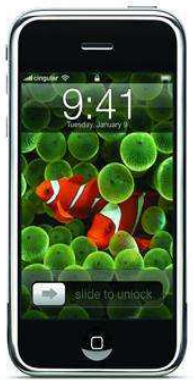
Video compression
MPEG-4 AVC/H.264

Frederic Dufaux
LTCI - UMR 5141 - CNRS
TELECOM ParisTech
frederic.dufaux@telecom-paristech.fr

Montpellier, April 30, 2013

Context and Background

Trends



- **More, more, more**
  - Higher spatial and temporal resolutions, higher pixel depths, higher number of views, more colors

Trends

- 300 million photos uploaded per day

- 60 hours of video uploaded every minute,
- more than 3 billion views per day,
- over 3 billion hours of video watched each month

Video Coding Standards: MPEG & H.26x

International Organization for Standardization

International Telecommunication Union

ISO/IEC JTC1/SC29 WG11 MPEG MOVING PICTURE EXPERTS GROUP

- MPEG-1(1992)
  - video on CD-ROM (1.5Mbit/s)
- MPEG-2 (1994)
  - Digital TV (5-10Mbit/s)
- MPEG-4 (1998-99)
  - Multimedia applications (10kbit/s-10Mbit/s)
- MPEG-4 AVC (2003)
  - Advanced video coding

- H.261 (1988)
  - Video-conferencing over ISDN (p*64 Kb/s)
- H.262 (1994)
  - Digital TV (5-10Mbit/s)
- H.263 (1996)
  - Video-telephony over POTS and Internet (8-64 Kb/s)
- H.264 (2003)
  - Advanced video coding

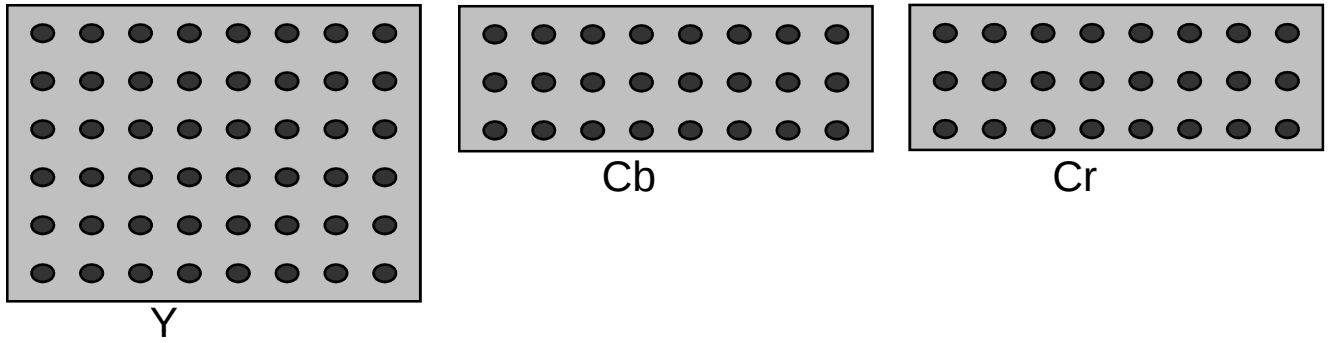
Under development: HEVC

Video Coding Standards: MPEG & H.26x

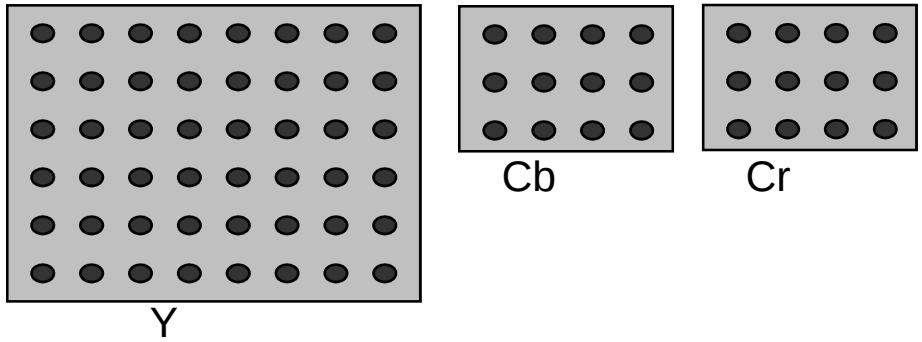
- **Compression efficiency**
  - Typically 50% gain every 5 years
  - Adding more efficient coding tools / modes
  - Functionalities such as scalability, error resilience, interactivity, low complexity, random access, …

Video formats

- ITU-R 601 (CCIR 601)
  - EU: 704 x 576, 25 f/s, int.; USA/Japan: 704 x 480, 30 f/s, int.
  - 4:2:2

