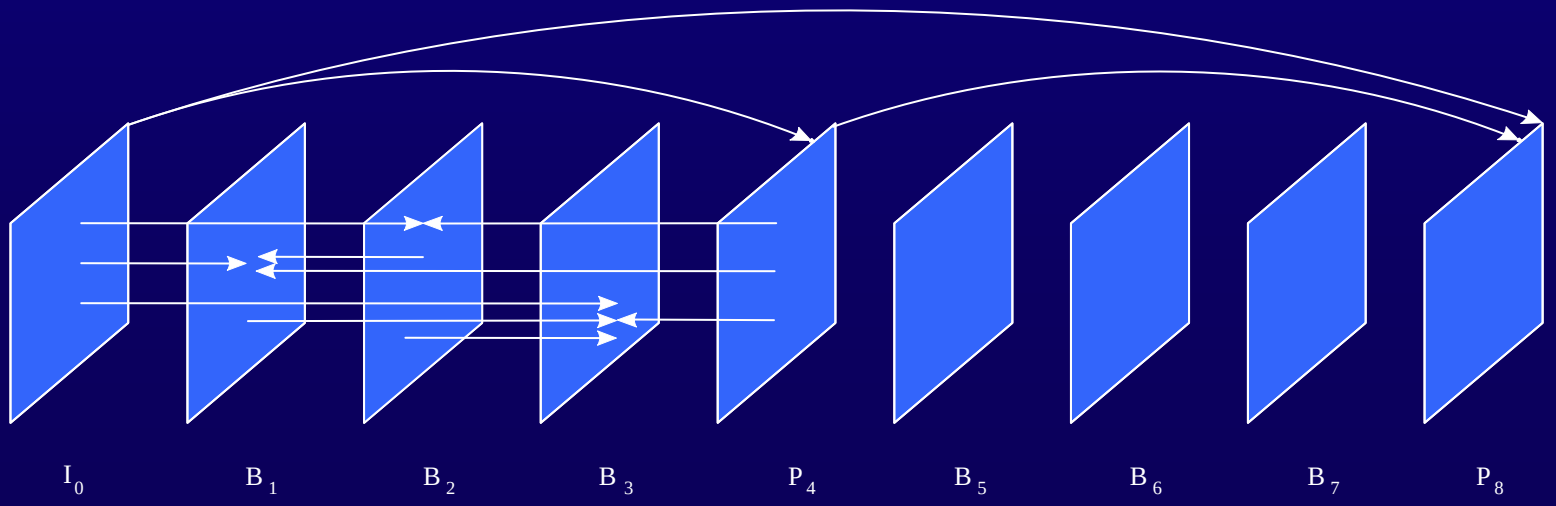
B-Pictures

- Advantages:
  - Improve coding efficiency
  - Provide temporal scalability
- 5 modes:
  - Direct mode: derived forward and backward MVs, none transmitted
  - Forward mode: prediction from a previous reference frame
  - Backward mode: prediction from a subsequent reference frame
  - Bi-directional mode: separate forward and backward MVs
  - Intra prediction mode
- MVs in direct mode:
  - $MV_F = (TR_B * MV)/TR_D$
  - $MV_B = (TR_B - TR_D) * MV/TR_D$

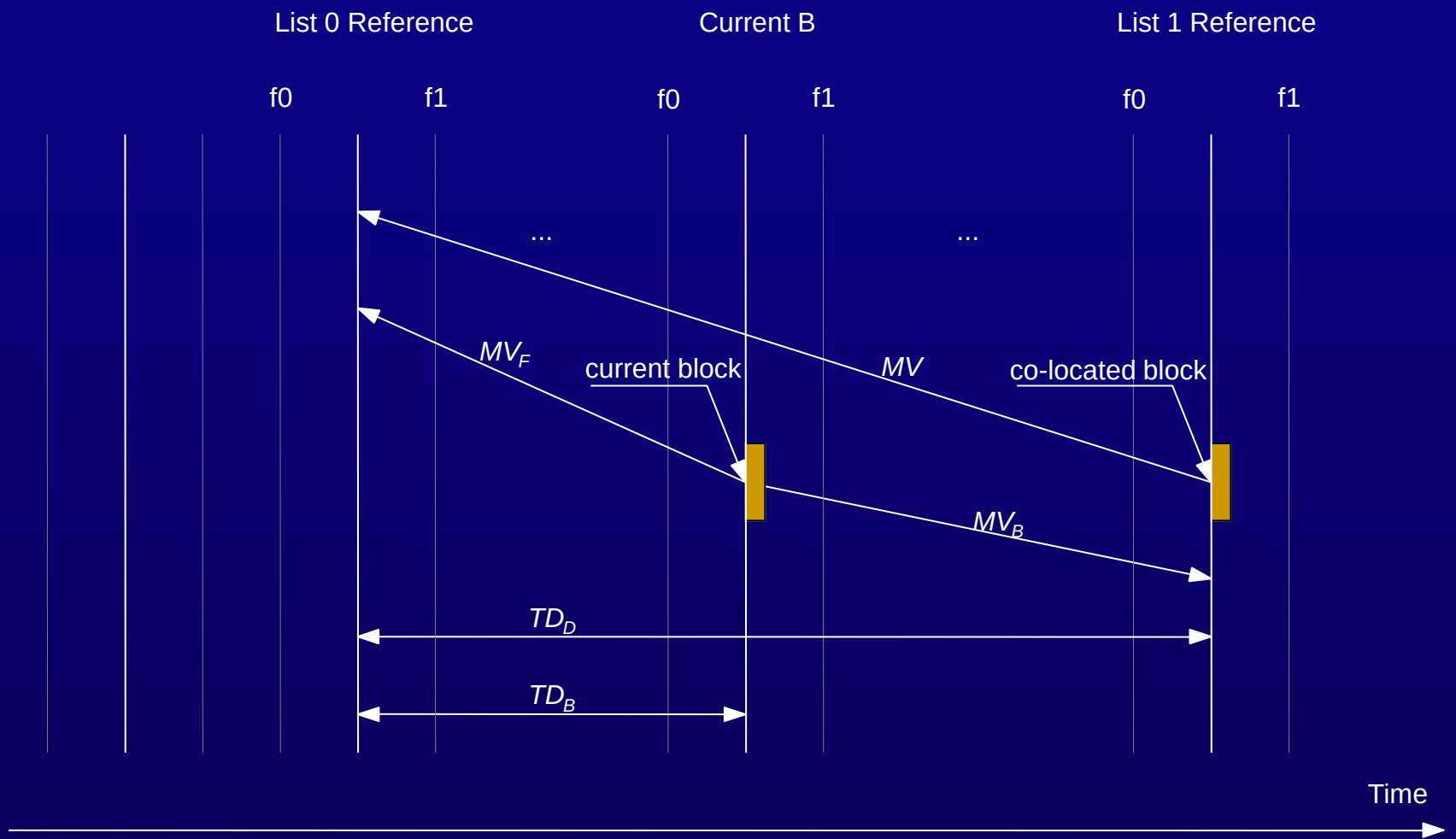


B-Pictures

- Direct Mode
  - No MV data is transmitted
  - Same block structure as co-located MB in temporally subsequent picture
  - MVs are computed as scaled version of corresponding MV of the co-located MB



B-Pictures



$Z = (TD_B \times 256) / TD_D$

$W = Z - 256$

$MV_F = (Z \times MV + 128) >> 8$

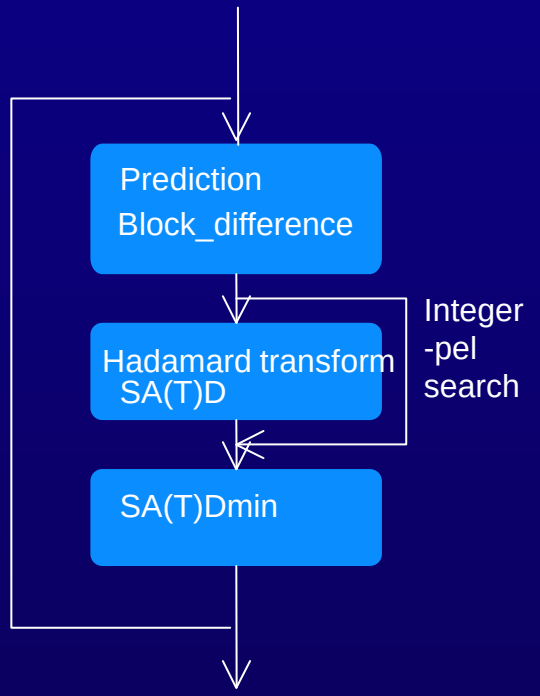
$MV_B = (W \times MV + 128) >> 8$

Mode Decision

- Block difference
  - *Diff(i,j) = Original(i,j) - Prediction(i,j)*
- SAD and SATD
  - *DiffT* means apply Hadamard transform to *Diff*

$$SAD = \sum_{i,j} |Diff(i,j)|$$

$$SATD = (\sum_{i,j} |DiffT(i,j)|)/2$$



Loop for prediction mode decision

Mode Decision

- Given the last decoded frames, Lagrange multipliers

  and the MB quantisation parameter QP.

  (Note: L_{MODE} for B or SP frame is 4 times as much as that for I or P frame.)

$$L_{MODE} = 0.85 \times 2^{QP/3},$$
$$L_{MOTION} = \sqrt{L_{MODE}},$$

Mode Decision

- Choose intra prediction modes for the Intra 4x4 macroblock mode by minimizing with

$$IMODE \in \{DC,\ HOR,\ VERT,\ DIAG\_D\,L, DIAG\_D\,R,\ VERT\_R, VERT\_L, HOR\_U, HOR\_D\}$$

- Determine the best Intra16x16 prediction mode by choosing the mode that results in the minimum *SATD*.

