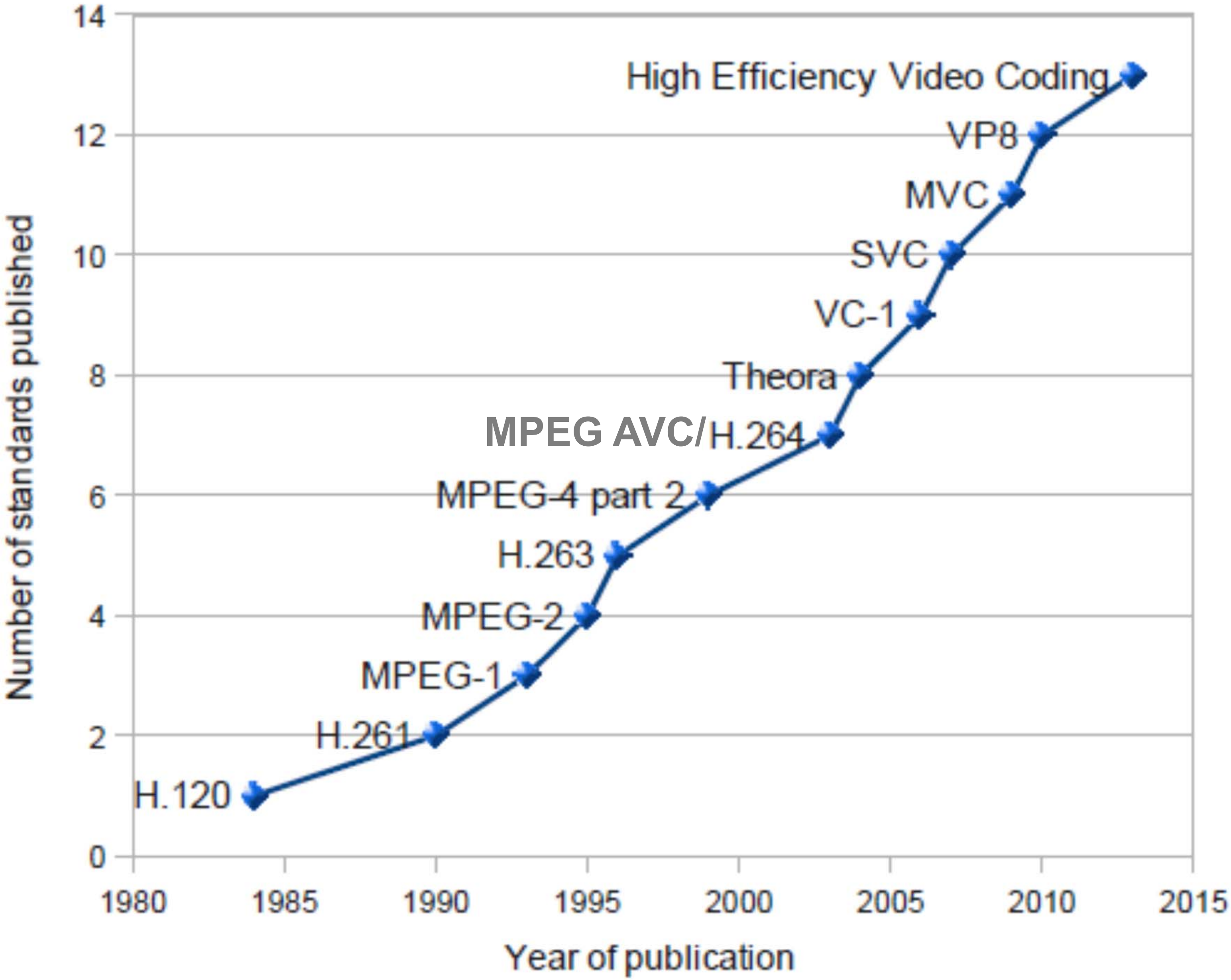
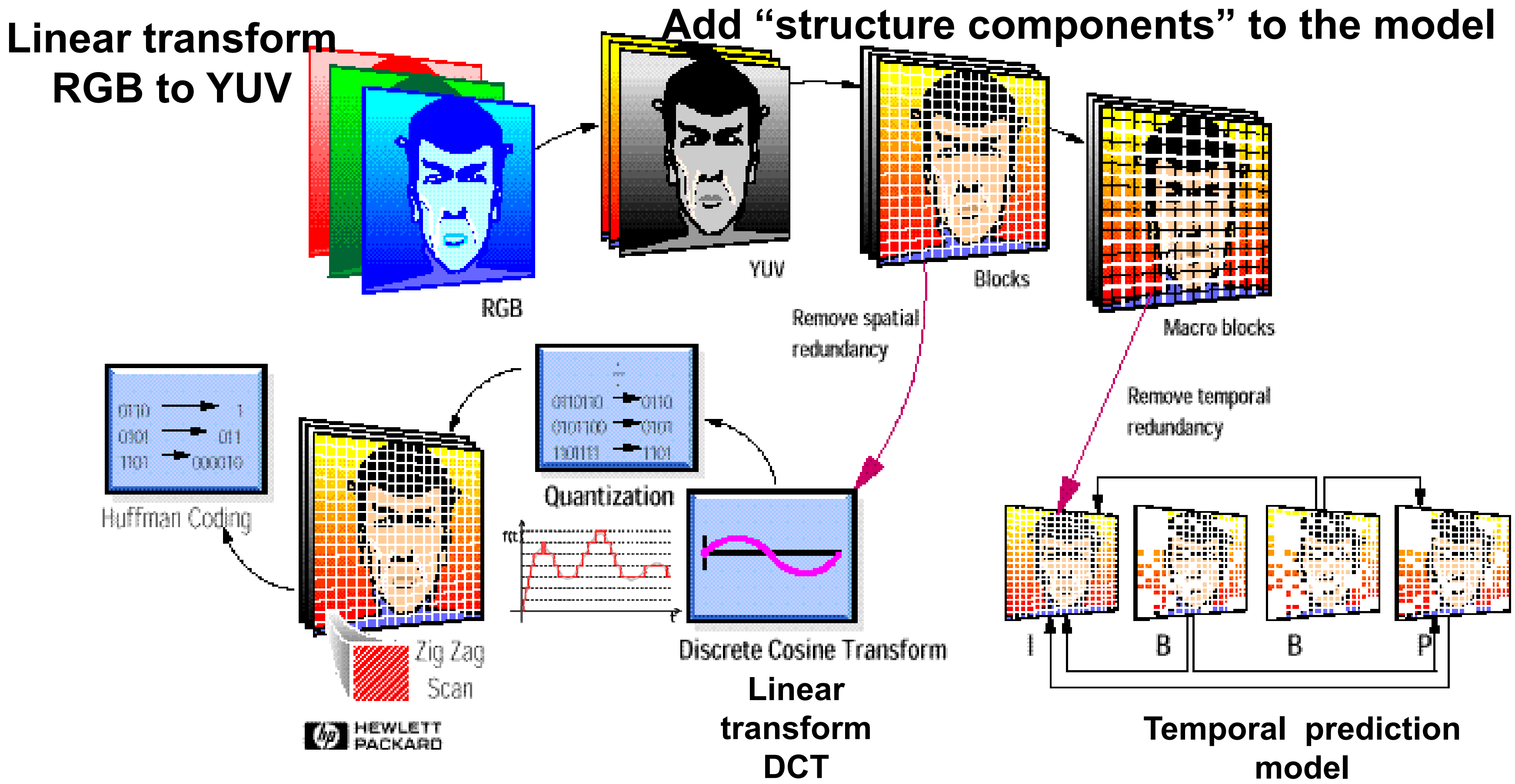
VIDEO COMPRESSION STANDARDS

- Family of standards: the evolution of the coding model state of the art (and implementation technology support):
  - H.261: videoconference x64 (1988)
  - MPEG-1: CD storage (up to 1.5 Mbit/s) (1991)
  - MPEG-2: digital TV and HDTV (1994)
    - 4-9 Mb/s TV, 20 Mb/s HDTV
  - H.263: very low bit rate (16 - 128 kb/s) (1996)
  - MPEG-4 Part 2 (arbitrary shaped objects) (1999)
  - MPEG-4 Part 10 (AVC/H.264) (Adv. Video Cod.) (2003)
  - MPEG-4 SVC (Scalable profile of Part 10 AVC) (2007)
  - MPEG-4 HEVC (High Efficiency Video Coding) (2013)
  - MPEG Future Video Coding (???)

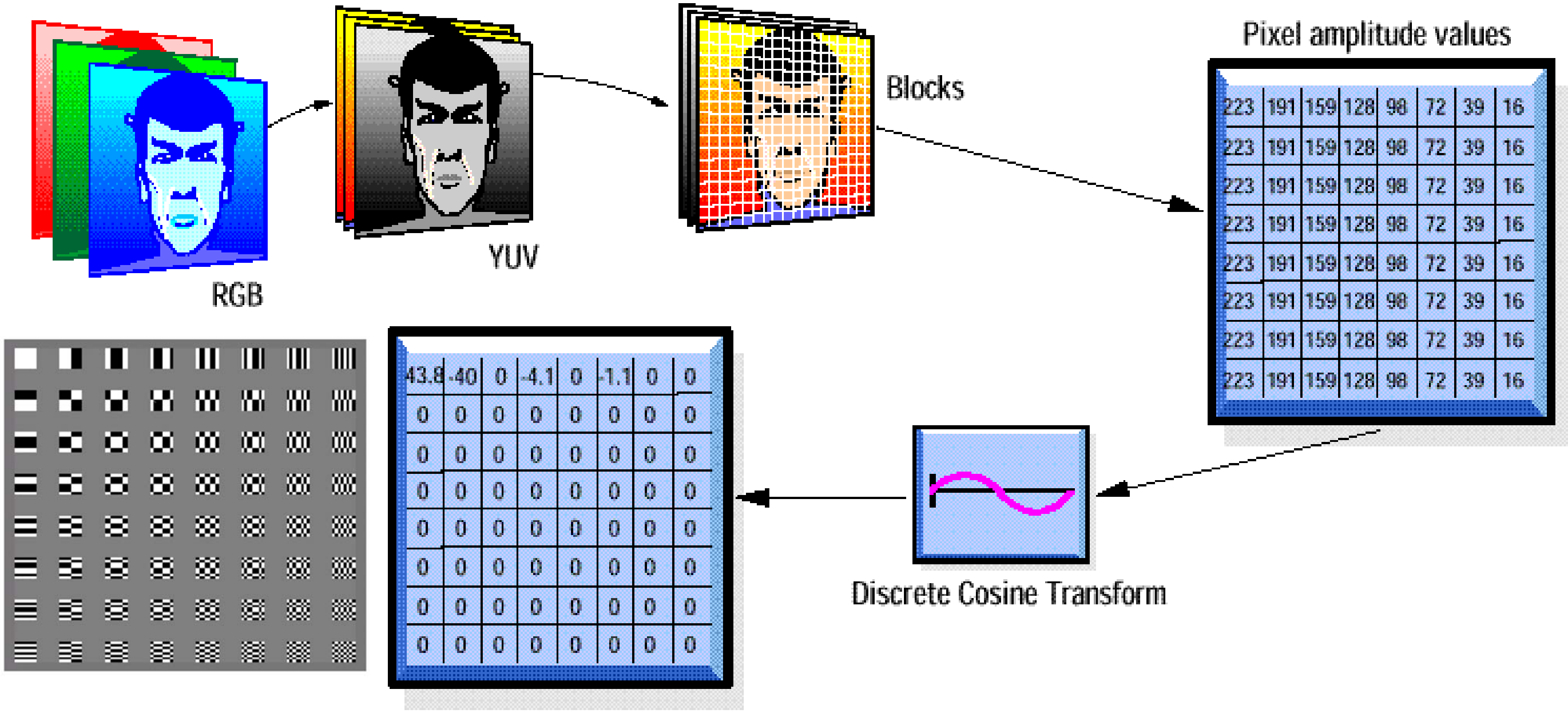
Time evolution of video standards



MPEG Video Coding

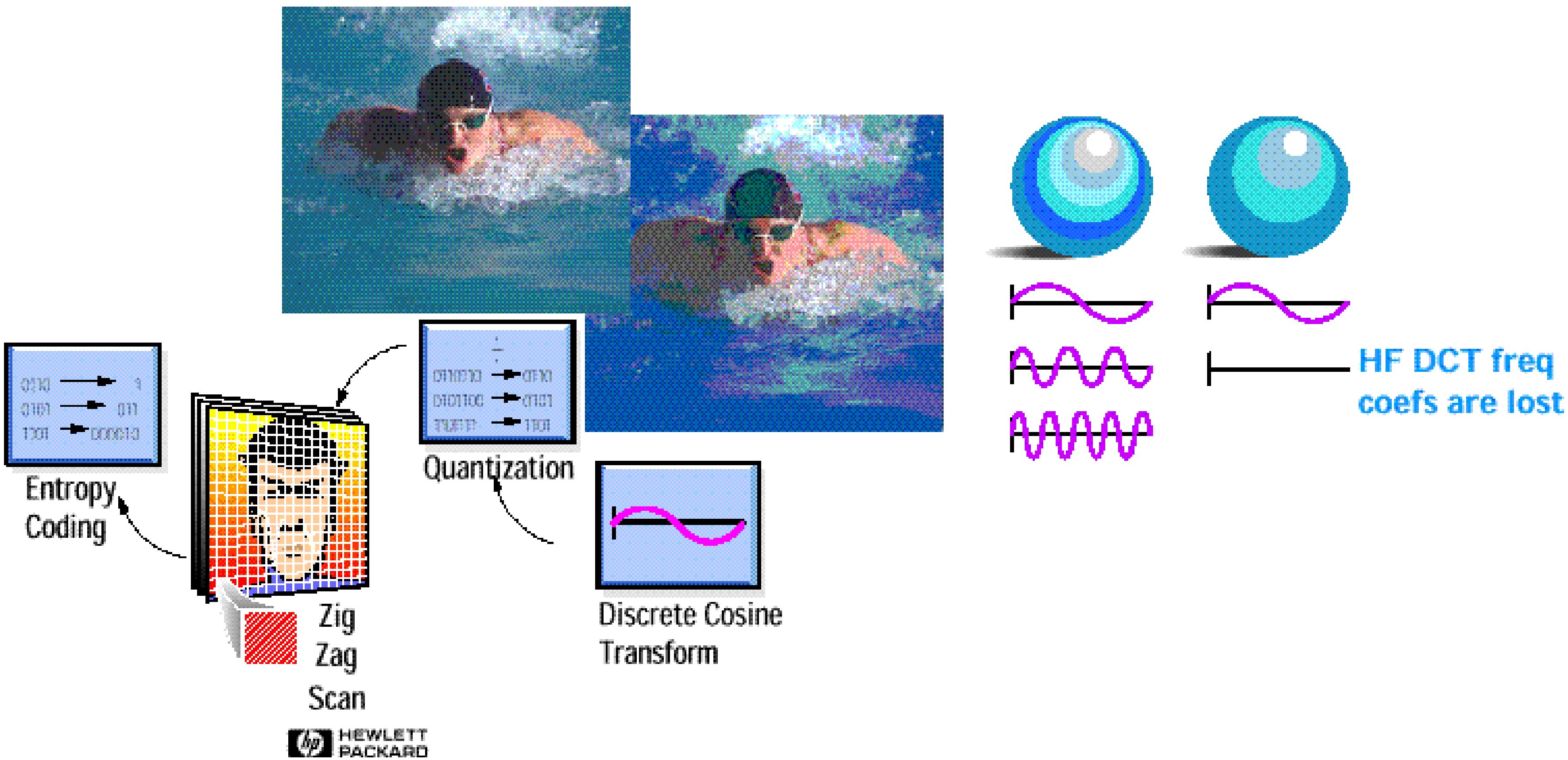


MPEG Video Coding (INTRA)