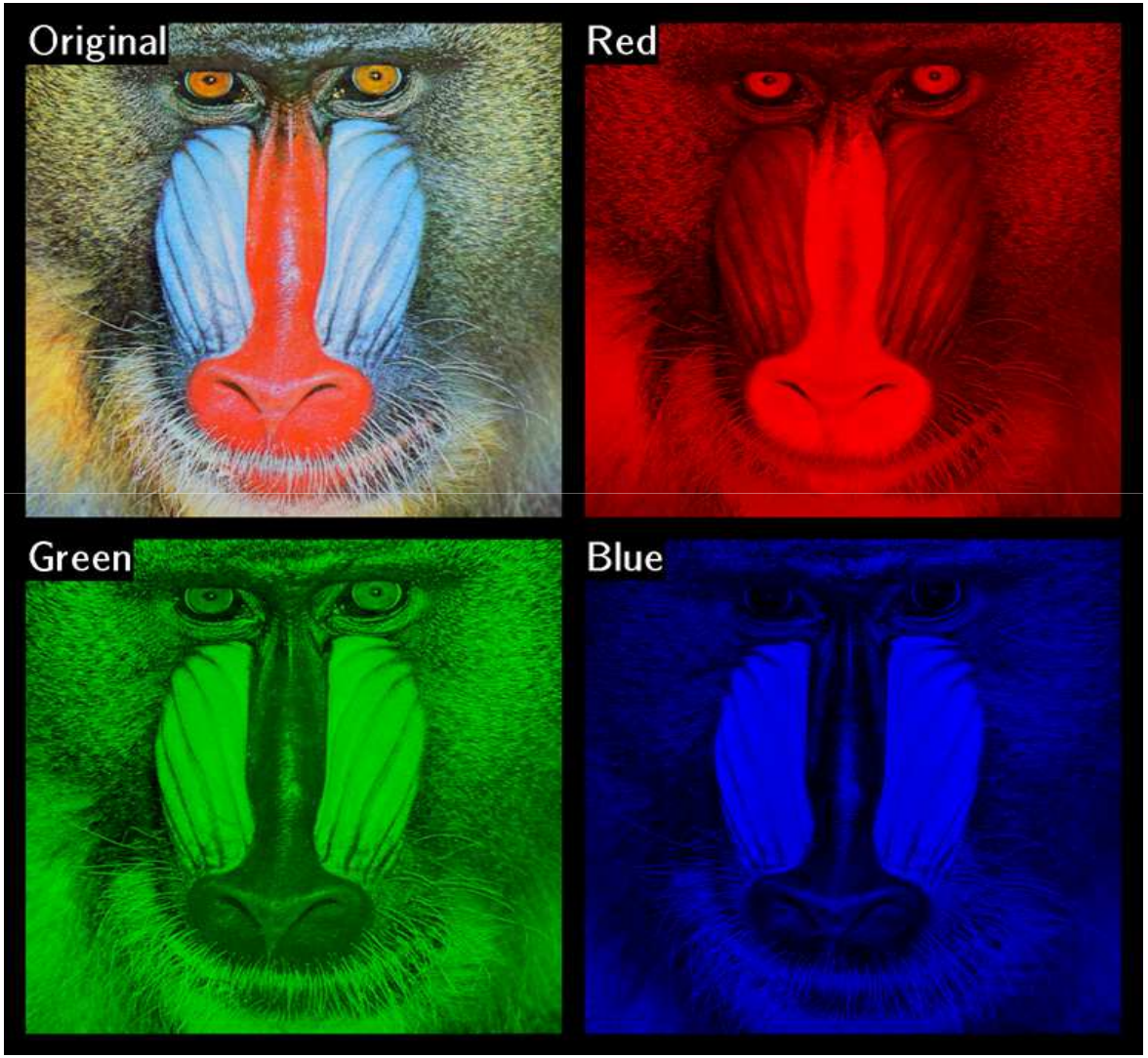


  - 4:2:0

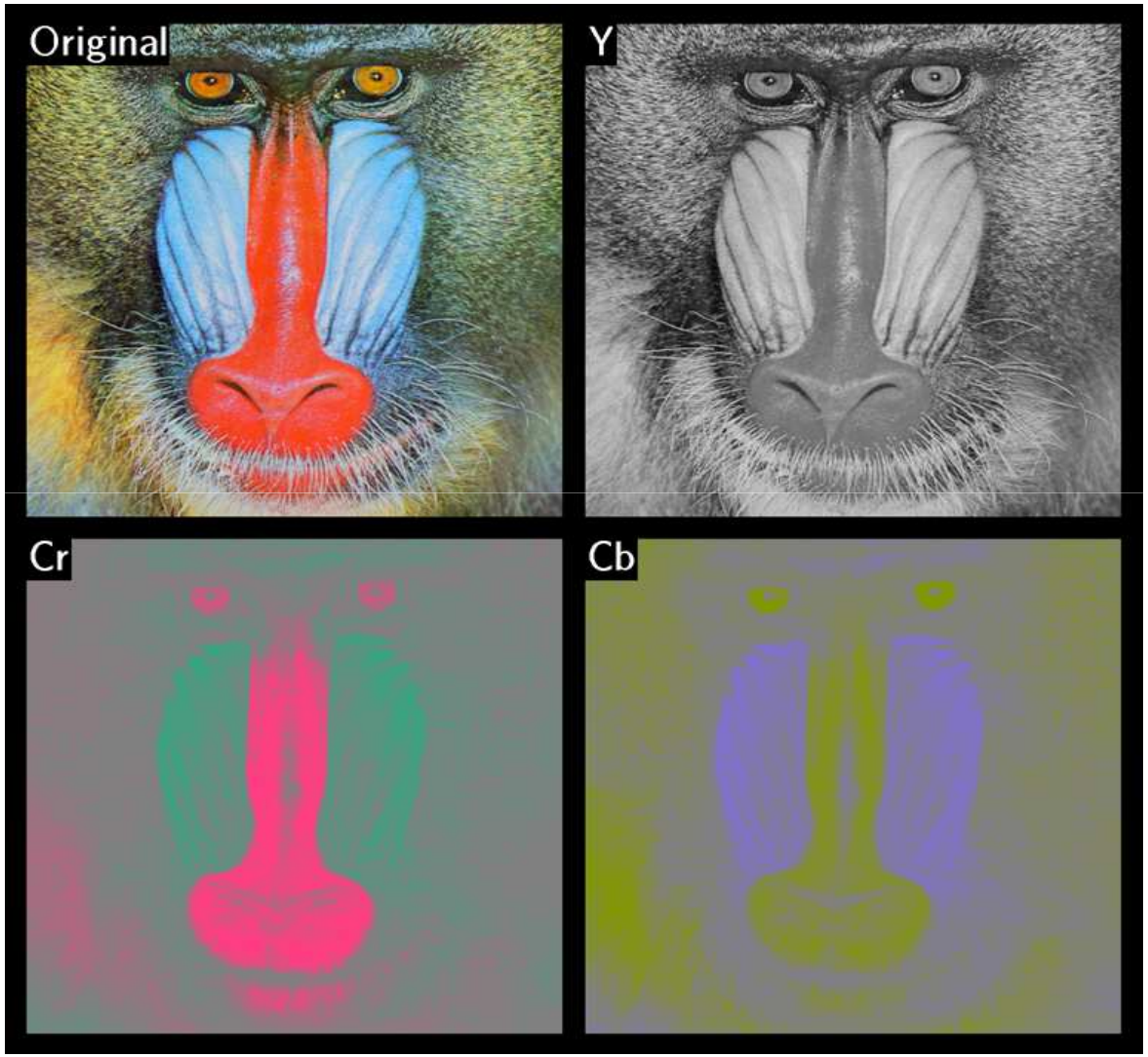


- CIF: 352 x 288, 30 f/s, YCbCr 4:2:0, prog.
- QCIF: 176 x 144, 30 f/s, YCbCr 4:2:0, prog.

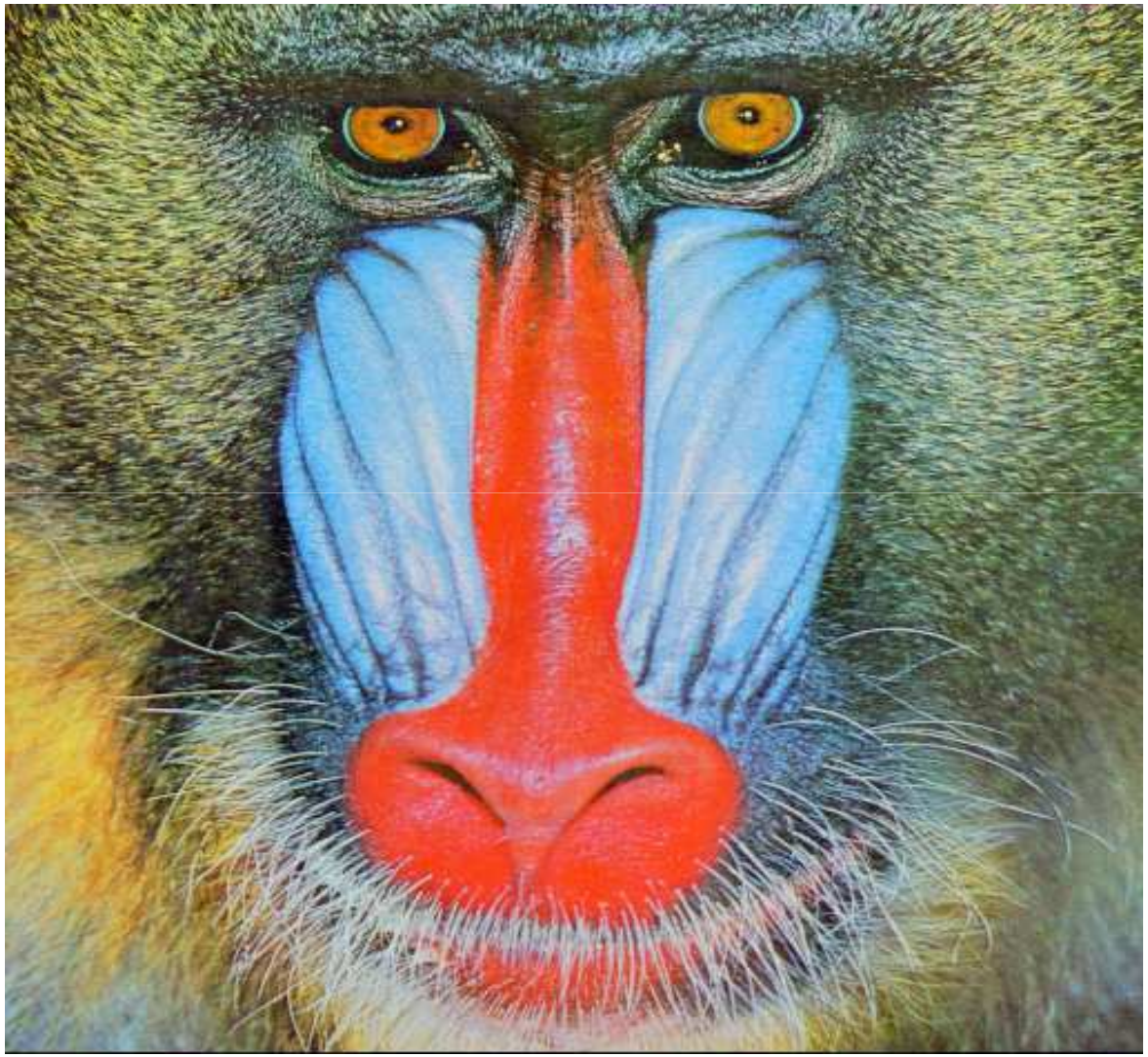
Video formats



Video formats



Video formats

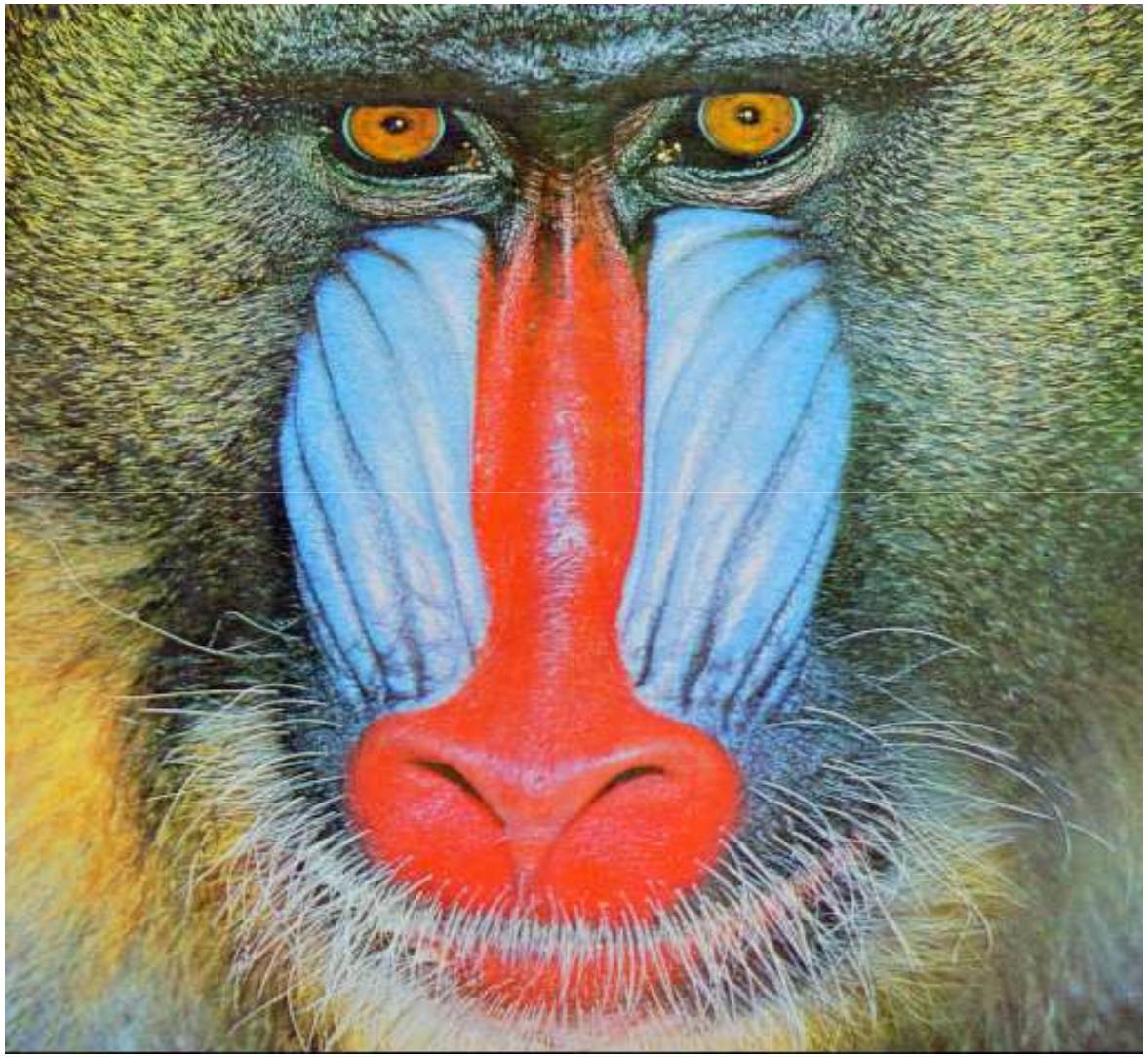


Video formats

Video formats

Video formats

Video formats



Video formats

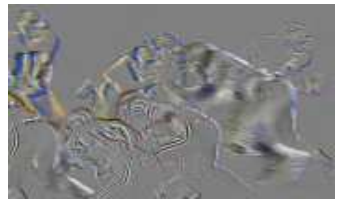
Video formats

Video formats

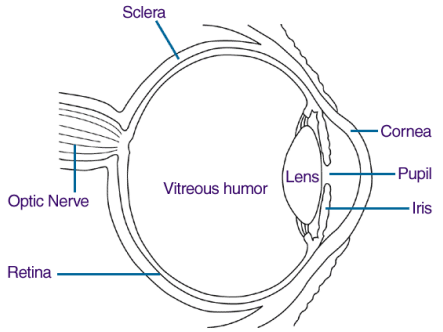
Fundamental concepts in video compression