Mode Decision

- For each 8x8 sub-partition
  - Perform motion estimation and reference frame selection by minimizing *SSD + L x Rate(MV, REF)*
  - B frames: Choose prediction direction by minimizing *SSD + L x Rate(MV(PDIR), REF(PDIR))*
  - Determine the coding mode of the 8x8 sub-partition using the rate-constrained mode decision, i.e. minimize *SSD + L x Rate(MV, REF, Luma-Coeff, block 8x8 mode)*
- Here the *SSD* calculation is based on the reconstructed signal after DCT, quantization, and IDCT

$$SSD(s,c,MODE|QP) = \sum_{x=1,y=1}^{16,16} (s_Y[x,y] - c_Y[x,y,MODE|QP])^2 + \sum_{x=1,y=1}^{8,8} (s_U[x,y] - c_U[x,y,MODE|QP])^2 + \sum_{x=1,y=1}^{8,8} (s_V[x,y] - c_V[x,y,MODE|QP])^2 ,$$

Mode Decision

- Perform motion estimation and reference frame selection for 16x16, 16x8, and 8x16 modes by minimizing

$$J(REF, \mathbf{m}(REF)|L_{MOTION}) = SA(T)D(s, c(REF, \mathbf{m}(REF))) + L_{MOTION} \cdot (R(\mathbf{m}(REF) - \mathbf{p}(REF)) + R(REF))$$

- B frames: determine prediction direction by minimizing

$$J(PDIR|L_{MOTION}) = SATD(s, c(PDIR, \mathbf{m}(PDIR))) + L_{MOTION} \cdot (R(\mathbf{m}(PDIR) - \mathbf{p}(PDIR)) + R(REF(PDIR)))$$

Mode Decision

- Choose the MB prediction mode by minimizing

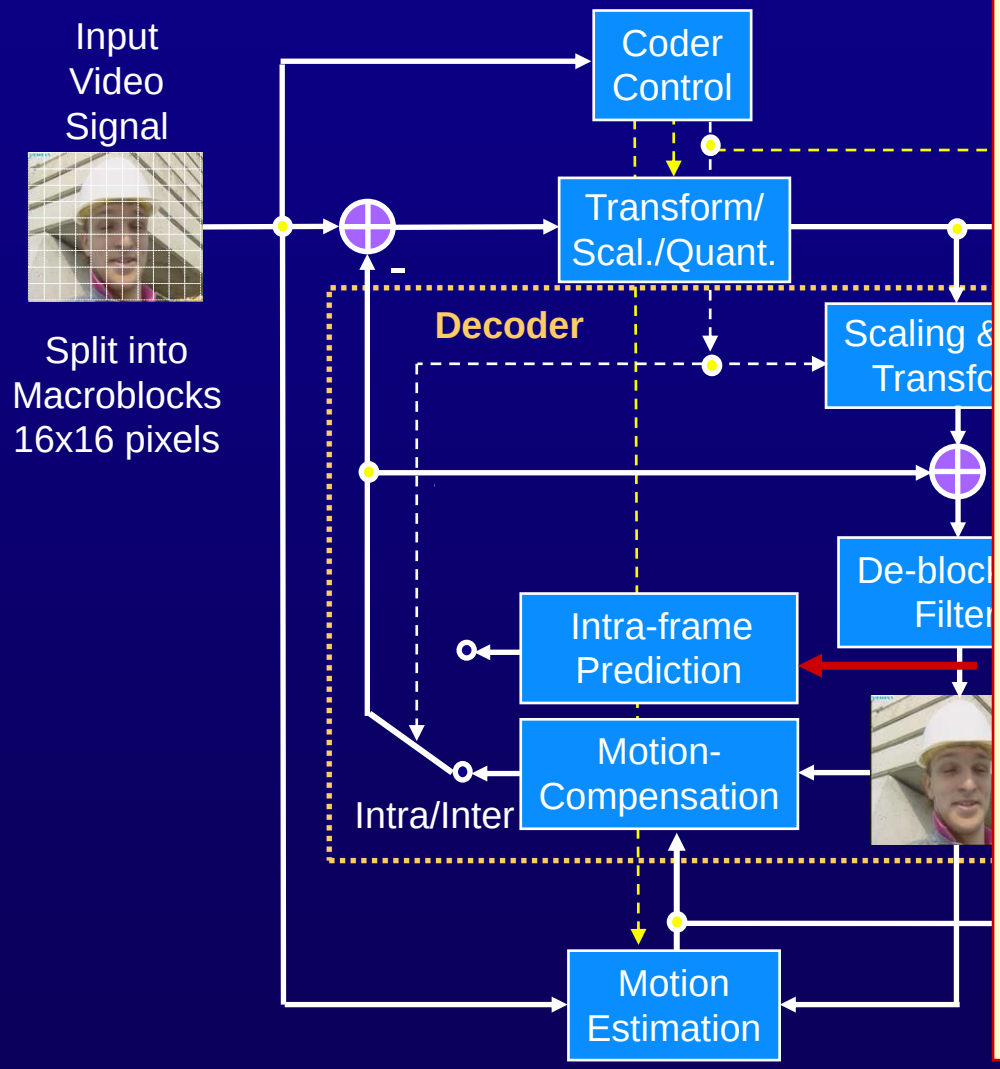
I: $$J(s,c,MODE|QP,L_{MODE}) = SSD(s,c,MODE|QP) + L_{MODE} \cdot R(s,c,MODE|QP)$$
$$MODE \in \{INTRA4\times 4,\ INTRA16\times 16\}$$

P: $$MODE \in \begin{Bmatrix} INTRA4\times 4,\ INTRA16\times 16,\ SKIP, \\ 16\times 16,\ 16\times 8,\ 8\times 16,\ 8\times 8, \end{Bmatrix}$$

B: $$MODE \begin{Bmatrix} INTRA4\times 4,\ INTRA16\times 16,\ DIRECT, \\ 16\times 16,\ 16\times 8,\ 8\times 16,\ 8\times 8 \end{Bmatrix}$$

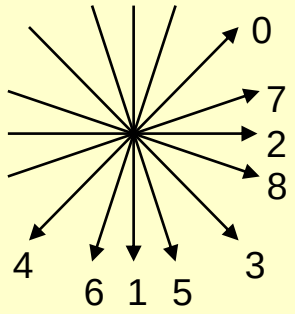
- "skip mode" refers to the 16x16 mode, where no motion and residual information are encoded

MPEG-4 AVC/H.264: Intra Prediction



- Directional spatial prediction (9 types for luma, 1 chroma)

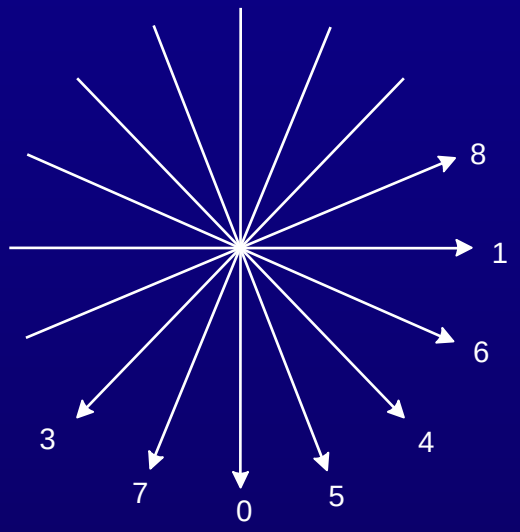
```
Q A B C D E F G H
I a b c d
J e f g h
K i j k l
L m n o p
M
N
O
P
```



- e.g., Mode 3:
  diagonal down/right prediction
  a, f, k, p are predicted by
  (A + 2Q + I + 2) >> 2

Intra Prediction: 4x4 Luma Blocks

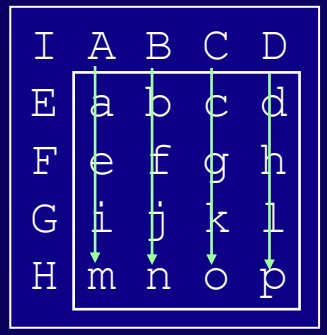
- Mode 0: vertical prediction
- Mode 1: horizontal prediction
- Mode 2: DC prediction
- Mode 3: Diagonal down/left prediction
- Mode 4: Diagonal down/right prediction
- Mode 5: vertical-left
- Mode 6: horizontal-down
- Mode 7: vertical-right
- Mode 8: horizontal-up