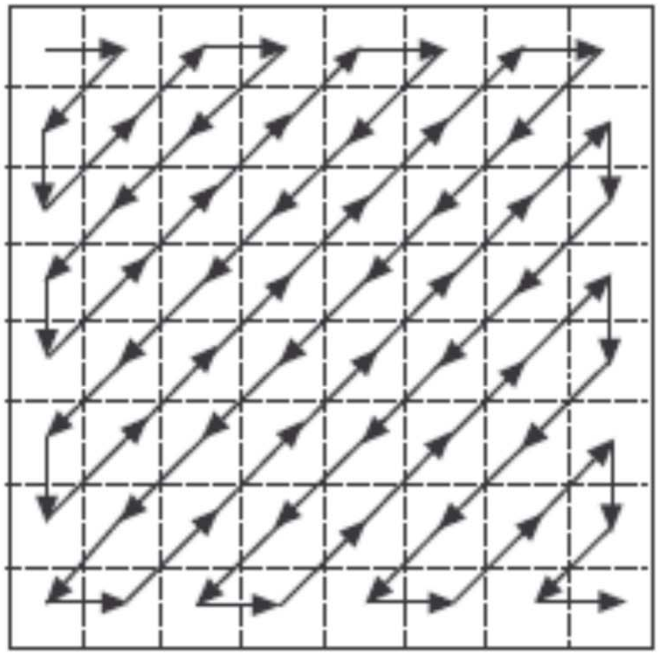


HEWLETT PACKARD

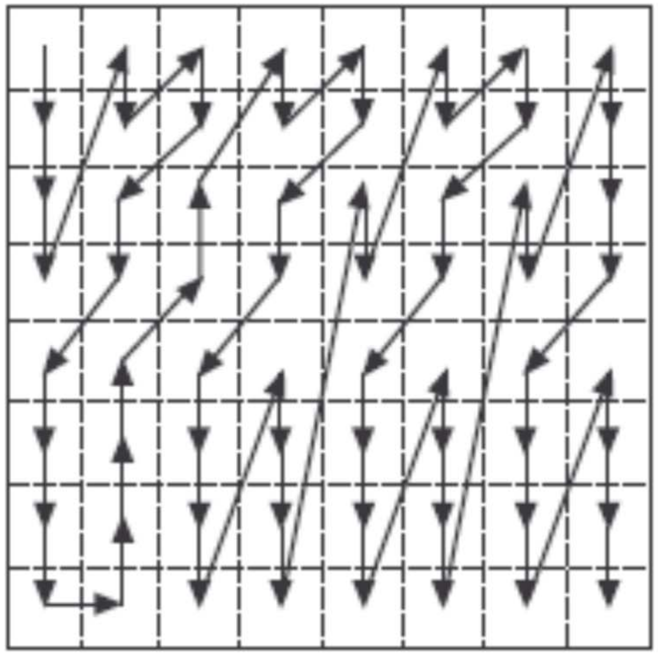
MPEG INTRA Picture Coding



Zig-Zag scan models

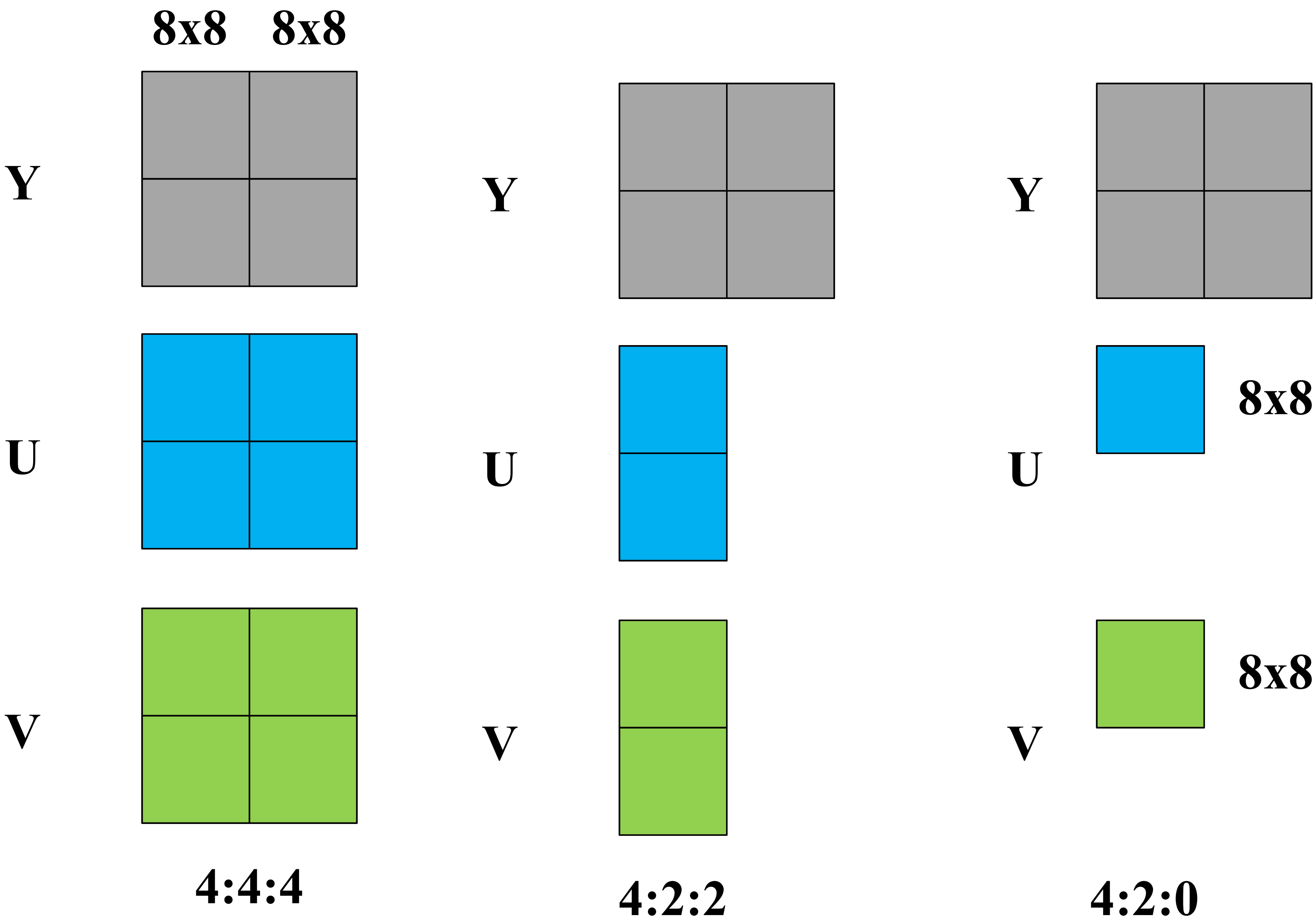


Zigzag or Classic (nominally for frames)
a)

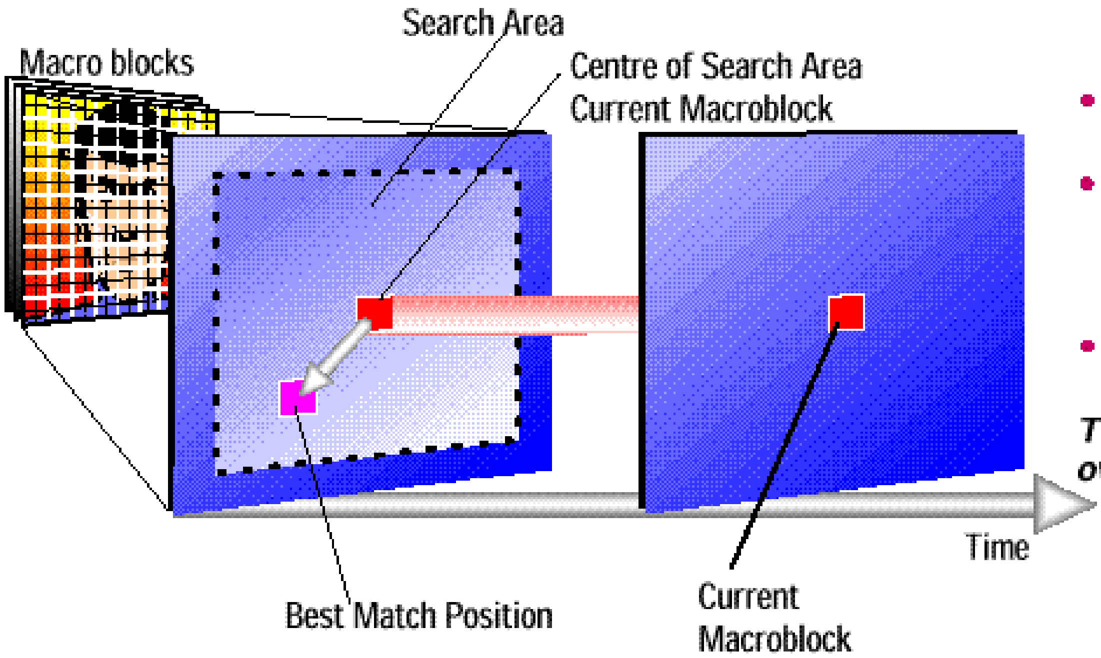


Alternate (nominally for fields)
b)

DCT patterns (MPEG-1 and MPEG-2)



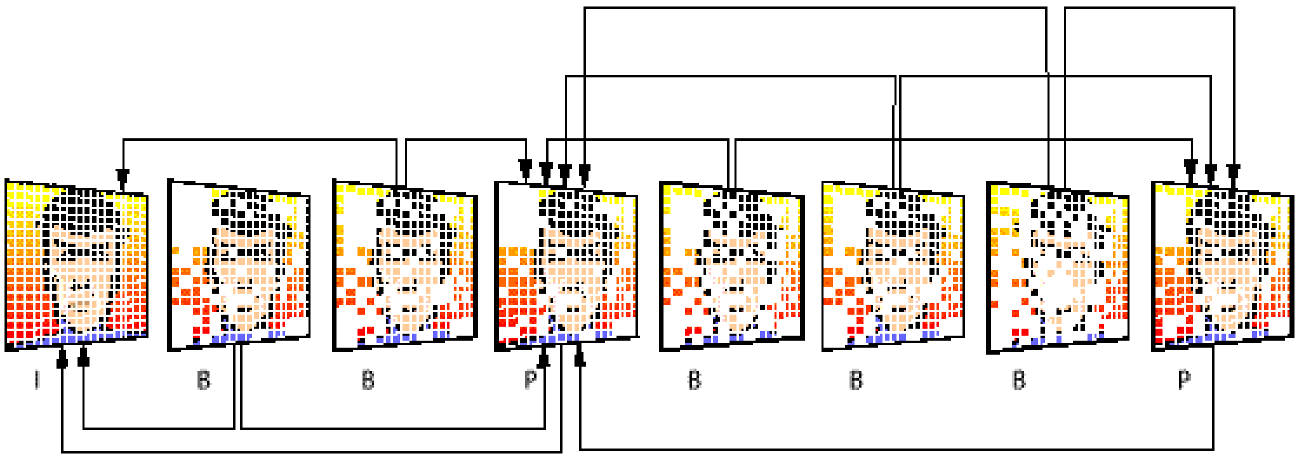
MPEG Motion prediction and compensation



- If adjacent frame macroblocks same, don't retransmit block.
- Search to see if same block exists somewhere else: if it does, just transmit its coordinates (motion vectors).
- Only Intra-code macroblocks which are completely new.

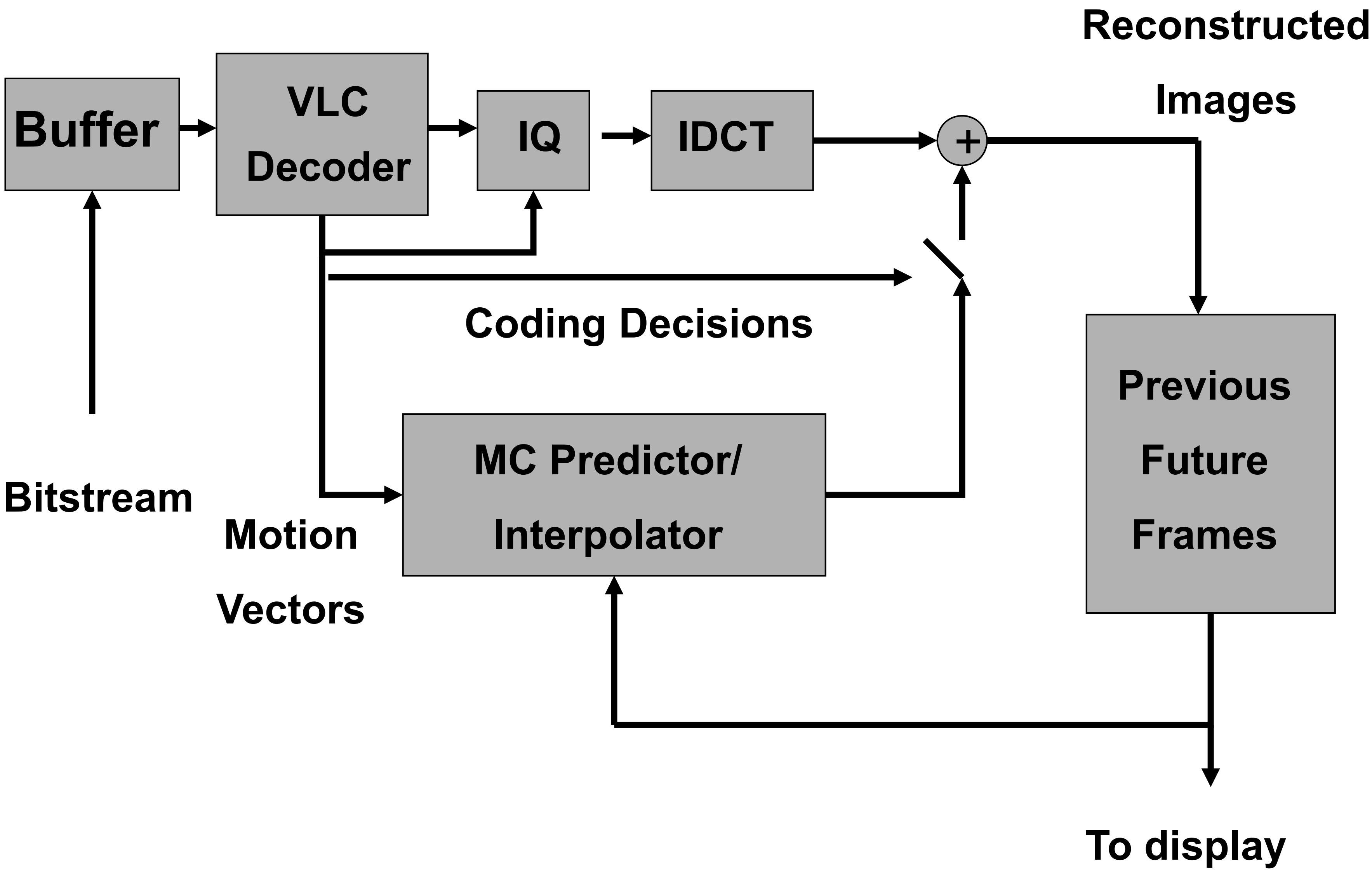
This process really drops the overall bit rate

MPEG Motion Prediction Scheme

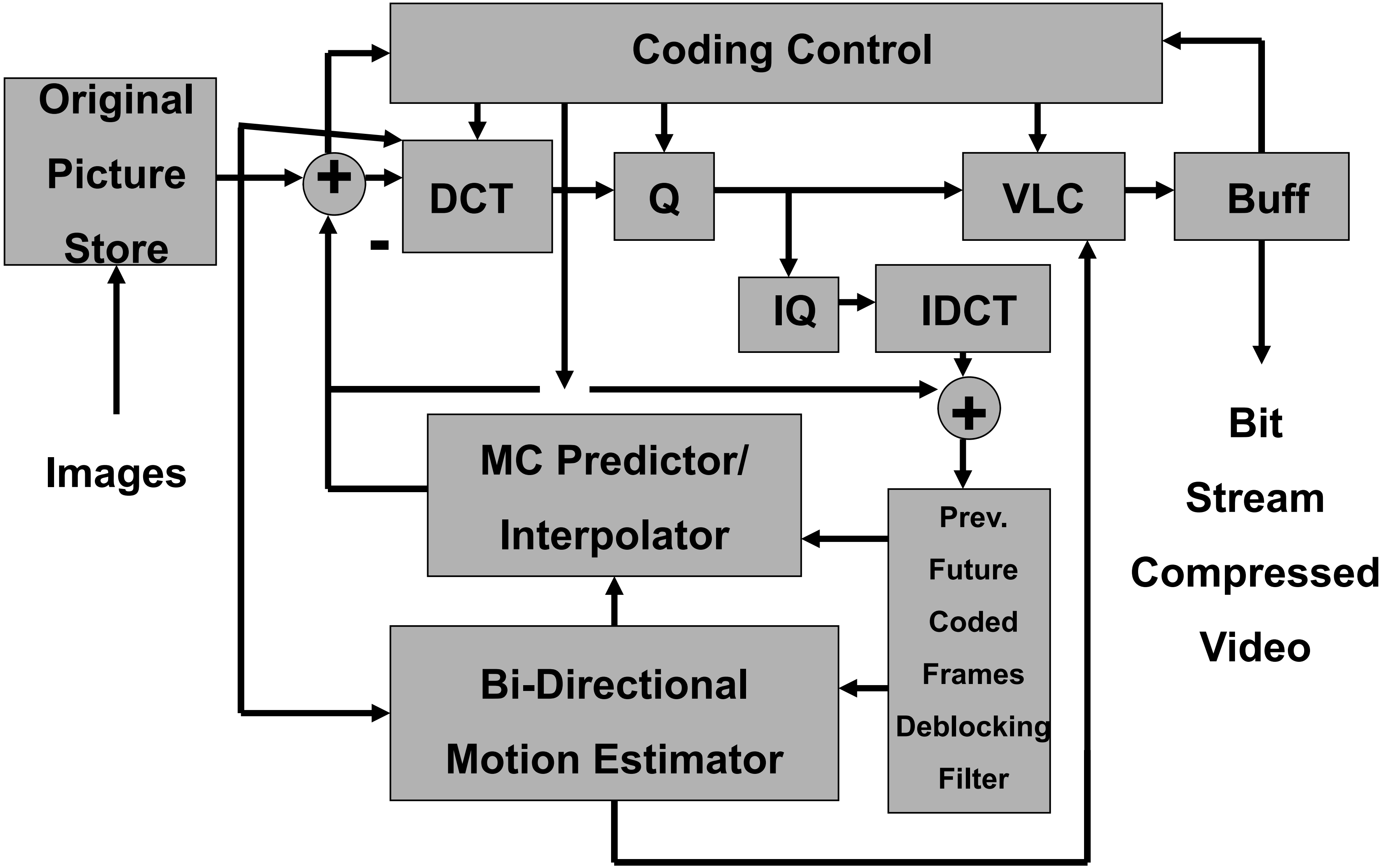


- I-frames: contain full picture information
- P-frames: predicted from past I or P frames
- B-frames: use past and future I or P frames
- Transmit I frames every 12 frames or so.

