University of Rostock
Institute of Communications Engineering

Image Compression

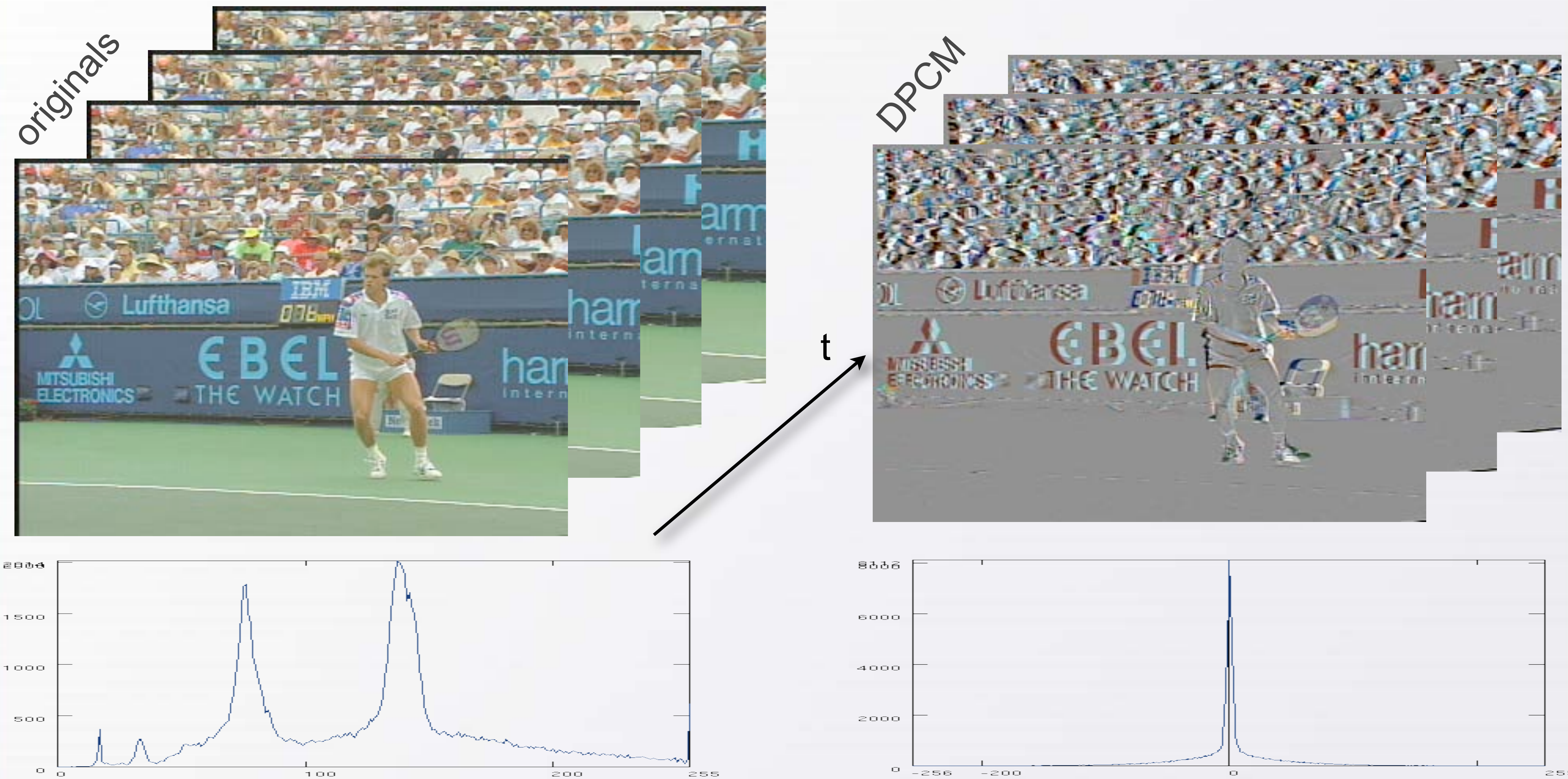
Universidad Politécnica de Madrid 2010

Dr.-Ing. Henryk Richter
henryk.richter@comlab.uni-rostock.de

University of Rostock
Institute of Communications Engineering
http://www.int.uni-rostock.de

basics

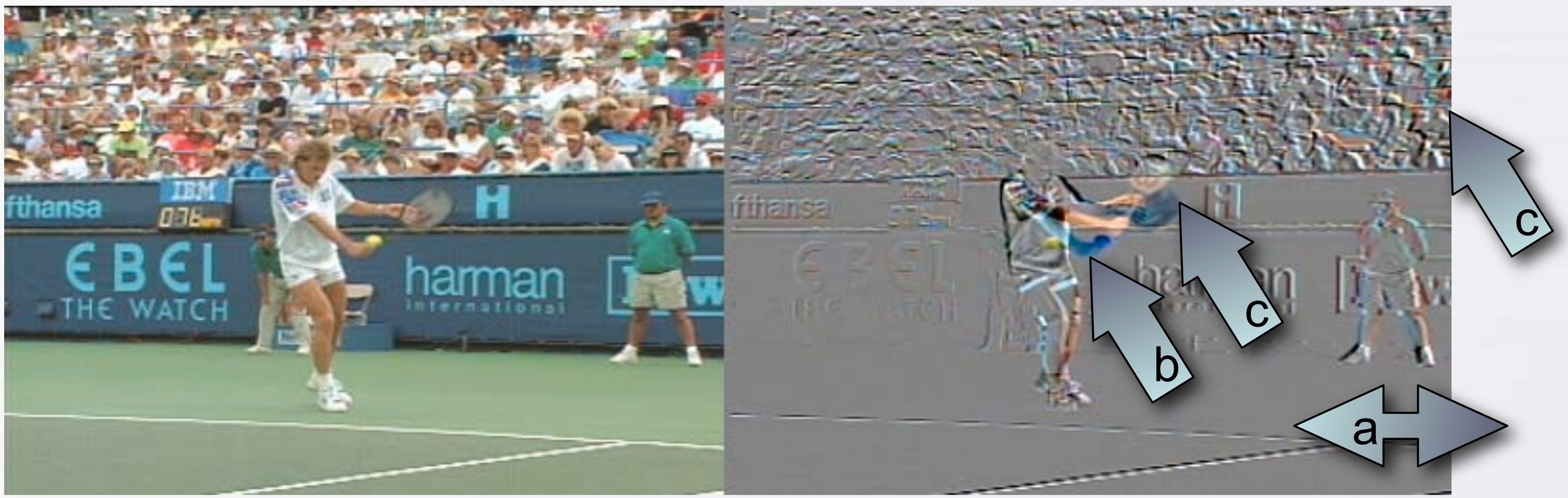
- correlation of neighboring images in video sequences induces redundancy in time direction ➠ inefficient single image coding



origins of differences between images

- complex motion of camera and objects (a)
- changes in illumination
- scene changes
- disturbances (noise, recording errors, aliasing)
- changes in object topology (b)
- revelations and occlusions on objects and frame borders (c)

general observation: video sequences are in-stationary



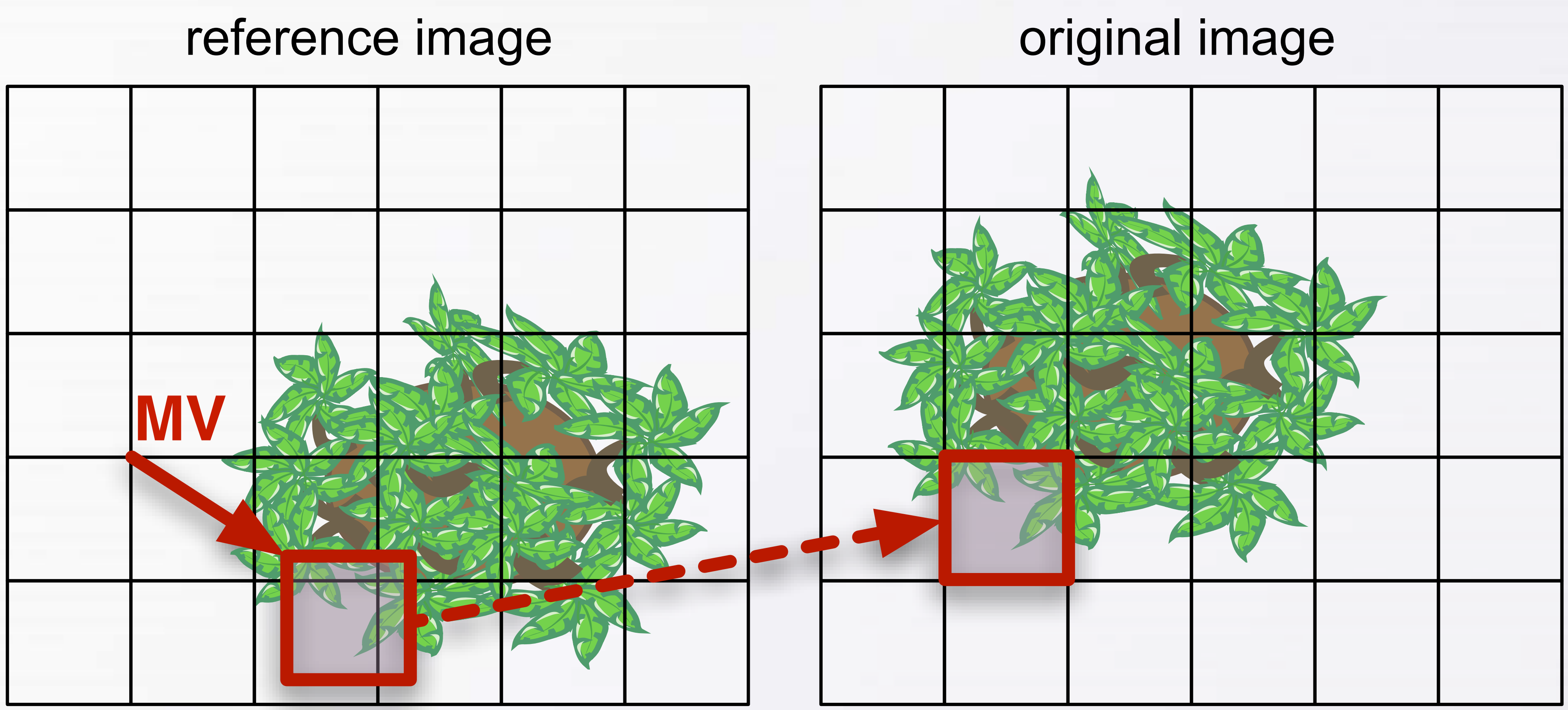
techniques for decorrelation along the time axis

- **spatial domain: corespondency methods**
  - DPCM (special case of motion compensation for stationary images)
  - block based local motion compensation
    - FSBM (fixed size block matching)
    - VSBM (variable size block matching)
    - OBMC (overlapping block motion compensation)
    - Object based block matching motion compensation
  - image based global motion compensation [Smolic 2001, Richter 2000]

- **transform domain:**
  - 3D Wavelet-transform [Schwarz 2000]
  - MCTF - motion compensated temporal filtering [Schäfer 2005]

translational block motion compensation

- each block receives a motion vector (MV)
- motion vector is a displacement of a block to predict in relation to that block‘s position within the reference image
- predicted block is gathered from the reference image and used as source for difference coding



block matching techniques (I)

- **various search metrics**
  - mean square error (MSE), sum of square error (SSE)
  - sum of absolute differences (SAD) or mean absolute difference (MAD)
  - discrete cross correlation

  - additionally: motion vector length

$$MAD(d_x, d_y) = \frac{1}{B_x \cdot B_y} \sum_{j=y}^{y+B_y-1} \sum_{i=x}^{x+B_x-1} \Big| I(i,j,t) - I(i-d_x, j-d_y, t-\Delta_t) \Big|$$

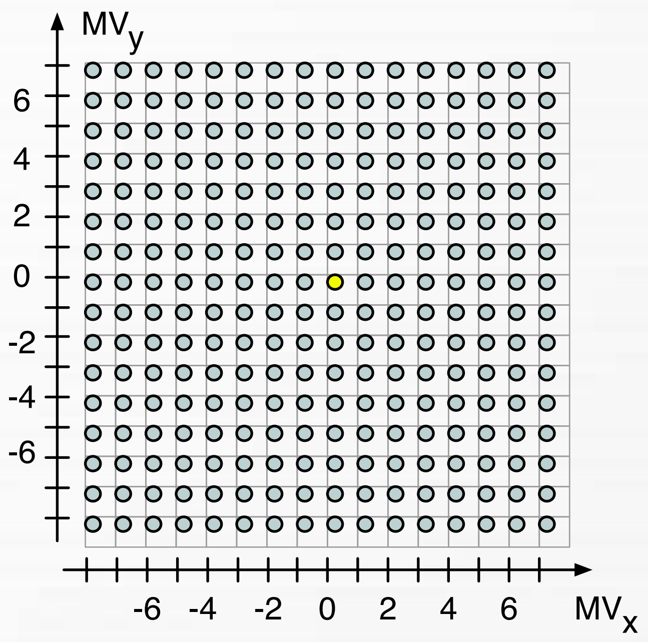
$$MSE(d_x, d_y) = \frac{1}{B_x \cdot B_y} \sum_{j=y}^{y+B_y-1} \sum_{i=x}^{x+B_x-1} \Big(I(i,j,t) - I(i-d_x, j-d_y, t-\Delta_t) \Big)^2$$

⇒ for constant block size, normalization (mean) not necessary

⇒ SAD/MAD favorable on architectures with hardware SAD support (e.g. Intel MMX2)

block matching techniques (2)

- **workflow**
  - for each candidate position
    - calculate displacement dependent prediction error metric
    - select best match



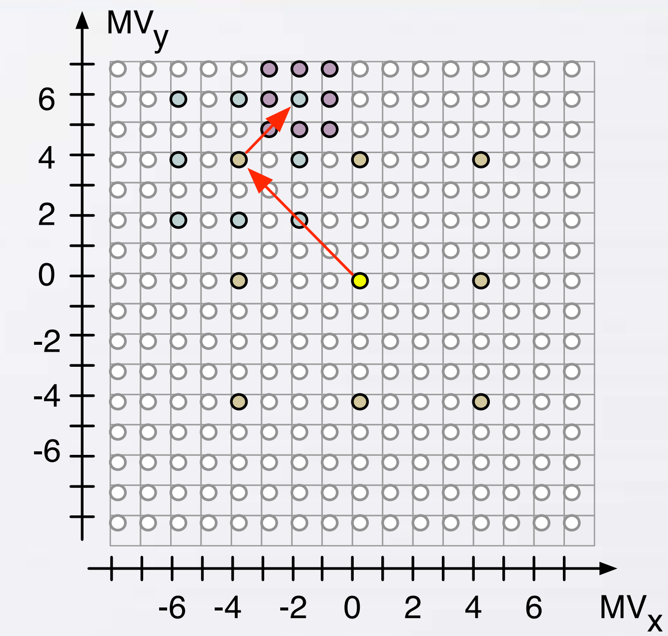
1) exhaustive search (“full search“)

- scan all candidate positions in area
- guarantees optimal result for large search window
- problem: computational complexity (here: 256 positions per compensated block)
- variation: spiral search

block matching techniques (3)

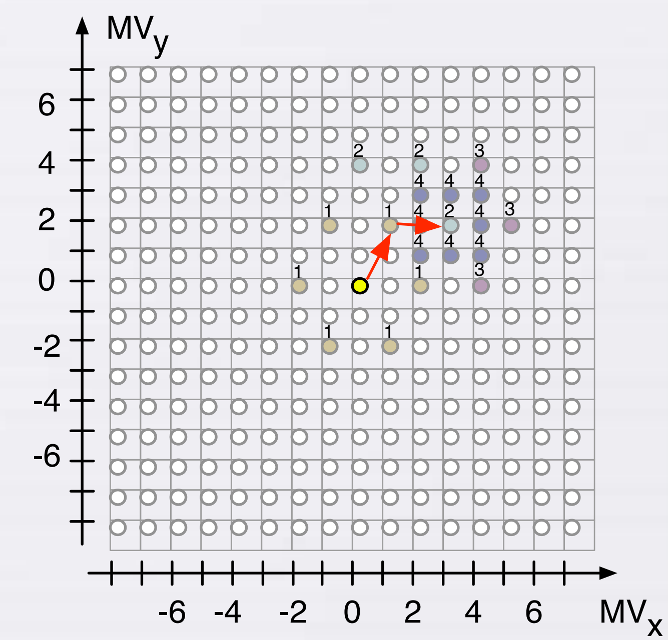
2) three step search (TSS)

- find best match in radius 4
- use best match for next iteration with radius 2
- use best match as start for radius 1 to obtain final result
- 25 search positions but unreliable, small search window
- variation: logarithmic search



3) hexagon search (similar to TSS)

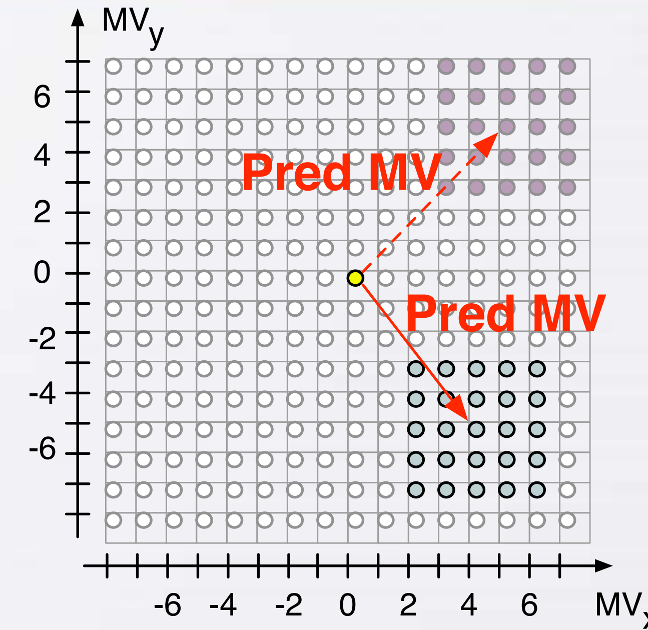
- as long as best match is not in the middle of the pattern, use best match as start for next iteration (ex. steps 1-3)
- in each iteration, only three new positions need to be calculated
- last step is radius 1 local search (ex. step 4)
- good small area results, heavily dependent on start MV, number of iterations not pre-determinated



block matching techniques (4)