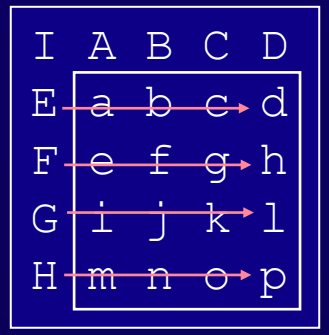
DC prediction:

pred(x, y) = Average of pixels A, B, C, D, E, F, G, and H

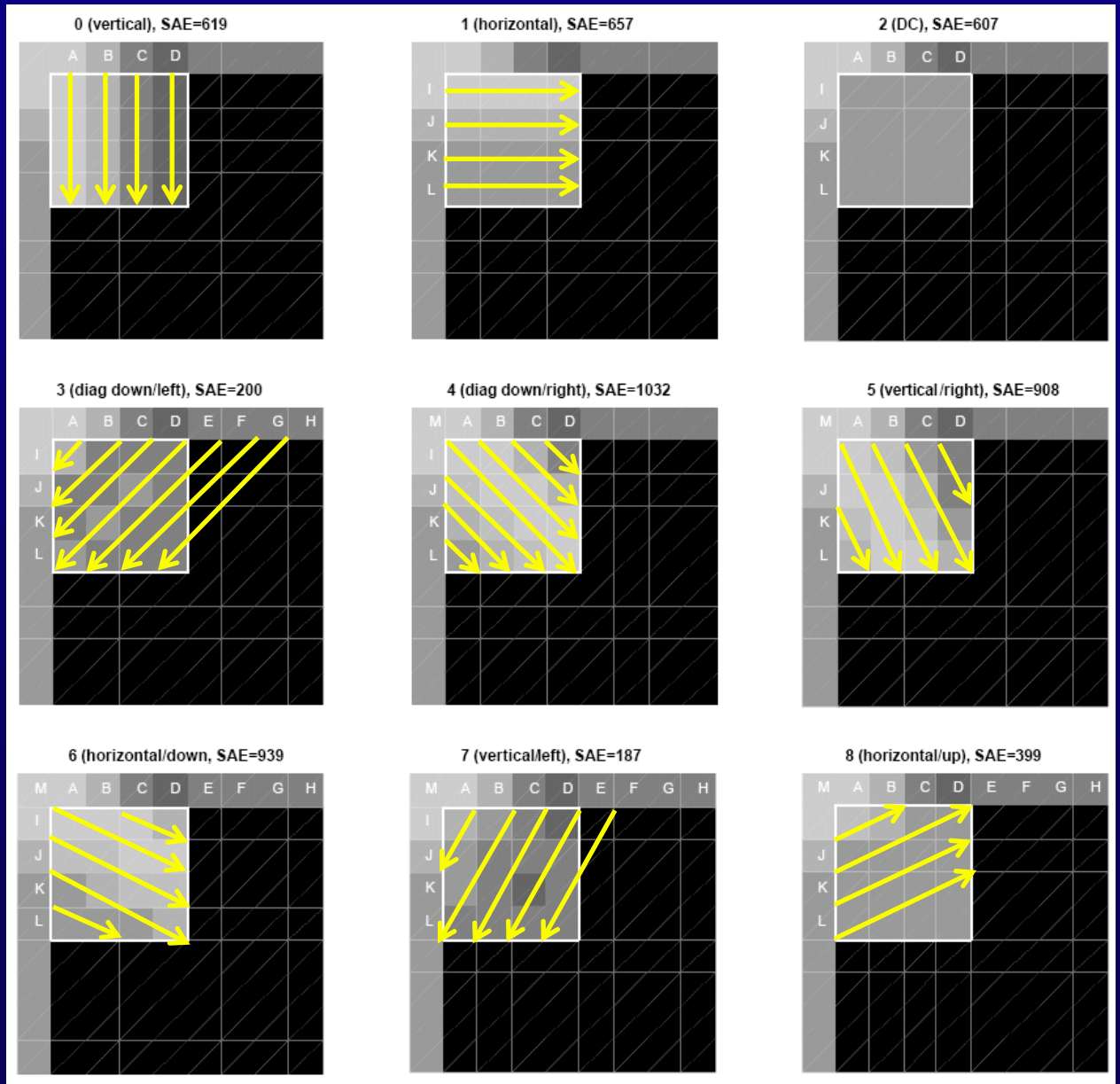
Mode 0

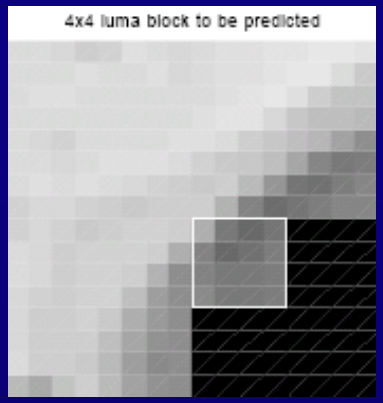


Mode 1



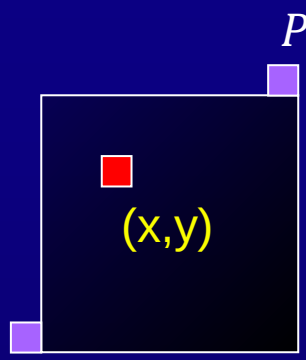
Intra Prediction: 4x4 Luma Prediction





Intra Prediction: 16x16 Luma Blocks

- Mode 0: vertical
- Mode 1: horizontal
- Mode 2: DC
- Mode 3: plane
  - Be used only if all neighboring samples are available



$$\mathbf{Pred}(x,y) = \mathit{Clip}(\ (a + b \cdot (x-7) + c \cdot (y-7) + 16)\ >>\ 5\ ),$$

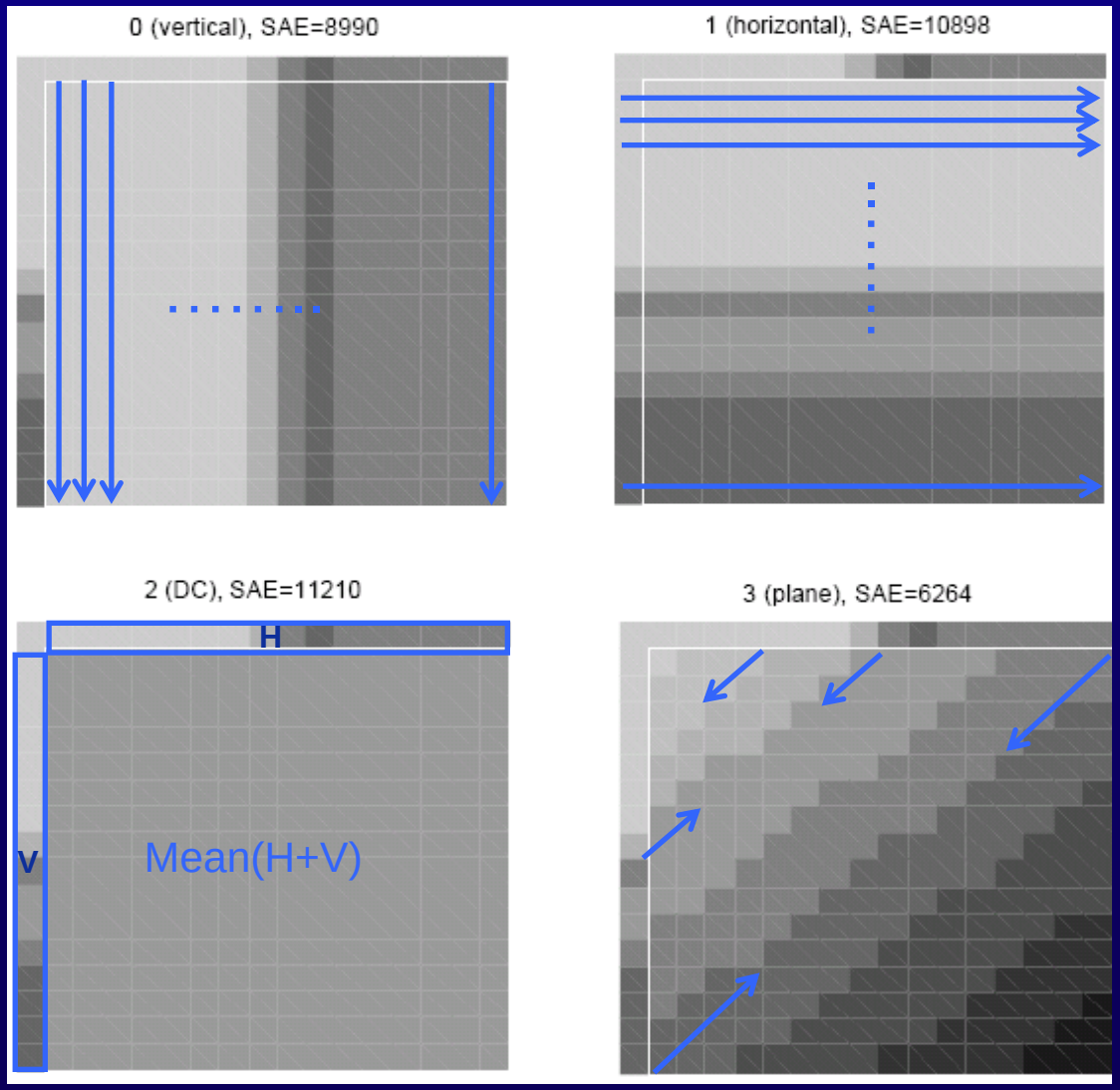
where

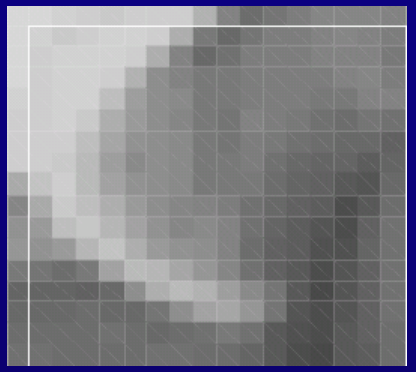
$$a = 16 \cdot \big(P(-1,15) + P(15,-1)\big)$$
$$b = (5 * H + 32) >> 6$$
$$c = (5 * V + 32) >> 6$$