MPEG-X DECODER



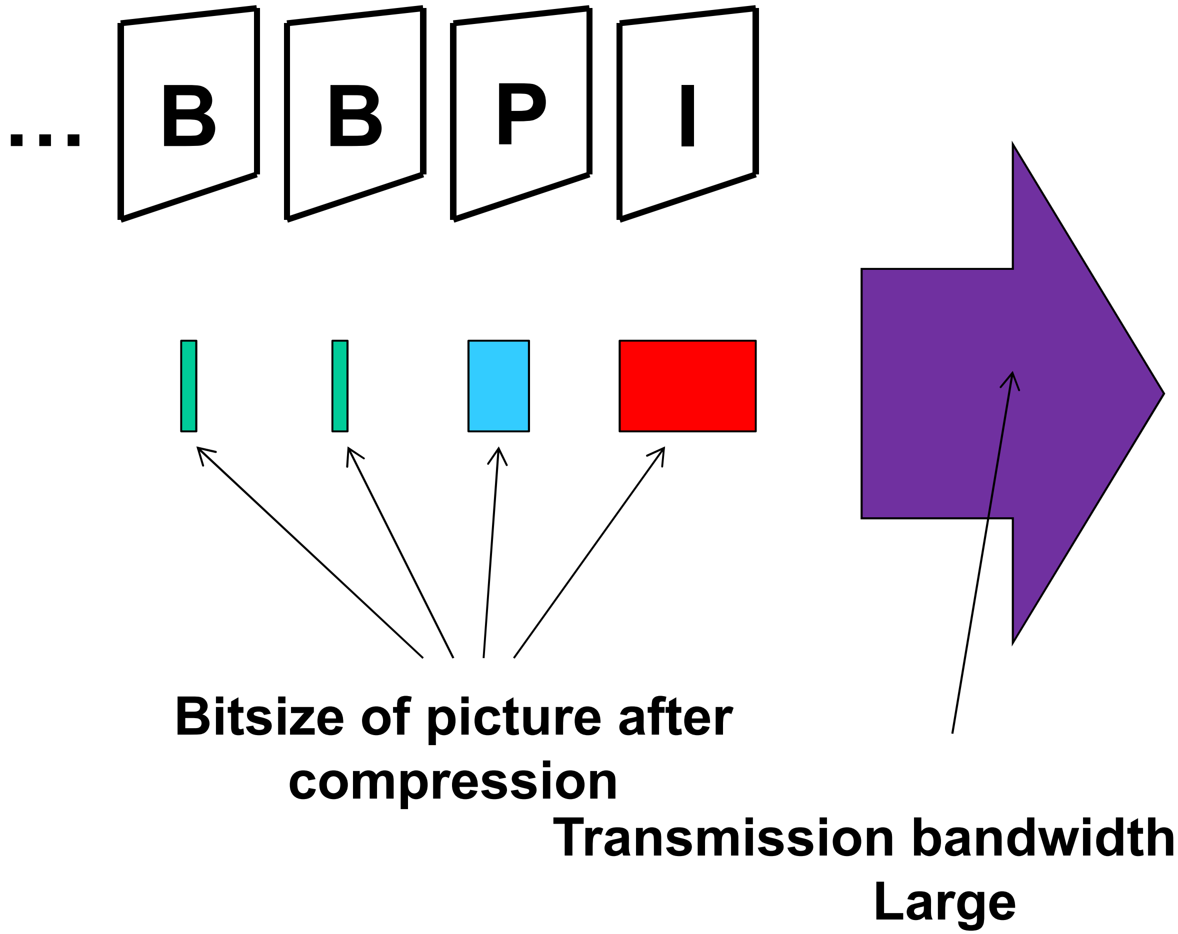
MPEG-X ENCODER



Fundamental ISO/IEC SC29WG11 (MPEG) Principles

- Decoder must be simple (focus on broadcast)
- Decoding syntax is completely specified
  - a decoder must be conformant (i.e. the full “encoder model” is known)
- Encoding “syntax” is fully specified
  - a video bit-stream must be conformant
- The encoder is not specified
  - Encoding algorithms are open to innovation
  - Encoder implementations are a competitive issue!!!

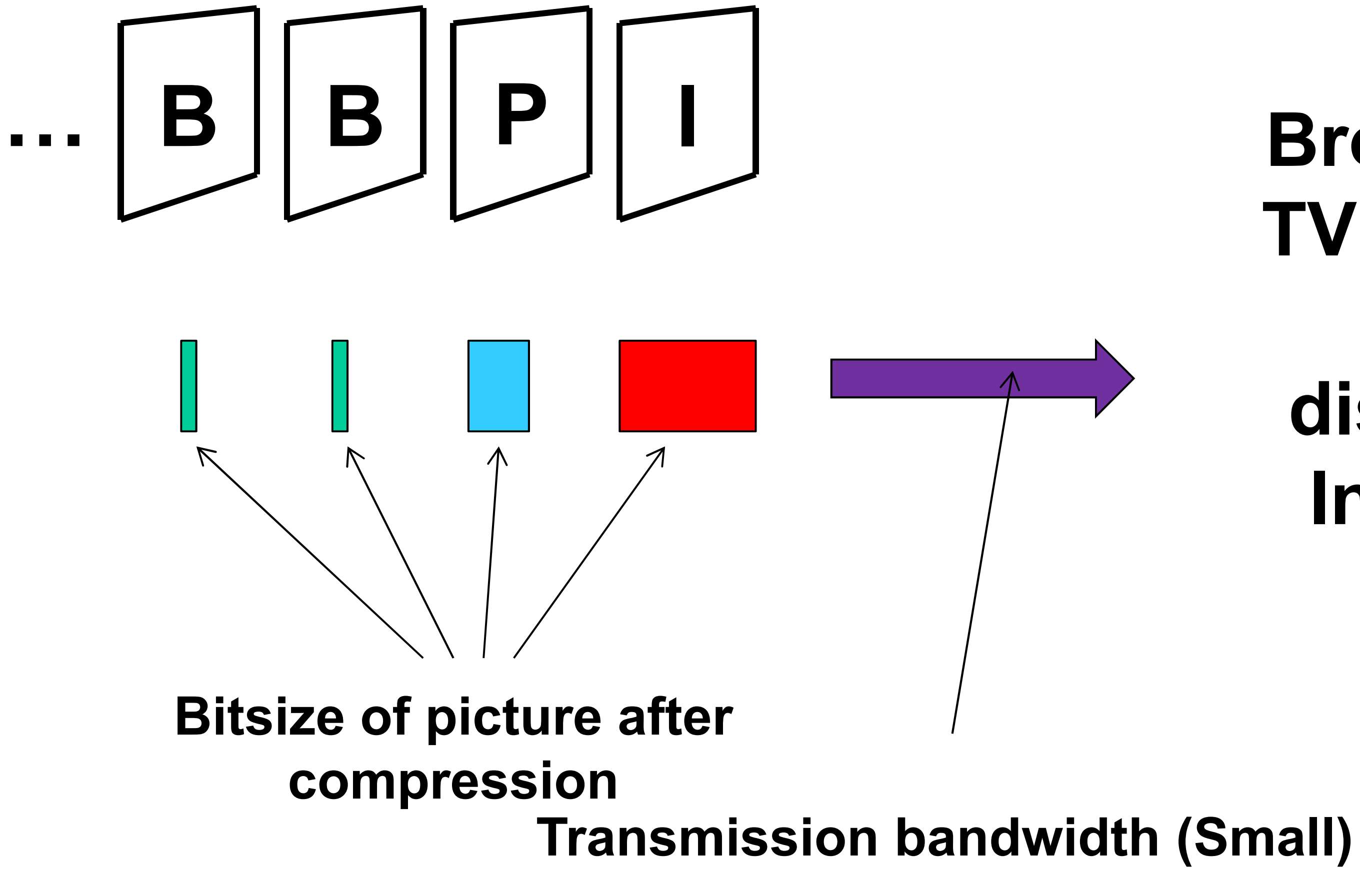
Constant and Variable Bitrate



Mass storage device Hard disk, Flash memory, DVD, etc, etc ….

- **Solution: fixed quantization step:**
  - **Constant quality and variable storage size**

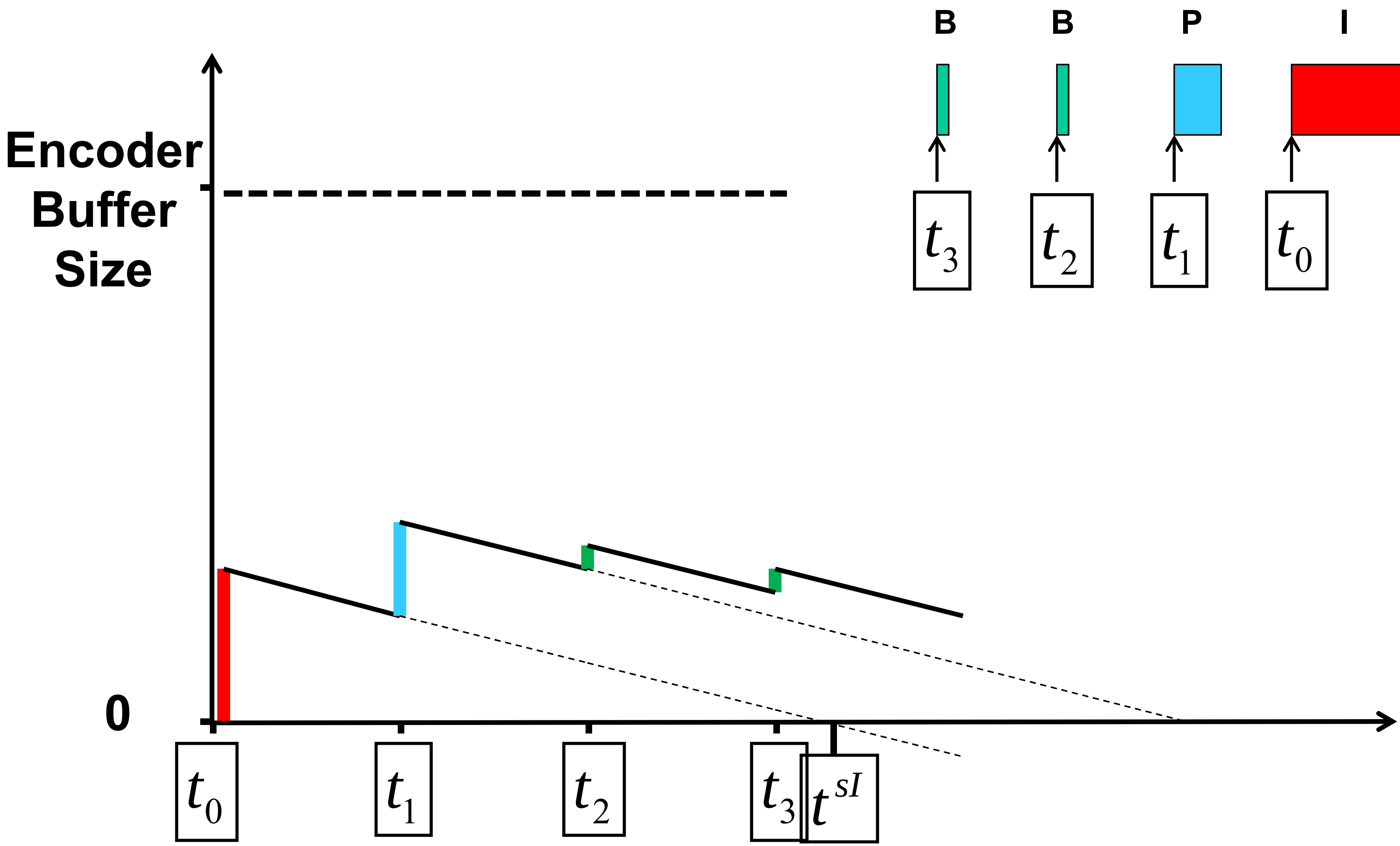
Variable Bitrate



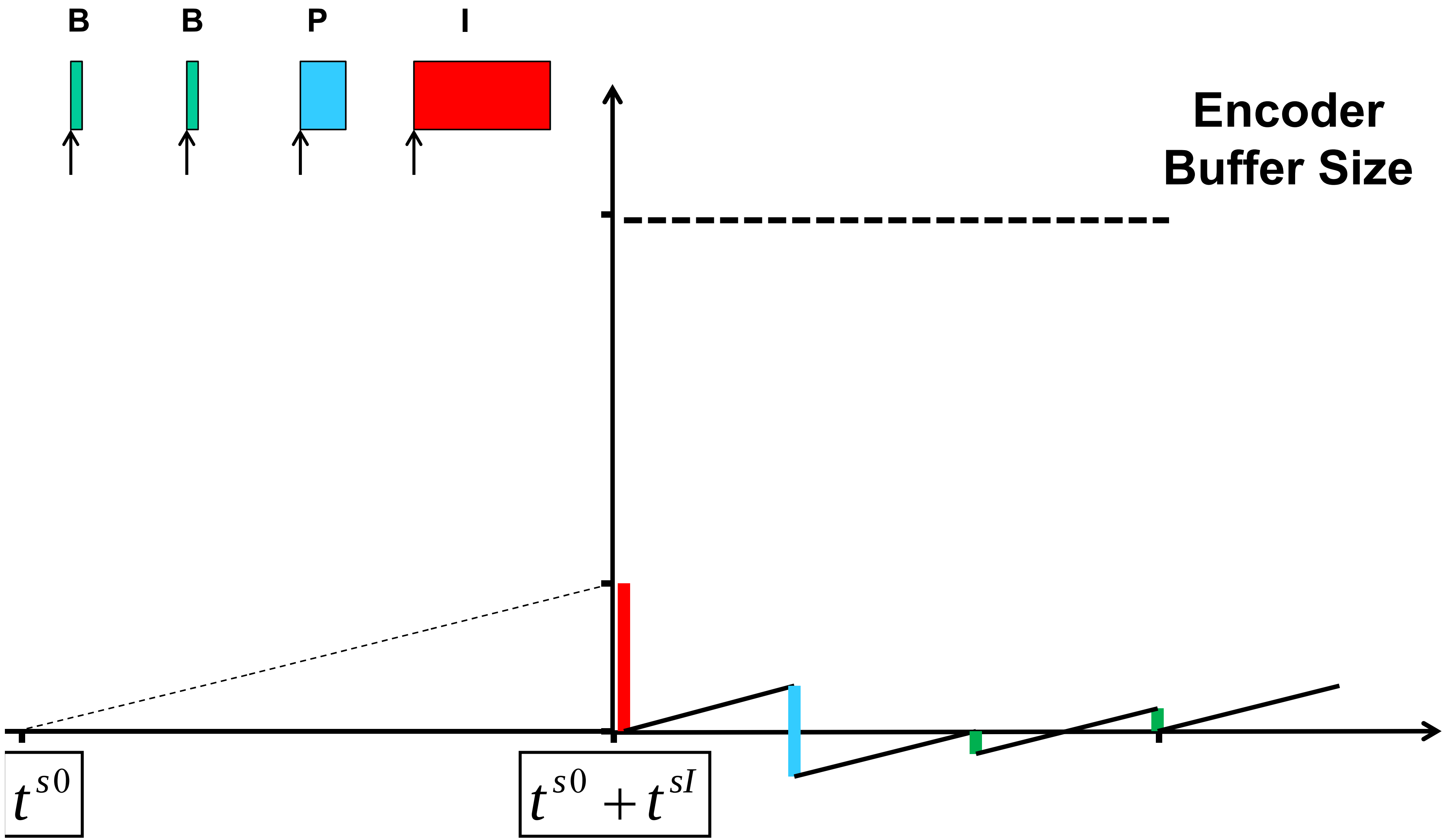
Broadcasting TV (DVB-T/S), Cable distribution, Internet TV