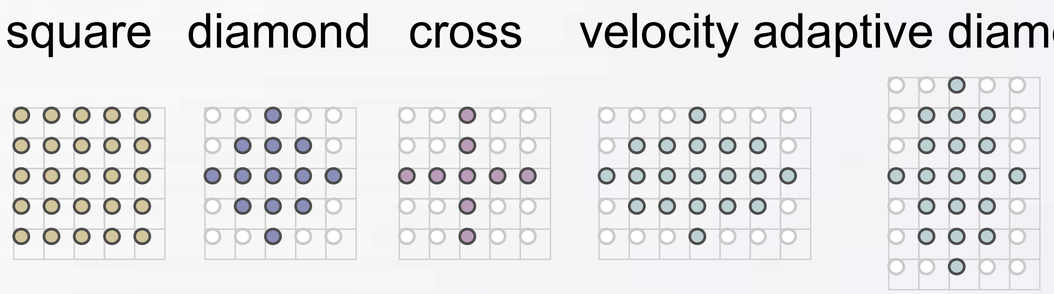
4) predictive search

- use predicted MV(s) from spatial/temporal neighbors
- perform small area exhaustive search around predicted MV(s)
- good complexity/reliability tradeoff, can cover large translations despite smaller search window
- variations: MVFAST, PMVFAST,



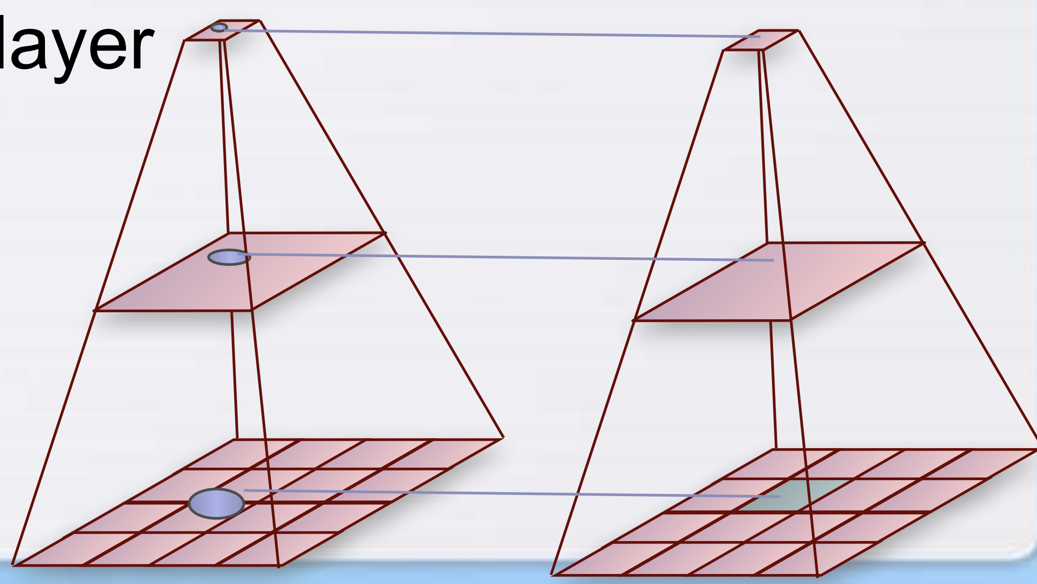
5) search patterns

- algorithms can be combined with different patterns of varying size



6) hierarchical block motion estimation

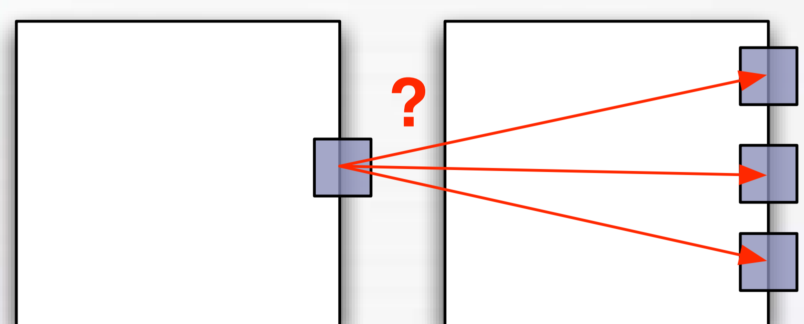
- calculate resolution pyramid
- use best vector from lowest layer for following layer



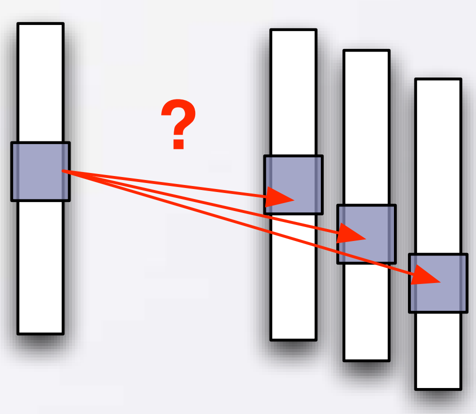
block motion compensation issues

- **additional problems:**
  - aperture problem
  - periodicity problem
  - block artifact problem, block structure not related to image content
- **in most cases, true motion not required, just a good match**
- **optimize for smooth motion vector field**

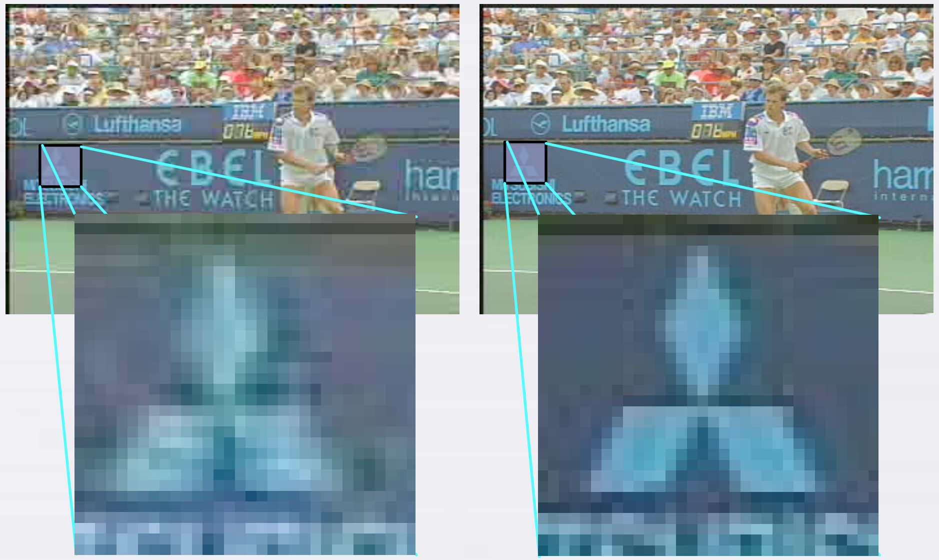
aperture problem



periodicity problem



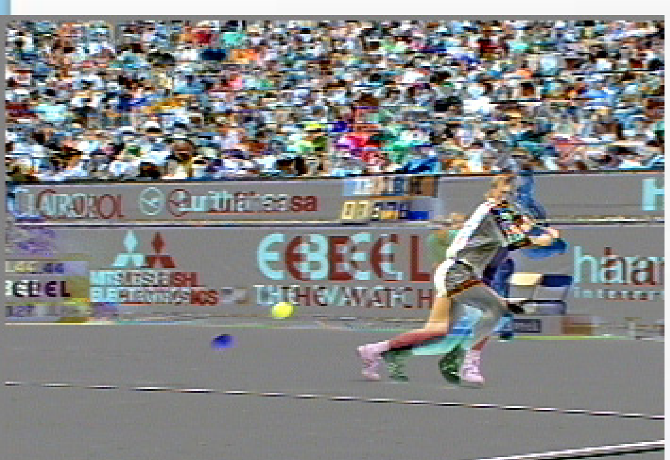
artifact problem



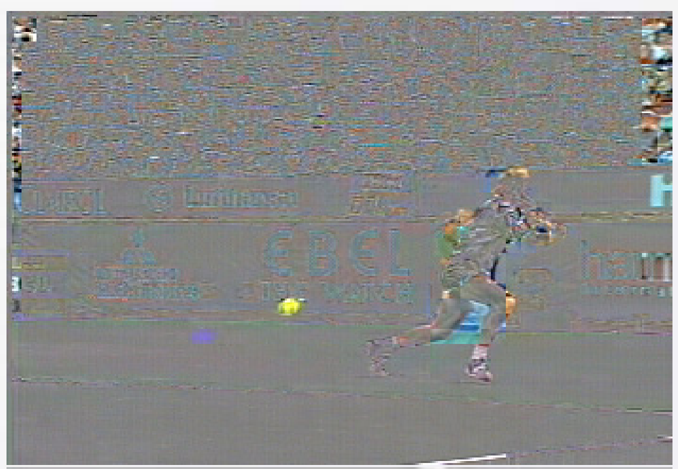
translational block motion compensation example

- **advantage of motion compensation is the drastic reduction of residual energy**
  - lower differences result in less significant data to be coded
- **issues:**
  - encoder complexity
  - in-stationary image parts, causality (e.g. frame borders)
  - motion vector transmission from encoder to decoder ➠ tradeoff

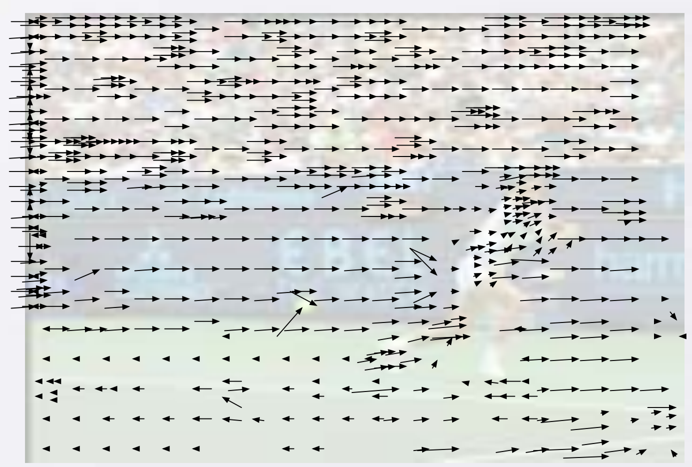
frame difference w/o MC



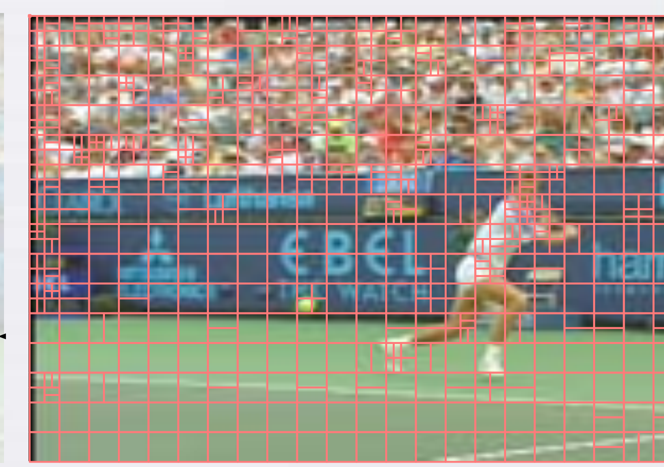
frame difference with MC



motion vector field

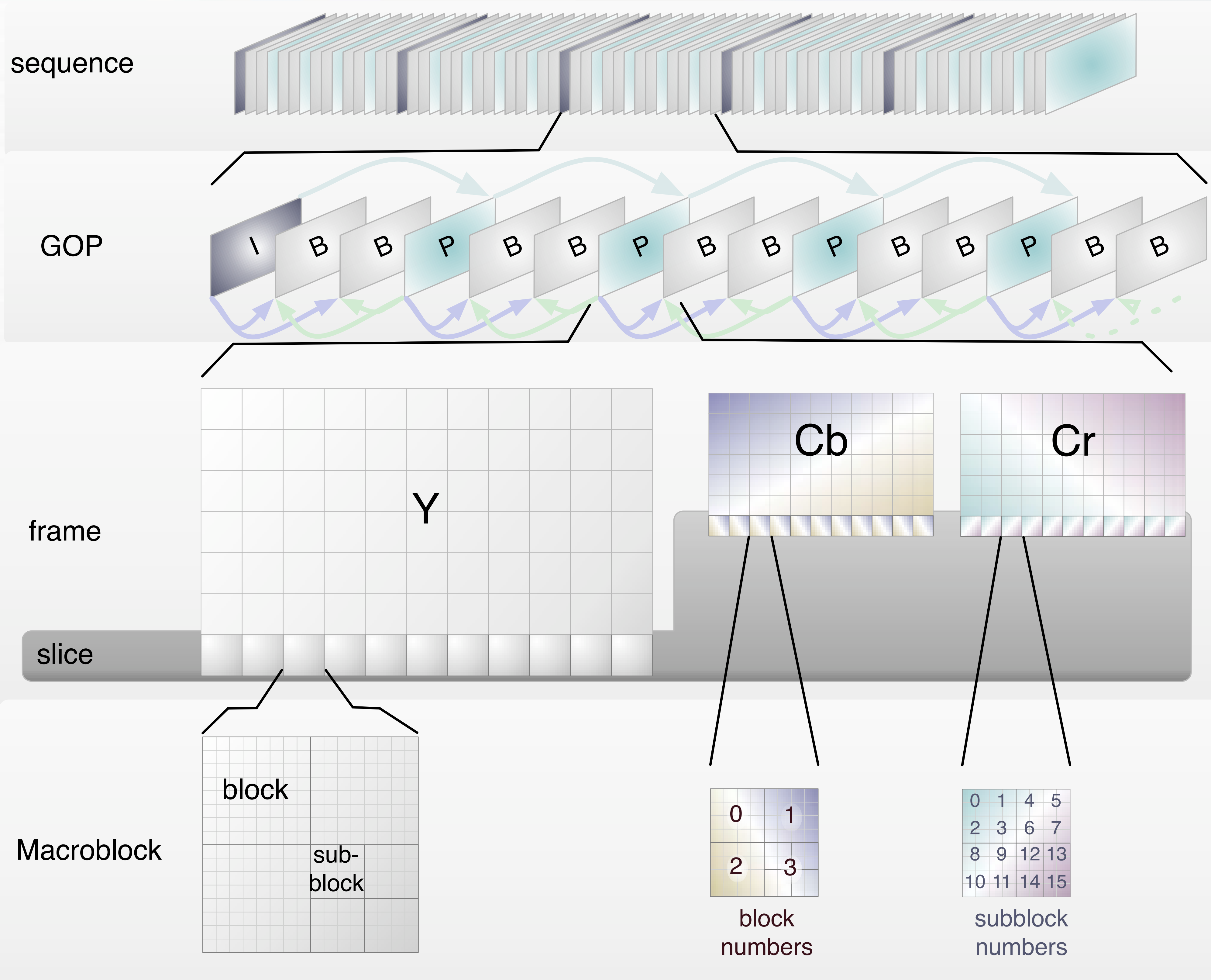


block partitioning



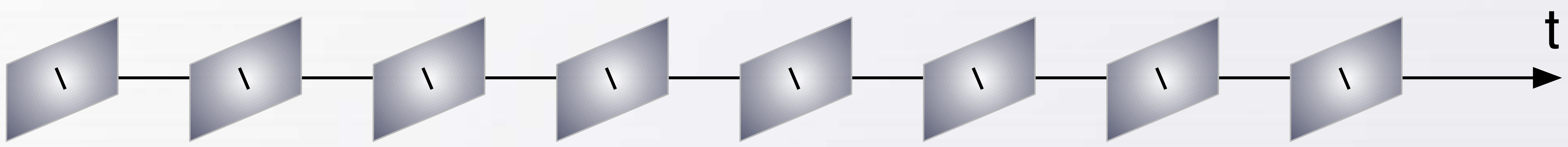
video coding standard semantics

- ➠ sequences are divided into groups of pictures, where frames depend on each other
- ➠ frames can be transmitted as partial packets, called slices
- ➠ basic processing uni of current video coding standards is the Macroblock, consisting of Y, Cb, Cr (ex.: 4:2:0)



basic frame types: I-frames

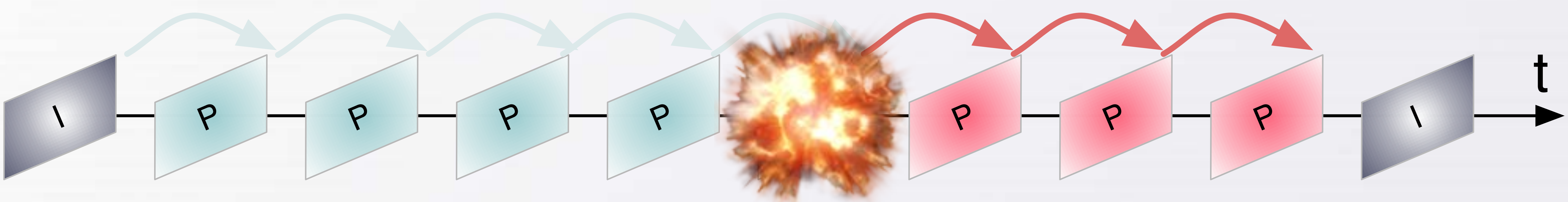
- **Intra-frames are independent intrinsic frames without reference to other frames**
  - (“I-frames“, „Intra-frames“ or „keyframes“)
- **required for coding/decoding start and random access to stream**
- **I-only sequences allow frame-accurate editing (DV, Motion-JPEG)**
- **low memory requirements in decoder**
- **drawback: low coding efficiency without temporal de-correlation**