$$H = \sum_{x=1}^{8} x \cdot (P(7+x,-1) - P(7-x,-1))$$

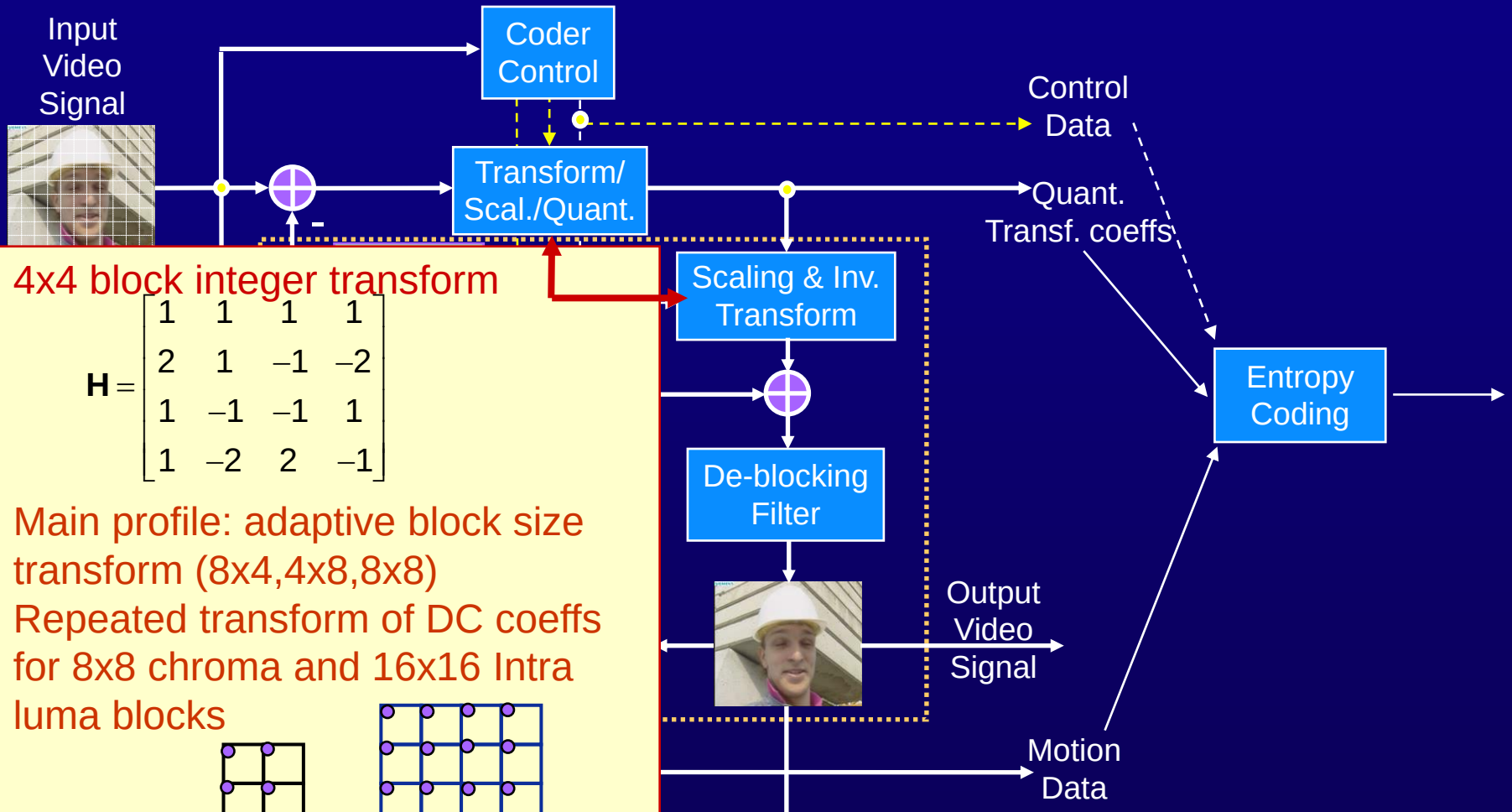
$$V = \sum_{y=1}^{8} y \cdot (P(-1,7+y) - P(-1,7-y))$$

Intra Prediction: 16x16 Luma Blocks





MPEG-4 AVC/H.264: Transform Coding



- 4x4 block integer transform

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}$$

- Main profile: adaptive block size transform (8x4,4x8,8x8)
- Repeated transform of DC coeffs for 8x8 chroma and 16x16 Intra luma blocks

Transform Coding: Luma DC

- Luma DC in Intra_16x16 MB using Hadamard transformation
  - Forward transform:

$$Y_D = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} X_{D00} & X_{D01} & X_{D02} & X_{D03} \\ X_{D10} & X_{D11} & X_{D12} & X_{D13} \\ X_{D20} & X_{D21} & X_{D22} & X_{D23} \\ X_{D30} & X_{D31} & X_{D32} & X_{D33} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \right) //2$$

  - Inverse transform:

$$X_{QD} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} y_{QD00} & y_{QD01} & y_{QD02} & y_{QD03} \\ y_{QD10} & y_{QD11} & y_{QD12} & y_{QD13} \\ y_{QD20} & y_{QD21} & y_{QD22} & y_{QD23} \\ y_{QD30} & y_{QD31} & y_{QD32} & y_{QD33} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$$

Transform Coding: Chroma DC

- Chroma DC in 8x8 block Hadamard transformation
  - Forward transform

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

  - Inverse transform

$$X_{QD} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} Y_{QD00} & Y_{QD01} \\ Y_{QD10} & Y_{QD11} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Transform: Luma and Chroma residual

- Luminance and chrominance 4x4 residual blocks

– Forward transform

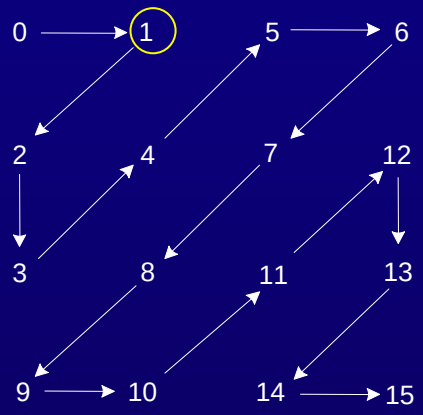
$$Y = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \begin{bmatrix} X_{00} & X_{01} & X_{02} & X_{03} \\ X_{10} & X_{11} & X_{12} & X_{13} \\ X_{20} & X_{21} & X_{22} & X_{23} \\ X_{30} & X_{31} & X_{32} & X_{33} \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 1 \\ 1 & 1 & -1 & -2 \\ 1 & -1 & -1 & 2 \\ 1 & -2 & 1 & -1 \end{bmatrix}$$

– Inverse Transform

$$X = \begin{bmatrix} 1 & 1 & 1 & 1/2 \\ 1 & 1/2 & -1 & -1 \\ 1 & -1/2 & -1 & 1 \\ 1 & -1 & 1 & -1/2 \end{bmatrix} \begin{bmatrix} y_{00} & y_{01} & y_{02} & y_{03} \\ y_{10} & y_{11} & y_{12} & y_{13} \\ y_{20} & y_{21} & y_{22} & y_{23} \\ y_{30} & y_{31} & y_{32} & y_{33} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1/2 & -1/2 & -1 \\ 1 & -1 & -1 & 1 \\ \frac{1}{2} & -1 & 1 & -1/2 \end{bmatrix}$$

Quantization/Dequantization (1/2)

- Scan order
  - 4x4 residual and 4x4 luma DC block



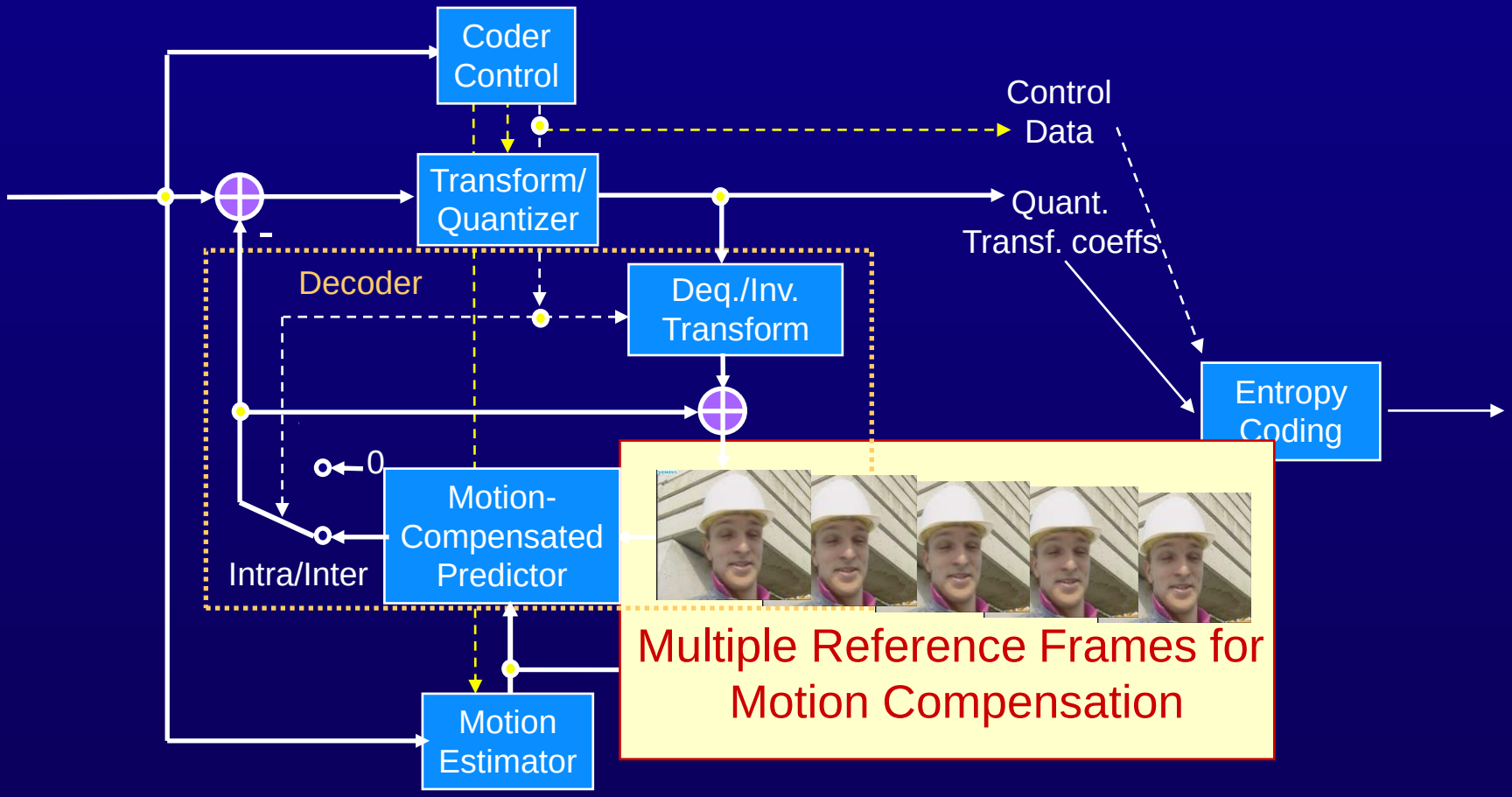
  - 2x2 chroma DC block
    - raster order