- **Solution: variable quantization:**
  - **Variable quality and fixed "average" bitrate**

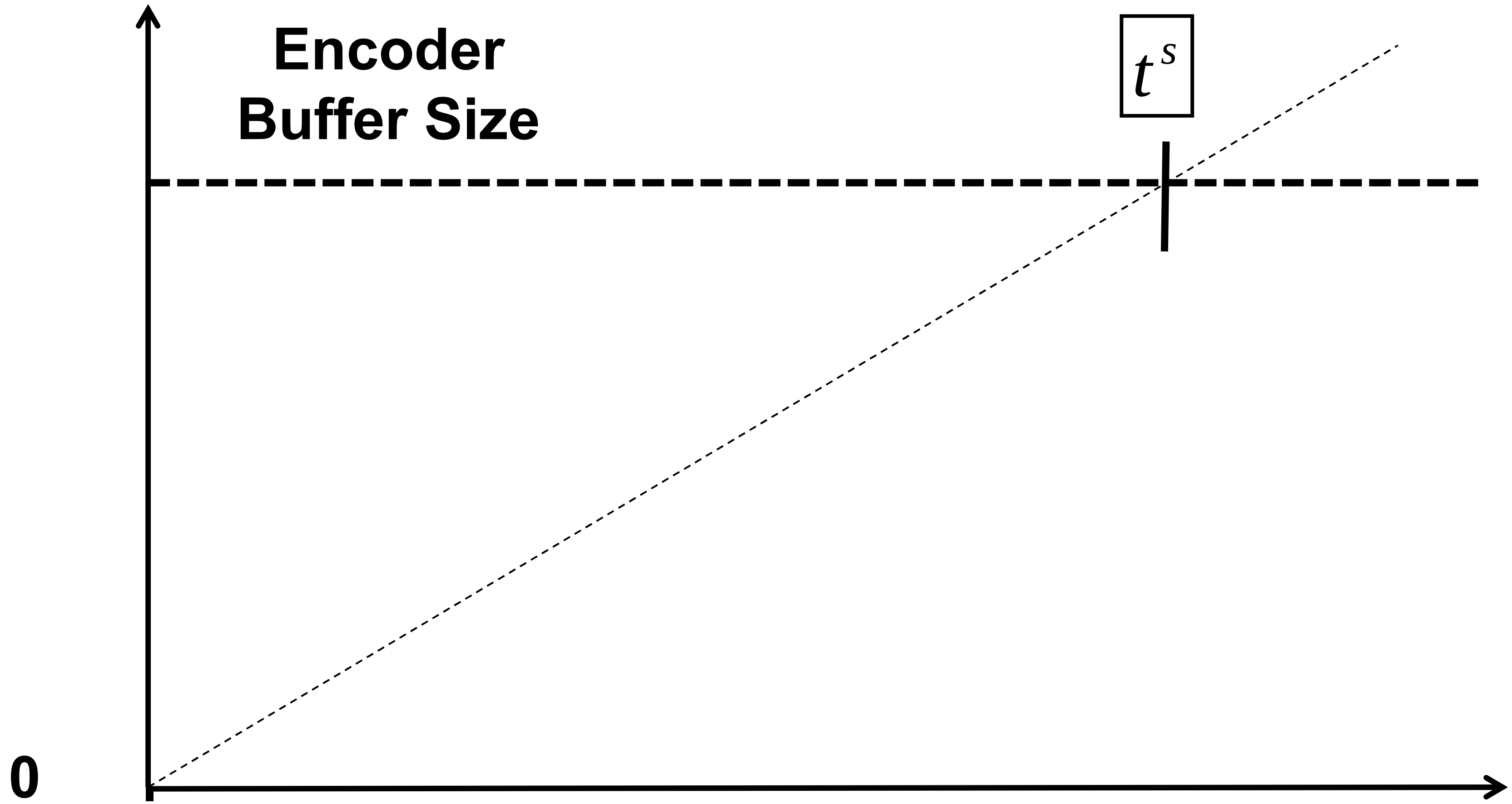
Encoder Buffer



Buffer Occupancy (Decoder)



End to End delay

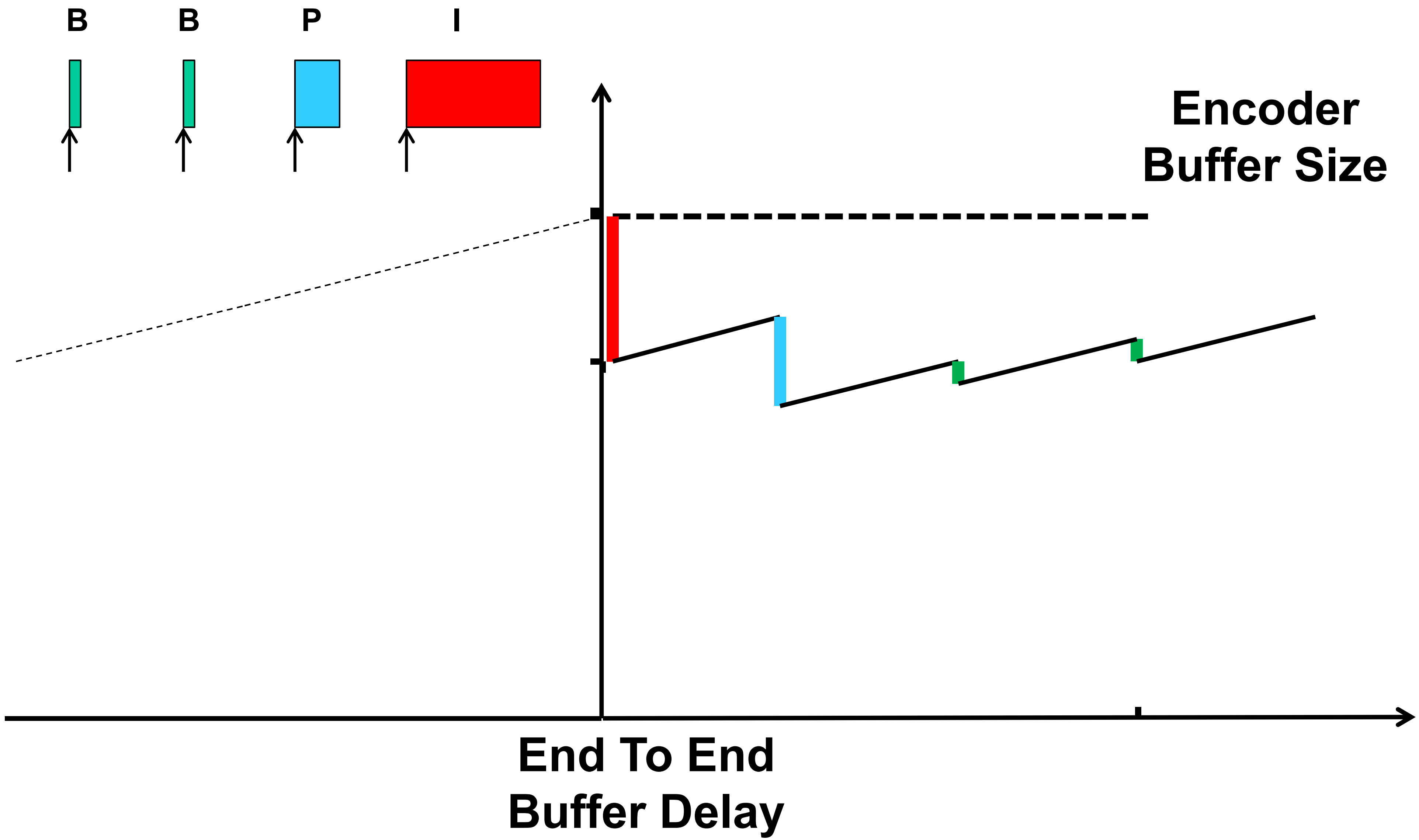


End To End Delay

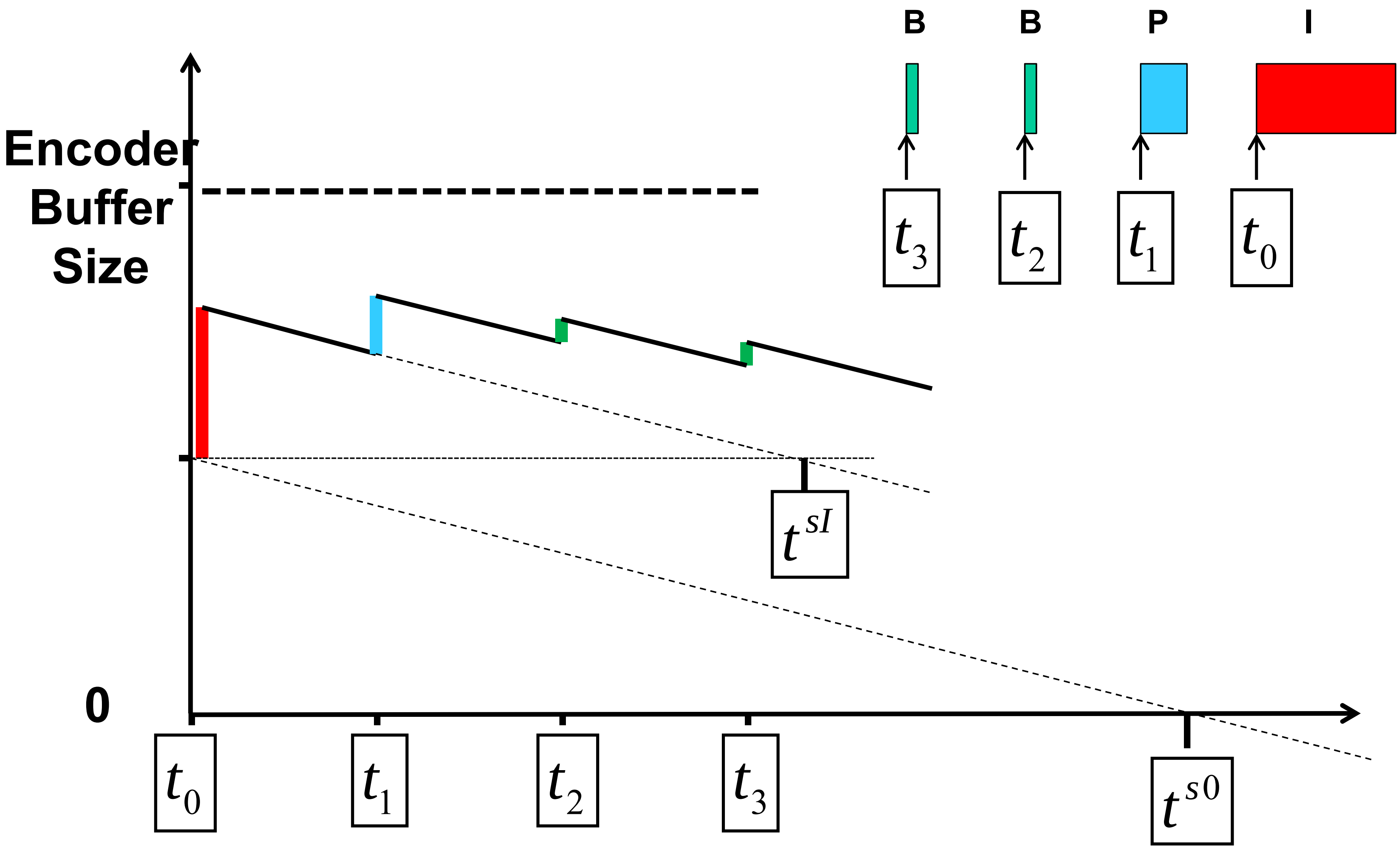
- End To End Delay (ETED) (constant):

$$ETED = \frac{Enc\_Buff\_Size(bit)}{Trans\_Bandwidth(bit/\sec)}$$

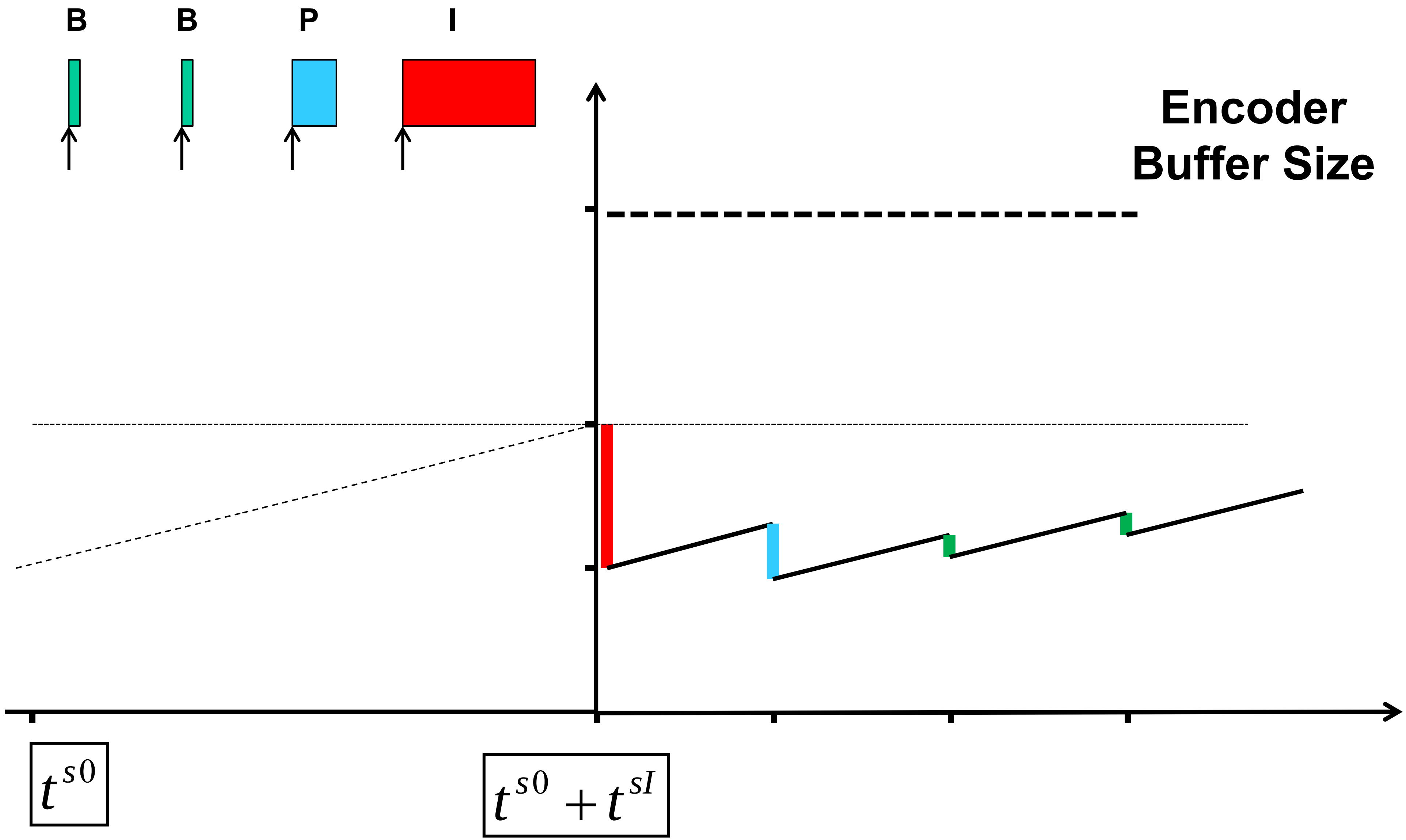
Buffer Occupancy (Decoder)



Buffer Control (encoder)

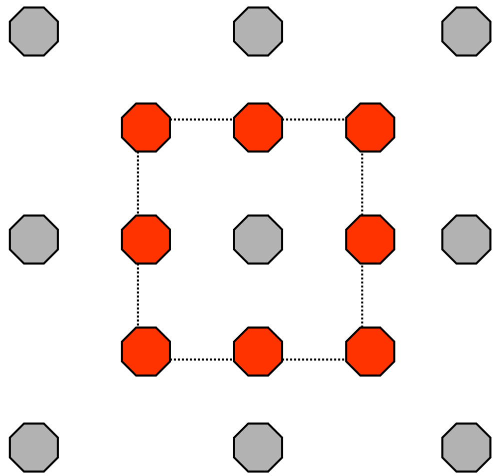


Buffer Occupancy (Decoder)