Compression: general concepts

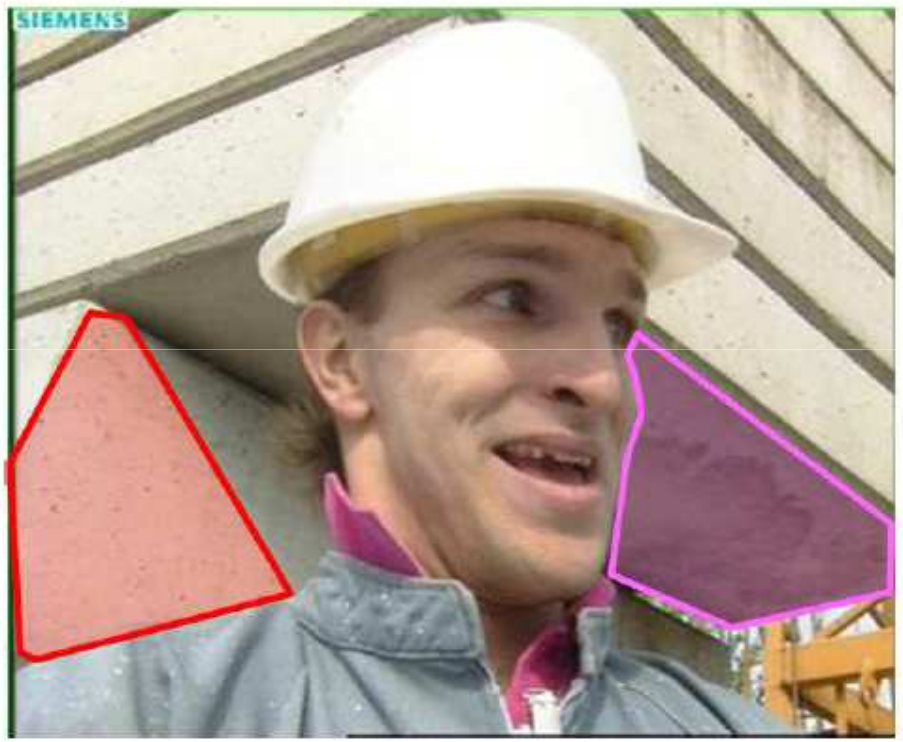
- Exploit the correlation in the data
  - Reduce redundancies
  - Lossless



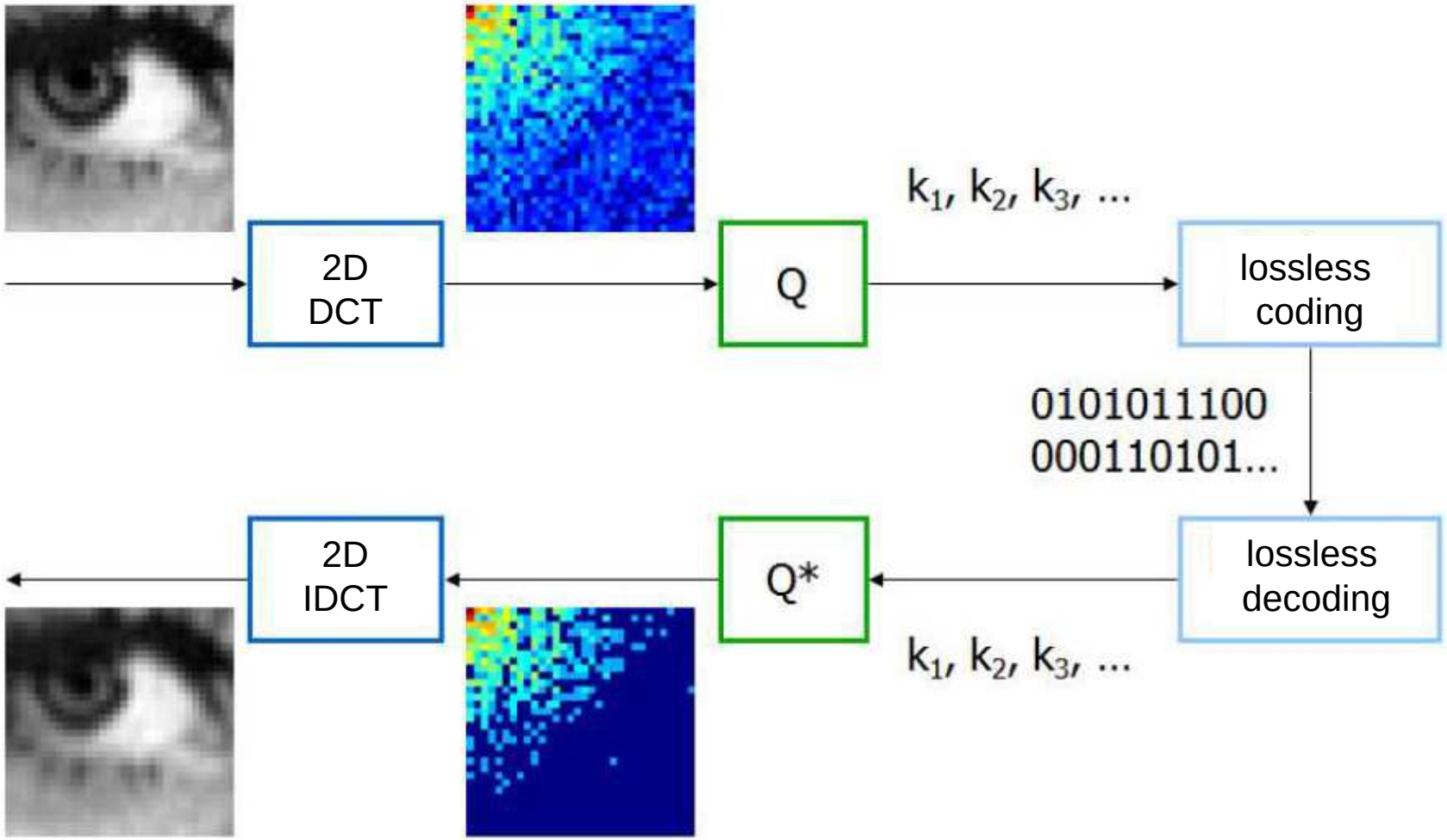
- Exploit the human visual system
  - Remove imperceptible data
  - Introduce (hardly) noticeable distortions



Spatial redundancies



Spatial transform coding

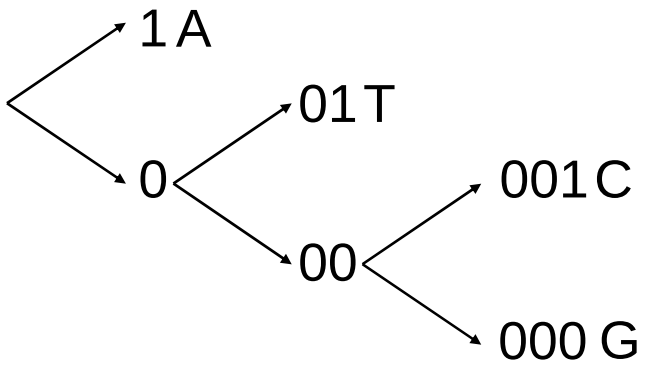


Entropy coding

- Modeling
  - Estimate probabilities of the symbols
- Coding
  - Produce a bit sequence based on these probabilities
  - Most common symbols use the shortest codes

- Huffman coding
- Arithmetic coding

$P_A=0.5$, $P_T=0.25$, $P_C=0.125$, $P_G=0.125$



Temporal redundancies

Temporal predictive coding

- **Prediction by frame difference:** $\hat{f}_k = \tilde{f}_{k-1}$

$$e_k = f_k - \hat{f}_k$$

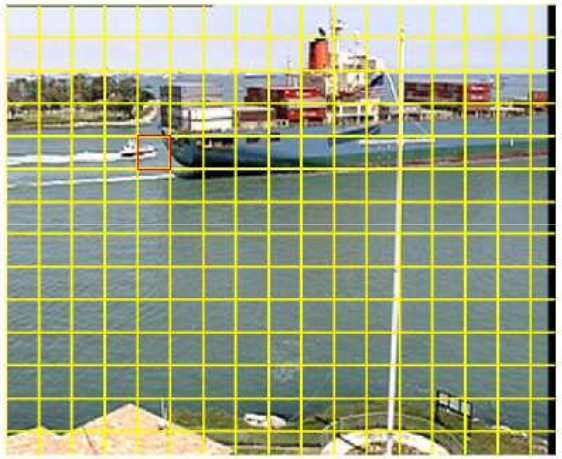
current frame

prediction error

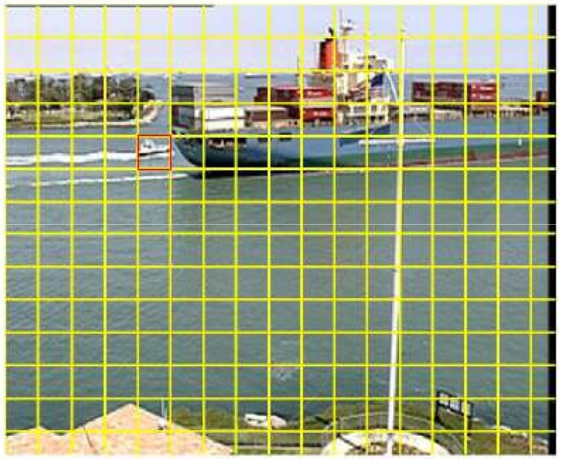
previous frame

Motion estimation and compensation

For every block in a current frame, the most similar block in a reference frame is searched for.



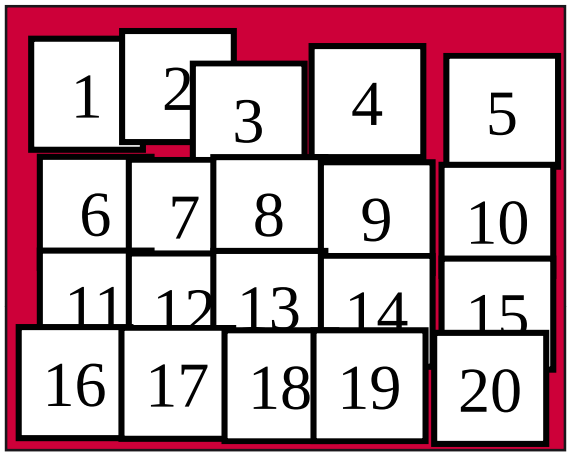
f(x,y,t -Δt)



f(x,y,t)

- Assumptions
  - The motion of pixels within the same block is uniform.