Quantization/Dequantization (2/2)

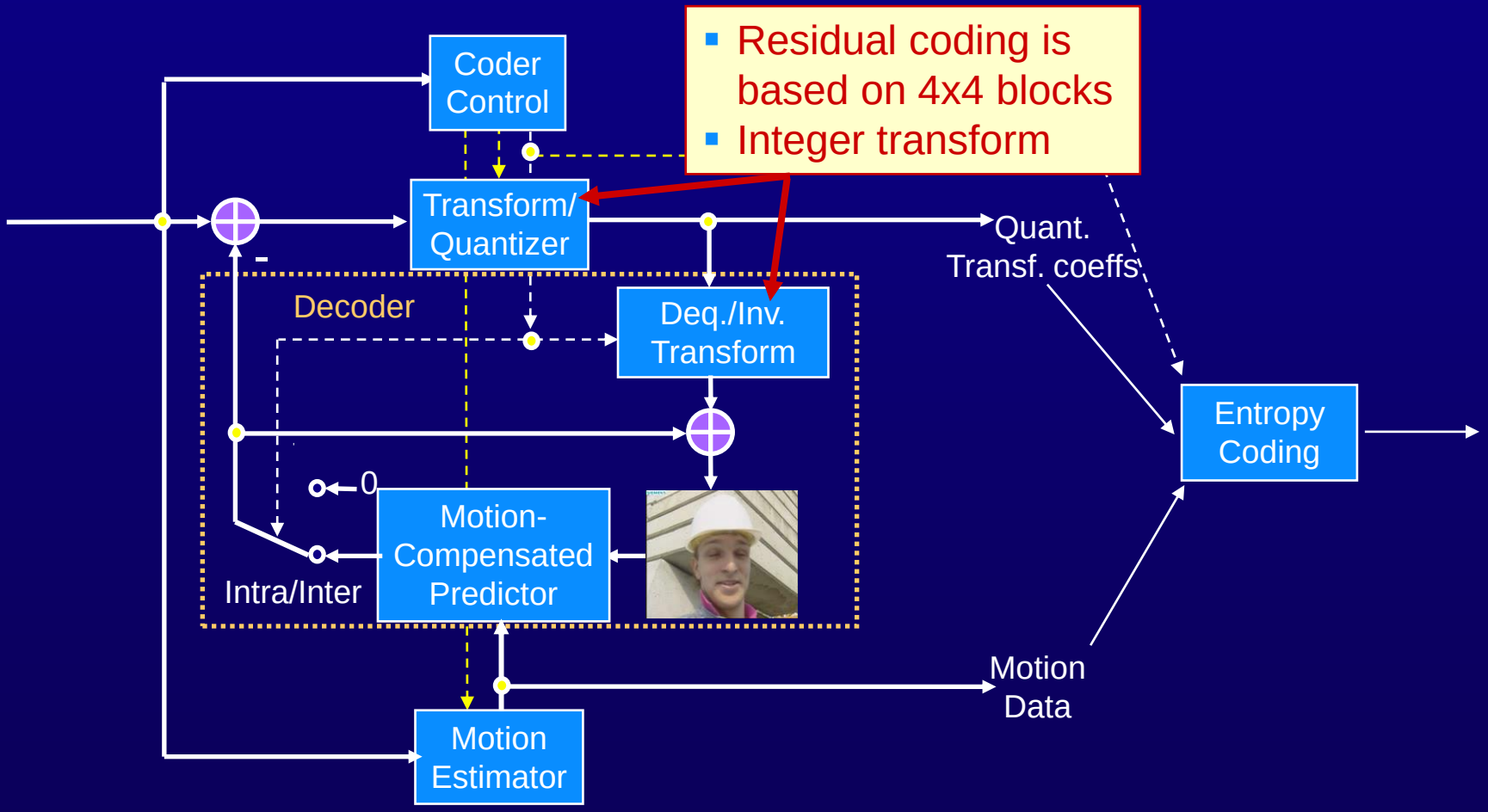
- QP: 0 ~ 51
- QP_Y: QP for luma coefficients
- QP_C: QP for chroma coefficients
  - QPC for chroma is determined from the current value of QP_Y

QP_Y	<30	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
QP_C	$=QP_Y$	29	30	31	32	32	33	34	34	35	35	36	36	37	37	37	38	38	38	39	39	39	39

MPEG-4 AVC/H.264: Multiple Reference Frames



MPEG-4 AVC/H.264: Residual Coding



Residual and Intra Coding

- **EXACT MATCH** simplified transform
  - Based primarily on 4x4 transform (all prior standards: 8x8)

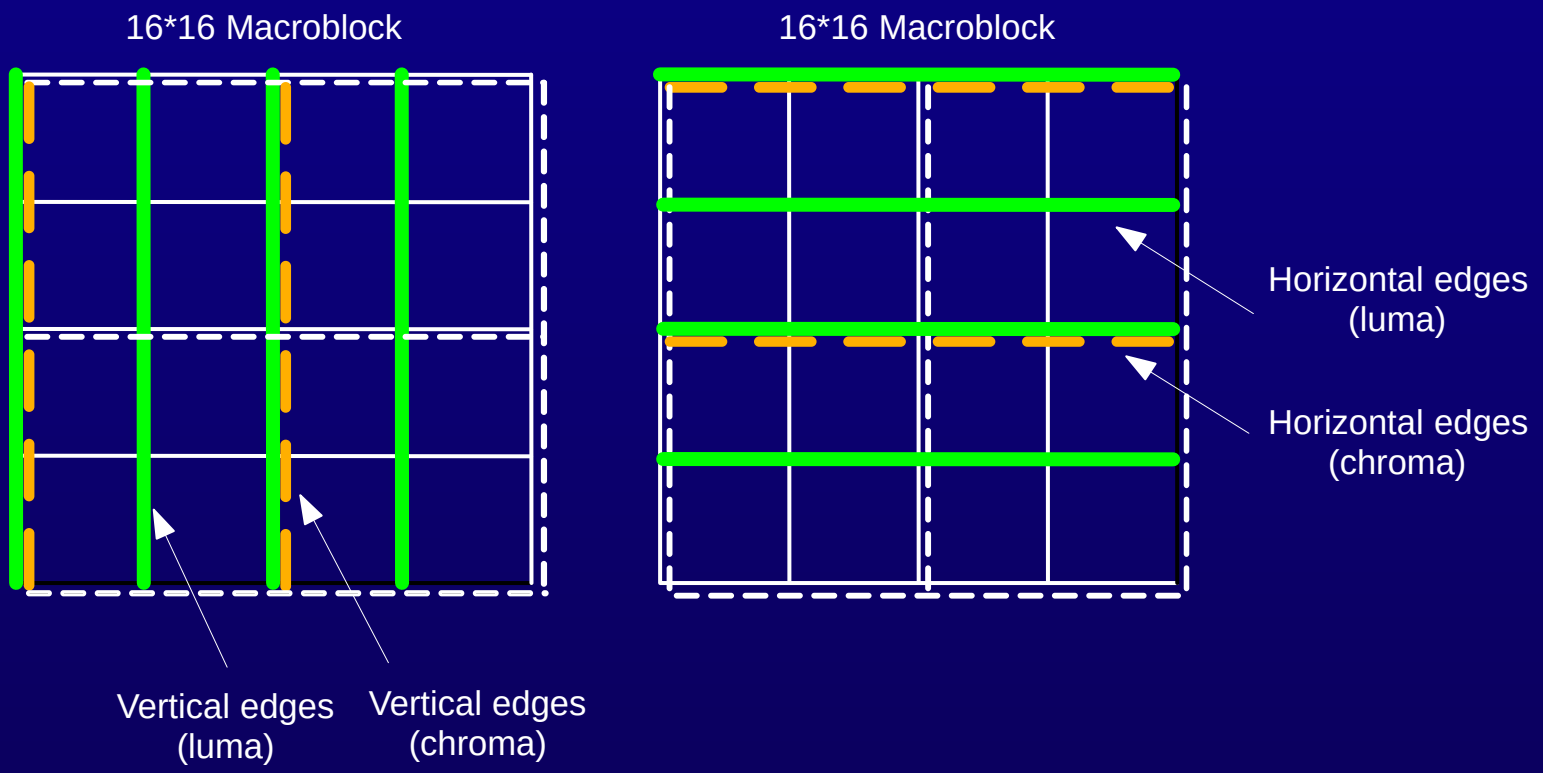
$$T=\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & 1 \end{bmatrix}$$