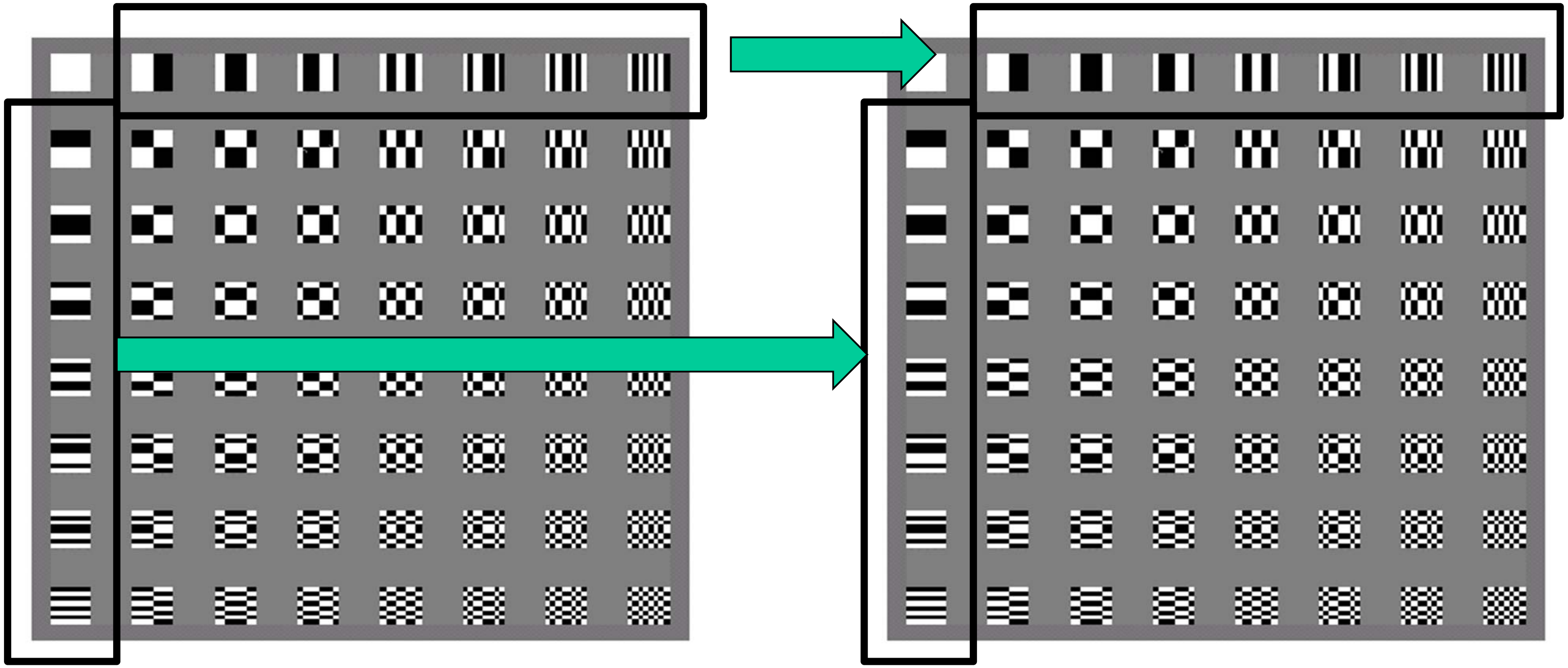
MPEG-2 CODING MODES EXTENSIONS

- Field / Frame prediction (all modes can be based on frames or separate fields)
- Low delay prediction modes
- Half - pixel interpolations for motion estimation.

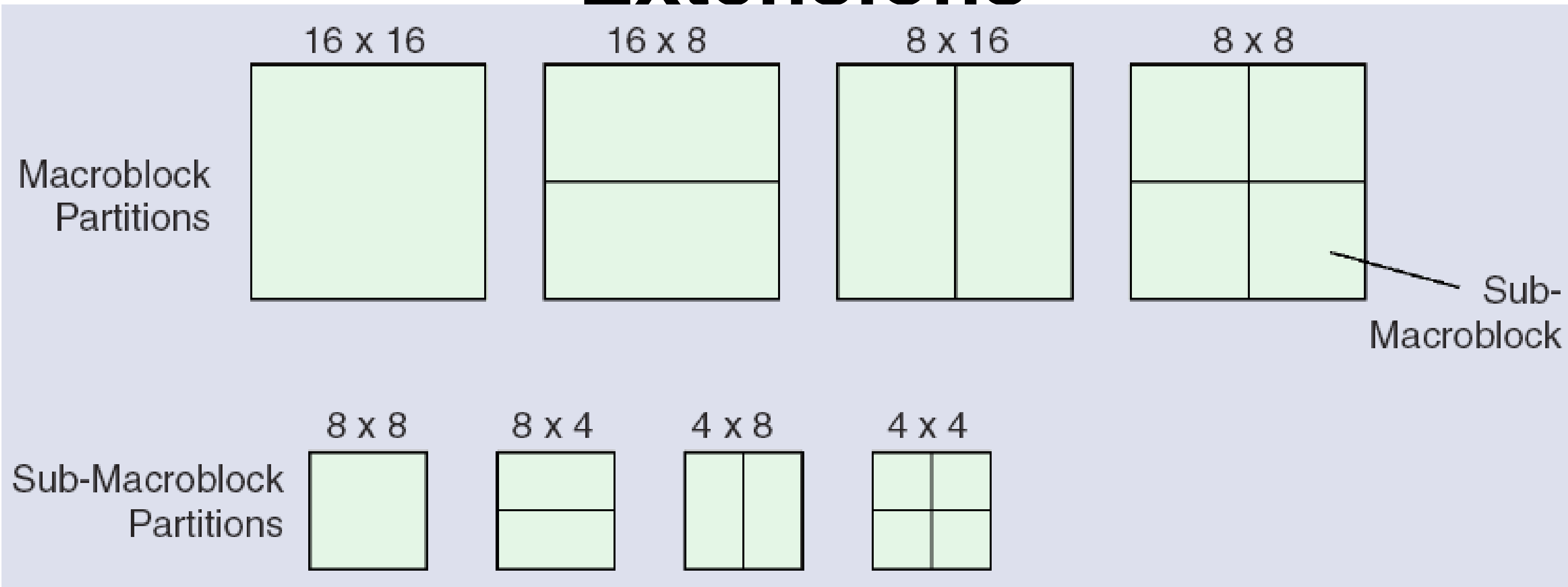


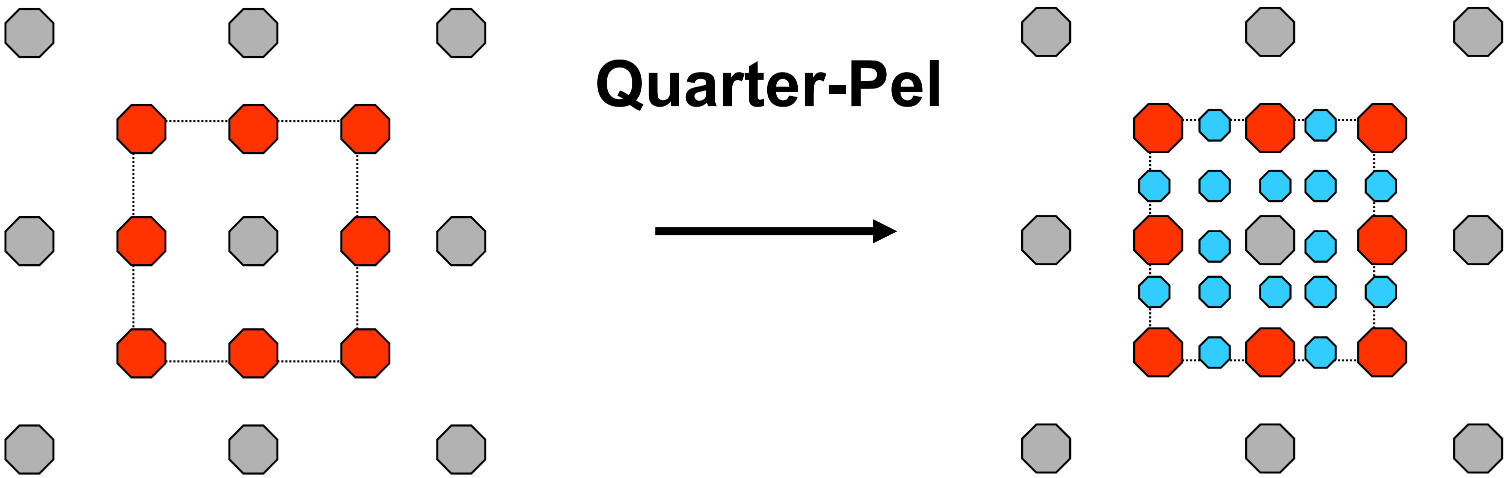
MPEG-4 Coding extensions

- AC-DC prediction for INTRA pictures (only DC was used in MPEG-1/2)



MPEG-4 (Part 2 and 10 AVC) Model Extensions

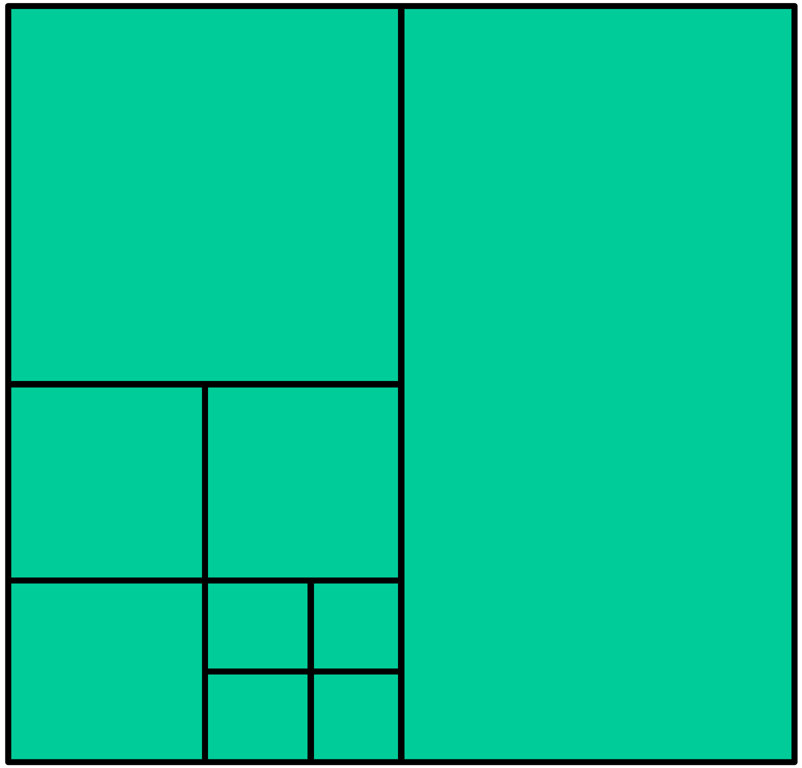




HEVC Coding Extensions

- IDCT: 8x8, 16x16, 32x32
- Any quadtree conf.
- Only Arithmetic Coding
- ………

32 x 32



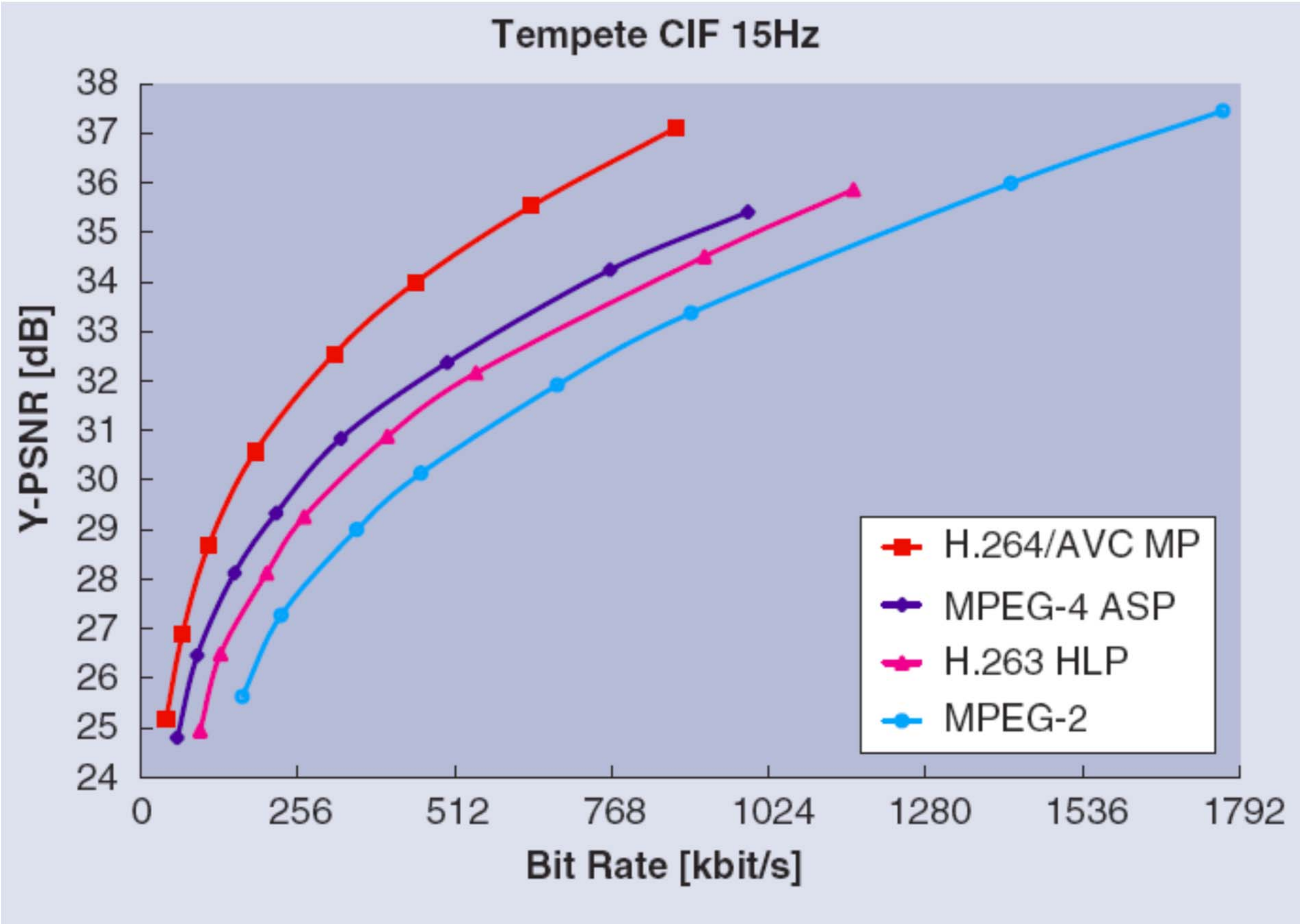
How to measure codec performance?