Backward motion compensation

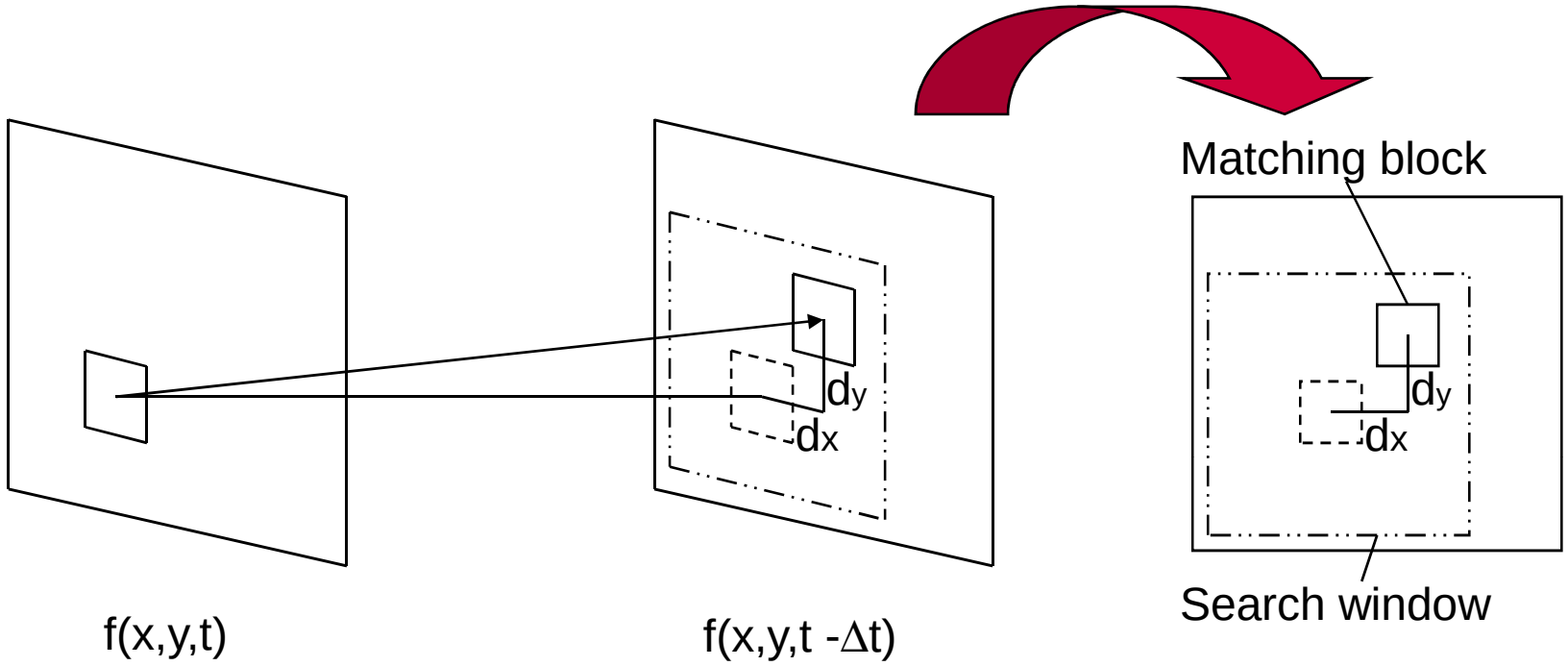


Previous frame

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20

Current predicted frame

Block matching motion estimation



- Parameters
  - Search strategy: number of candidate blocks, search order, maximum displacement
  - Matching function
  - Block size

Block matching motion estimation

- **Similarity measure**

$$\min \sum_{B} \left\| f(x,y,t) - f(x-d_x, y-d_y, t-\Delta t) \right\|$$

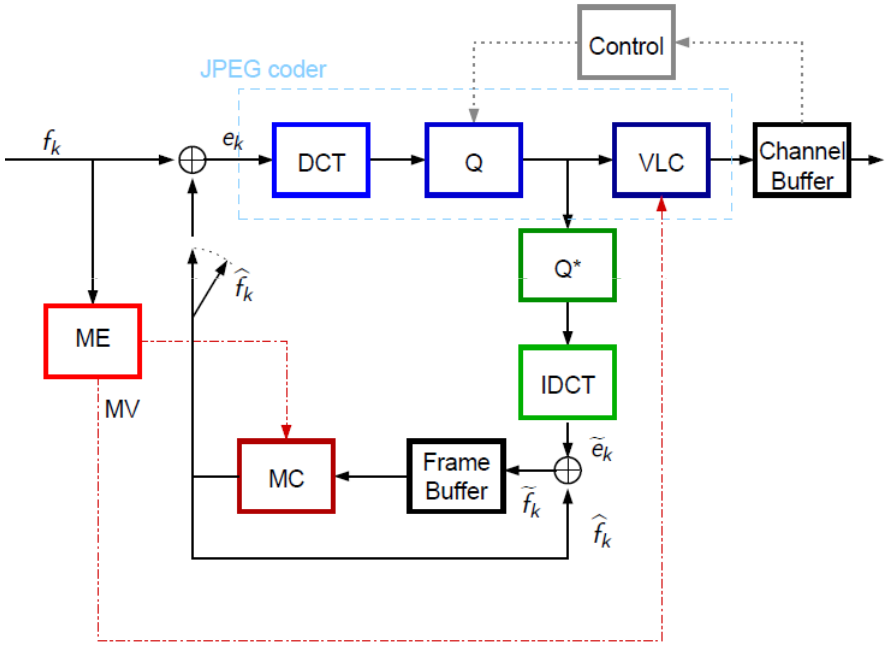
- **Search methods**
  - full-search
  - logarithmic
  - N-step
  - conjugate
  - pyramidal
  - …

Block matching motion estimation

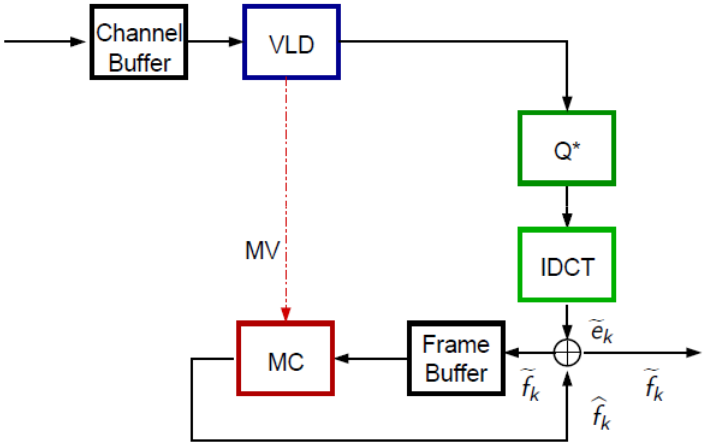
- **Advantages**
  - Explicit minimization of prediction errors
  - Regular structure (specially in full search): suitable for VLSI implementation
  - Low overhead (one motion vector per block)
  - Robustness to noise
- **Drawbacks**
  - Complex
  - High prediction errors in the border regions of moving objects
  - Blocking artifact

Hybrid video coding - Encoder

input frame
split in 16x16 MB
and 8x8 Block



Hybrid video coding - Decoder

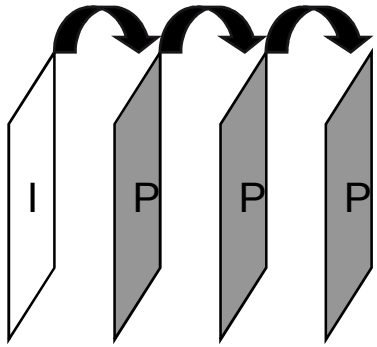


Hybrid video coding

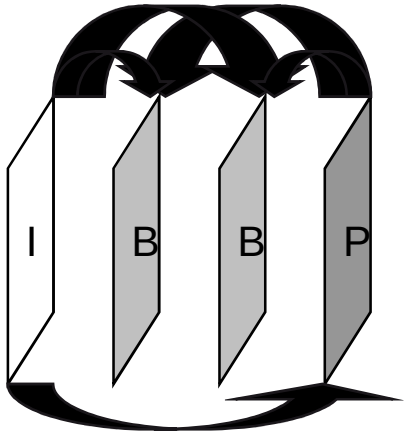
- Hybrid motion compensated DCT
  - 8x8 block DCT
  - 16x16 Macroblock motion compensation with ½ pixel accuracy
  - Scalar quantization
  - Zig-zag scan
  - Variable-length coding and run-length coding

I-/P-/B-frames

- Intra-frames (I)
  - Independently coded (JPEG-like)
  - Low compression, low complexity
  - Random access
  - Error robustness



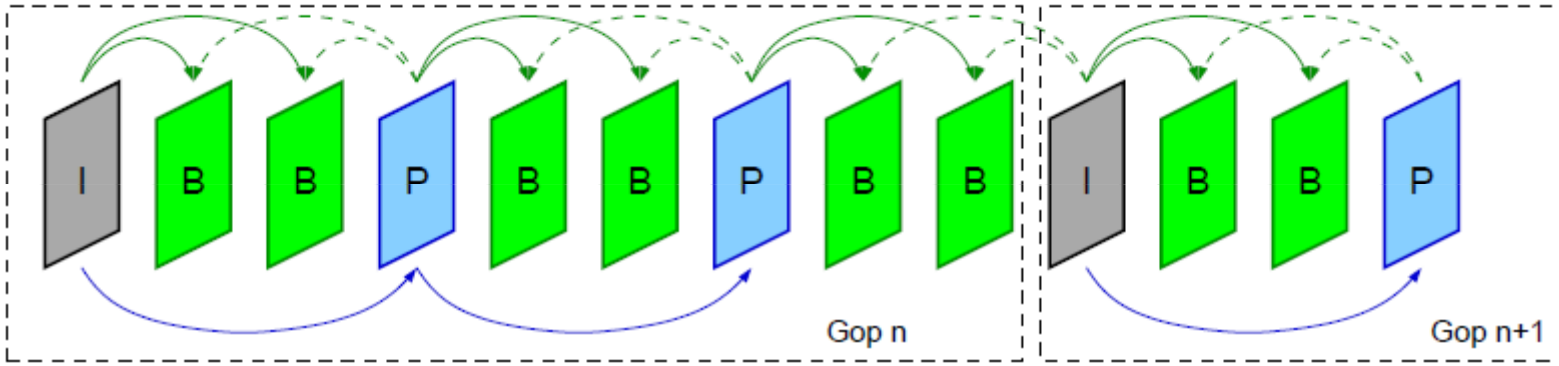
- Predicted-frames (P)
  - Predicted from previous I/P frame
  - High compression, high complexity