  - Requires only **16 bit** arithmetic (including intermediate values)
  - Expanded to 8x8 for chroma by 2x2 transform of the DC values
  - Easily extensible to 10-12 bits per component
- Adaptive block transform sizes for main profile
- Intra coding structure
  - Directional spatial prediction (10 types luma, 1 chroma)
  - Expanded to 16x16 for luma intra by 4x4 transform of the DC values

Quantization and Deblocking

- Quantization of transform coefficients
  - Logarithmic step size control
  - Extended range of step sizes
  - Smaller step size for chroma (per H.263 Annex T)
  - Table-driven
- Reconstruction is 16-bit multiply, add, shift
- In-loop deblocking filter

Deblocking Filter



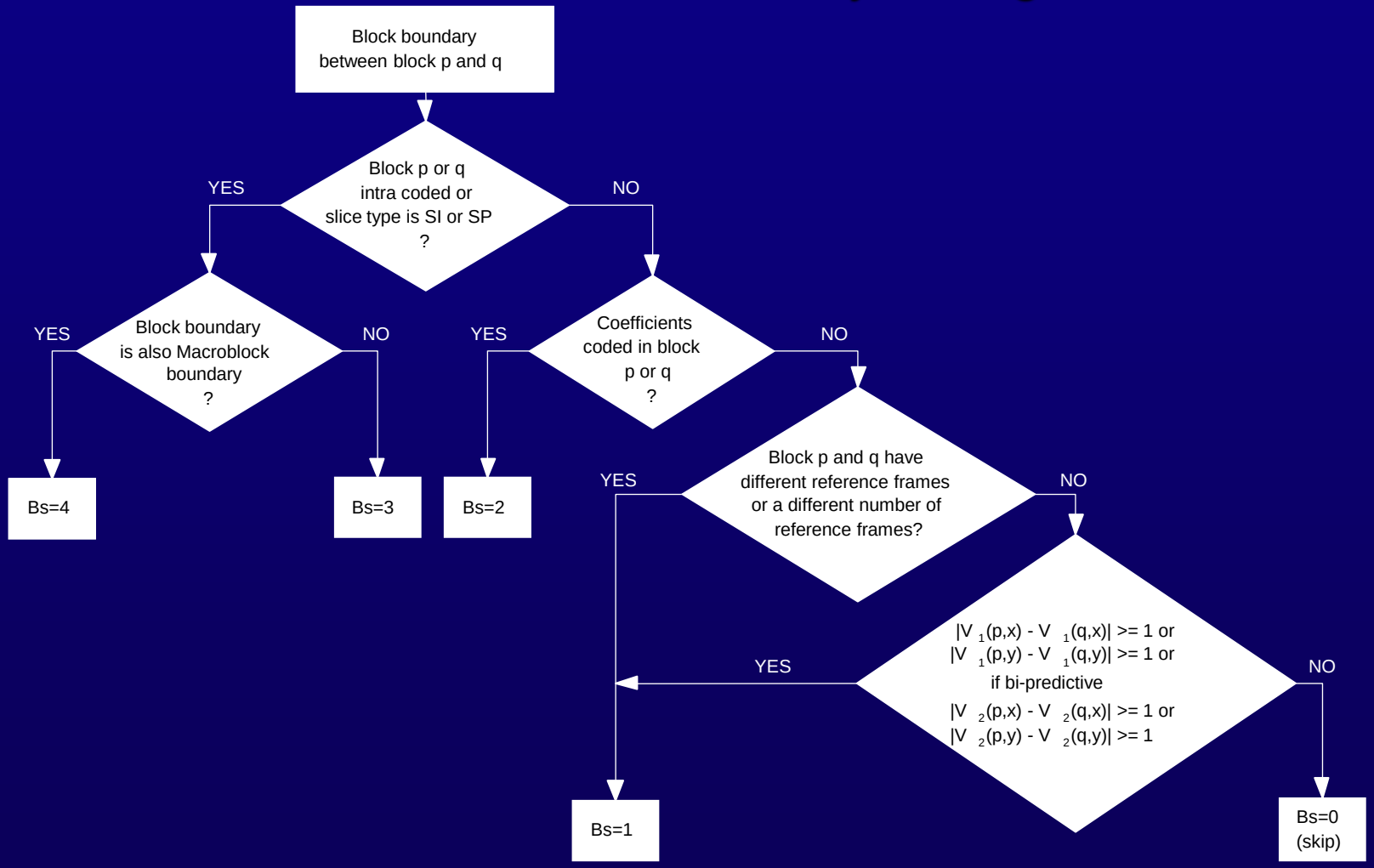
- Boundaries in a macroblock to be filtered (luma boundaries shown with solid lines and chroma boundaries shown with dotted lines)

Deblocking Filter

- **Content dependent boundary filtering strength**
  - For each boundary between neighbouring 4x4 luma blocks, a “boundary strength” ***Bs*** is assigned
  - If ***Bs=0***, filtering is skipped for that particular edge
  - In all other cases, filtering is dependent on the local sample properties and the value of ***Bs***

Deblocking Filter

- Flowchart to determine boundary strength *Bs*



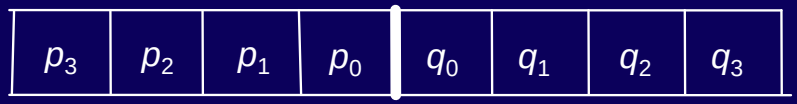
Deblocking Filter

- Thresholds for each block boundary
  - Set of samples across this edge are only filtered if *Bs* ≠ *0* **&&** $|p_0 - q_0| < \alpha$ **&&** $|p_1 - p_0| < \beta$ **&&** $|q_1 - q_0| < \beta$
  - α and β are determined by IndexA and IndexB, respectively, where

    IndexA = *Clip3*(0, 51, QPav + Filter_Offset_A)

    IndexB = *Clip3*(0, 51, QPav + Filter_Offset_B)
  - Filter_Offset_A and Filter_Offset_B used to modify filter characteristics

$$Clip3(a, b, c) = \begin{cases} a, & c < a \\ b, & c > b \\ c, & otherwise \end{cases}$$



Deblocking Filter: *Bs* < 4

- $\Delta = Clip3\left(-C, C, \left(\left((q_0 - p_0) \ll 2 + (p_1 - q_1) + 4\right) \gg 3\right)\right)$
- $P_0 = Clip1(p_0 + \Delta)$
- $Q_0 = Clip1(q_0 - \Delta)$
  - $ap = |p_2 - p_0|$
  - $aq = |q_2 - q_0|$
  - If $ap < \beta$, $P_1 = p_1 + Clip3(-C_0, C_0, \left(p_2 + (p_0 + q_0) \gg 1 - (p_1 \ll 1)\right) \gg 1)$
  - If $aq < \beta$, $Q_1 = q_1 + Clip3(-C_0, C_0, \left(q_2 + (p_0 + q_0) \gg 1 - (q_1 \ll 1)\right) \gg 1)$
  - C_0 is determined by *IndexA* and *Bs*
  - $Clib1(x) = clip3(0{,}255, x)$

Deblocking Filter: *Bs* = 4

- Left/upper side
- If $ap < \beta$ && $|p_0 - q_0| < ((\alpha >> 2) + 2)$ (A)

  $P_0 = (p_2 + 2*p_1 + 2*p_0 + 2*q_0 + q_1 + 4) >> 3$

  $P_1 = (p_2 + p_1 + p_0 + q_0 + 2) >> 2$
- In the case of luma filtering,

  $P_2 = (2*p_3 + 3*p_2 + p_1 + p_0 + q_0 + 4) >> 3$
- Otherwise, if the condition of Eq. (A) does not hold,

  $P_0 = (2*p_1 + p_0 + q_1 + 2) >> 2$

Deblocking Filter: *Bs* = 4

- Right/lower side
- If $aq < \beta$ **&&** $|p_0 - q_0| < ((\alpha >> 2) + 2)$ (B)

  $Q_0 = (p_1 + 2*p_0 + 2*q_0 + 2*q_1 + q_2 + 4) >> 3$

  $Q_1 = (p_0 + q_0 + q_1 + q_2 + 2) >> 2$
- In the case of luma filtering,

  $Q_2 = (2*q_3 + 3*q_2 + q_1 + q_0 + p_0 + 4) >> 3$
- Otherwise, if the condition of Eq. (B) does not hold,