- Subjective tests of perceived quality:
  - Young students (sight and attention)
  - Statistical double blind methods
- Objective measurements:

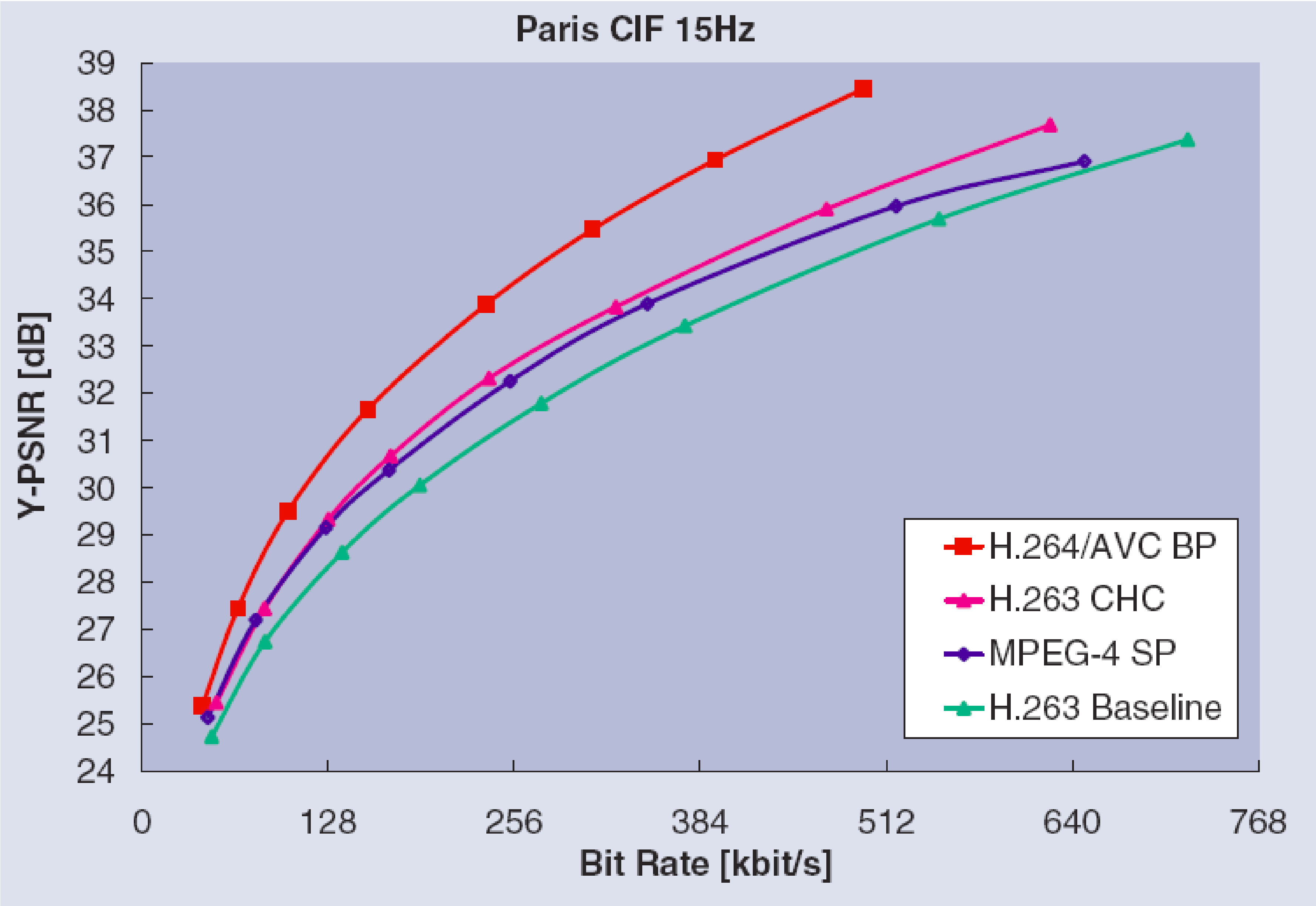
$$PSNR = -10 Log_{10}\left(\frac{\frac{1}{N}\sum_{0}^{N-1}(Y_{orig} - Y_{decoded})^2}{255^2} \right)$$

- Attention: same PSNR very different visual results:
  - Errors in visible (uniform, shapes) or in less visible elements (edges, textures)

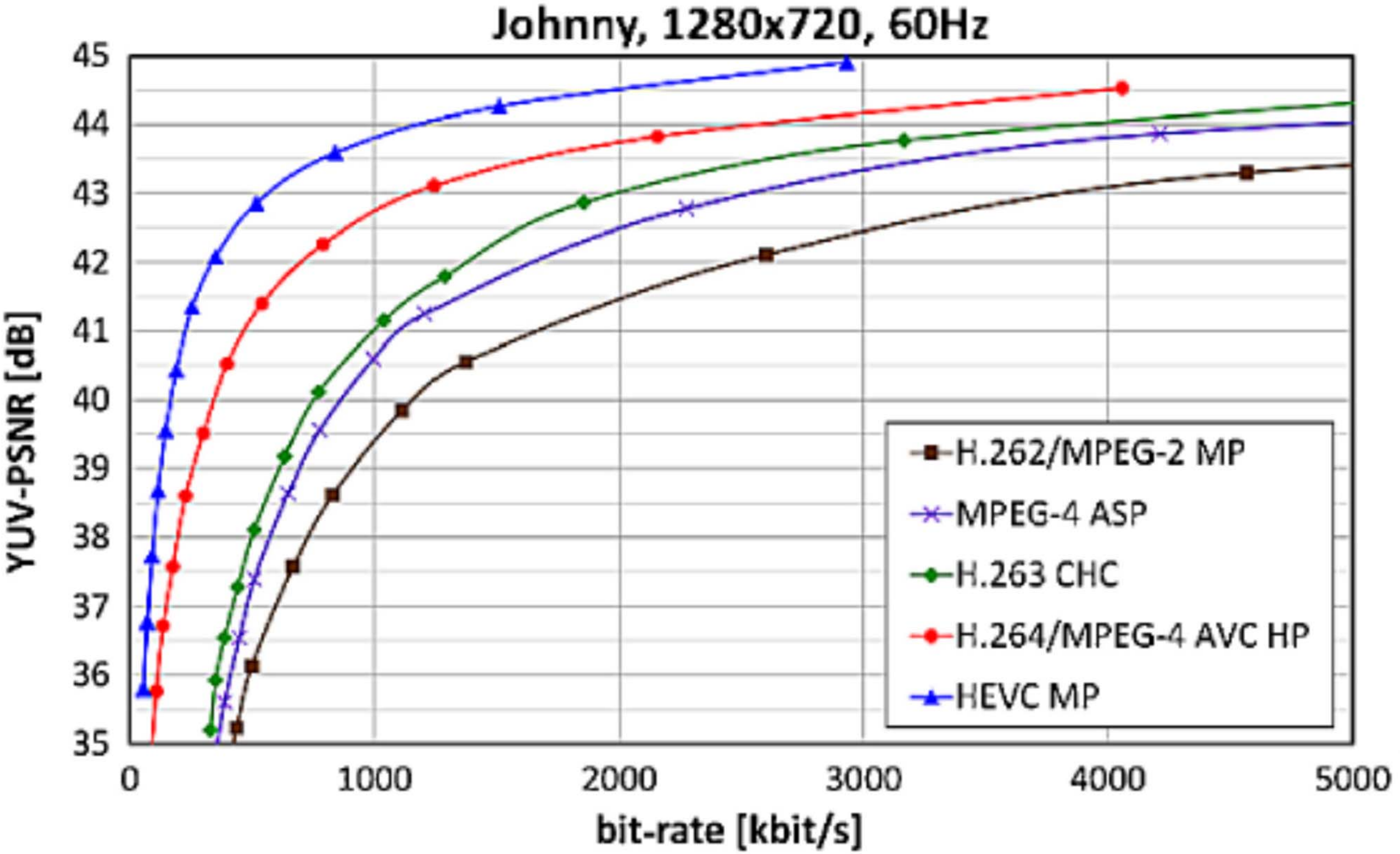
PSNR Performance: Video Streaming Appl.



Performance Videoconference Appl.

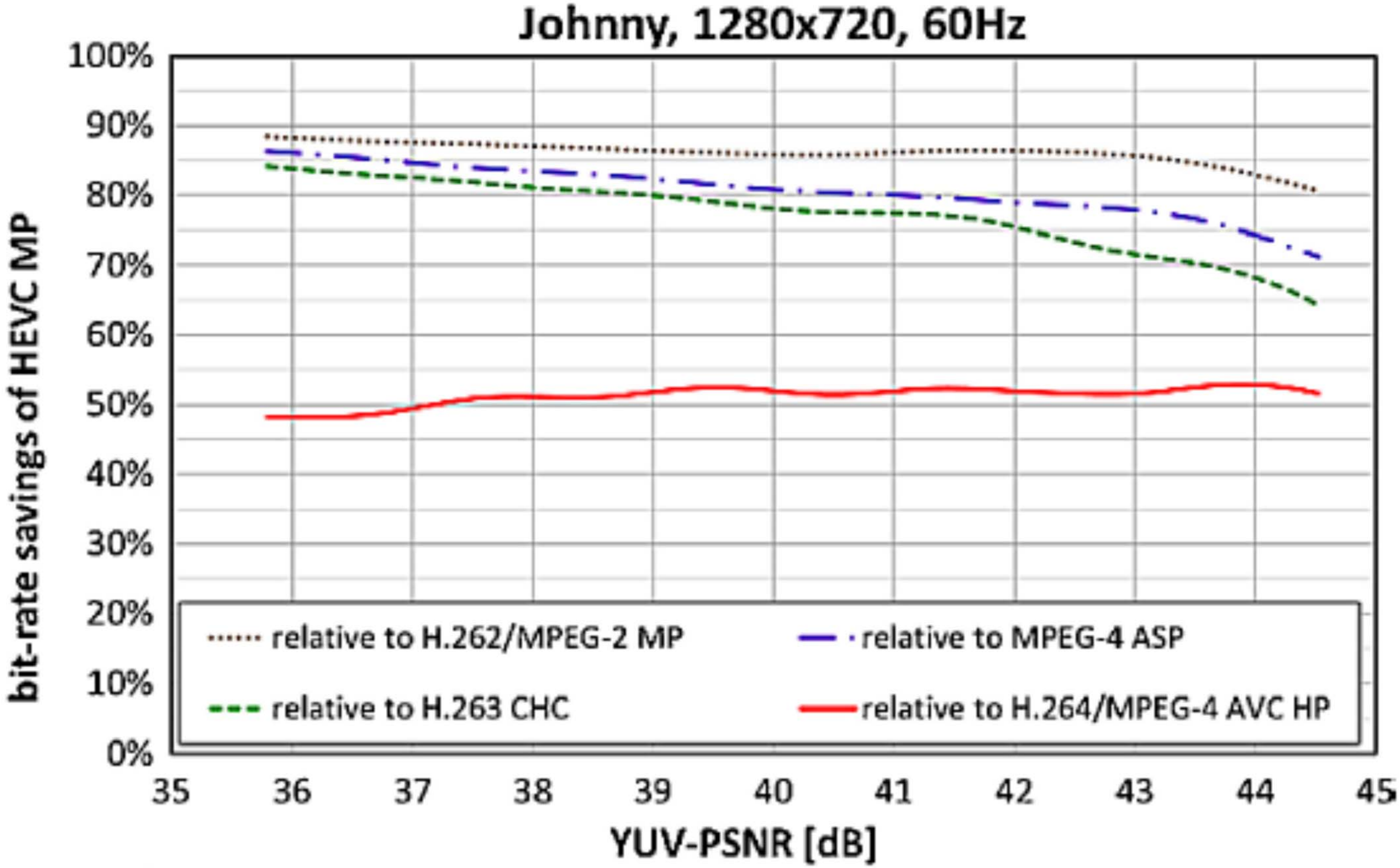


PSNR Performance (HEVC): HDTV Video 720P



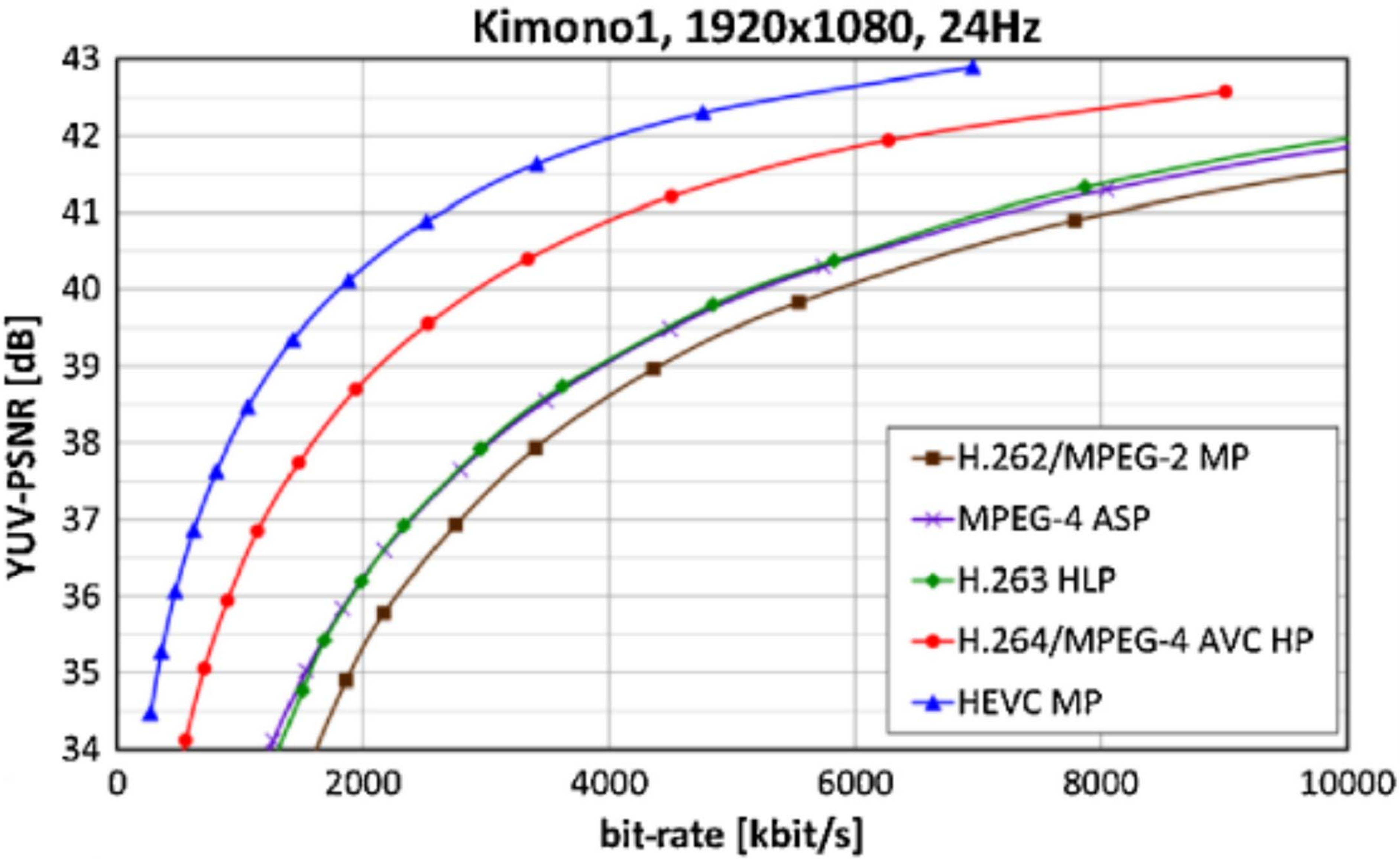
J. Ohm et al, "Comparison of the coding efficiency of video coding standards including High Efficiency Video Coding, IEEE Trans on CSVT, Vol 22, No 12, Dec 2013.

Bitrate Saving Performance (HEVC): HDTV Video 720P



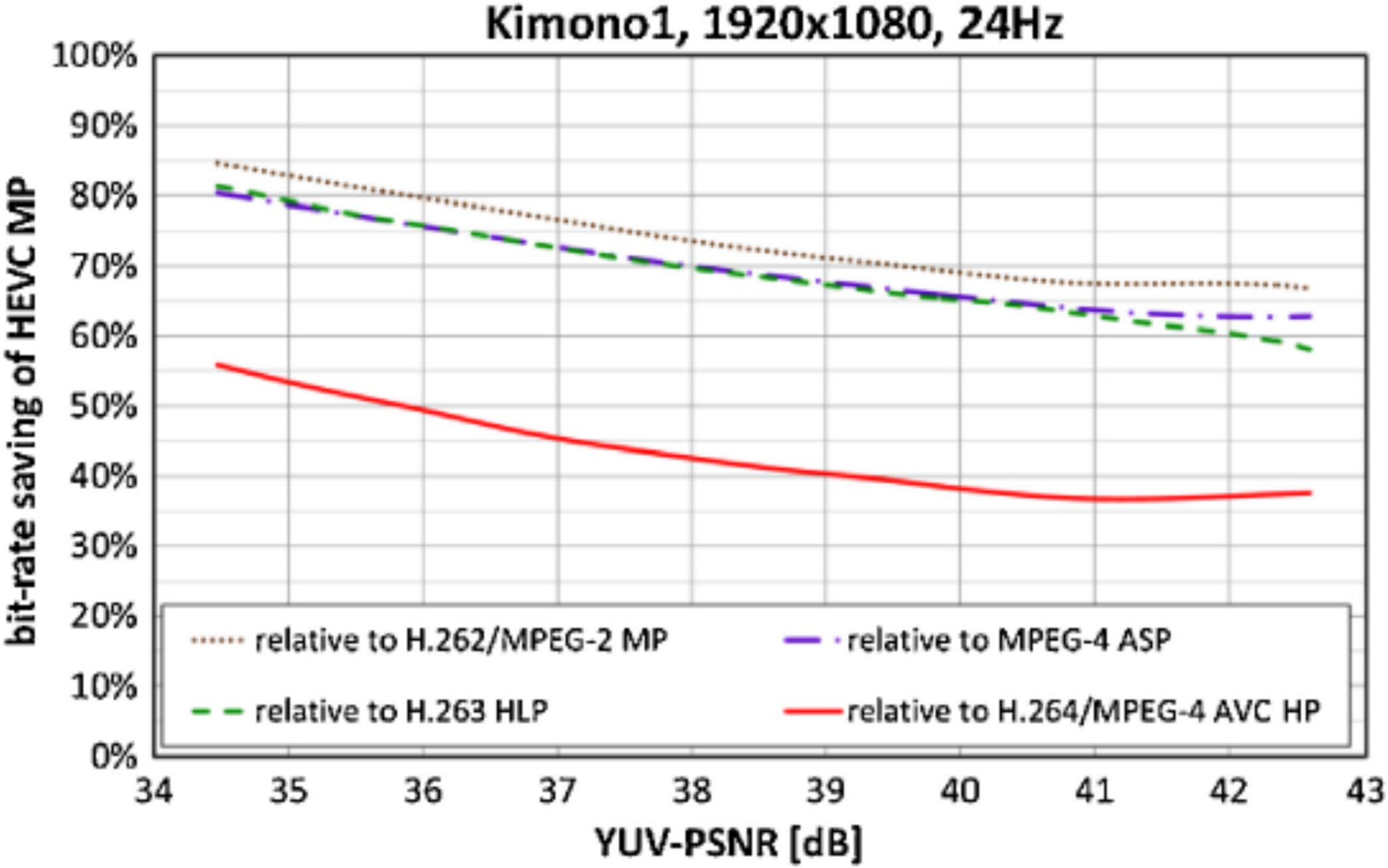
J. Ohm et al, “Comparison of the coding efficiency of video coding standards including High Efficiency Video Coding, IEEE Trans on CSVT, Vol 22, No 12, Dec 2013.