- Bidirectional-frames (B)
  - Predicted from previous and next I/P frames
  - Coding delay

Group of Pictures (GOP)

- Start with an I-frames

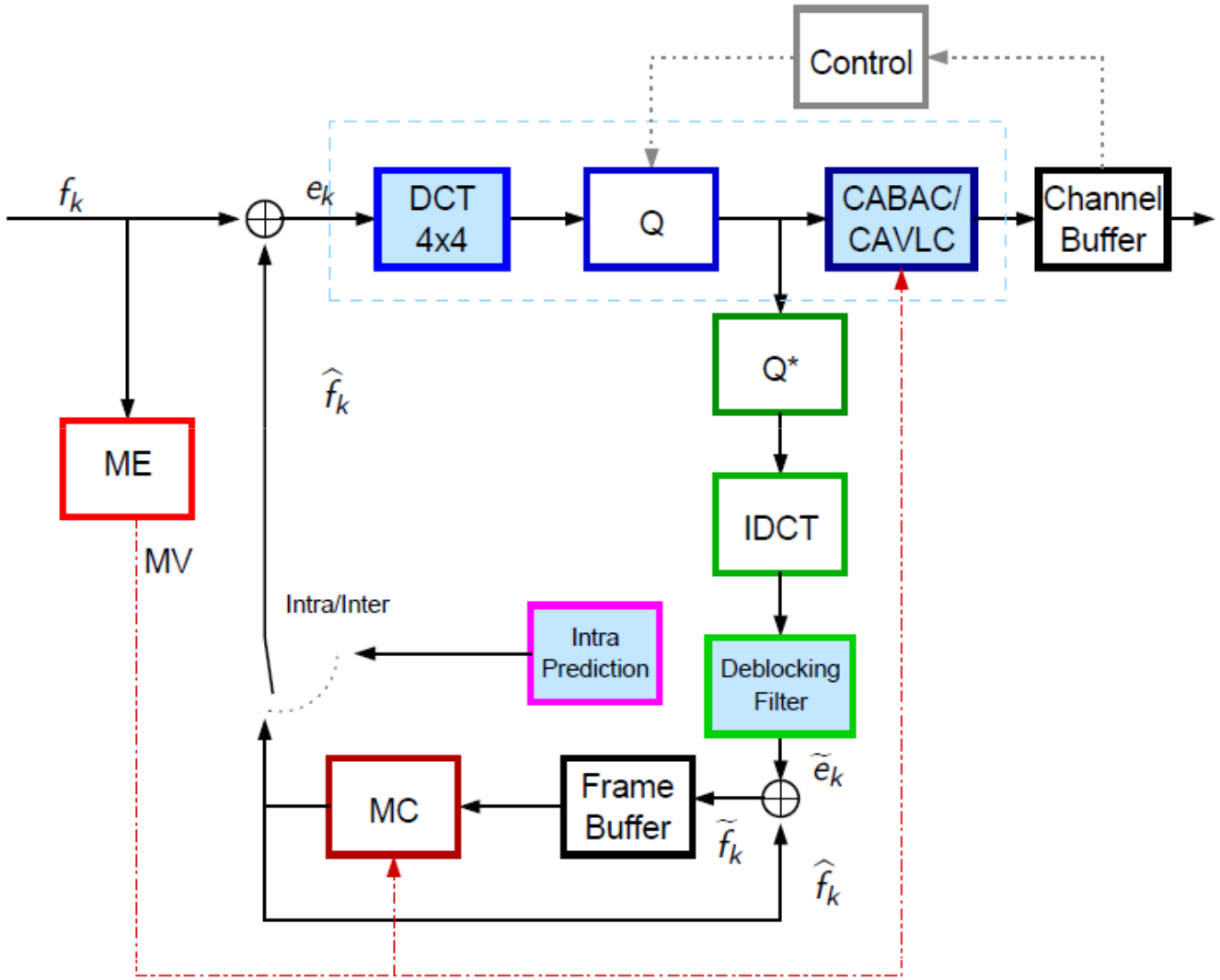


MPEG-4 AVC / H.264

H.264/AVC

- **Advanced Video Coding**
  - State-of-the-art video coding performance
  - Also known as MPEG-4 Part 10, AVC, H.264
- **International standard since 2003**
- **Better compression efficiency when compared to earlier standards (H.263, MPEG-2, MPEG-4)**
  - Mainly due to a better prediction
- **But more complex**
  - Remains reasonable especially as decoder

Block Diagram: H.264/AVC Encoder



4x4 integer transform

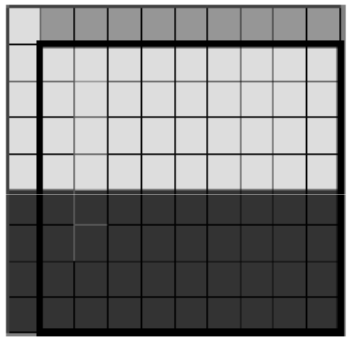
- **4x4 DCT-like transform**
  - Efficient implementation (additions, subtractions, bit shifts)
  - Avoid mismatch between encoder and decoder
  - 2nd level transform on the 16 DC coefficients of a 16x16 MB

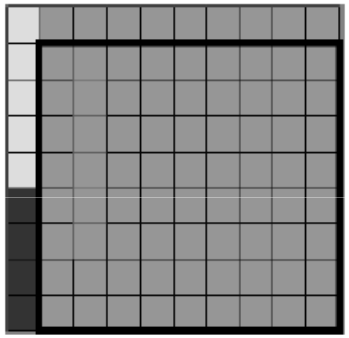
$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}$$

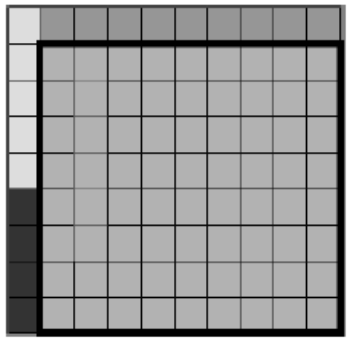
- **Quantization**
  - Non-linear step size control

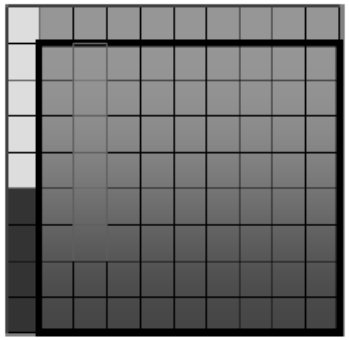
Intra prediction

- **Directional spatial prediction**
  - 16x16 (luma, 4 modes)


horizontal


vertical

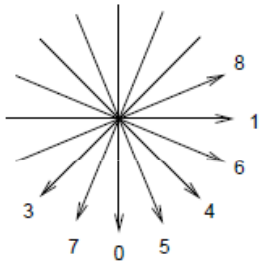

average


planar

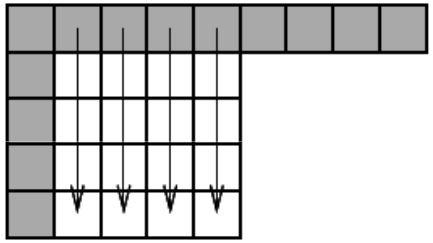
Intra prediction

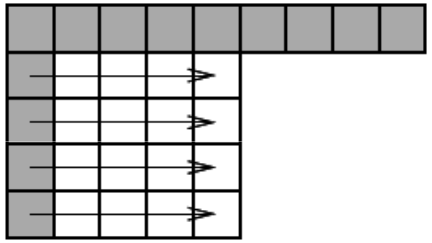
- **Directional spatial prediction**
  - 4x4 (luma, 9 modes)

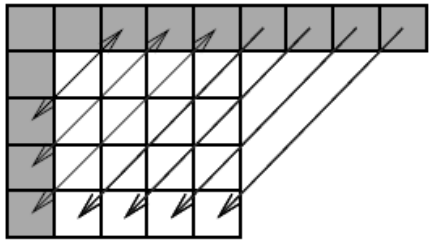
Q	A	B	C	D	E	F	G	H
I	a	b	c	d				
J	e	f	g	h				
K	i	j	k	l				
L	m	n	o	p				
M								
N								
O								
P								

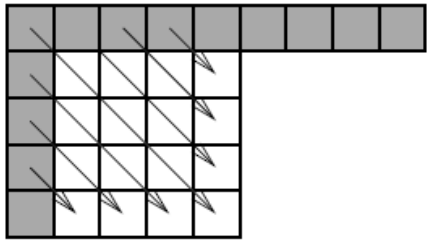


Directions


Mode 0

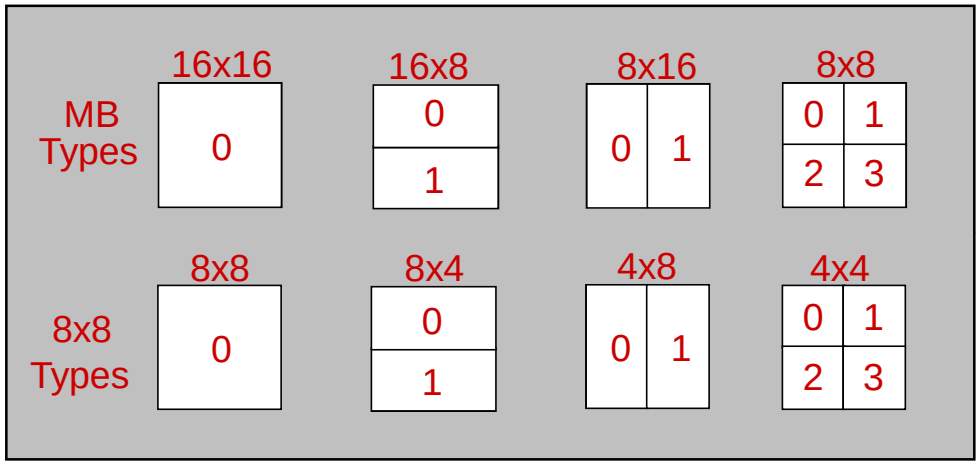

Mode 1


Mode 3


Mode 4

Temporal prediction

- **Flexible block MC**
  - Variable motion block sizes



  - Motion estimation at ¼ pixel accuracy (6-tap filter) (1/8 sample bilinear for chroma)