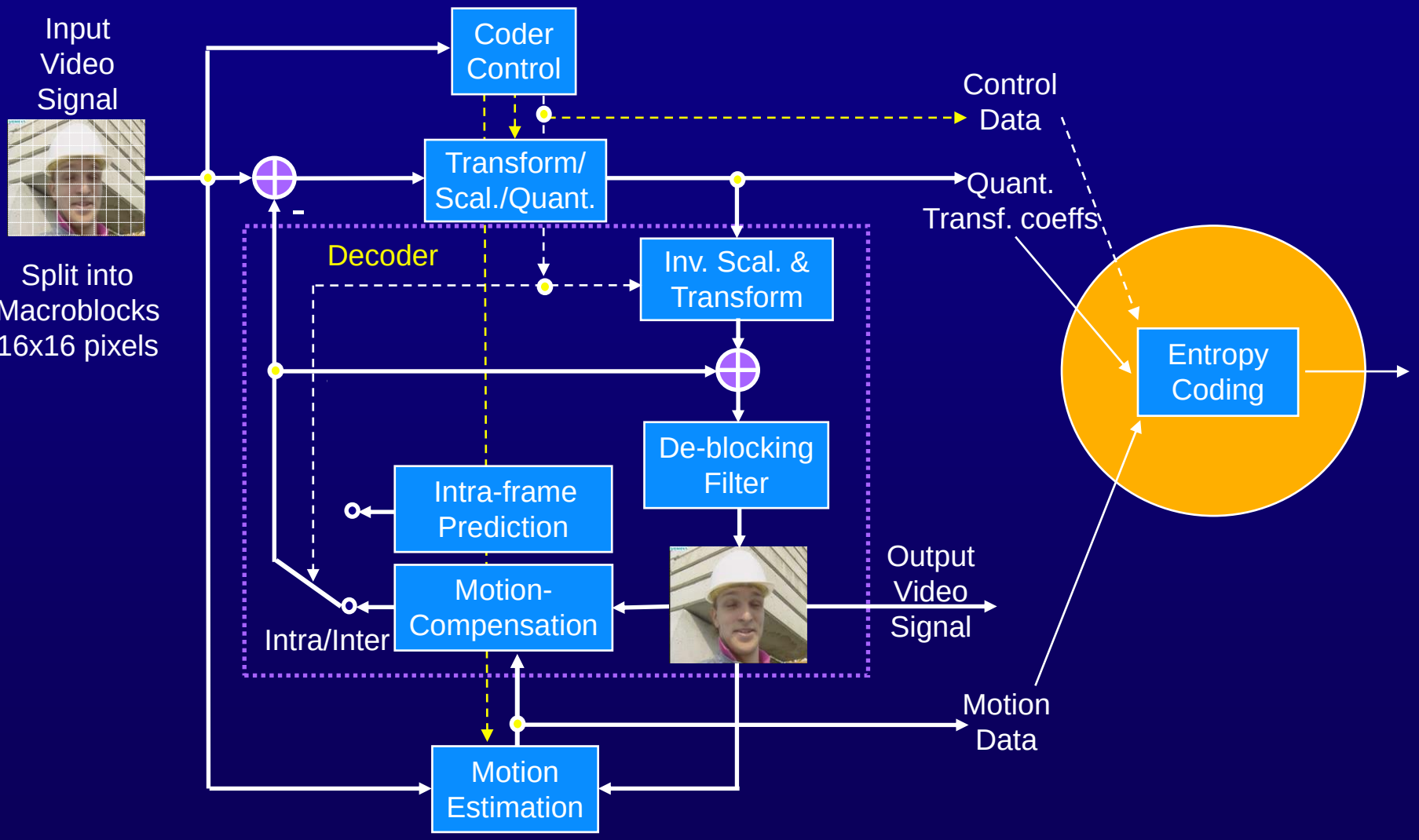
  $Q_0 = (2*q_1 + q_0 + p_1 + 2) >> 2$

Deblocking Filter

1) without filtering

2) with H264/AVC deblocking

Entropy Coding



Variable Length Coding

- Exp-Golomb code is used for all symbols except for transform coefficients
- Context adaptive VLCs coding of transform coefficients
  - No end-of-block, but number of coefficients is decoded
  - Coefficients are scanned backwards
  - Contexts are built dependent on transform coefficients

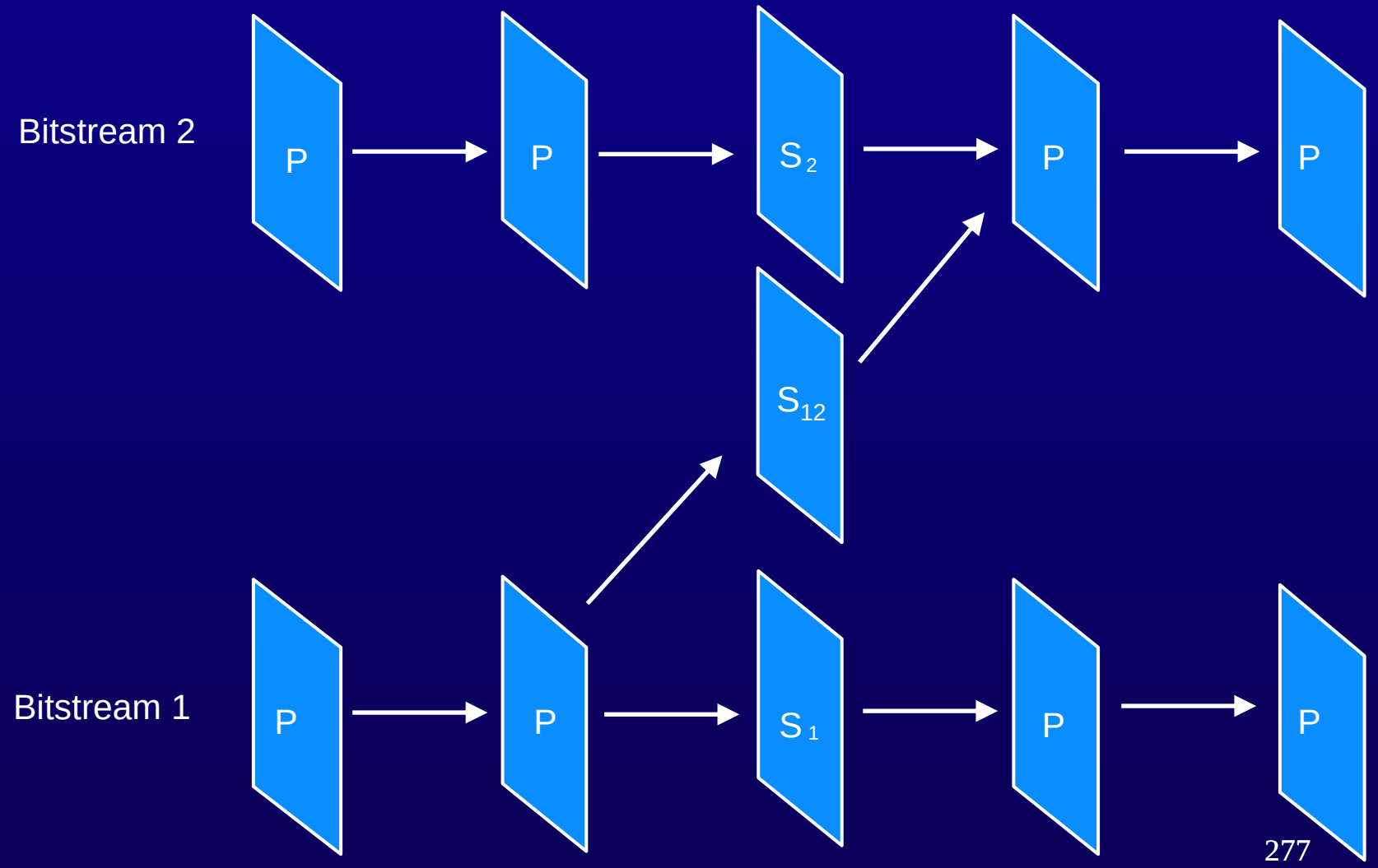
Content-based Adaptive Binary Arithmetic Coding (CABAC)

- Usage of adaptive probability models for most symbols
- Exploiting symbol correlations by using contexts
- Restriction to binary arithmetic coding
  - Simple and fast adaptation mechanism
  - Fast binary arithmetic codec based on table look-ups and shifts only
- Average bit-rate saving over CAVLC 10-15%