Henryk Richter

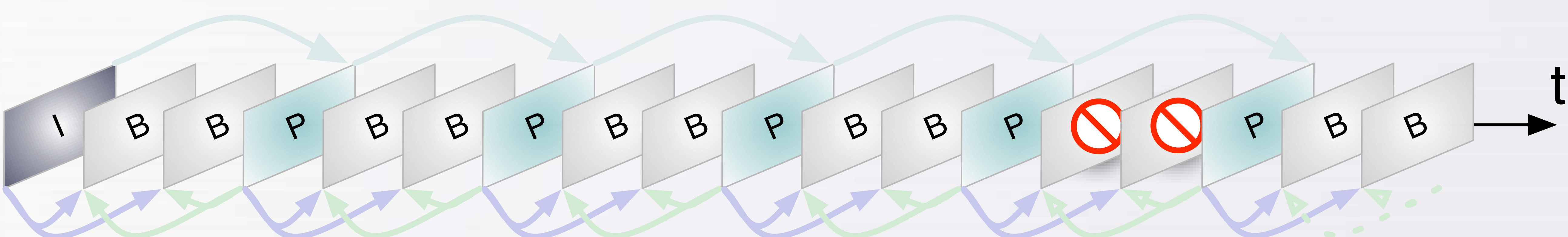
basic frame types: P-frames

- P-frames (Predicted frames) are motion compensation predicted from frames transmitted earlier within the current GOP
- motion compensation limited to single direction per block ("forward prediction")
- contain mostly difference information related to previous frames
- no random access to P-frames
- decoding error invalidates all subsequent predictive frames (until next I-frame)



basic frame types: B-frames

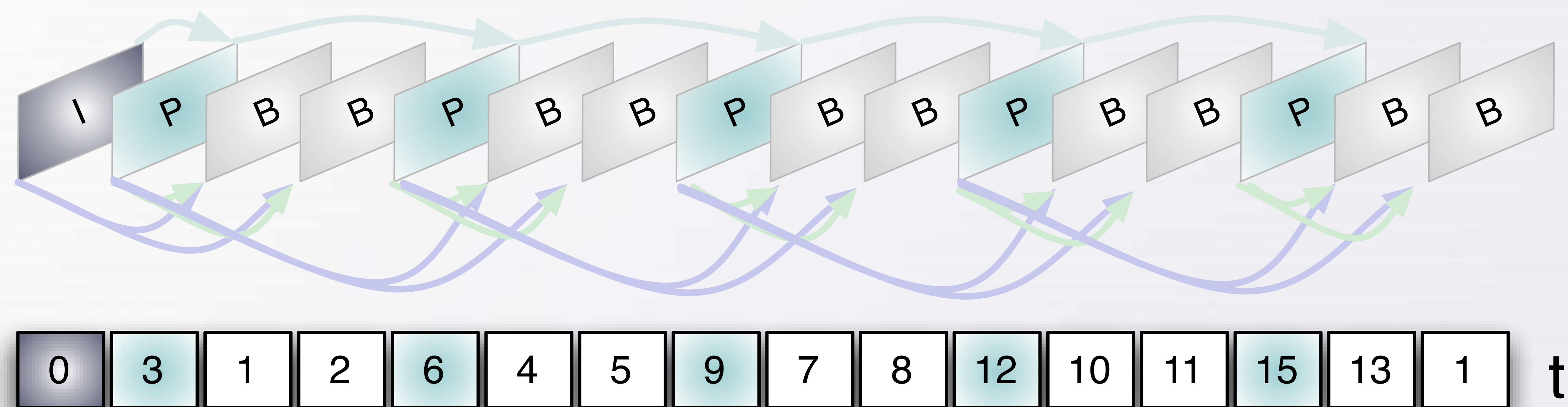
- **B-frames (Bi-directional frames) are motion compensation predicted in two directions**
- **motion compensation either forward, backward or bidirectional**
  - bidirectional by averaging forward and backward predictors
- **contain mostly difference information related to previous frames**
- **in most standards not used as reference frames (disposable)**
  - provide cost-effective means of fast-forward, random-access
  - coded with less residual information



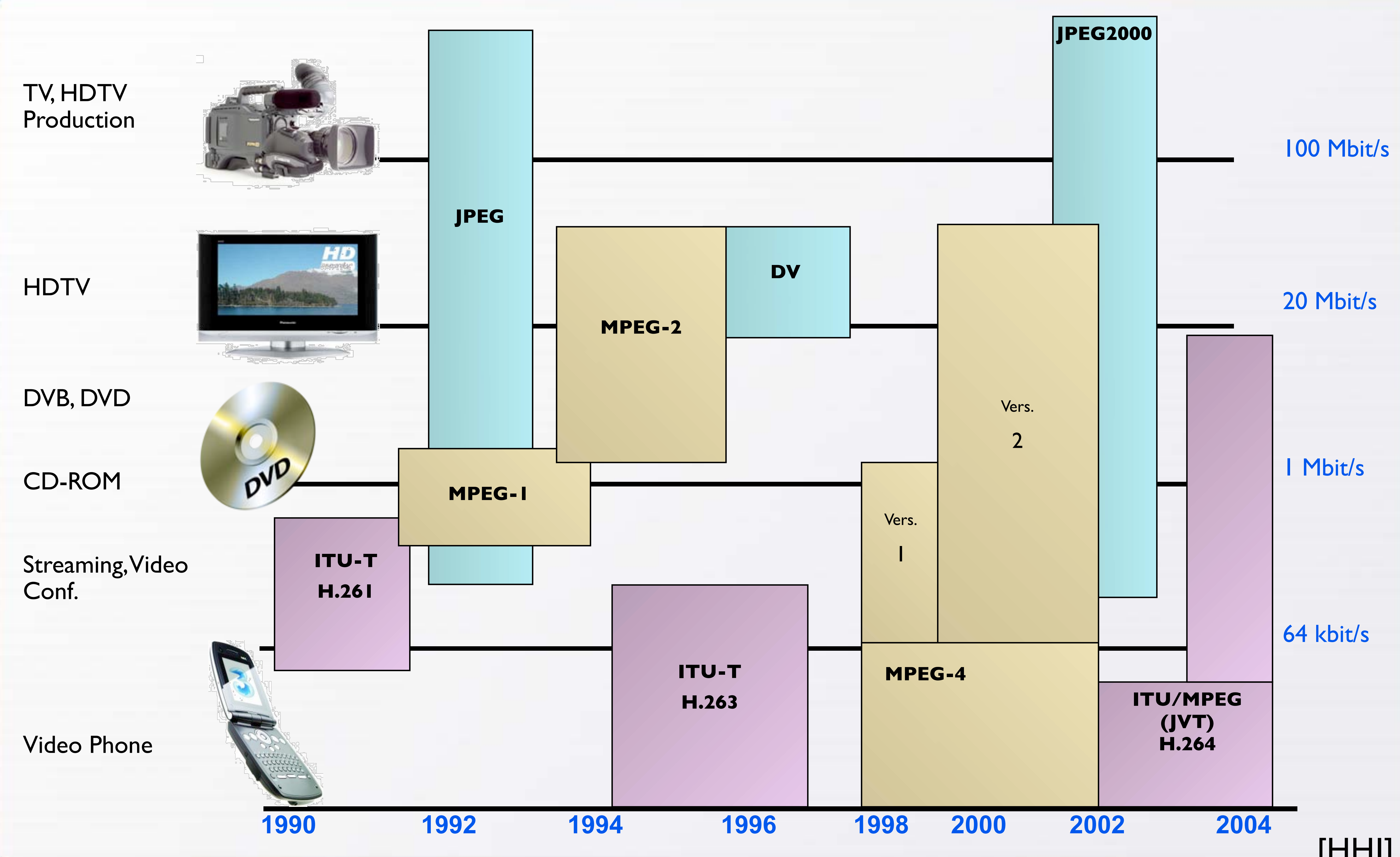
*display order

basic frame types: B-frames (2)

- **drawbacks:**
  - to ensure causality, display order and coding order are different
  - residual information bits in B-frames do not contribute to other frames
  - temporal distance between anchor P-frames leads to less correlation and higher bitrate in P-frames (➠ successive B-frame limit)
  - additional frame delay in encoder and decoder
    - (bad for streaming, conversational applications)
  - increased encoder and decoder computational complexity and memory requirements



video coding standards overview



video coding standards

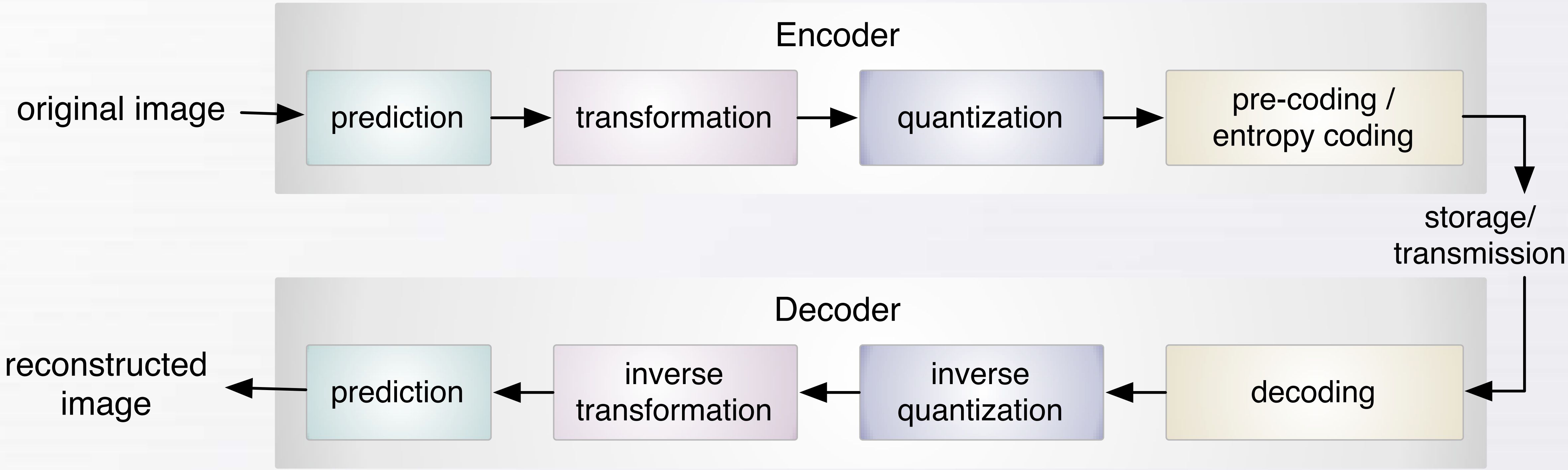
- **scope**
  - data stream representation (systems, audio, video)
  - mandatory decoder operations, **NOT** encoder
- **constraints**
  - video streams
    - image resolution and representation
    - frame rate
  - audio streams
    - sampling rates
    - bits per sample and sampling format
    - number of channels
  - general
    - data rates
    - allowed tools

MPEG

- **Motion Picture Experts Group (http://www.mpeg.org)**
  - MPEG-1 (1993)
    - audio/video coding, bitrate ≤ 1.5 MBit/s
  - MPEG-2 (1996)
    - audio/video coding, initial bitrate ≤ 15 MBit/s, various profiles
    - digital TV (DVB-S,-C,-T), DVD
  - MPEG-3 (skipped)
    - planned as HDTV standard, features incorporated into MPEG-2
  - MPEG-4 (2000, version 2)
    - audio/video coding for internet applications, mobile devices
    - object based coding, low bitrate coding
    - not limited to PCM data (structured audio, 3D objects)
  - MPEG-7
    - media descriptions, meta information, search/retrieval
  - MPEG-21
    - planned universal media framework

Hybrid Video Coding

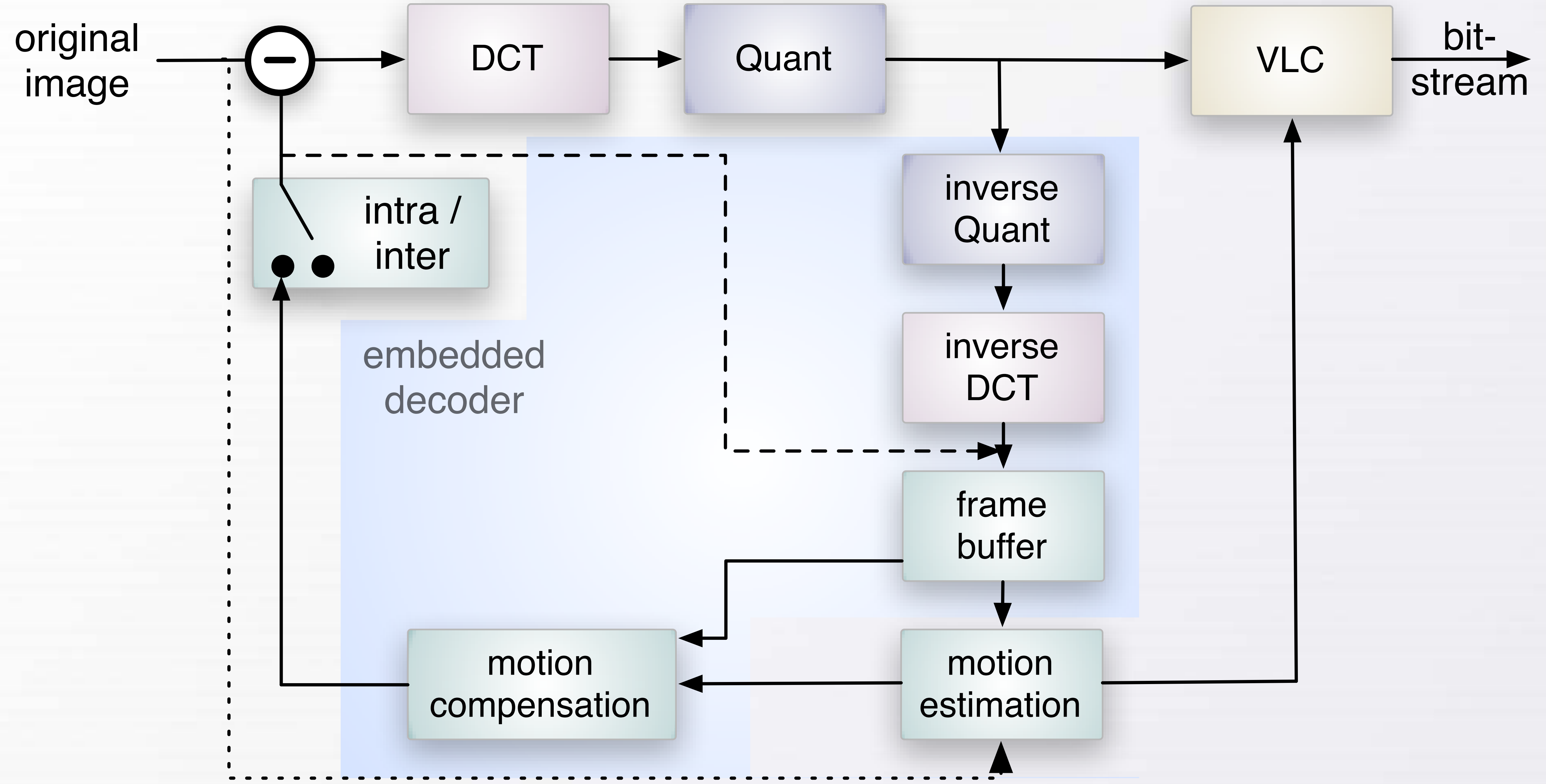
- **combination of motion compensated prediction and transform coding of the residual**
  - lossless
    - prediction step where the prediction error is retained
    - transformation step
    - coding of remaining data
  - lossy
    - quantization



MPEG-1

- **Finalized 1993 (ISO/IEC 11172)**
- **first MPEG-Standard for digital video**
  - primary target Video-CD, CD-I
    - ~1.5 MBit/s overall data rate, 64...192 kBit/s Audio
    - image format 352x288 pixel @25 FPS, 352x240 pixel @ 30FPS
    - up to 74 min video on single-speed CD
- **parts**
  - Part1: Systems (packet stream, a/v multiplex)
  - Part2: Visual (MPEG-1 video coding)
  - Part3: Audio (MPEG-1 Audio Layers I,II,III)
  - Part4: Compliance testing
  - Part5: Reference Software Model

MPEG-1 schematic



„closed loop coding“:
- decoder already part of encoder
- reference frames in encoder buffer match the frames available to the decoder