PSNR Performance (HEVC): HDTV Video 1080P



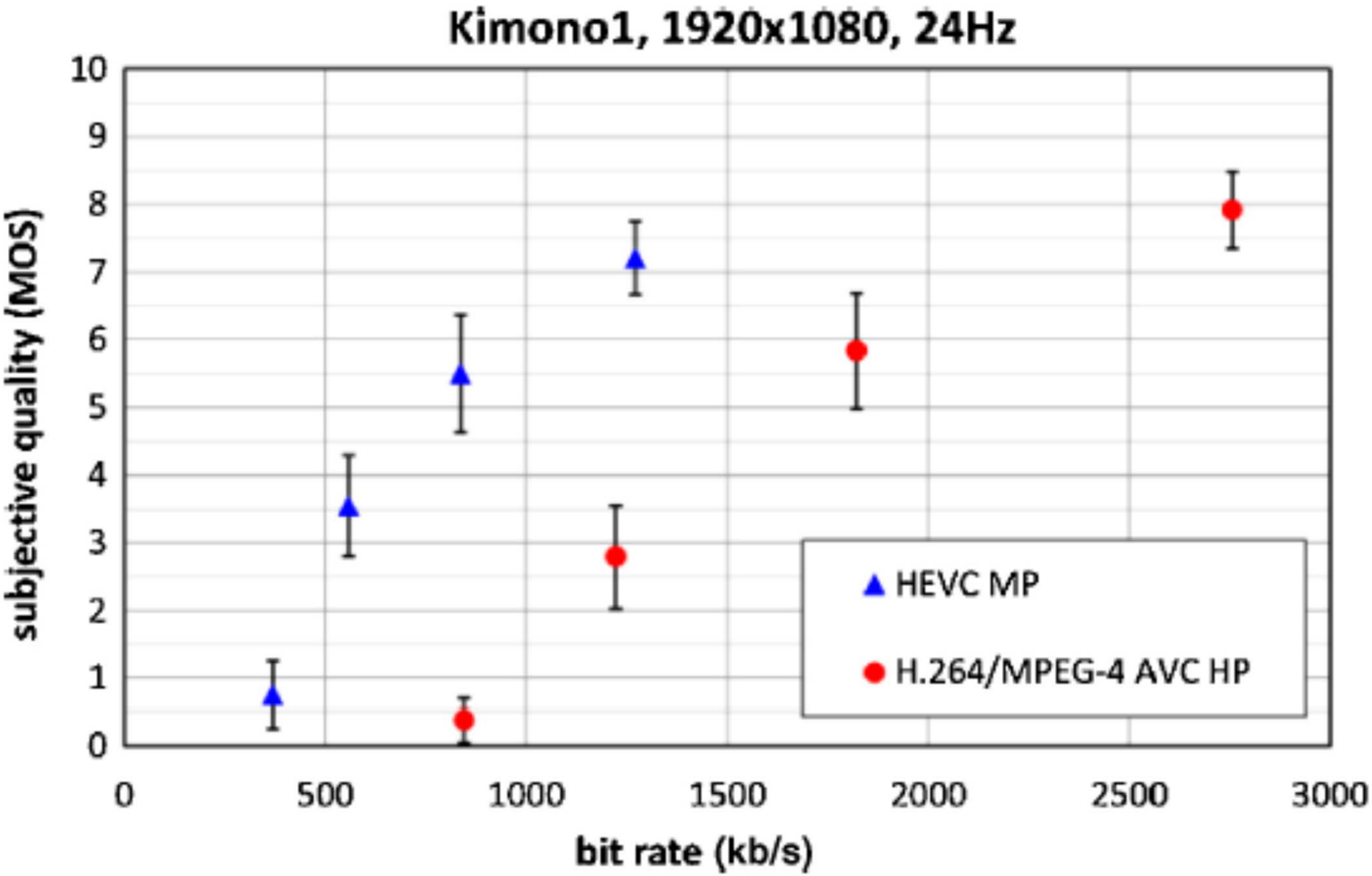
J. Ohm et al, “Comparison of the coding efficiency of video coding standards including High Efficiency Video Coding, IEEE Trans on CSVT, Vol 22, No 12, Dec 2013.

Bitrate Saving Performance (HEVC): HDTV Video 1080P



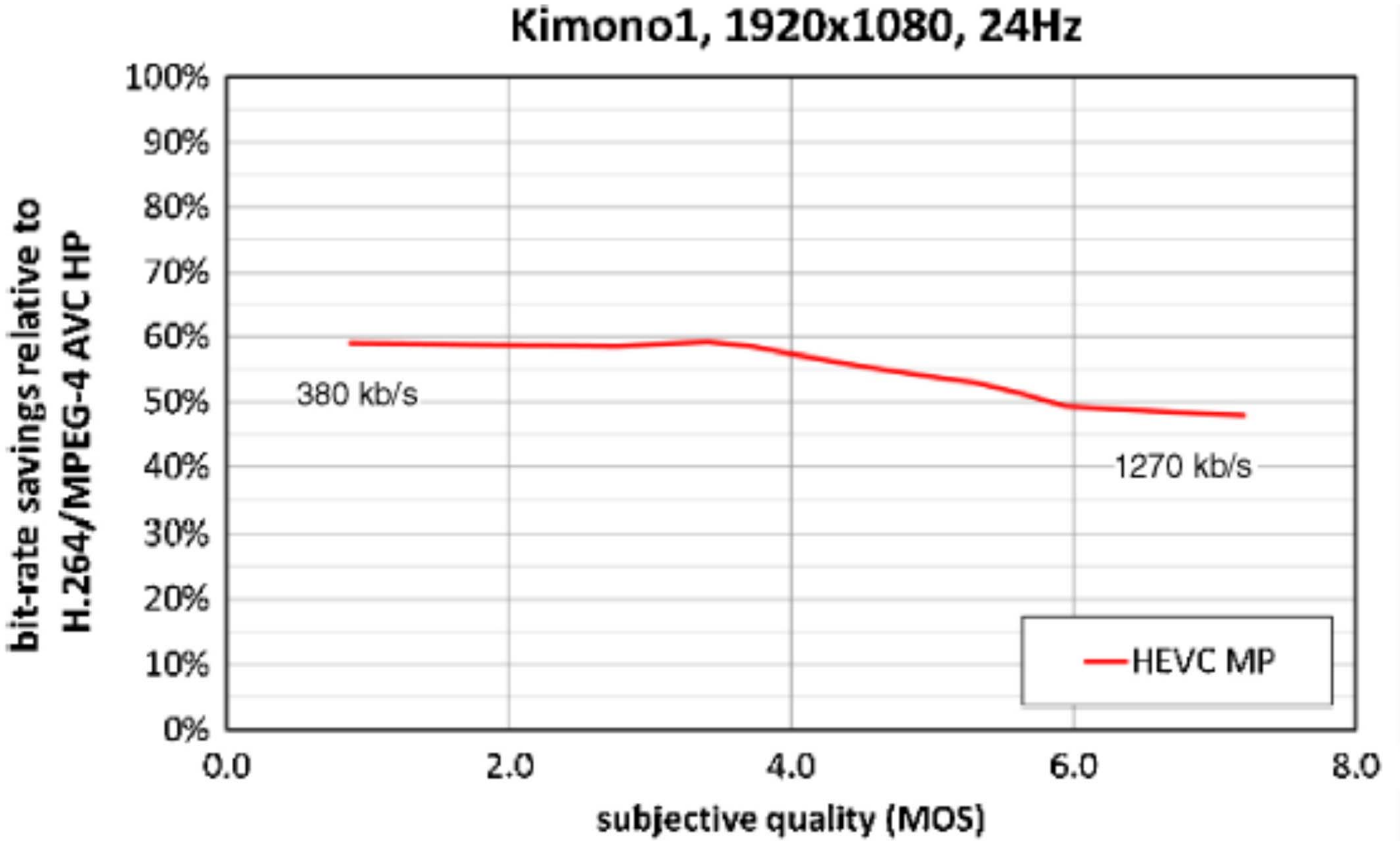
J. Ohm et al, "Comparison of the coding efficiency of video coding standards including High Efficiency Video Coding, IEEE Trans on CSVT, Vol 22, No 12, Dec 2013.

Perceptual Quality (HEVC vs AVC): HDTV Video 1080P



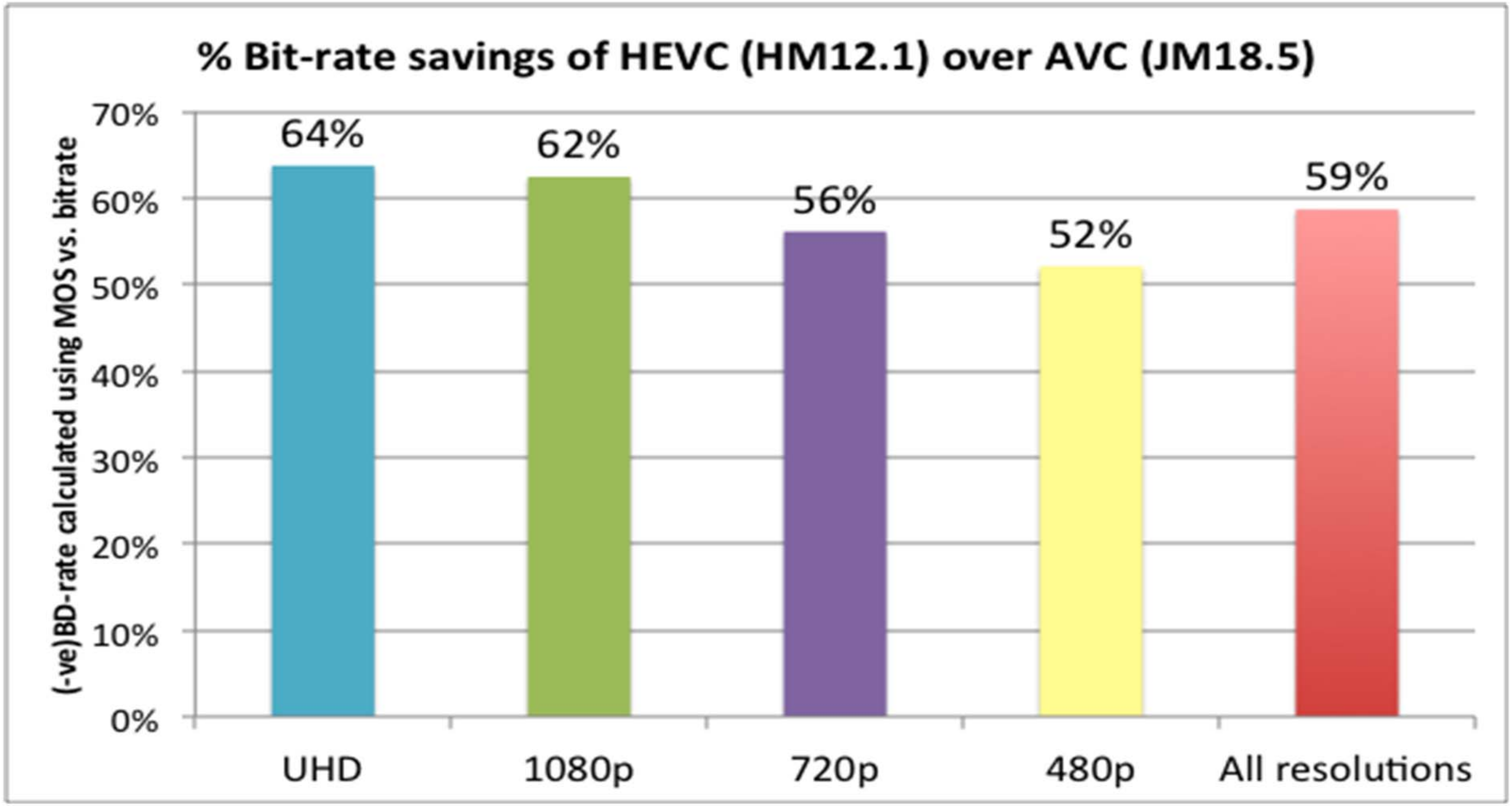
J. Ohm et al, “Comparison of the coding efficiency of video coding standards including High Efficiency Video Coding, IEEE Trans on CSVT, Vol 22, No 12, Dec 2013.

Bitrate Paving Perf. (HEVC vs AVC): HDTV Video 720P



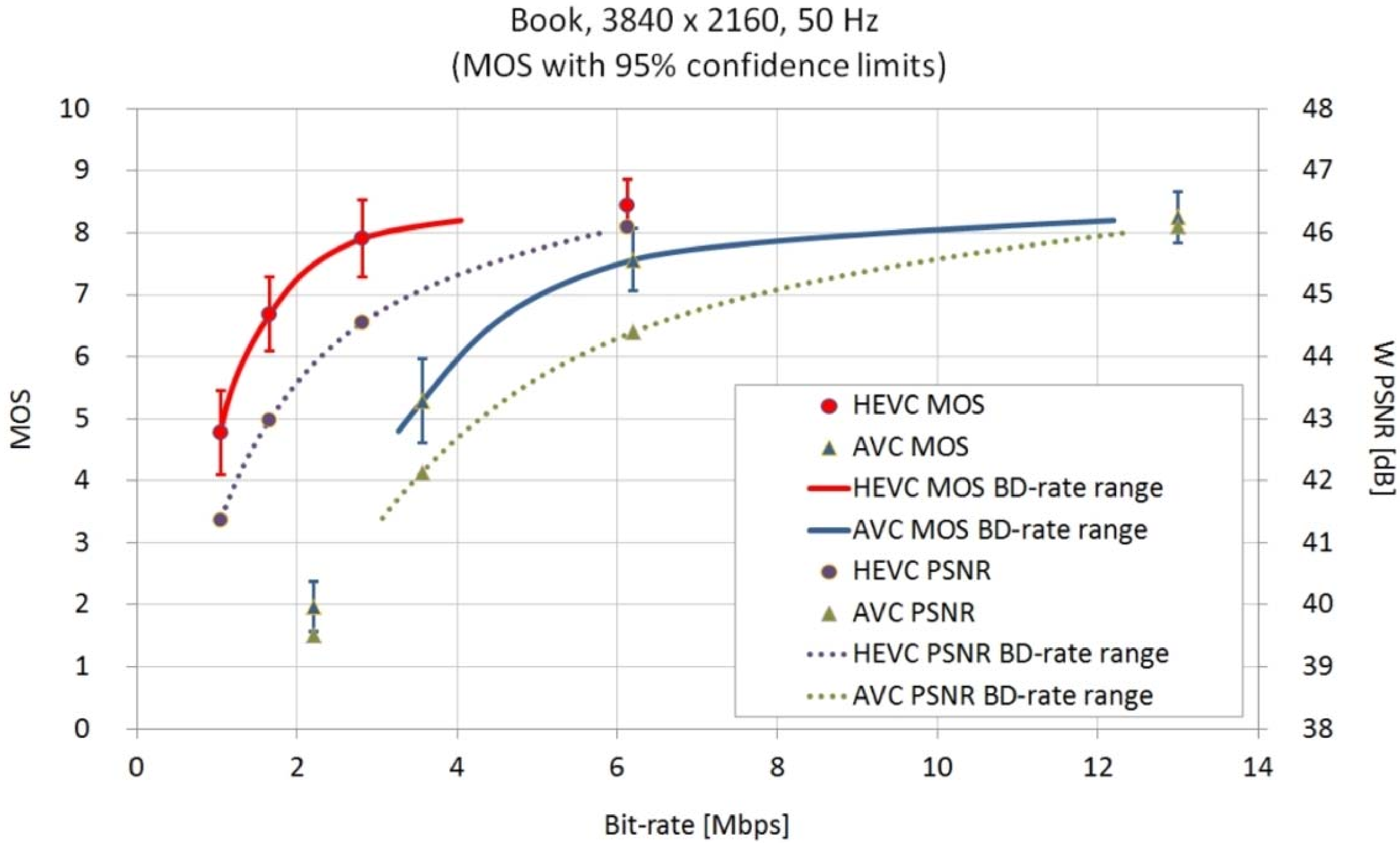
J. Ohm et al, “Comparison of the coding efficiency of video coding standards including High Efficiency Video Coding, IEEE Trans on CSVT, Vol 22, No 12, Dec 2013.