Temporal prediction and mode selection

- **Lagragian optimization**
  - Mode selection:

$$J_{\text{mode}} = D + \lambda_{\text{mode}} R$$

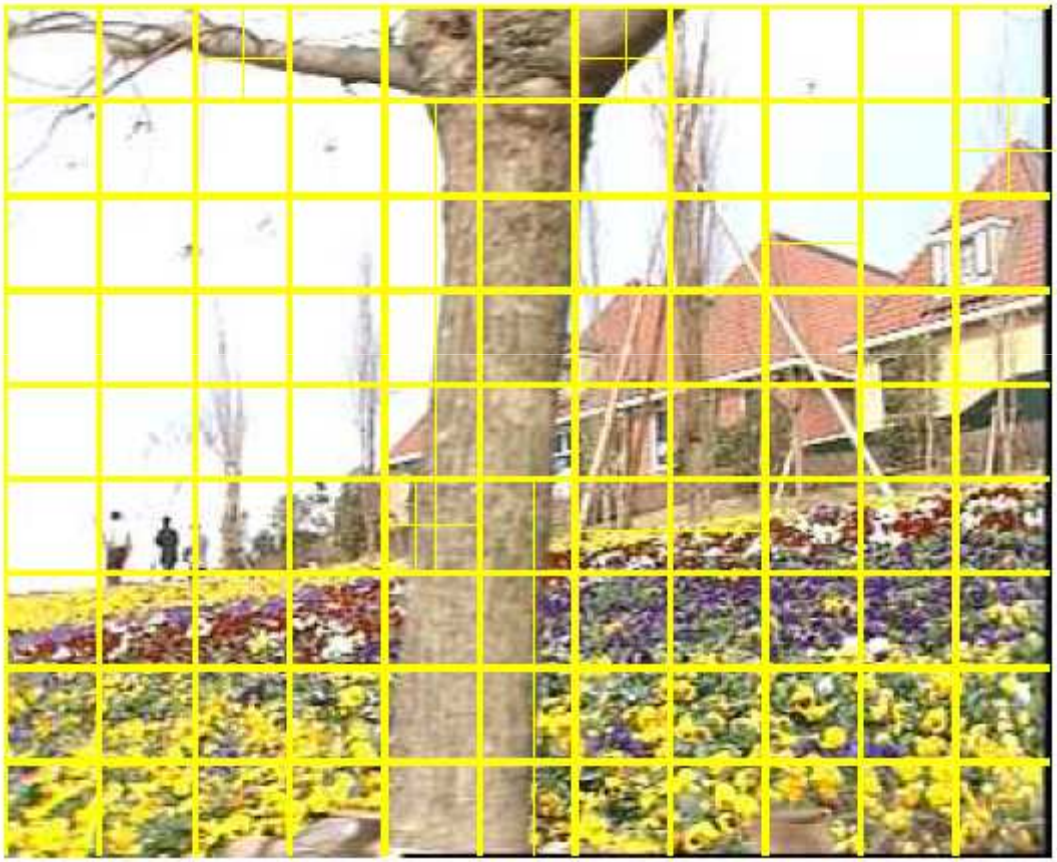
    - D: Bloc distortion (e.g. MSE), include prediction, transform and quantization
    - R: associated bitrate (MV, residual, signaling)
  - Motion estimation:

$$J_{\text{ME}} = D(\mathbf{v}) + \lambda_{\text{ME}} R(\mathbf{v})$$

    - D: e.g. SAD from block matching
    - R: bitrate to encode motion vectors

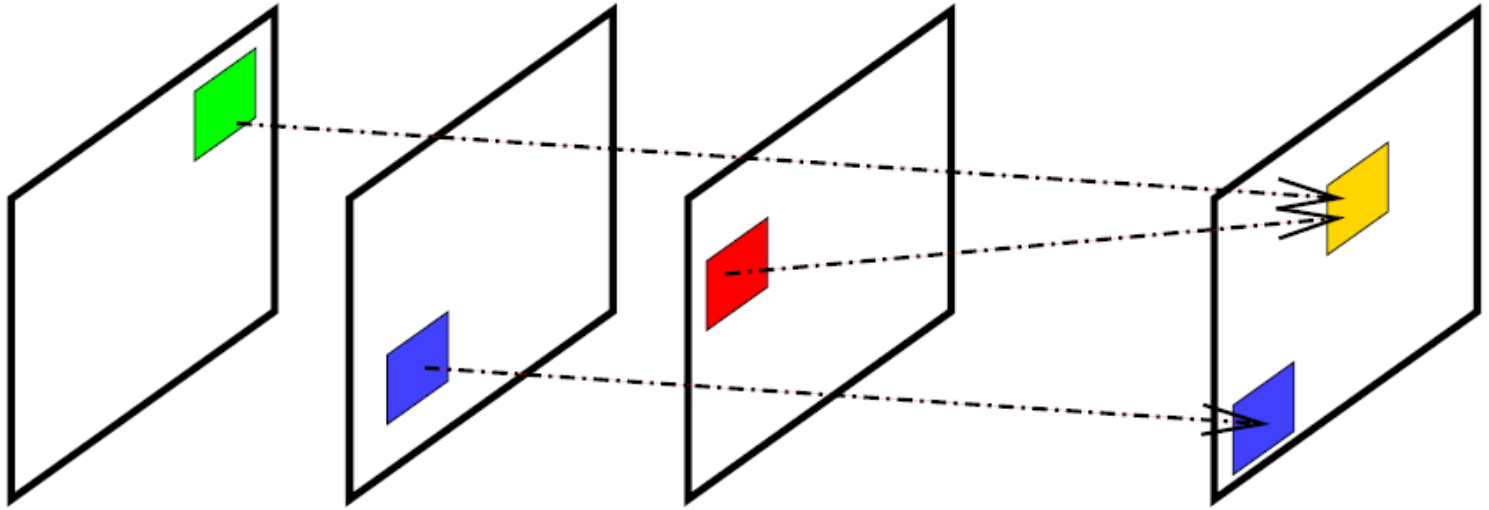
  - (This process is not normative!!)

Temporal prediction



Generalized P- and B- frames

- **Generalized P- and B-frames**
  - Flexible choice of reference frame for motion compensation
  - Multiple reference frames (up to 5)
  - List of reference frames
  - B-frames with arbitrary weights



In-loop Deblocking

- **Post-filtering**
  - Independent of the encoding/decoding
  - Out of scope of the standard
  - Flexible design

- **In-loop deblocking**
  - Filtering is applied in the prediction loop
  - Process is normalized
  - Better performances (subjective and objective)

In-loop Deblocking

Two entropy codes: CAVLC and CABAC

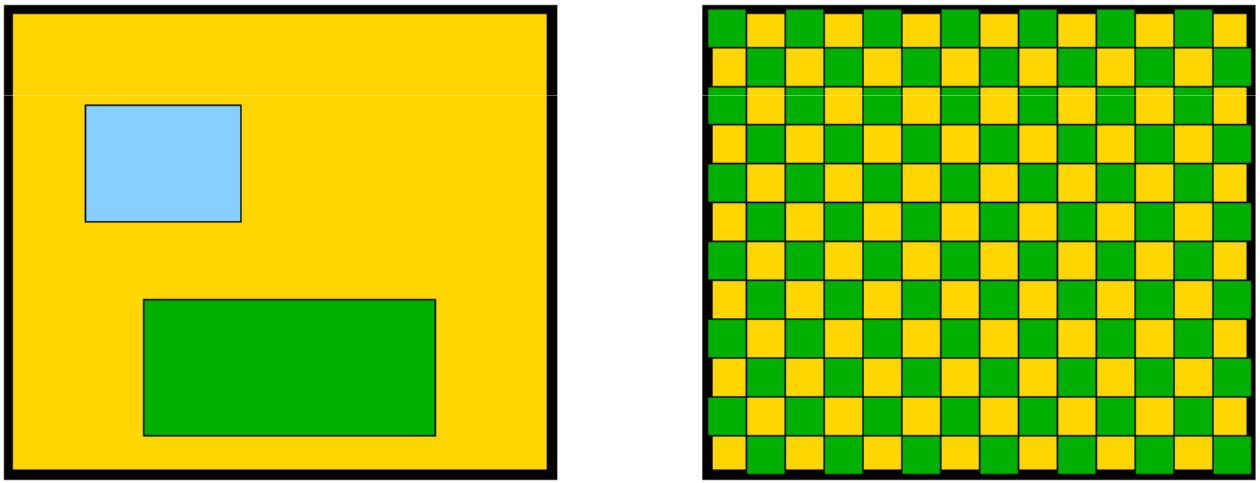
- **Entropy coding**
  - CAVLC: Context-Adaptive Variable Length Coding
  - CABAC: Context-Adaptive Binary Arithmetic Coding
  - Binary arithmetic coding similar to JPEG 2000
  - More complex but 5~10 % better compression efficiency

Weighted Prediction

- **Weighted prediction**
  - Inter-picture prediction with scaling and offset
  - Encoder-specified or timing-derived weights
  - Multiple weights per picture supported
  - Tremendous advantage for fade-in, fade-out, and cross-fades

Flexible Marcoblock Ordering (FMO)