MPEG-1 quantization of DCT coefficients

- **Intra DC coefficient**
  - fixed (de-)quantizer $F'(0,0) = F_Q(0,0) \cdot 8$
  - reason: block artifacts
- **AC coefficients**
  - intra quantizer matrix W(u,v) perceptually modeled
  - inter quantizer matrix flat
  - linear quantizer scale q_{scale} = 1,2,...,31
  - quantization formula
    $$F_Q(u,v) = \big(16\,F(u,v) + l\big)/(2\,W(u,v)\,q_{scale}\big)$$
    - intra: $l = sgn(F(u,v) \cdot W(u,v) \cdot q_{scale}$, inter: $l=0$
  - reconstruction formula
    $$F'(u,v) = \big(2\,F_Q(u,v) + k\big) \cdot W(u,v) \cdot q_{scale}\,/\,16$$
    - intra: $k=0$, inter: $k=sgn(F_Q(u,v))$

intra block matrix W_i

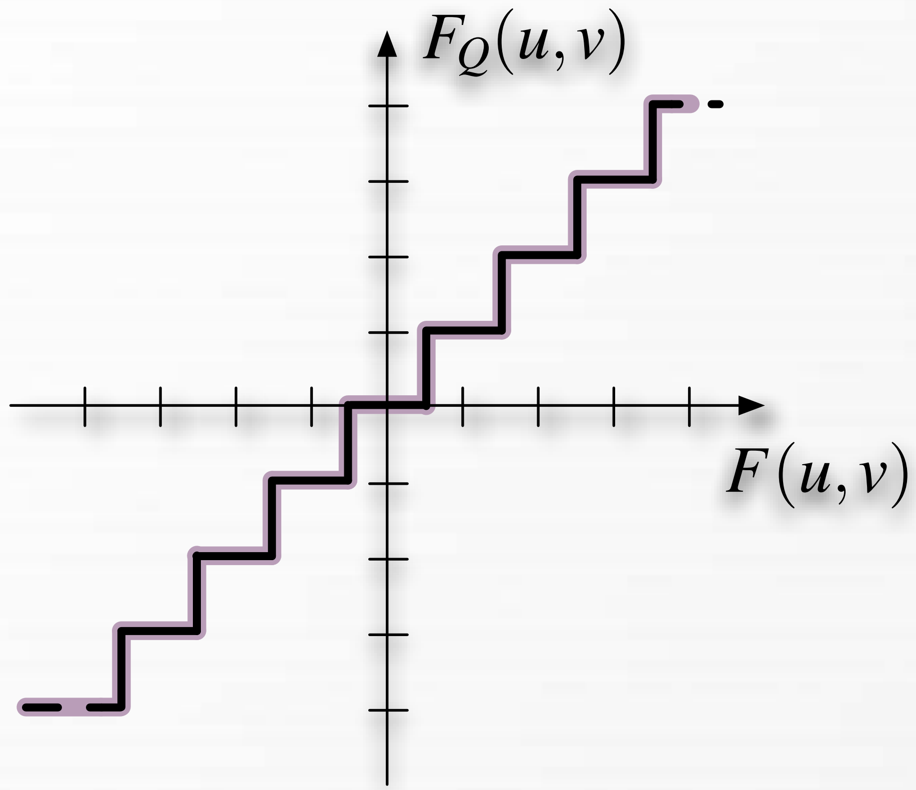
08	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

inter block matrix W_P

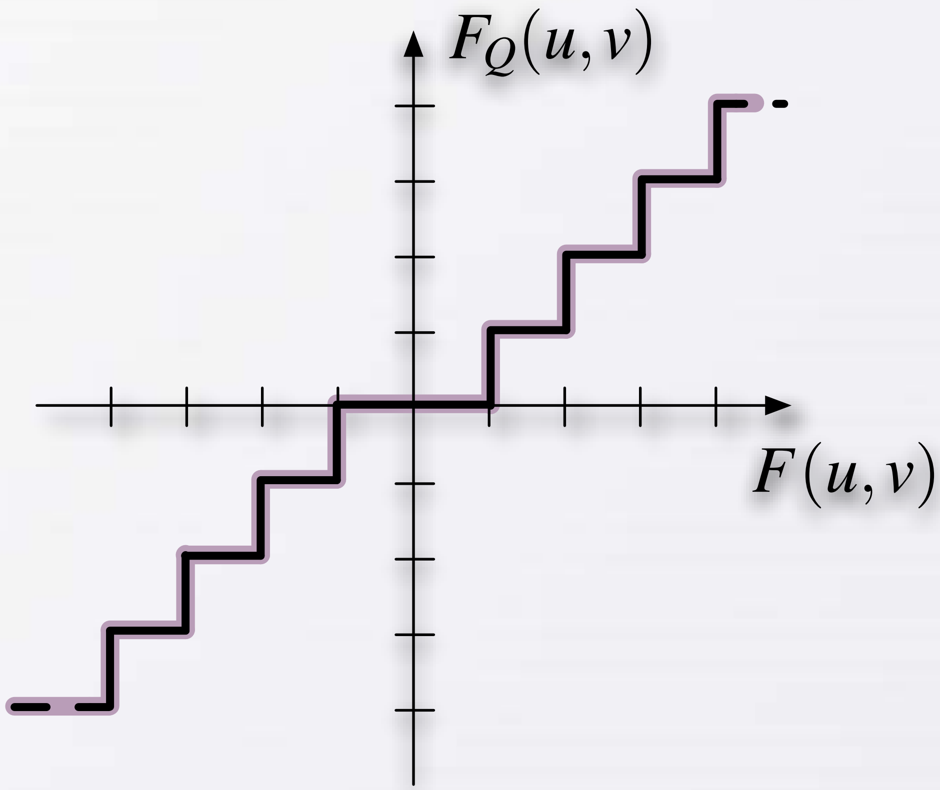
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

MPEG-1 quantization (2)

intra: rounding



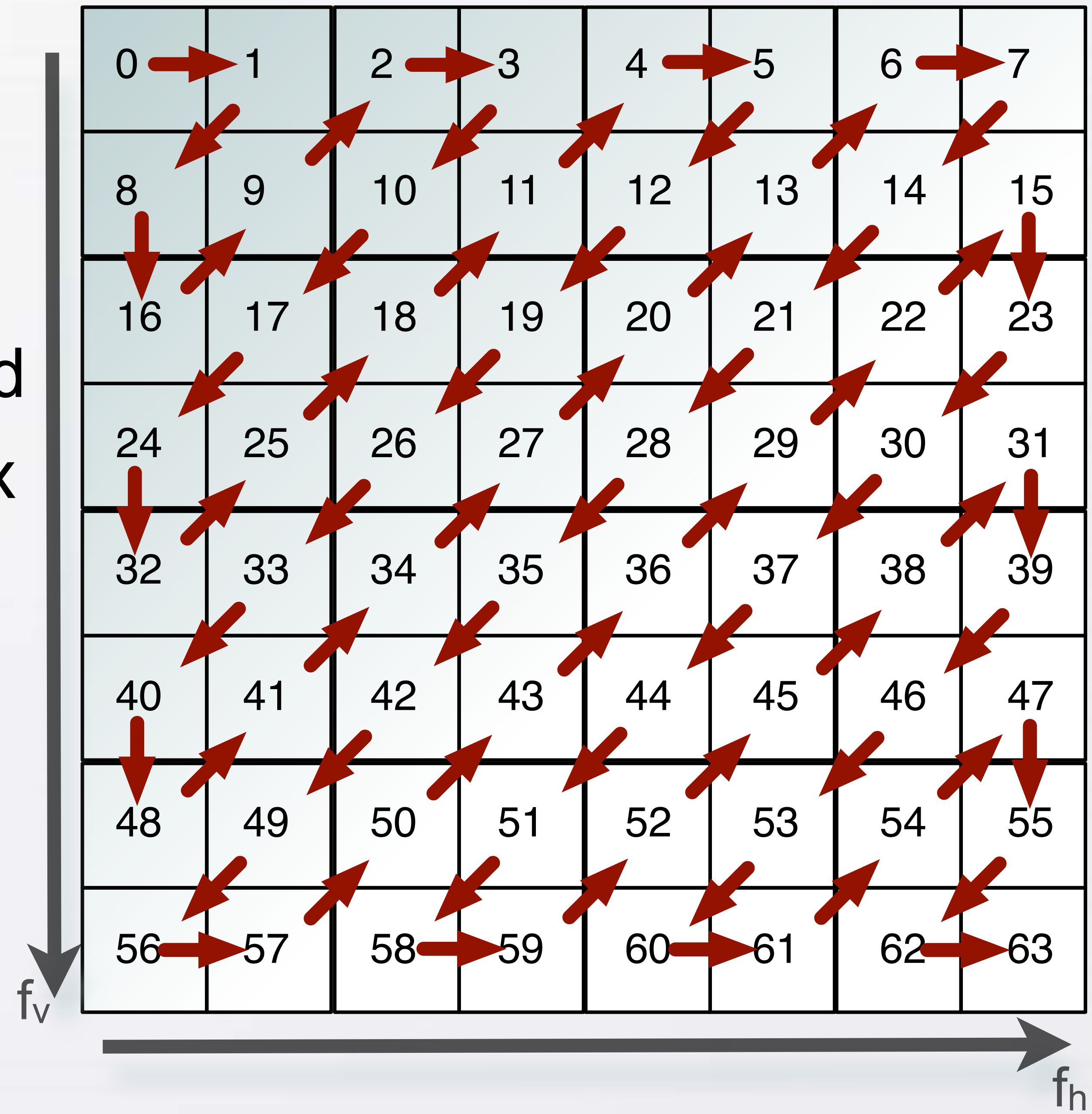
inter: truncation (dead zone)



- q_{scale} **is the main parameter for rate control**
- q_{scale} **can be changed between macroblocks (diff. transmission)**
- **intra quantization is tuned for accuracy**
- **inter quantization is tuned for suppression of small coefficients**

MPEG-1 block scanning

- **frequency progressive zigzag scanning**
  - DC coefficient coded separately
    - difference to previously coded DC
  - key concept: diagonal frequency progression
  - low frequency coefficients prioritized
  - compare to intra quantization matrix



MPEG-1 VLC

- **group symbols into „run“ and „level“**
  - „run“ is the number of zeroes before each significant coefficient in block-scan order
- **jointly transmit run and level as complex code**
  - uniform distribution of sign assumed and appended to all codes
  - finish transmitted block with EOB (end of block) marker
  - unlikely symbols prefixed with ESCAPE code and sent as FLC

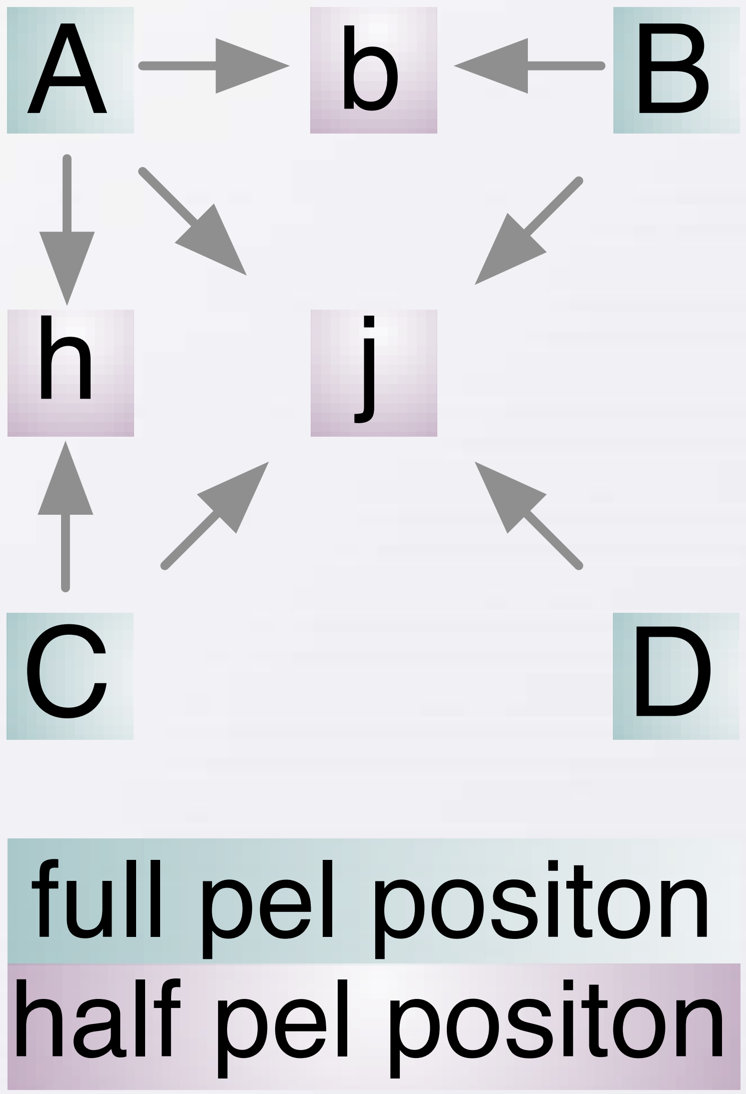
code	run	level
10	EOB	-
1s*	0	1
11s*	0	1
011s	1	1
0100s	0	2
0101s	2	1
00101s	0	3
00110s	4	1
00111s	3	1
000110s	1	2
000111s	5	1
1	ESCAPE	-

symbol-code-mapping (exemplary)

*different code of first symbol in inter blocks

MPEG-1 Motion Compensation

- **16x16, half pel resolution**
  - *virtual* half pixel positions gained by interpolation
  - rules for virtual pixels b,h,j
    - b = (A+B+1)/2
    - h = (A+C+1)/2
    - j = (A+B+C+D+2)/4
- **algorithm applied to luma and chroma, latter vectors scaled**
- **differential coding of motion vectors**
  - predictor from last macroblock
  - VLC



MPEG-1 Mode Decision

- **I-frames**
  - Intra Macroblocks only
- **P-frames**
  - Forward (last P/I-frame) 16x16 + MV
  - Intra
- **B-frames**
  - Forward (last P/I-frame) 16x16 + MV
  - Backward (next P/I-frame) 16x16 + MV
  - Bi-directional (last and next frame) 16x16 + 2 MV
  - Intra
- **optimal mode not trivial: rate-distortion problem**

MPEG-2

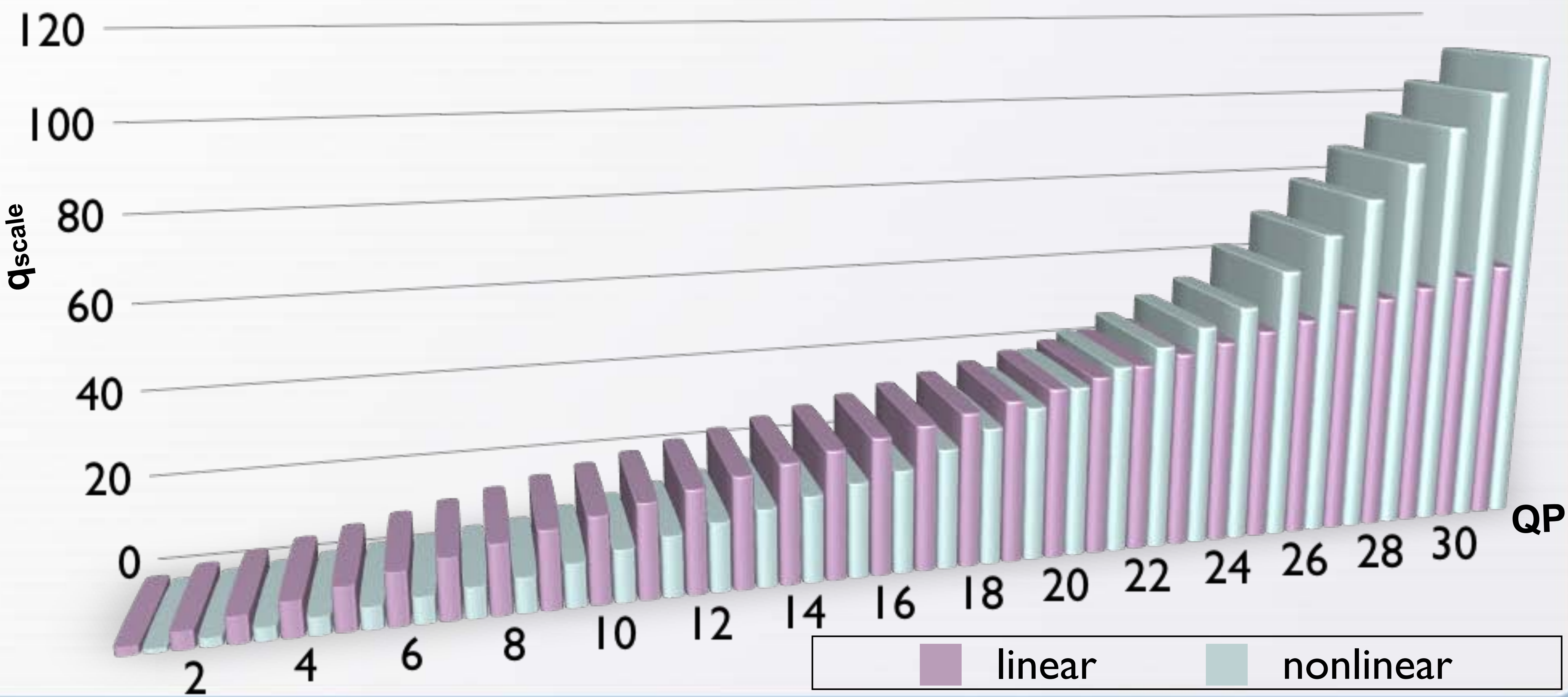
- **Finalized 1996 (ISO/IEC 13818), joint recommendation ITU-T H.262**
- **primary application digital TV**
  - from MPEG-1 up to HDTV (~1 MBit/s ... 80 MBit/s)
  - initial target: CCIR 601 interlaced video at 3-9 MBit/s (DVB, DVD)
- **parts**
  - Part1: Systems (program stream, transport stream),a.k.a. ITU-T H.222
  - Part2: Visual (MPEG-2 video coding)
  - Part3: Audio (MPEG Audio Layers I,II,III, multi-channel extensions)
  - Part4: Compliance testing
  - Part5: Reference Software Model
  - Part6: Digital Storage Media Command and Control (DSM-CC)
  - Part7: Advanced Audio Coding (AAC), formally added in 1997
  - Part9: Extension for real time interfaces
  - Part10: Compliance extensions for DSM-CC

MPEG-2 specifics

- **decoders required to decode legacy MPEG-1 streams**
  - extended syntax compared to MPEG-1, mostly unchanged semantics
- **main changes in video coding** [ISO00b] [MPFL97]
  - range extension for width, height, bitrate, etc.
  - interlace coding
    - picture structure (frame/field,adaptive frame/field)
    - motion compensation
    - DCT interlacing, coefficient scanning
  - scalable extension
  - error concealment
  - profiles and levels
  - extended aspect ratio information for non square pixels, „anamorphic“
  - extended IDCT mismatch control
  - stuffing macroblocks
  - AC run/level syntax

MPEG-2 specifics (2)