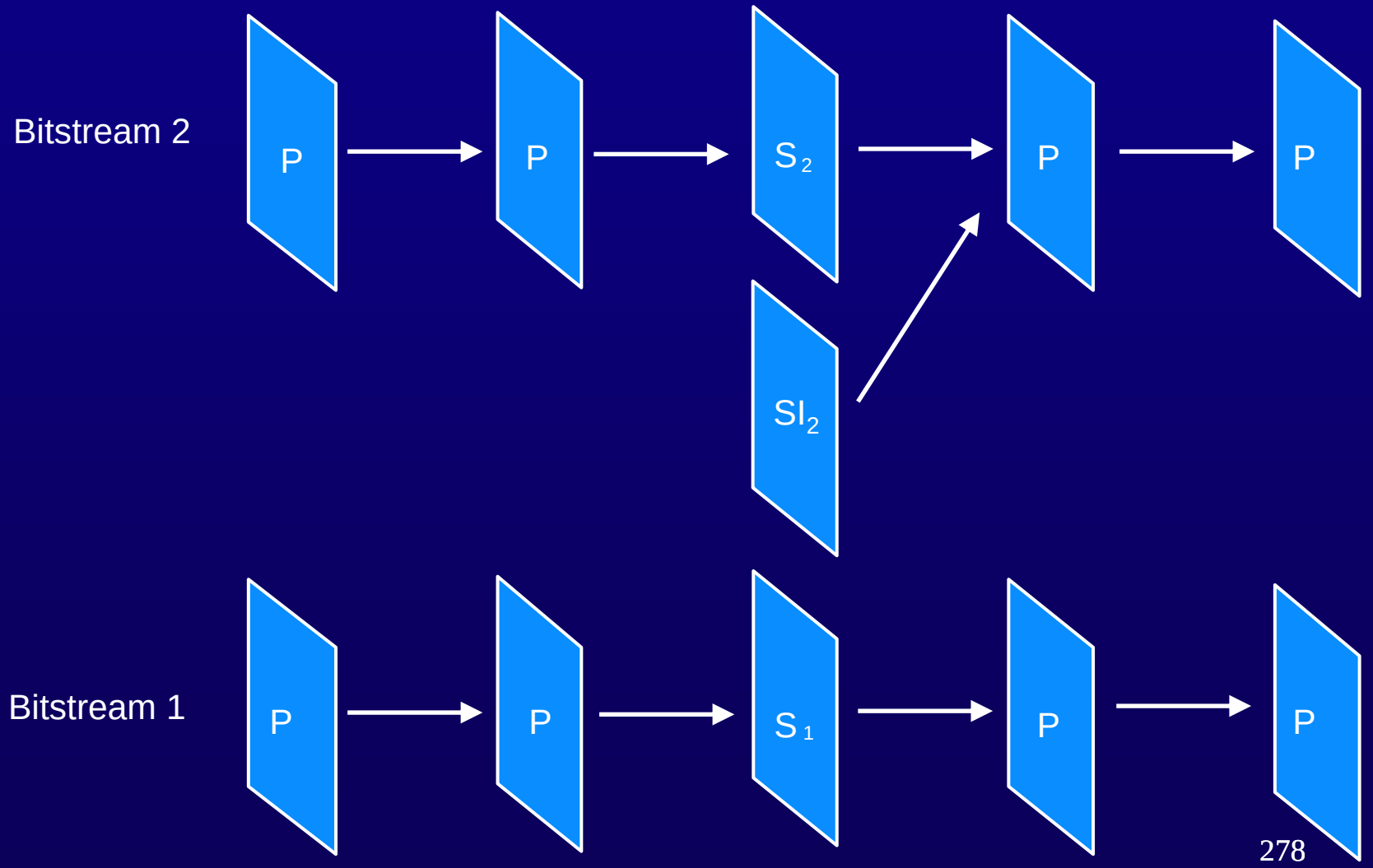
SP/SI Frame

- SP frame
  - motion-compensated predictive coding
  - similar to P
  - SP allows identical reconstruction even when different reference pictures are being used
- SI frame
  - spatial prediction
  - similar to I
  - SI allows identical reconstruction to a corresponding SP
- provide functionalities for bitstream switching, splicing, random access, VCR functionalities such as fast-forward, and error resilience/recovery

SP/SI Frame: Bitstream Switching



SP/SI Frame: Bitstream Splicing



Profiles

- Baseline profile
- Extended profile
- Main profile

Baseline Profile

- I and P picture types
- In-loop deblocking filter
- 1/4-sample motion compensation
- VLC-based entropy coding: CAVLC
- 4:2:0 Chrominance format
- Field pictures (for Level 2.1 and above)
- use 15 or fewer reference frames
- have a compression ratio per picture of 4:1 or greater

Extended Profile

- Bi-predictive slices
- SP and SI slices
- Weighted prediction
- All features in baseline profile are included

Main Profile

- CABAC
- Interlaced pictures
- All features in baseline profile are included

Level Definitions

Level #	Max Picture Size (MBs)	Max Video Bitrate (1000 bits/sec)	Horizontal MV Range (full pels)	Vertical MV Range (full pels)	Minimum luma Bi-predictive block size
1	99	64	[-2048, 2047.75]	[-64, 63.75]	8x8
1.1	396	128	[-2048, 2047.75]	[-128, 127.75]	8x8
1.2	396	768	[-2048, 2047.75]	[-128, 127.75]	8x8
2	396	2000	[-2048, 2047.75]	[-128, 127.75]	8x8
2.1	792	4000	[-2048, 2047.75]	[-256, 255.75]	8x8
2.2	1620	4000	[-2048, 2047.75]	[-256, 255.75]	8x8
3	1620	8000	[-2048, 2047.75]	[-256, 255.75]	8x8
3.1	3600	20000	[-2048, 2047.75]	[512, 511.75]	8x8
3.2	5120	20000	[-2048, 2047.75]	[512, 511.75]	8x8
4	8192	20000	[-2048, 2047.75]	[512, 511.75]	8x8
5	19200	TBD	[-2048, 2047.75]	TBD	8x8

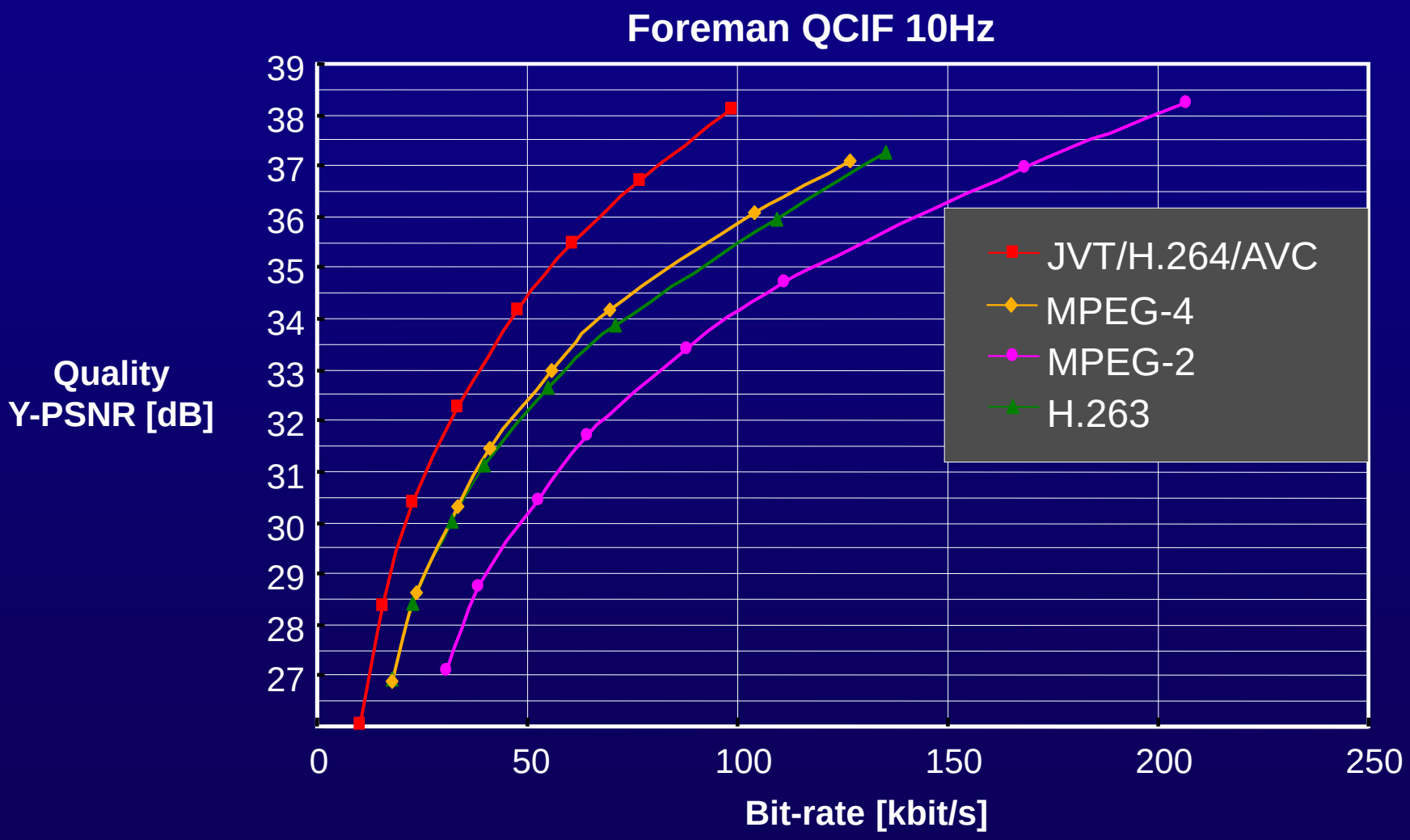
Complexity of H.264 Codec Design

- Codec design includes relaxation of traditional bounds on complexity (memory & computation) – rough guess 2-3x decoding power increase relative to MPEG-2, 3-4x encoding
- Problem areas:
  - Smaller block sizes for motion compensation (cache access issues)
  - Longer filters for motion compensation (more memory access)
  - Multi-frame motion compensation (more memory for reference frame storage)
  - More segmentations of macroblock to choose from (more searching in the encoder)
  - More methods of predicting intra data (more searching)
  - Arithmetic coding (adaptivity, computation on output bits)

Performance Comparison

- Test of different standards
- Using same rate-distortion optimization techniques for all codecs
- Streaming test: high-latency (included B frames)
- Real-time conversation test: no B frames
- Several video sequences for each test
- Compare four codecs:
  - MPEG-2
  - H.263 MPEG-4
  - JVT/H.26L/AVC (with & without B pictures)

Coding Efficiency Comparison



Conclusions

- Video coding layer is based on hybrid video coding and similar in spirit to other standards
- New key features are:
  - enhanced motion compensation
  - small blocks for transform coding
  - improved deblocking filter
  - enhanced entropy coding
- Bit-rate savings generally 50% or better against any otherstandard for the same perceptual quality
- Increased complexity relative to prior standards
- Standard of both ITU-T VCEG and ISO/IEC MPEG