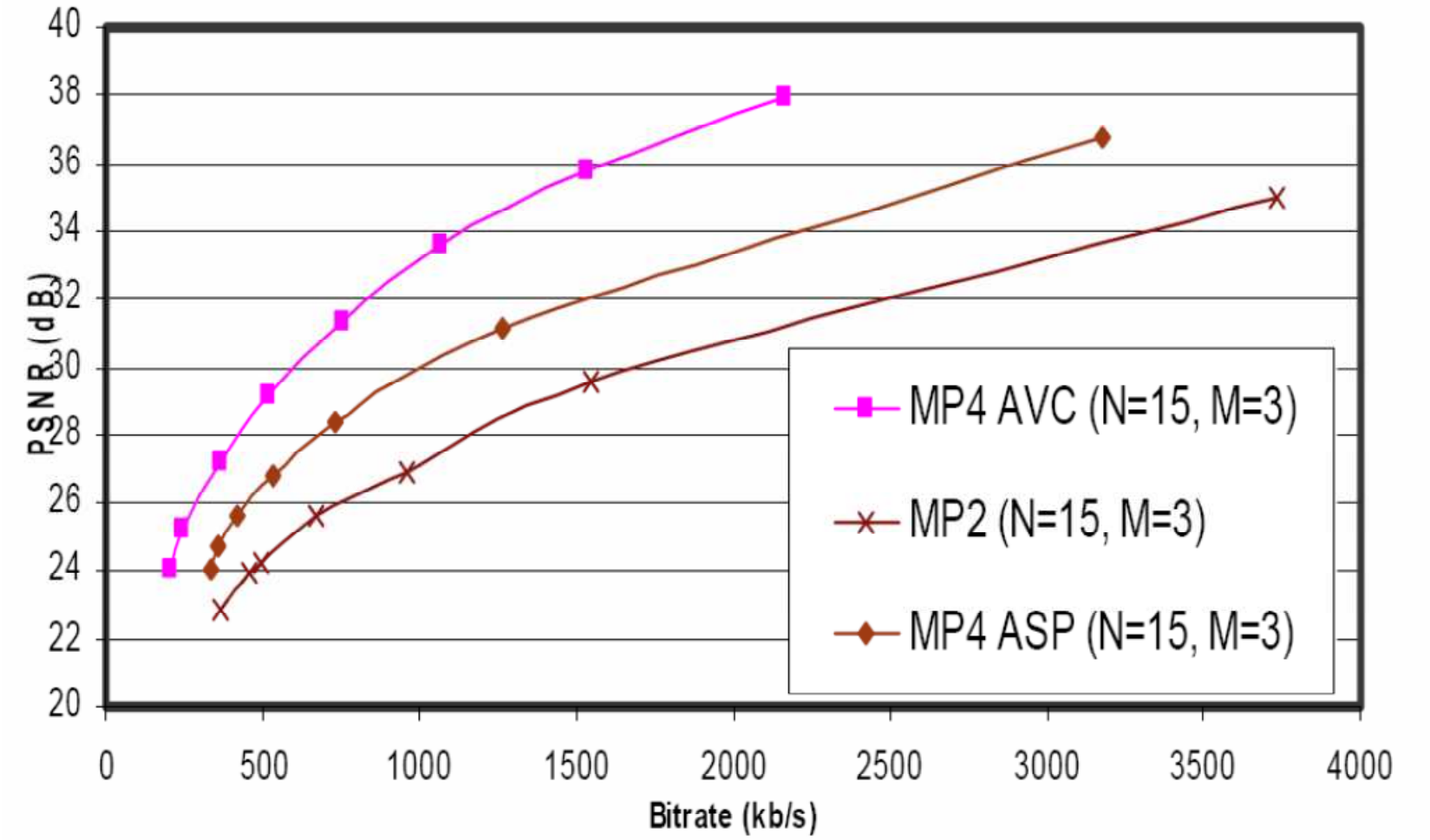
- **FMO**
  - Simple type of Multiple Description Coding
  - Improve error robustness



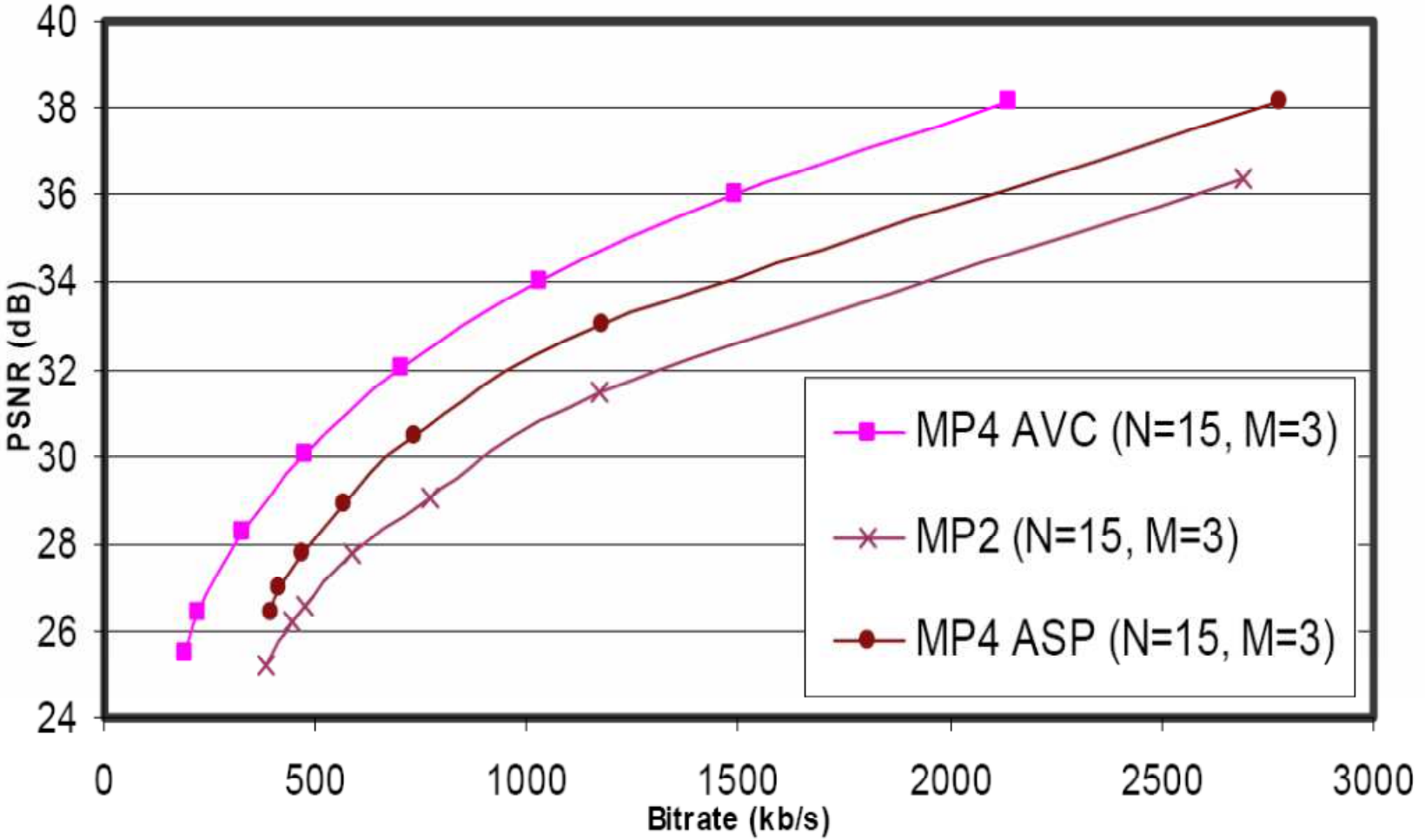
Profiles / levels

- **Profiles**
  - Set of tools targeting specific classes of applications
    - Baseline Profile: low-cost applications, error robustness
    - Main Profile: digital TV broadcast
    - Extended Profile: streaming video
    - High Profile: HDTV, blu-ray
    - High 10: 10 bits per sample
    - High 4:2:2: 4:2:2 chroma sampling, up to 10 bits
    - High 4:4:4: 4:4:4 chroma sampling, up to 14 bits
    - (21 profiles!)
- **Levels**
  - Set of constraints: maximum picture resolution, frame rate, bit rate, etc…

Test: Mobile & Calendar (CIF)

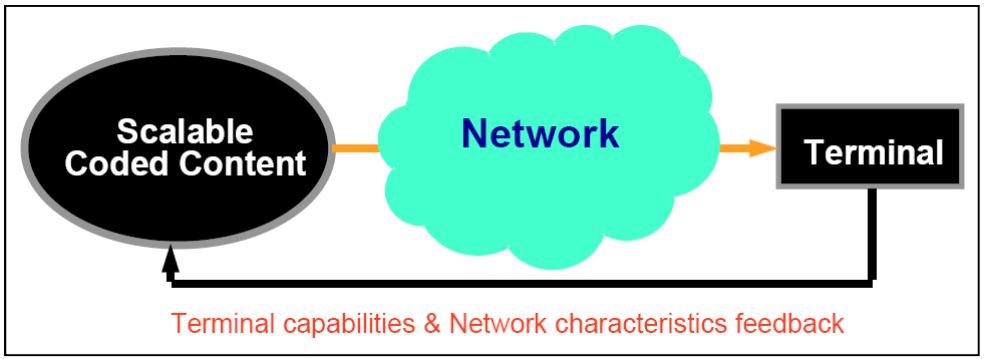


Test: Bus (CIF)


PSNR (dB)
40
38
36
34
32
30
28
26
24
0
500
1000
1500
2000
2500
3000
Bitrate (kb/s)
MP4 AVC (N=15, M=3)
MP2 (N=15, M=3)
MP4 ASP (N=15, M=3)

Scalable Video Coding (SVC)

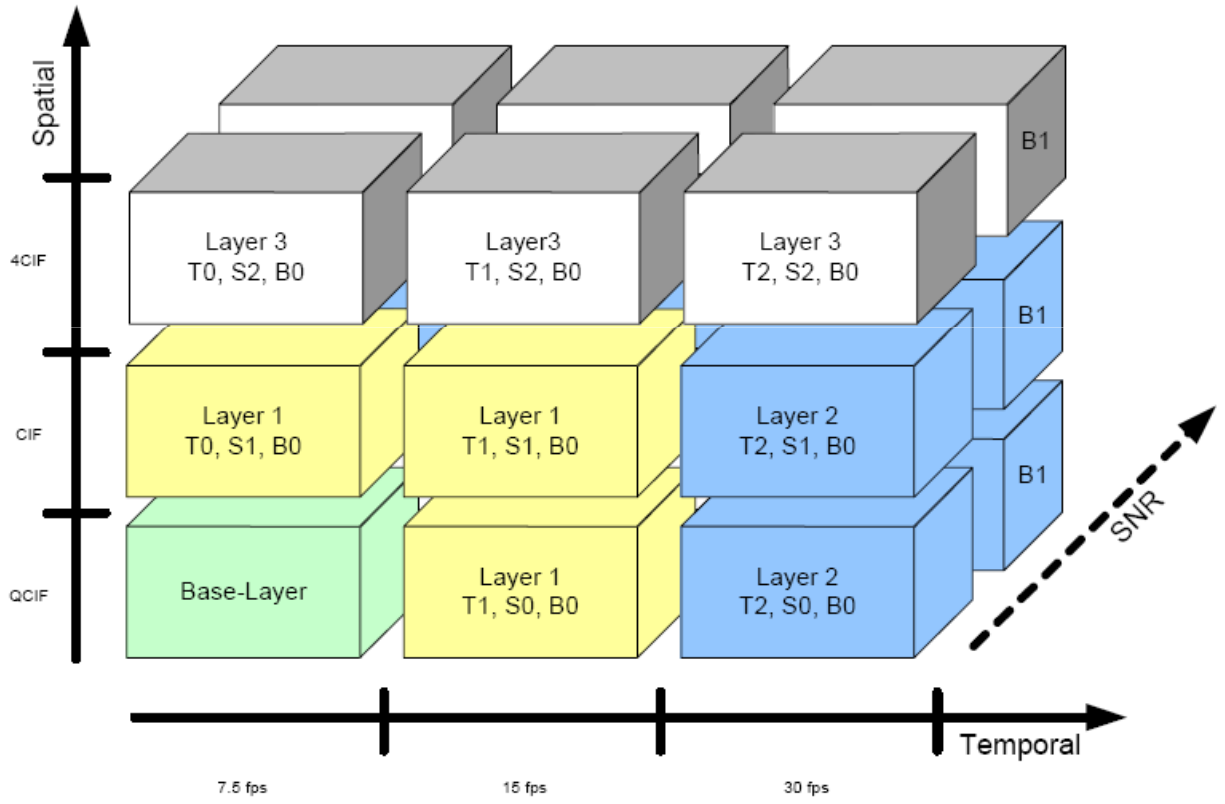
- Extension of AVC
  - Encode once, decode many times
  - Universal Media Access (UMA)