- **main changes (cont)**
  - intra DC quantization: dividers 8,4,2,1 for precision 8,9,10,11 Bit
  - AC quantization with doubled accuracy, linear or nonlinear QP scale
  - separate quantization tables for luma/chroma, custom matrices



MPEG-2 Profiles / Levels

profile	simple	main	4:2:2	high	spatial scalable	SNR scalable
short desc.	SP	MP	422P	HP	SCP	SNR
B-frames	no	yes				
scalable	no	no	no	SNR or spatial		SNR
subsampling	4:2:0	4:2:0	4:2:2	4:2:0 4:2:2	4:2:0	4:2:0

level	Low Level	Main Level	High 1440	High Level
short desc.	LL	ML	H-14	HL
width (pixels)	352	720	1440	1920
height (pixels)	288	576	1152	1152
framerate (Hz)	30	30	30	30
bitrate (MBit/s)	4	15	60	80

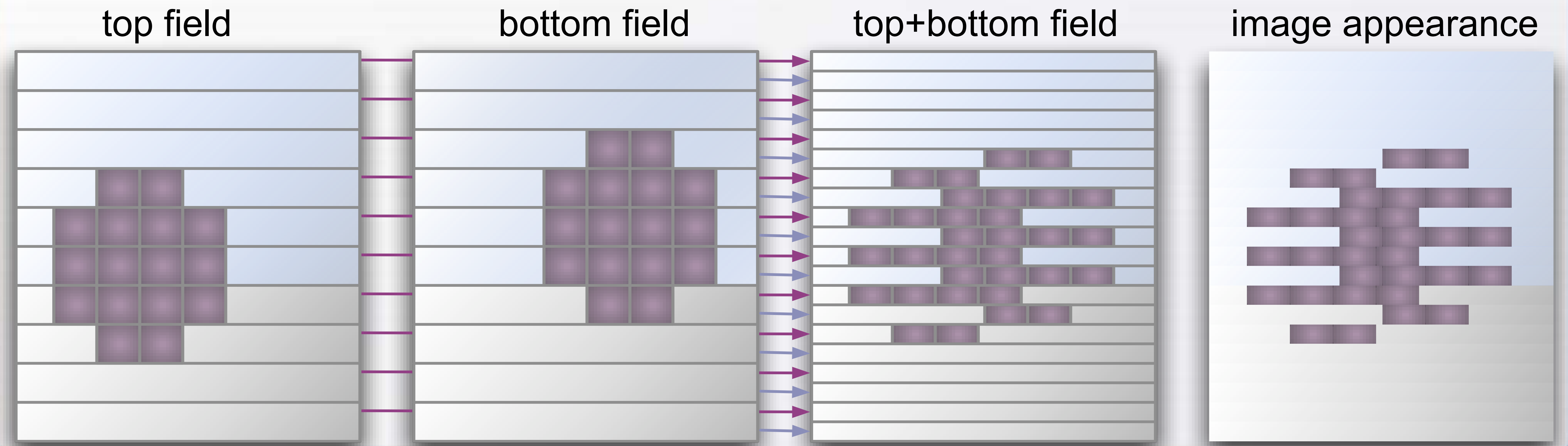
MPEG-2 Profile @ Level combinations

profile@level	resolution (pixels)	max. framerate	sub-sampling	bitrate (MBit/s)	application
SP@LL	176x144	15	4:2:0	0.096	mobile phones
SP@ML	352x240	24	4:2:0	0.384	PDA, streaming
	352x288	15			
MP@LL	352x288	30	4:2:0	4	(S)VCD
MP@ML*	720x576	25	4:2:0	15	DVD, DVB
	720x480	30			
MP@H-14	1440x1080	30	4:2:0	60	HDV (25 MBit/s)
	1280x720	30			
MP@HL	1920x1080	30	4:2:0	80	ATSC, HD-DVB
	1280x720	60			
422P@ML	720x576	25	4:2:2	50	Sony IMX
	720x480	30			

*Main Level note: for PAL/NTSC compatibility: either 720x576@25 Hz or 720x480@29,97 Hz

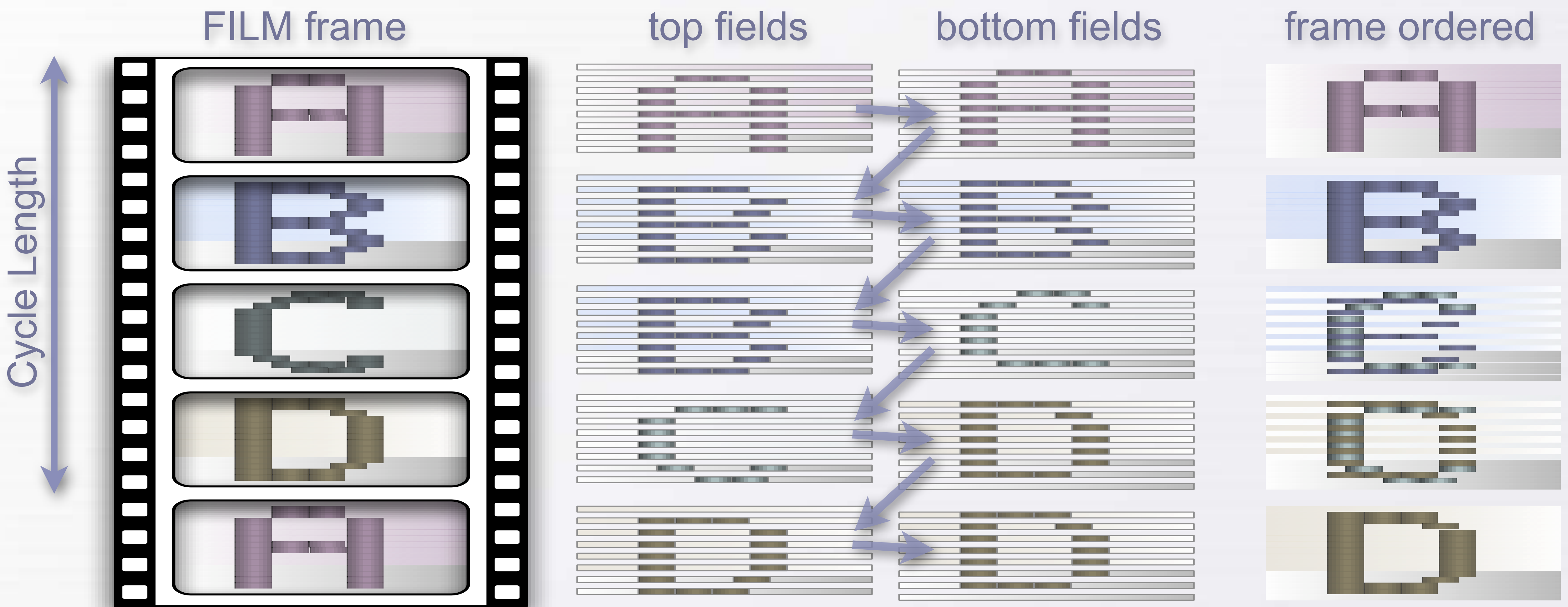
MPEG-2 Interlace

- **analog TV is based on half frames**
  - one line offset between „top“ field and „bottom“ field
  - 50/60 fields per second, temporal offset between fields is 1/50, 1/60 s
  - trade-off between smooth motion and vertical flickering, constrained by bandwidth
- **➠ compatibility of MPEG-2 with interlaced video is key feature for digital TV with existing TV-sets**



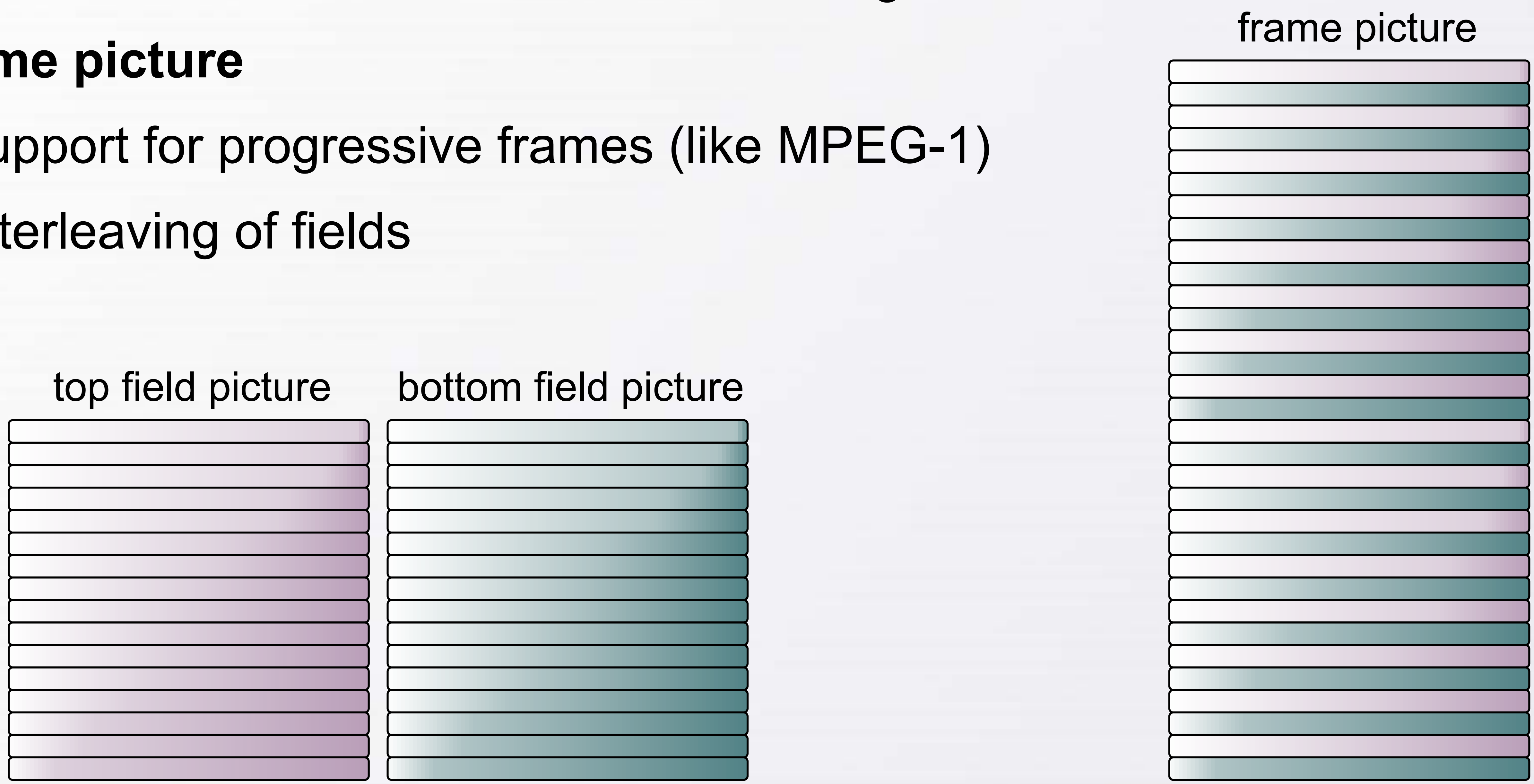
MPEG-2 3:2 pulldown

- **TV cameras provide interlaced material, cinema cameras do not**
- **24 Hz progressive cinema film conversion**
  - 29.97 Hz NTSC DVD by 3:2 pulldown after 0.1% deceleration
  - 25 Hz PAL DVD by 4% acceleration, two fields from one frame



MPEG-2 frame pictures / field pictures

- **two picture structures allowed in MPEG-2**
- **field picture**
  - fields coded as frames of half overall height
- **frame picture**
  - support for progressive frames (like MPEG-1)
  - interleaving of fields

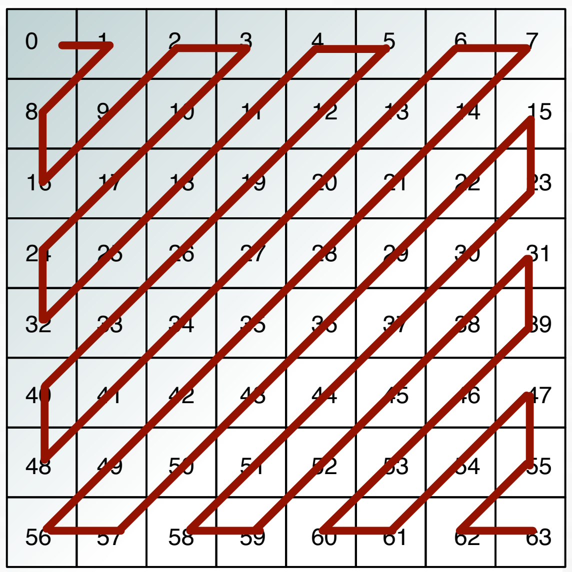


MPEG-2 interlaced DCT support

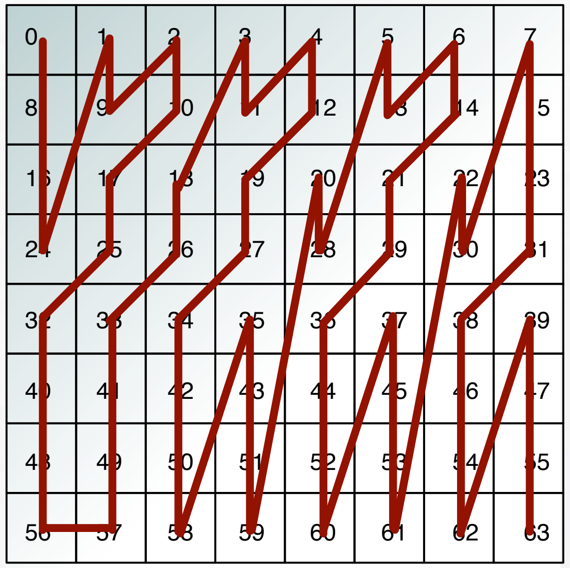
- alternate scan order on frame level, interlace DCT on macroblock level

coefficient scan order

normal scan order

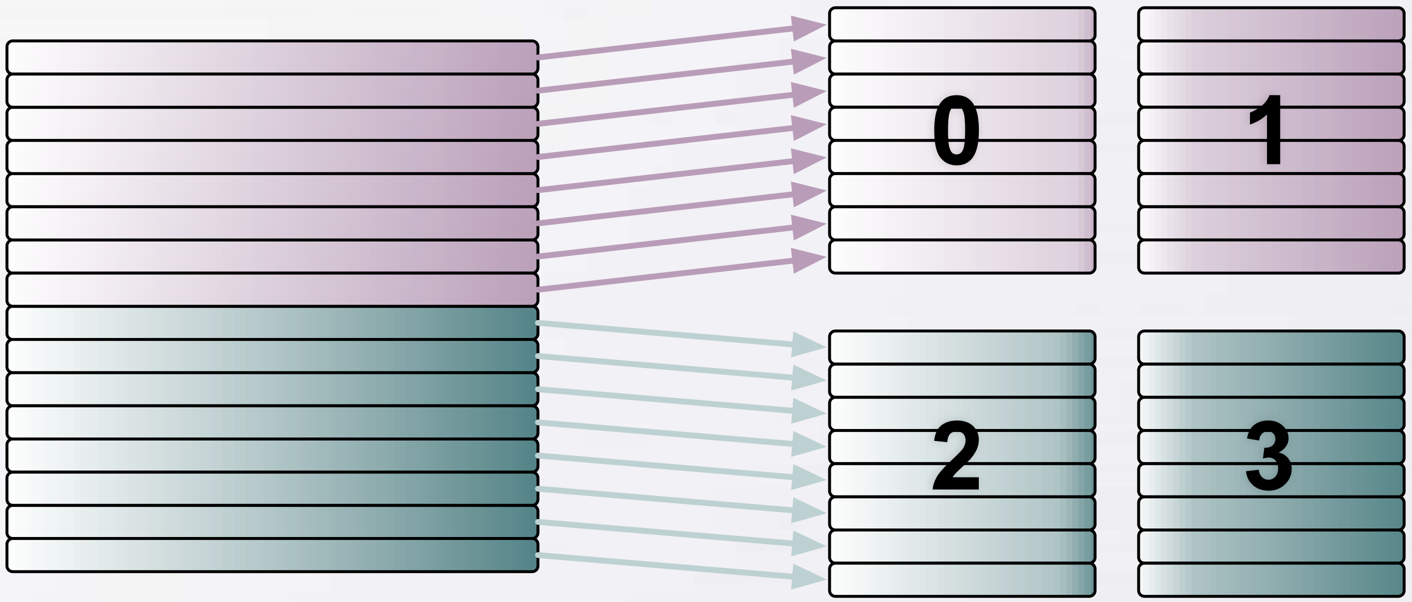


alternate scan order

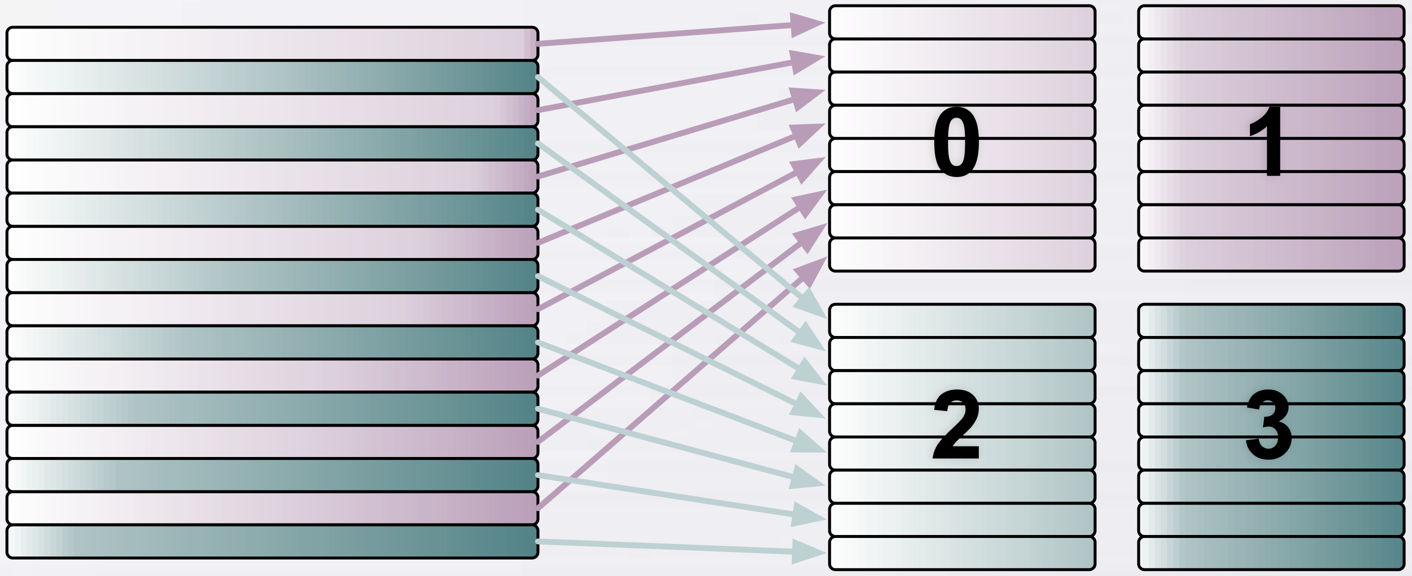


macroblock to DCT block mapping

progressive DCT permutation



interlace DCT permutation



MPEG-2 Scalability

- **graceful degradation of image quality in error-prone environments**
- **singular stream support for different transmission bandwidths**
- **singular support for devices of different capabilities**
  - SNR scalability (related to data partitioning)
    - base layer with coarse quantization
    - enhancement layer(s) with finer quantization
  - Spatial scalability
    - base layer at low spatial resolution
    - enhancement layer with higher spatial resolution
  - Temporal scalability
    - lower frame rate in base layer
  - Hybrid scalability

ITU-T Rec. H.263

- „Video coding for low bit rate communication“
- H.263 (version 1) finalized in 1996 [H263]
- H.263+ (version 2) finalized in 1998
- H.263++ (version 3) finalized in 2000

- Key design choices/requirements
  - use of available technology
  - low complexity (for low cost devices)
  - interoperability and/or coexistence with other standards (H.320,H.261)
  - error robustness
  - hardware or software implementation
  - QOS parameters: resolution, delay, frame rate, color performance
  - application: video phone or multimedia terminal

ITU-T Rec. H.263 (2)

- **H.263 version 1 provides the base syntax for MPEG-4 Part 2**
  - H.263 baseline profile must be supported by all compliant MPEG-4 video decoders
  - after adoption of H.263v1, divergent development of H.263 and MPEG-4
- **popular applications (today)**
  - video functionality of mobile phones
    - 3GPP standard: MP4 container, H.263 content
  - video streaming
    - initial video codec for Adobe® Flash Video
- **H.263+, H.263++ added features like quarter pel motion estimation, long term prediction, scalability**
  - effectively testbed for new technologies ⇒ H.26L project (JVT/H.264)

Institute of Communications Engineering

H.263 Profiles

profile	0	1	2	3	4	5	6	7	8
unrestricted motion compensation		✓	✓	✓	✓	✓	✓	✓	✓
extended motion vector range						✓	✓	✓	✓
arithmetic coding									
8x8 motion compensation		✓	✓	✓	✓	✓	✓	✓	✓
overlapped block motion compensation (OBMC)			✓			✓	✓	✓	✓
PB-frames									
advanced intra coding		✓		✓	✓	✓	✓	✓	✓
deblocking filter		✓		✓	✓	✓	✓	✓	✓
slice structured coding (without submodes)				✓	✓		✓		✓
slice structured coding with arbitrary slice order							✓		✓
Temporal (B-frames), Temporal, SNR, Spatial Scalability, B-frames for temporal scalability									✓
data partitioning					✓				
reference picture resampling (warping)									✓
Enhanced Reference Picture Selection						✓	✓	✓	✓

H.263 | H.263+ (version 2) | H.263++ (version 3)

H.263 Levels

level	10	20	30	40	45	50	60	70
Max picture format	QCIF 176x144	CIF 352x288	CIF 352x288	CIF 352x288	QCIF 176x144 CPFMT for H.263+	CIF, CPFMT	720x288 CPFMT	720x576 CPFMT
Min. picture interval	2002/(30000)s	2002/(30000)s for CIF 1001/(30000)s for QCIF,sub-QCIF	1001/(30000) s	1001/(30000)s	2002/(30000)s CPCFC for H.263+	1/50s for CIF or lower 1001/(60000)s at 352x240 or smaller CPCFC	1/50s at 720x288 or lower 1001/(60000)s at 720x240 or smaller CPCFC	1/50s at 720x576 or lower 1001/(60000) s at 720x480 or smaller CPCFC
Max bit rate in kBit/s	64	128	384	2048	128	4096	8192	16384

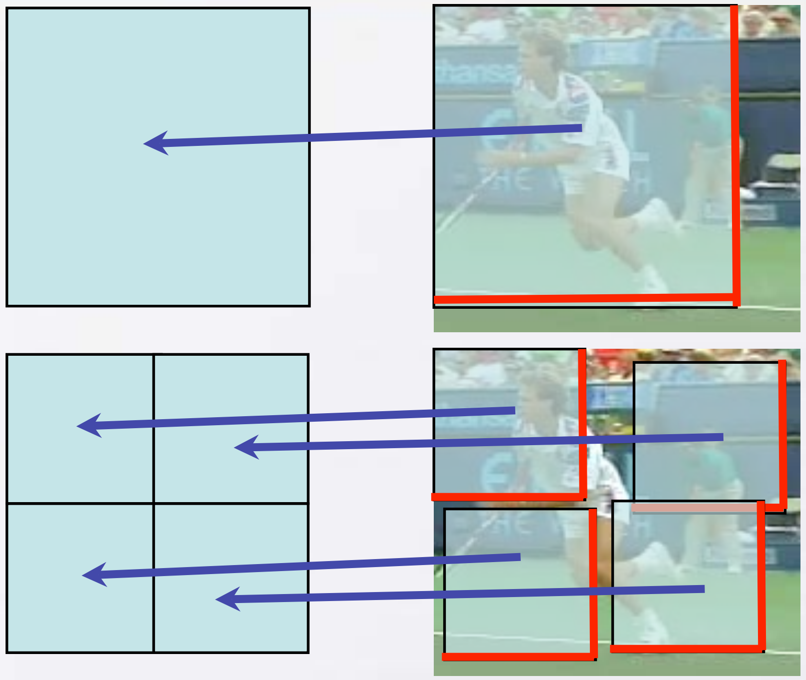
H.263+, H.263++

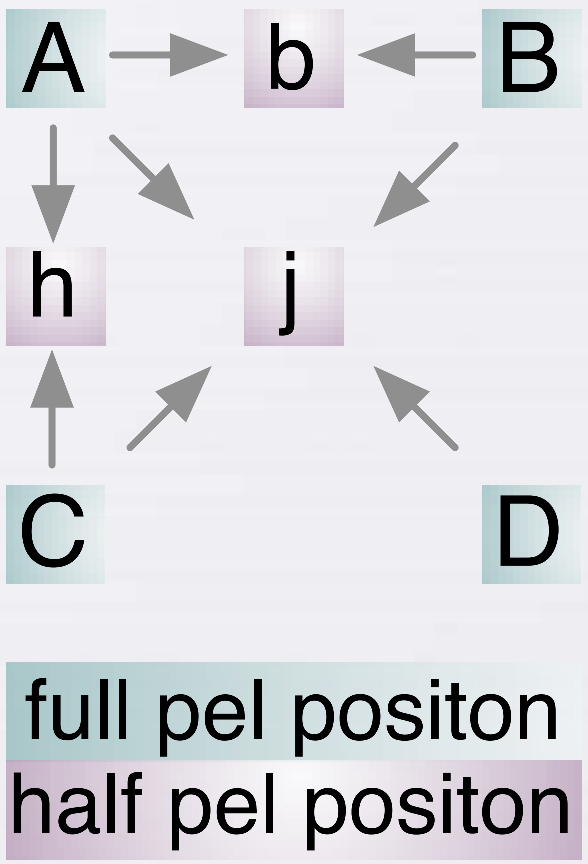
CPCFC = Custom Picture Clock Frequency   CPFMT = Custom Picture Format

- **predefined set of spatial resolutions**
  - sub-QCIF, QCIF, CIF, 4CIF, 16CIF
  - custom sizes only in H.263+, H.263++

H.263 Motion Compensation (I)

- **16x16, half pel resolution**
  - one vector per macroblock
- **8x8, half pel resolution (APM)**
  - four vectors per macroblock
- **interpolation rounding control**
  - traditional MPEG-1/-2 method biased towards „+1“
  - rules for virtual pixels b,h,j
    - $b = (A+B+1-RC)/2$
    - $h = (A+C+1-RC)/2$
    - $j = (A+B+C+D+2-RC)/4$
  - RC recommended to alternate values 0 and 1 between consecutive P-frames





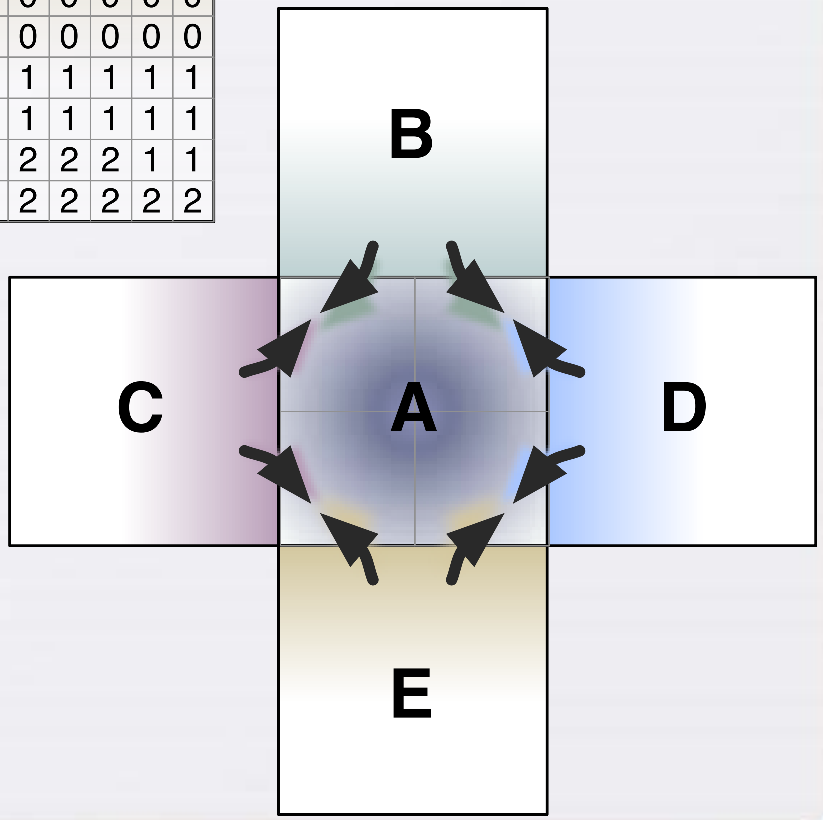
H.263 OBMC

- **overlapped block motion compensation**
  - motion vectors of surrounding blocks applied in addition to current block
  - each pixel predicted from 3 sources
  - luma only
- **prediction weighting**
  - prediction weights depend on distance to block center
  - overlapping 4 pixels deep
- **goals**
  - reduced motion uncertainty, quantization noise
  - reduced block artifacts

$$W_A = \frac{1}{8}\begin{bmatrix} 4&5&5&5&5&5&5&4\\ 5&5&5&5&5&5&5&5\\ 5&5&6&6&6&6&5&5\\ 5&5&6&6&6&6&5&5\\ 5&5&6&6&6&6&5&5\\ 5&5&6&6&6&6&5&5\\ 5&5&5&5&5&5&5&5\\ 4&5&5&5&5&5&5&4 \end{bmatrix} \quad W_B = \frac{1}{8}\begin{bmatrix} 2&2&2&2&2&2&2&2\\ 1&1&2&2&2&2&1&1\\ 1&1&1&1&1&1&1&1\\ 1&1&1&1&1&1&1&1\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0 \end{bmatrix}$$

$$W_C = \frac{1}{8}\begin{bmatrix} 2&1&1&1&0&0&0&0\\ 2&1&1&1&0&0&0&0\\ 2&2&1&1&0&0&0&0\\ 2&2&1&1&0&0&0&0\\ 2&2&1&1&0&0&0&0\\ 2&2&1&1&0&0&0&0\\ 2&1&1&1&0&0&0&0\\ 2&1&1&1&0&0&0&0 \end{bmatrix} \quad W_D = \frac{1}{8}\begin{bmatrix} 0&0&0&0&1&1&1&2\\ 0&0&0&0&1&1&1&2\\ 0&0&0&0&1&1&2&2\\ 0&0&0&0&1&1&2&2\\ 0&0&0&0&1&1&2&2\\ 0&0&0&0&1&1&2&2\\ 0&0&0&0&1&1&1&2\\ 0&0&0&0&1&1&1&2 \end{bmatrix}$$

$$W_E = \frac{1}{8}\begin{bmatrix} 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 1&1&1&1&1&1&1&1\\ 1&1&1&1&1&1&1&1\\ 1&1&2&2&2&2&1&1\\ 2&2&2&2&2&2&2&2 \end{bmatrix}$$



H.263 unrestriced motion compensation

- **in MPEG-1/-2 and H.263 baseline, all motion compensated blocks must point inside image area**
- **unrestricted MC allows motion vectors point partially or completely beyond image borders**

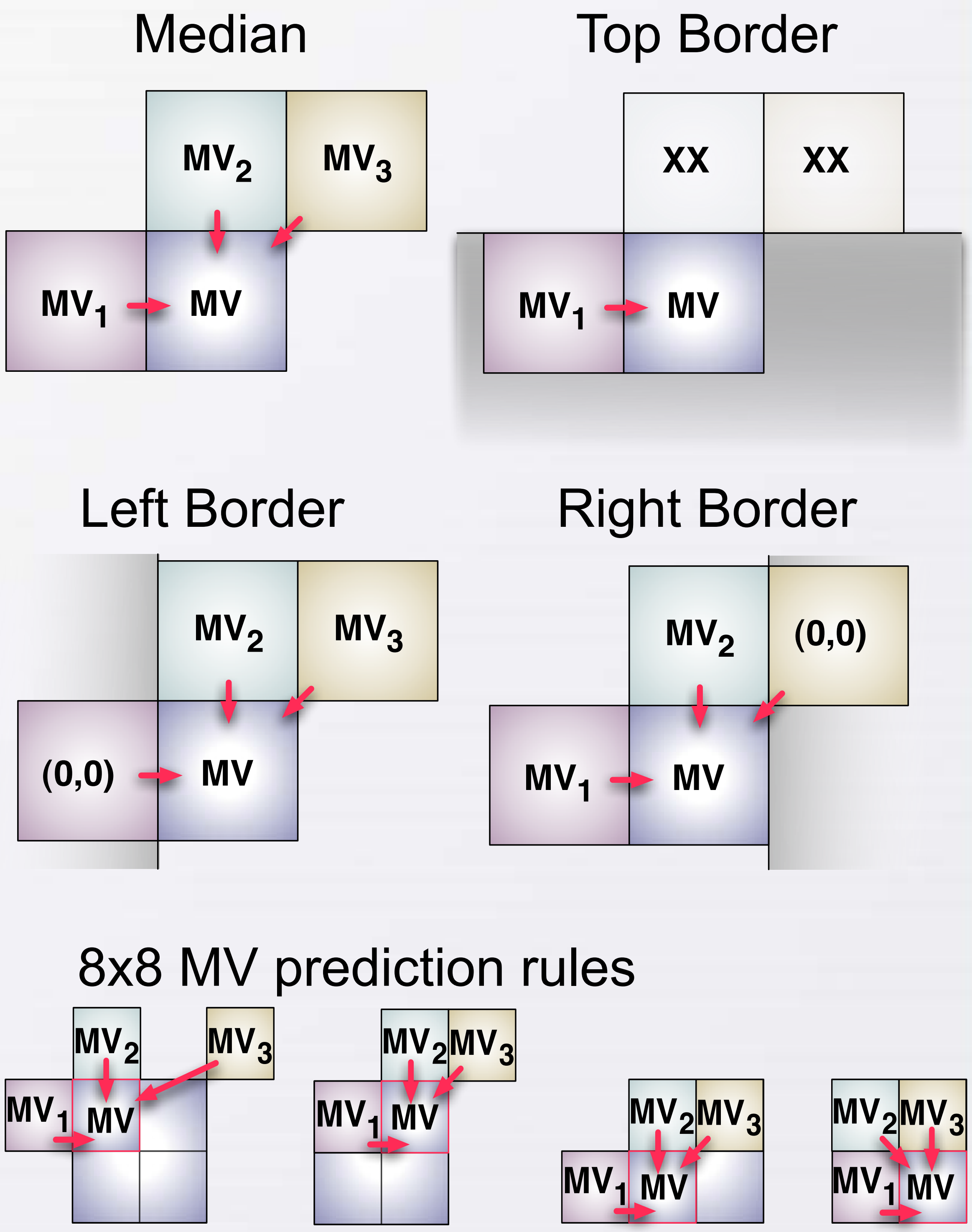
basic rule for any referenced pixel outside the image area:

replicate nearest border pixel of image

➠ main use: camera motion

H.263 motion vector prediction

- **motion vectors predicted from surrounding blocks**
- **component-wise (*x,y*) median of three predictors**
- **difference to median prediction is MVD, coded in bitstream**
- **special rules when macroblock is on image or GOB (group of blocks) border**
- **special rules for INTER4V (8x8) mode**



$$\mathrm{MEDIAN}(a,b,c) = a + b + c - \mathrm{MAX}(a,b,c) - \mathrm{MIN}(a,b,c)$$