Scalable Video Coding (SVC)

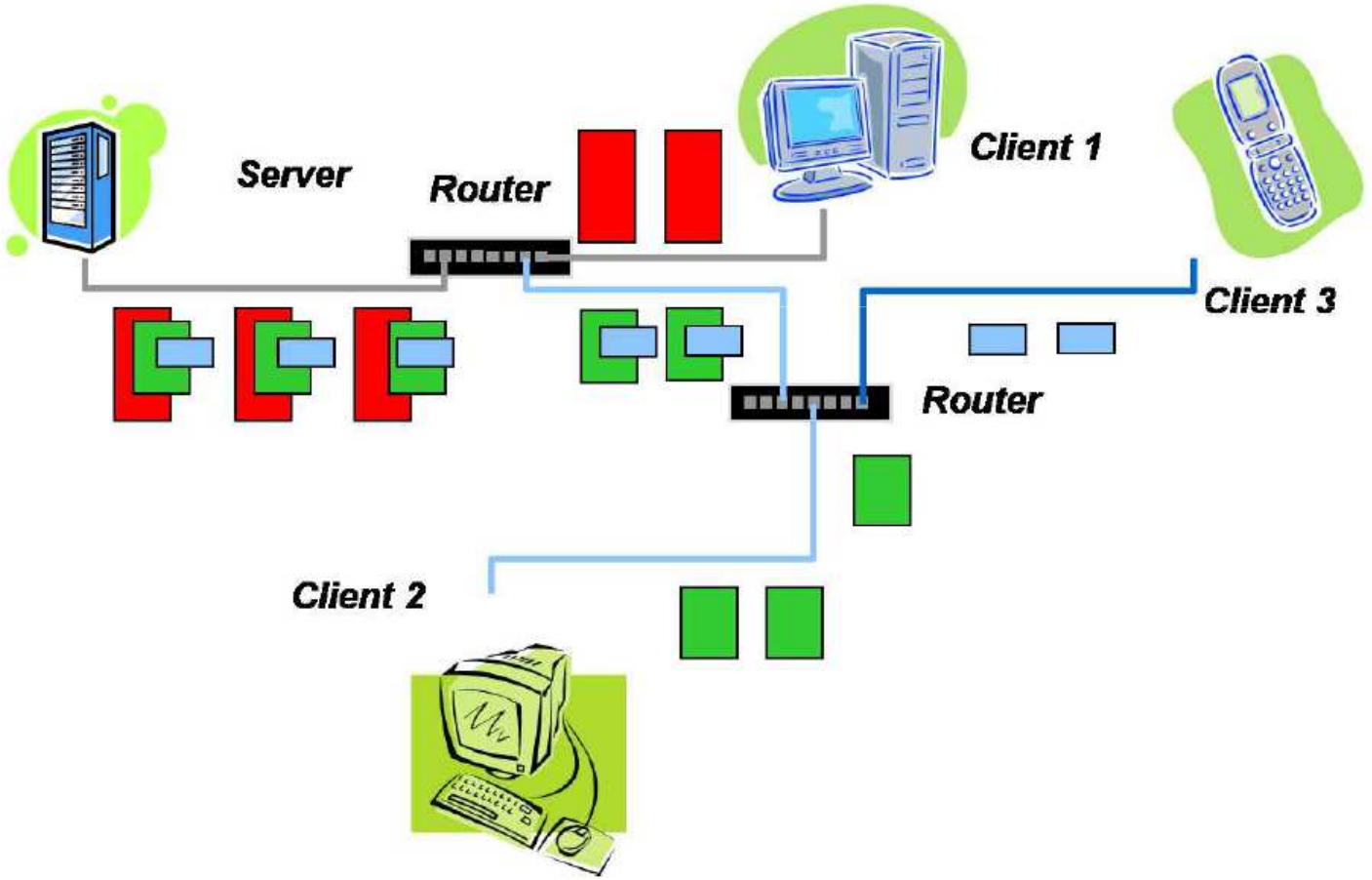
- Layered coding
  - Base layer
    - *H.264/AVC compatible*
    - *can be decoded by all H.264/AVC decoders*
  - Enhancement layers:
    - *Useless without the base layer*
    - *Spatial resolution*
    - *Temporal resolution*
    - *Quality*

Scalable Video Coding (SVC)



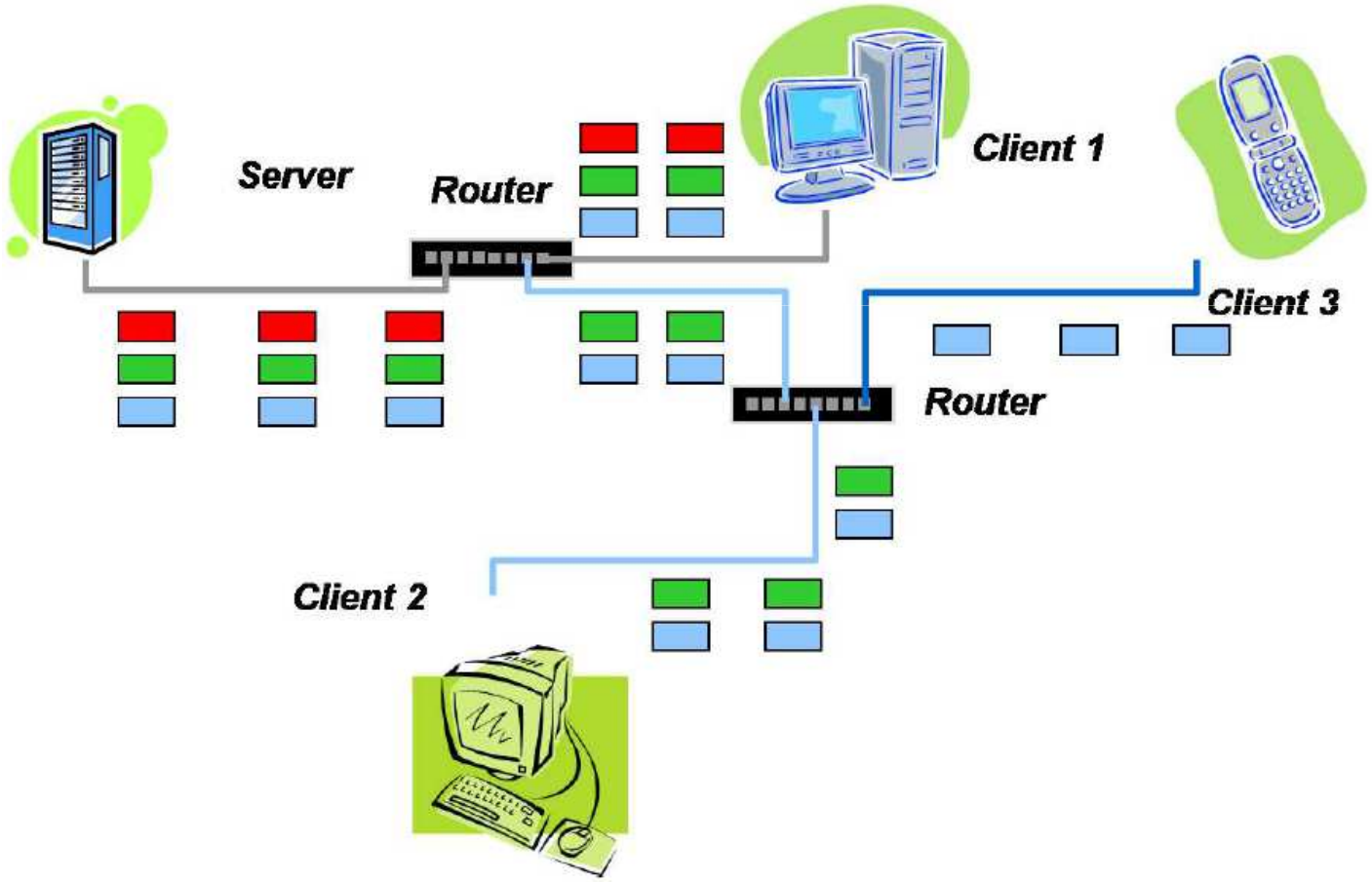
Scalable Video Coding (SVC)

- Video transmission **without** scalability



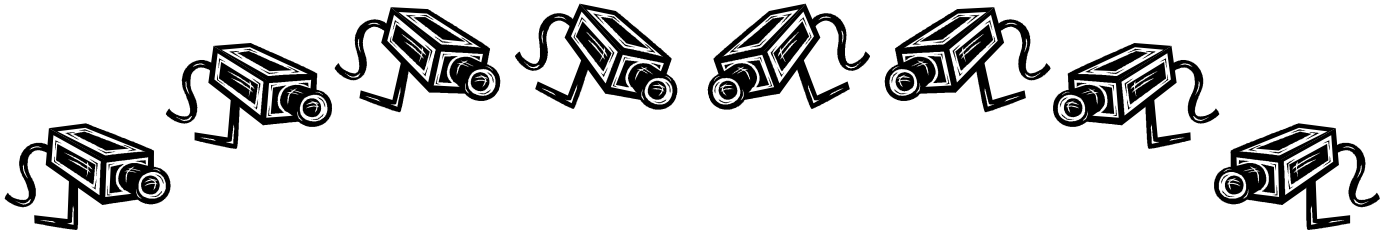
Scalable Video Coding (SVC)

- Video transmission **with** scalability



Multiview Video Coding (MVC)

– Multiple cameras capturing overlapped images from the same scene with different viewing position



- Extension of AVC
  - Predictive coding along time and across views
  - Very complex encoder
  - Cameras have to communicate