H.263 quantization

- **Intra DC coefficient**
  - fixed (de-)quantizer $F'(0,0) = F_Q(0,0) \cdot 8$
- **AC coefficients**
  - uniform quantization matrix
  - quantizer indices QP = 1,2,3,...,31
  - even quantizer scale q_{scale} = 2,4,6,...,62
  - quantization formula

$$F_Q(u,v) = F(u,v)/q_{scale} \quad 0 \leq u,v \leq 7$$

  - reconstruction formula

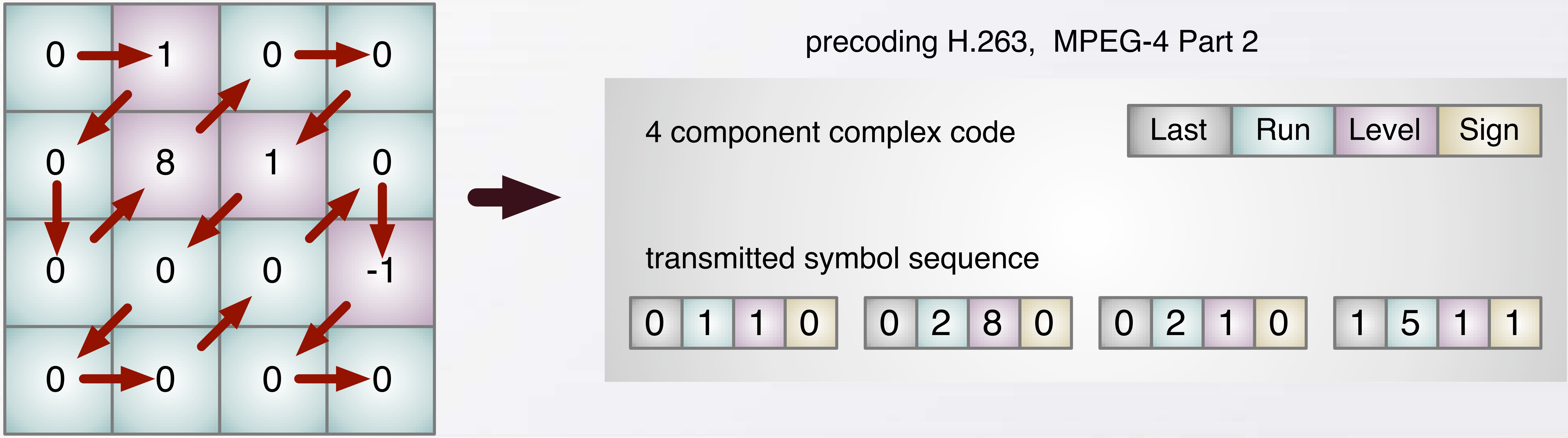
$$F'(u,v) = sgn\left(F_Q(u,v)\right) \cdot \frac{\left(2\left|F_q(u,v)\right| + 1\right) \cdot q_{scale}}{2} \quad 0 \leq u,v \leq 7$$

- ➠ **low complexity quantization with dead-zone around zero**

H.263 coefficient coding

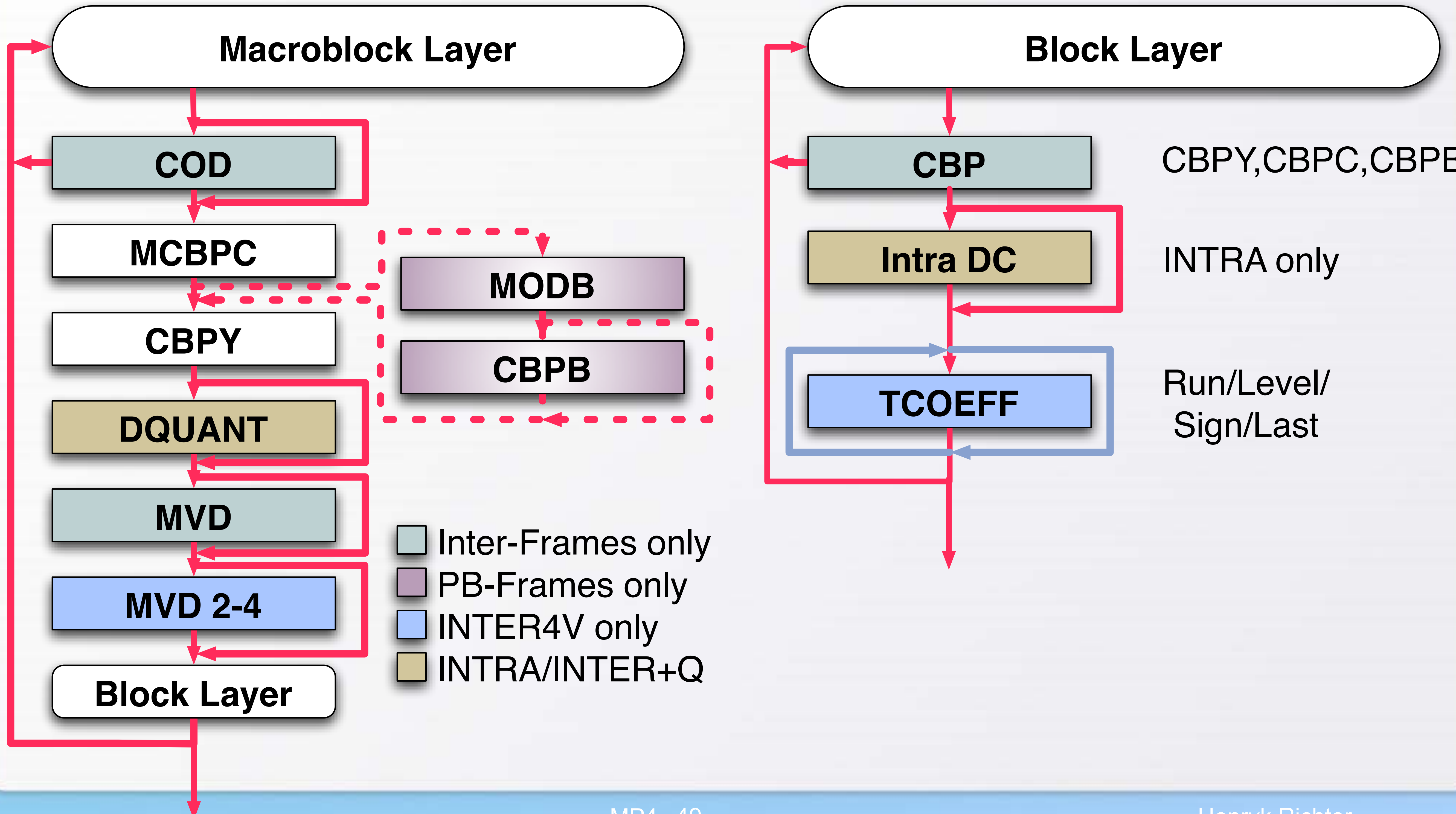
- **significant coefficients (level) associated with number of zeroes (run) between them**
- **same 8x8 zigzag scan order as in JPEG, MPEG-1**
- **symbols grouped into complex codes of run, absolute level, sign and indicator if last significant coefficient in block**
- **entropy coding by VLC tables (distinct Intra/Inter tables, ESCAPE symbol for improbable combinations via FLC)**

(exemplary) 4x4-block with coefficients



H.263 macroblock flow

- improved coding efficiency through skip flag and conditional elements
- complex codes instead of flag bits allow optimized VLC tables



H.263 MB types

picture type	MB type	Name	COD	MCBPC	CBPY	DQUANT	MVD	MVD_{2-4}
INTRA	3	INTRA		✓	✓			
INTRA	4	INTRA+Q		✓	✓	✓		
INTRA	Stuffing	-		✓				
INTER	Not coded	SKIP	✓					
INTER	0	INTER	✓	✓	✓		✓	
INTER	1	INTER+Q	✓	✓	✓	✓	✓	
INTER	2	INTER4V	✓	✓	✓		✓	✓
INTER	3	INTRA	✓	✓	✓			
INTER	4	INTRA+Q	✓	✓	✓	✓		
INTER	5	INTER4V+Q	✓	✓	✓	✓	✓	✓
INTER	Stuffing	-	✓	✓				

- mode decision optimization essential for good coding efficiency
- in low complexity image areas, COD flag reduces overall data amount for the respective macroblock to a single bit

MPEG-4

- **Title**
  - Information technology — Coding of audio-visual objects

- **Application areas**
  - Internet Multimedia
  - Interactive Video Games
  - Interactive Storage Media
  - Multimedia Mailing
  - Networked Database Services
  - Remote Emergency Systems
  - Remote Video Surveillance
  - Wireless Multimedia
  - IPC Interpersonal Communications (videoconferencing, videophone)

MPEG-4 Parts

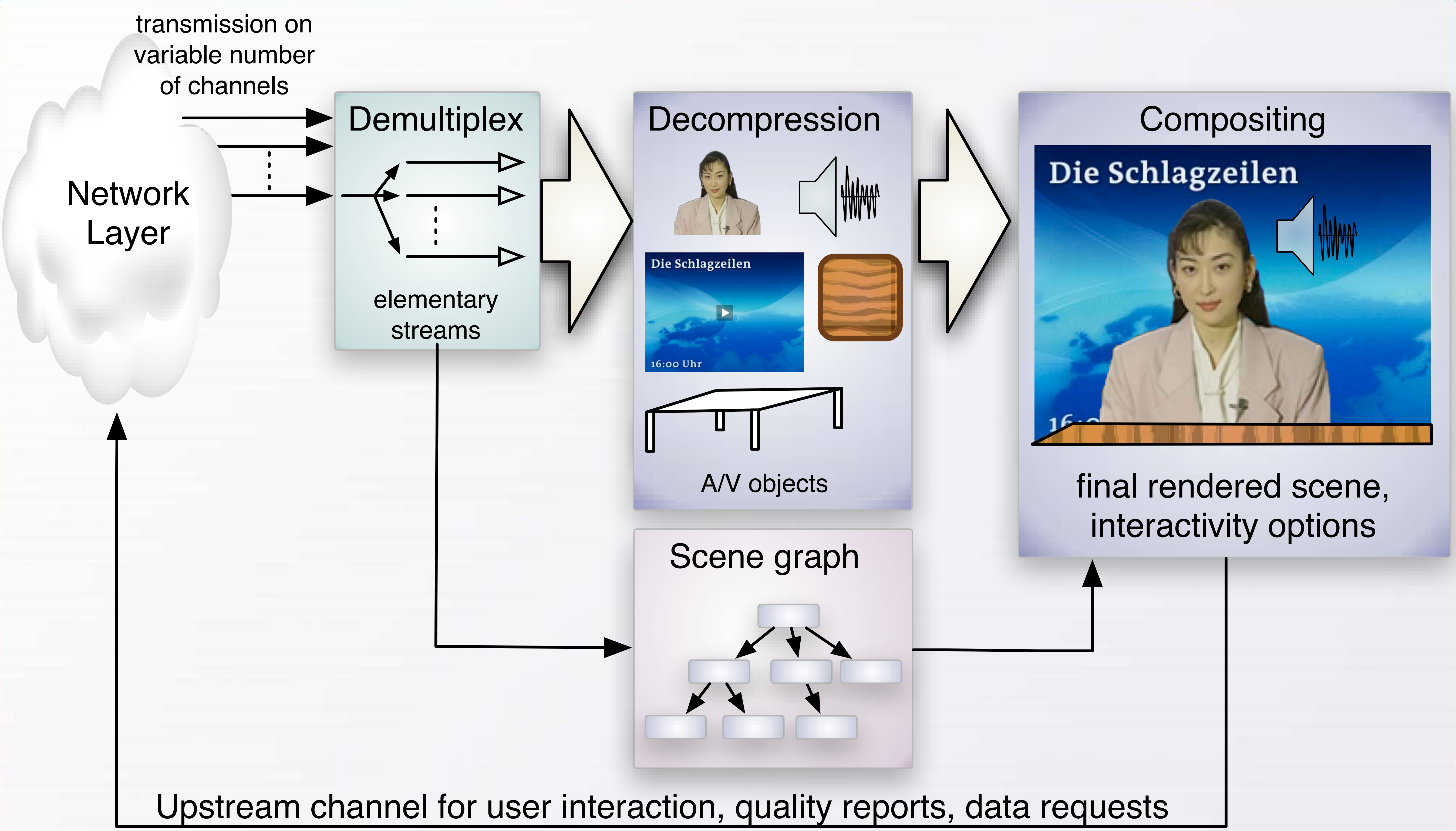
- **MPEG4 (ISO/IEC 14496)**
  - Part1 Systems (1999) (includes extensions to MPEG transport streams)
  - Part2 Visual (2001) (incorporates large parts from H.263)
  - Part3 Audio (2001) (includes new modes to AAC, CELP, HVXC speech codecs)
  - Part4 Conformance (2001)
  - Part5 Reference Software (2001)
  - Part6 Delivery Multimedia Integration Framework (2000)
  - Part 7 Optimized Visual Reference Software
  - Part8 Carriage of MPEG-4 Contents over IP (2002) (esp. RTP)
  - Part9 Reference Hardware Description (2003)
  - Part10 Advanced Video Coding (2003) (MPEG-4 AVC a.k.a. ITU-T H.264)
  - Part11 Scene description and Application engine, (BIFS)
  - Part12 ISO Base Media File Format: A file format for storing media content.
  - Part13 Intellectual Property Management and Protection (IPMP)
  - Part14 MPEG-4 File Format (based on Part12, derived from Quicktime file format)
  - Part15-23: Text representation/compression, texture compression etc.

MPEG-4 concepts

- **intented as generic audiovisual framework including**
  - audio
  - text
  - still images
  - video
  - 2D/3D mesh objects

- **visual representation**
  - scene graph, based on elementary objects
  - client-side compositing
  - arbitrary shaped objects
    - non-rectangular natural images and video, optionally with alpha-channel
    - synthesized objects (VRML)
  - interactivity, back channel information

MPEG-4 terminal, receiver side



MPEG-4 Visual Profiles

profile / tools	Simple	Core	Main	Advanced Real Time Simple	Advanced Coding Efficiency	Advanced Simple	Fine Granularity Scalable
Basic (I-VOP, P-VOP, AC/DC Prediction,4MV, unrestricted MV)	✓	✓	✓	✓	✓	✓	✓
Error resilience (Slice Resync, Data Partitioning, RVLC)	✓	✓	✓	✓	✓	✓	✓
Short Header (H.263)	✓	✓	✓	✓	✓	✓	✓
B-VOP		✓	✓		✓	✓	✓
H.263 + MPEG quantization		✓	✓		✓	✓	✓
P-VOP based temporal scalability		✓	✓		✓		
Binary Shape		✓	✓		✓		
Grey Shape (Alpha channel)			✓		✓		
Interlace			✓		✓	✓	✓
Sprite			✓				✓
Dynamic Resolution Conversion				✓			
NewPred				✓			
Global Motion Compensation					✓	✓	
1/4 Pel Motion Compensation					✓	✓	
SA-DCT					✓		
Fine Granularity Scalability							✓
FGS Temporal Scalability							✓

- Visual object types for raster image sources (i.e. natural images)
- additional object types for synthetic images
  - (2D Mesh, Texture, Face Coding)

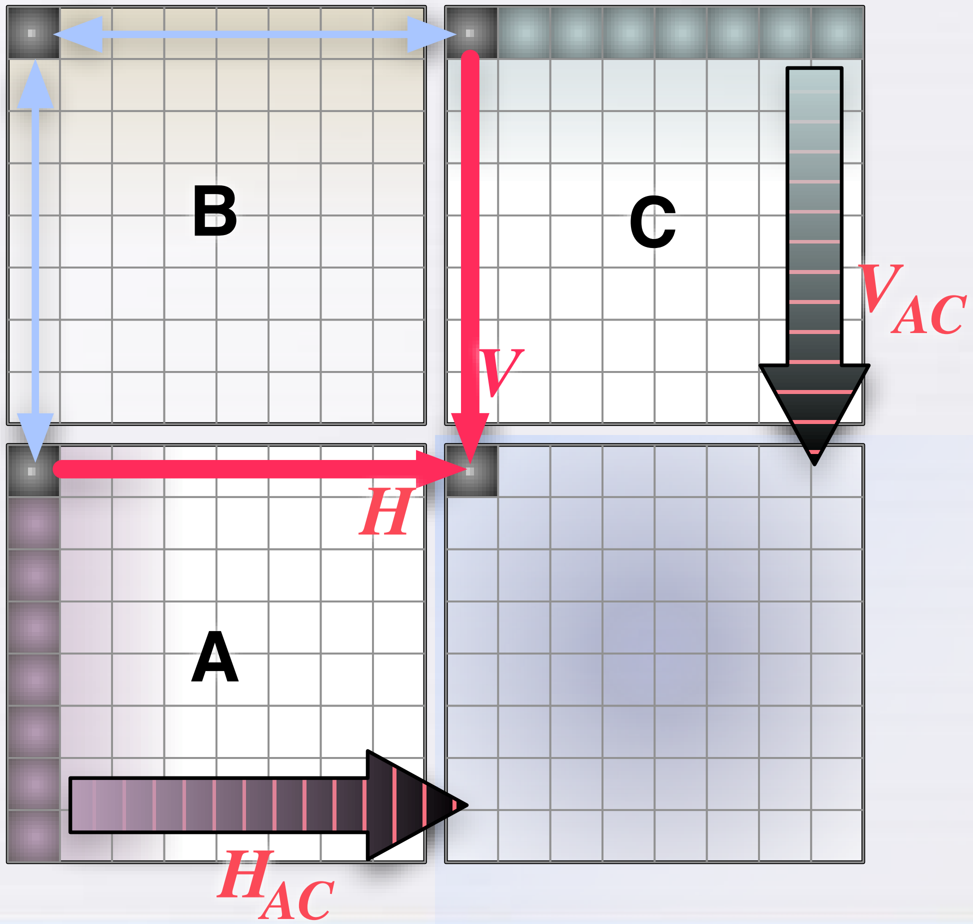
MPEG-4 popular implementations / derivates

- **DivX :-) 3.11**
  - Microsoft proposal MPEG-4 V3, rejected by ISO due to late entry
  - Microsoft released that codec as „MP43“ within their own media framework, artificially limited to ASF streams as target systems layer
  - codec is <u>not</u> MPEG-4 compliant, though lossless conversion feasible
  - feature set comparable to H.263, stunning image quality at that time
  - codec was hacked to allow inclusion into AVI files
  - DivX name originated from „Digital Video Express“, a rental service with limited disc lifetime or limited play counts, respectively
- **DIVX 4.0+, XVID, 3IVX**
  - MPEG-4 V.2 ASP codec framework, initially based on reference software „MoMuSys“, open-source release by DIVX forked into XVID
  - bitstreams are MPEG-4 Part 2 compliant, although proper interoperability is version-dependent with regressions in early versions
- **RealVideo (certain versions), Quicktime 6**

MPEG-4 AC/DC prediction

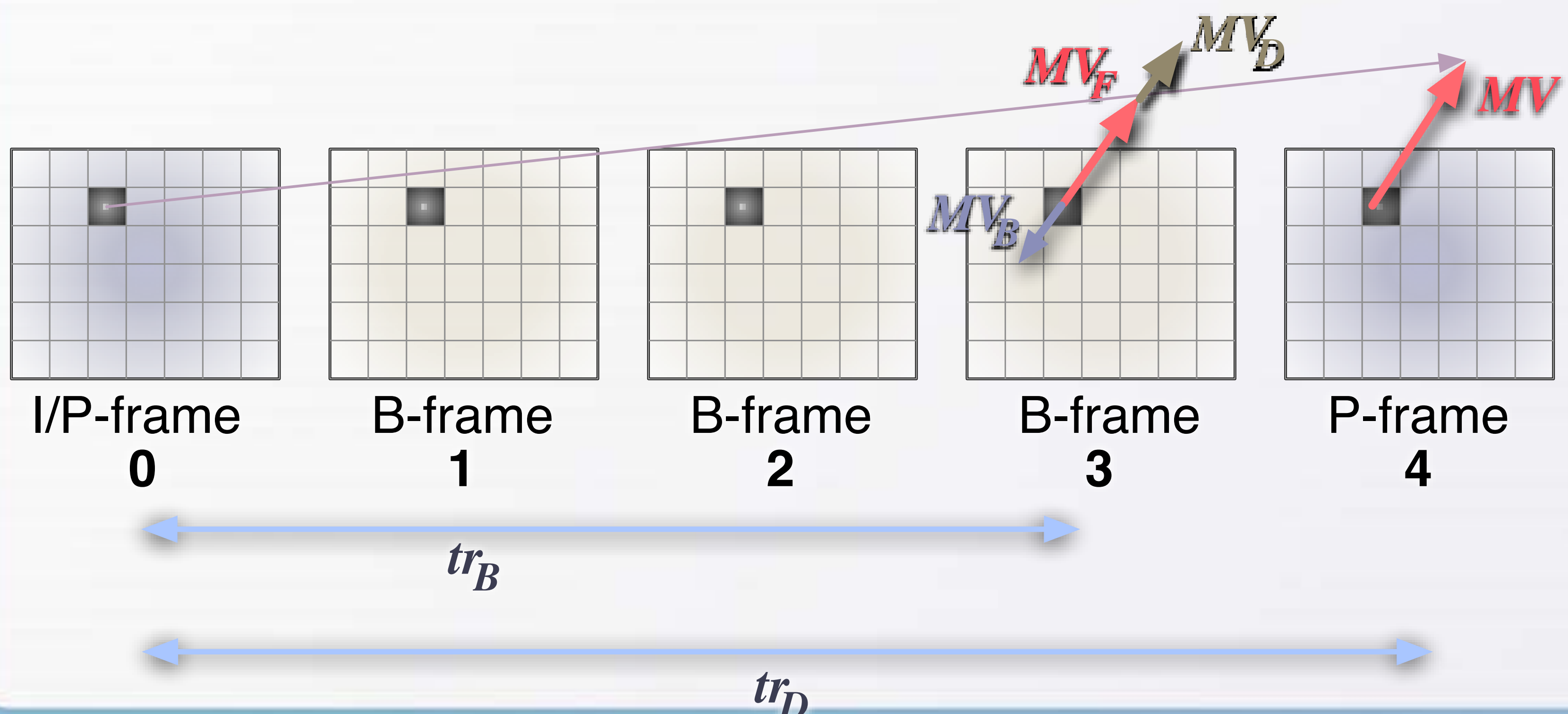
- **improved spatial de-correlation by prediction in frequency domain**
- **adaptive decision whether vertical or horizontal DC prediction**
- **flag for optional AC prediction**
  - direction derived from DC prediction decision
  - either first row or column of AC coefficients from upper or left block
  - predictor coefficients scaled to match QP or current MB
- **effect in spatial domain:**
  - horizontal/vertical block average copied to current block

$$
\begin{aligned}
&if\left(|F_B(0,0) - F_A(0,0)| < |F_B(0,0) - F_C(0,0)|\right)\\
&\qquad PRED = V\\
&else\\
&\qquad PRED = H
\end{aligned}
$$



MPEG-4 B-Frames

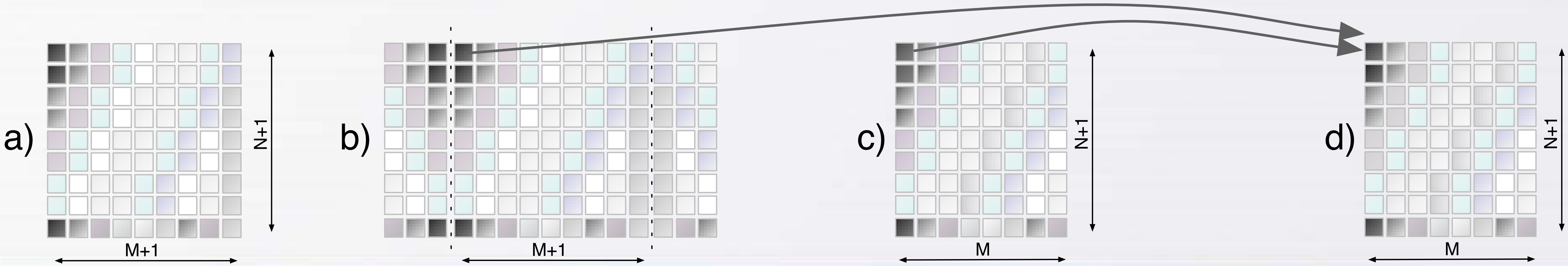
- **DIRECT Mode**
  - in addition to forward, backward and bi-directional prediction
  - predict bi-directional motion vector pairs (MV_B, MV_P) from
    - motion between backward reference I/P-frame and future reference P-frame
    - co-located macroblocks in the reference frames
    - temporal distance between the reference frames
    - implicitly assume 8x8 blocks using the default accuracy of 1/4 or 1/4 pel
  - optionally MV prediction error as difference motion vector (MV_D) transmitted



MPEG-4 quarter-pel MC

- new feature to MPEG-4 is 1/4 pixel motion compensation accuracy
- 8-tap filter for approximating half pel positions
- high accuracy operations limited to 8-tap filter stage only, downscaling/clipping directly after half pel approximation
- quarter pel positions calculated by averaging half pel positions and nearest raster pixel
- mirrored block extension limits picture data bandwidth to the same range as in conventional interpolation methods

block from reference frame (M+1) x (N+1) pixels → mirroring of three border pixels → half pel filtering using 8-tap -8 24 -48 160 160 -48 24 -8 → quarter pel resolution by averaging half pel positions and adjacent raster pixels

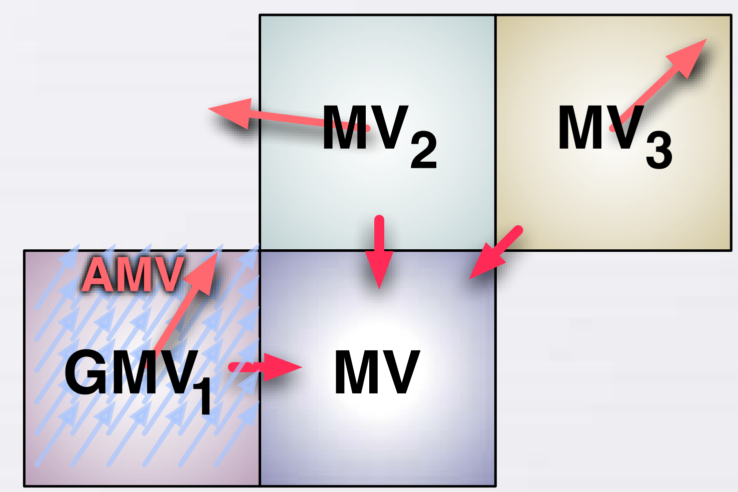


MPEG-4 GMC

- **Global Motion Compensation**
  - allowed in S-frames (sprite frames) only, no bidirectional prediction
  - special case of sprites (static, one frame life time)
  - 1 extra bit per macroblock to signal global/local motion model
  - global model
    - either 2, 4 or 6 parameters (translation,scaling,affine)
    - mapping on 16x16 macroblocks, per-macroblock global/local decision
    - warping parameters transmitted with points of pre-defined reference location
    - bilinear interpolation, accuracy up to 1/64
  - residual coding identical to other motion compensated macroblocks

2 warp points 4 warp points

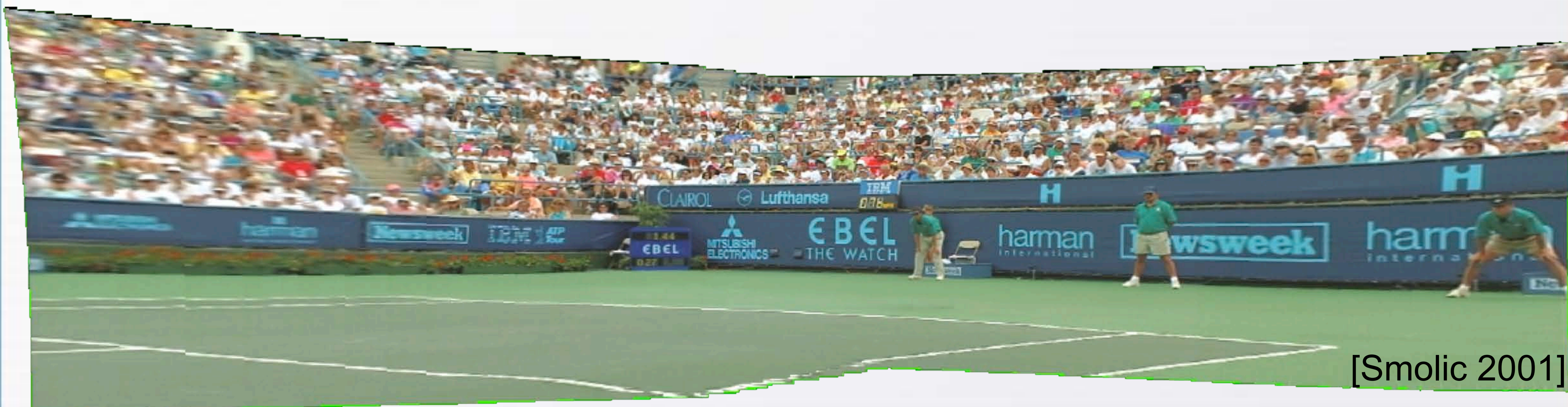
1 warp point 3 warp points (regular sprites only, not GMC)



average MV calculated for local MC

MPEG-4 Sprites

- **Sprites**
  - non-rectangular objects
  - intended for panoramic background images, where foreground motion is added as individual video objects
  - for displayed scenes, a parameter set is transmitted for the visible part
  - incremental sprite update possible
  - up to 8 warping parameters (perspective model) applied



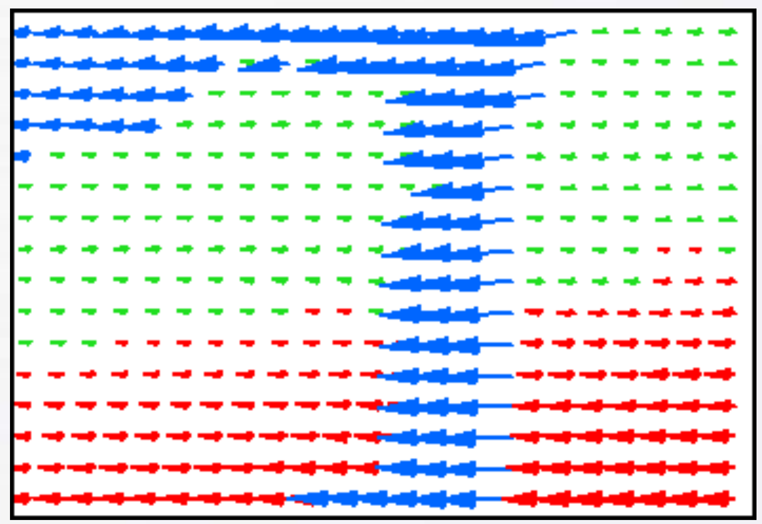
[Smolic 2001]

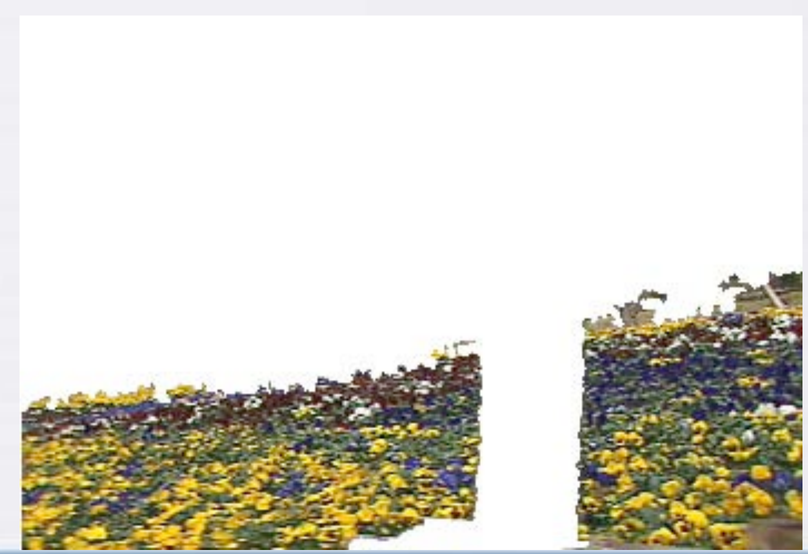
MPEG-4 segmentation example

- **motion based video segmentation**
  - affine parametric motion fields for object detection and segmentation
  - intended for 3D wavelet coding [Schwarz 2000]
  - useful for MPEG-4 shape coding as well

original image

motion field





References

- [Schwarz 2000] H. Schwarz: Untersuchungen zur objektbasierten Videocodierung mit einer 3D-Wavelet-Transformation, Dissertation, Uni. Rostock, 2000, ISBN 3-89820-173-2
- [Smolic 2001] A. Smolic: Globale Bewegungsbeschreibung und Video Mosaiking unter Verwendung parametrischer 2D-Modelle, Schätzverfahren und Anwendungen, Dissertation, RWTH Aachen, 2001
- [Richter 2000] H. Richter: Entwicklung echtzeitfähiger Verfahren zur globalen Bewegungsschätzung für MPEG-4, Diplomarbeit, Uni. Rostock, 2000
- [Schäfer 2005] R. Schäfer, H. Schwarz, D. Marpe, T. Schierl, T. Wiegand: MCTF and Scalability Extension of H.264/AVC and its Application to Video Transmission, Storage, and Surveillance, VCIP 2005 (http://ip.hhi.de/imagecom_G1/assets/pdfs/VCIP05_SVC.pdf)

References

- [ISO00b] ISO/IEC MPEG and ITU-T VCEG. Information technology - Generic coding of moving pictures and associated audio information - part2: Video. ISO/IEC 1318-2:2000 | ITU-T Rec. H.262), December 2000
- [MPFL97] J. L. Mitchell, W. B. Pennebaker, C. E. Fogg, D. J. LeGall. MPEG Video Compression Standard. Chapman and Hall, New York, USA, 1997
- [H263] International Telecommunication Union. Video coding for low bit rate communication - ITU-T Recommendation H.263, March 1996, http://www.itu.int/rec/T-REC-H.263/e

References

- [Schwarz 2000] H. Schwarz: Untersuchungen zur objektbasierten Videocodierung mit einer 3D-Wavelet-Transformation, Dissertation, Uni. Rostock, 2000, ISBN 3-89820-173-2
- [Smolic 2001] A. Smolic: Globale Bewegungsbeschreibung und Video Mosaiking unter Verwendung parametrischer 2D-Modelle, Schätzverfahren und Anwendungen, Dissertation, RWTH Aachen, 2001
- [Richter 2000] H. Richter: Entwicklung echtzeitfähiger Verfahren zur globalen Bewegungsschätzung für MPEG-4, Diplomarbeit, Uni. Rostock, 2000
- [Schäfer 2005] R. Schäfer, H. Schwarz, D. Marpe, T. Schierl, T. Wiegand: MCTF and Scalability Extension of H.264/AVC and its Application to Video Transmission, Storage, and Surveillance, VCIP 2005 (http://ip.hhi.de/imagecom_G1/assets/pdfs/VCIP05_SVC.pdf)

References

- [ISO00b] ISO/IEC MPEG and ITU-T VCEG. Information technology - Generic coding of moving pictures and associated audio information - part2: Video. ISO/IEC 1318-2:2000 | ITU-T Rec. H.262), December 2000
- [MPFL97] J. L. Mitchell, W. B. Pennebaker, C. E. Fogg, D. J. LeGall. MPEG Video Compression Standard. Chapman and Hall, New York, USA, 1997