MPEG HEVC VS AVC

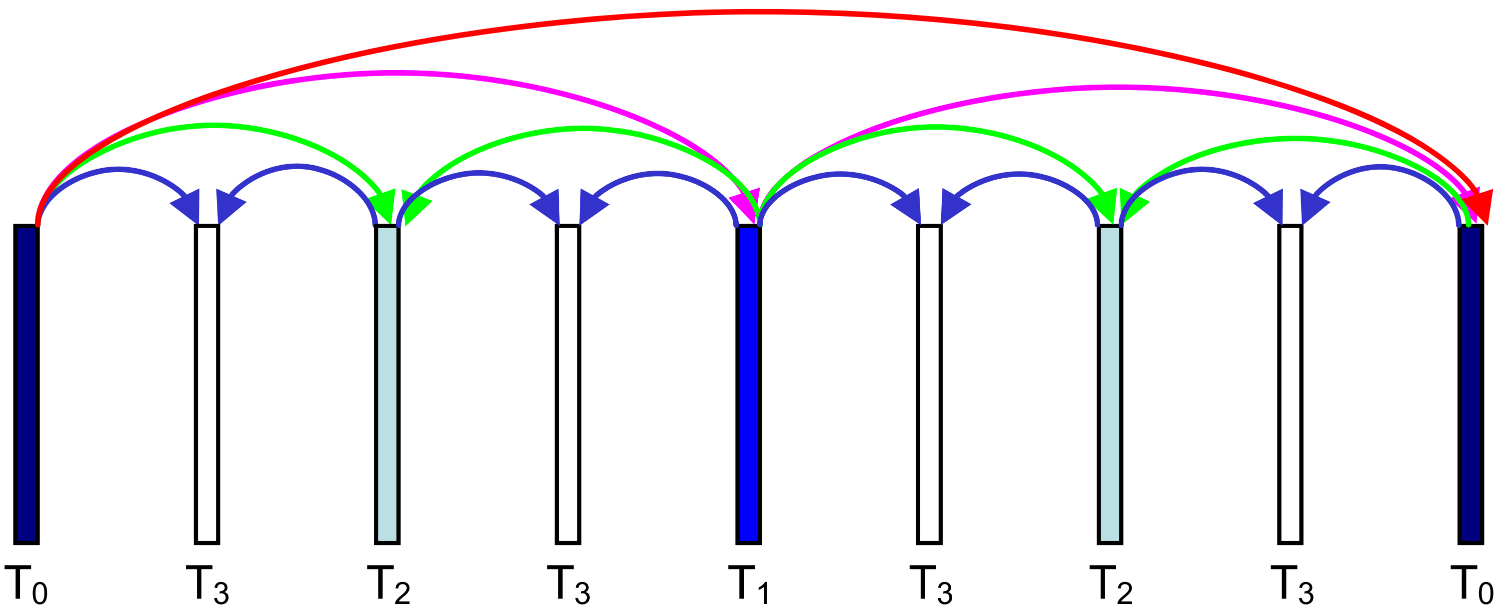


Bitrate saving for same perceived video quality of HEVC versus AVC/H.264. Source: document number m33340 JCTVC-Q0204r4 - "HEVC verification test results" presented at the April 2014 MPEG-JCTVC Valencia meeting.

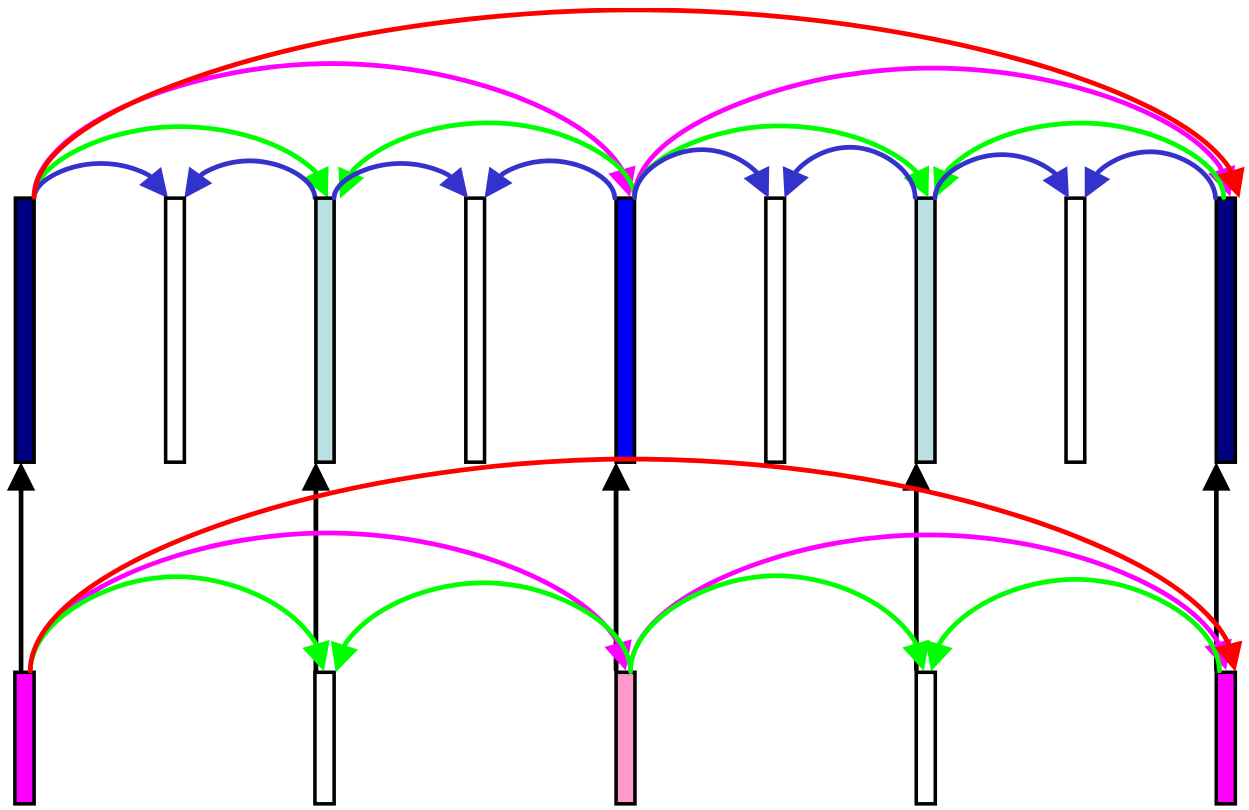
MPEG HEVC VS AVC



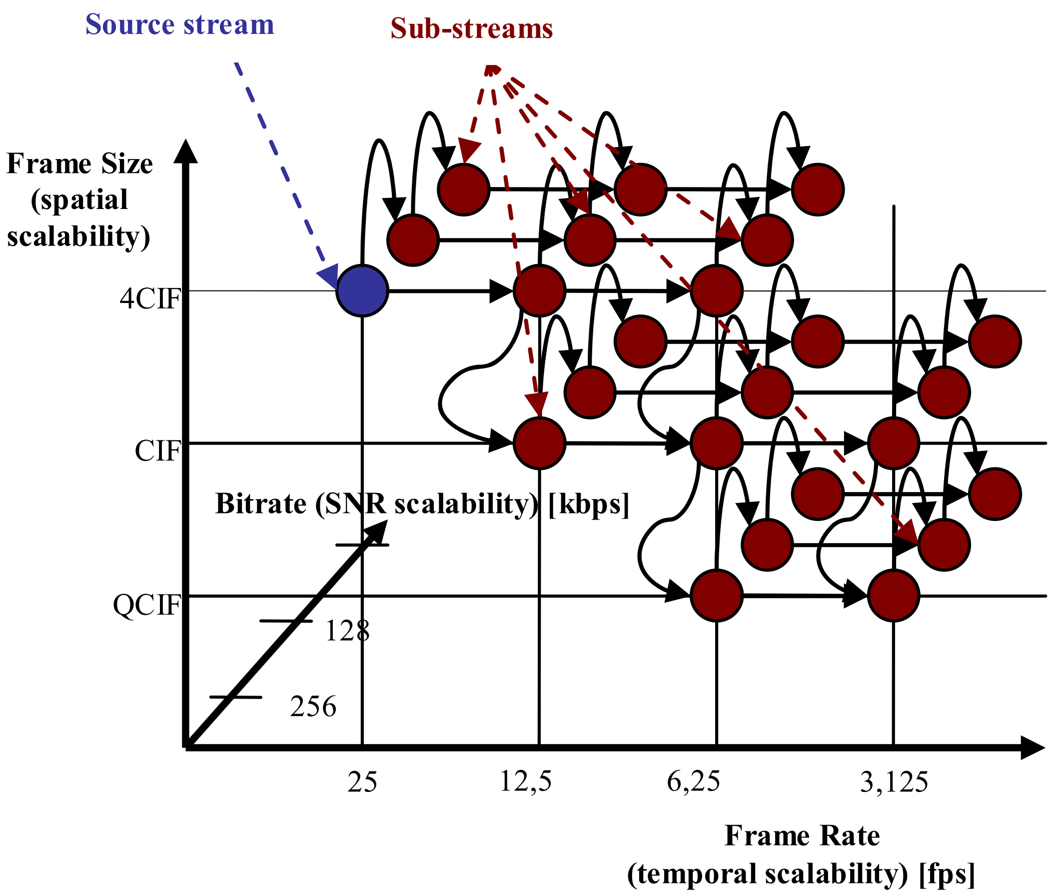
MPEG-4 Hierarchical Prediction



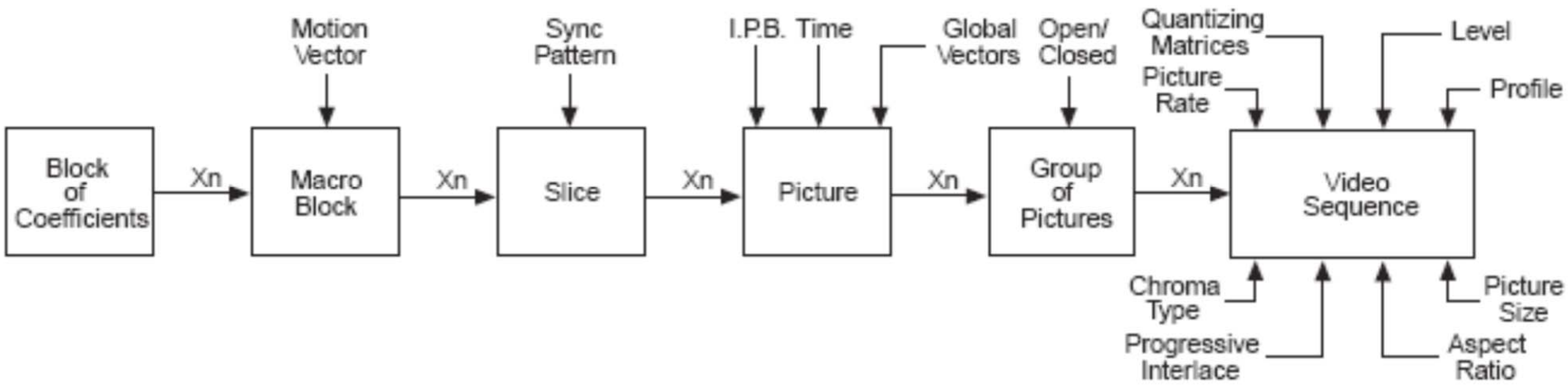
MPEG Spatial and Temporal Scalability



MPEG SCALABILITY



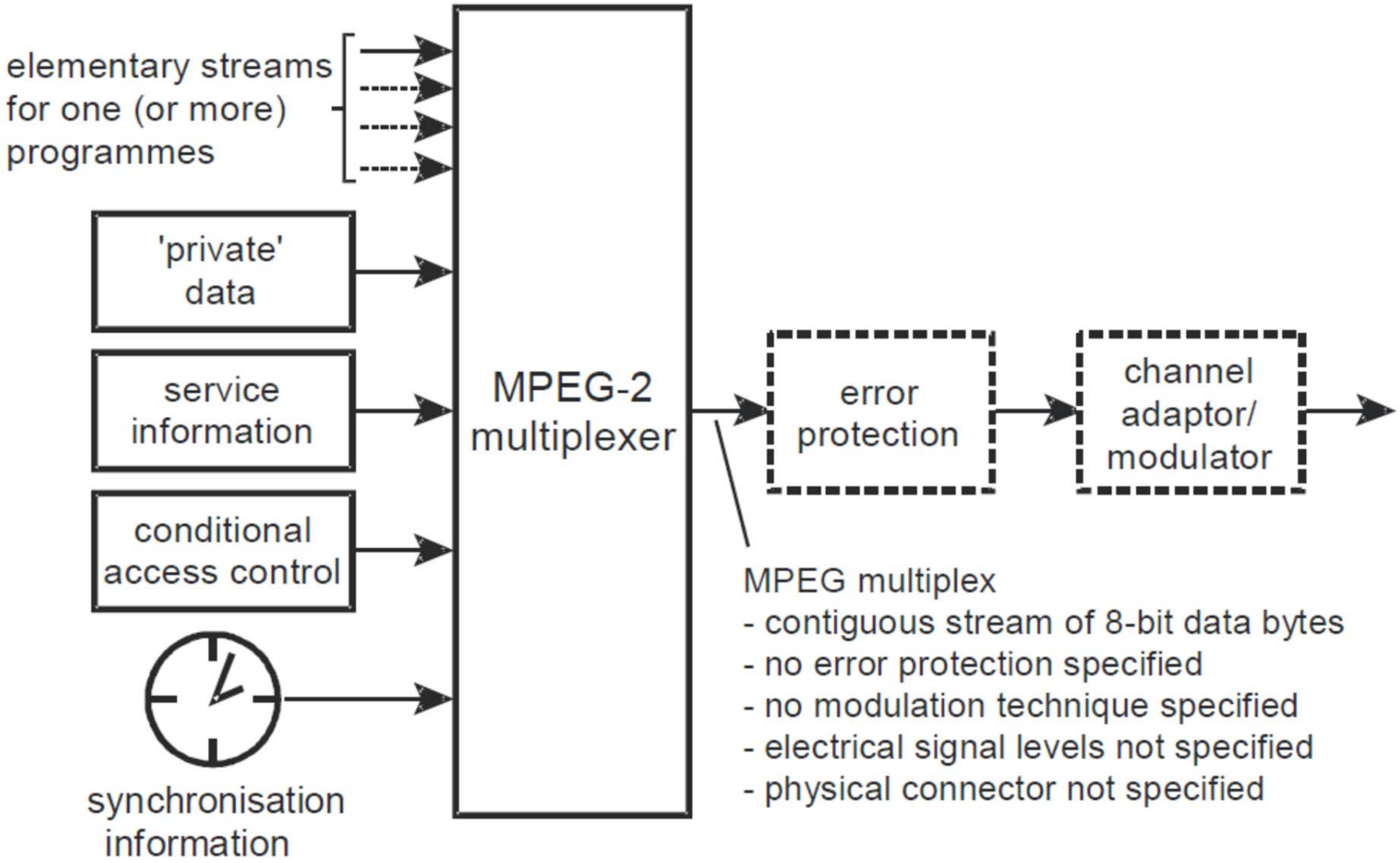
MPEG-2 Syntax Hierarchy



MPEG-2 Systems

- MPEG-2 Systems Objective:
  - How MPEG-compressed video and audio data streams may be multiplexed together with other data to form a single data stream suitable for digital transmission or storage.
- Three main elements:
  - The Multiplexes Structure (Elementary and Program Streams),
  - The service information that may be present;
  - The system of time stamps and clock references used to synchronize related components of a program at the decoder

MPEG-2 Systems Multiplexer



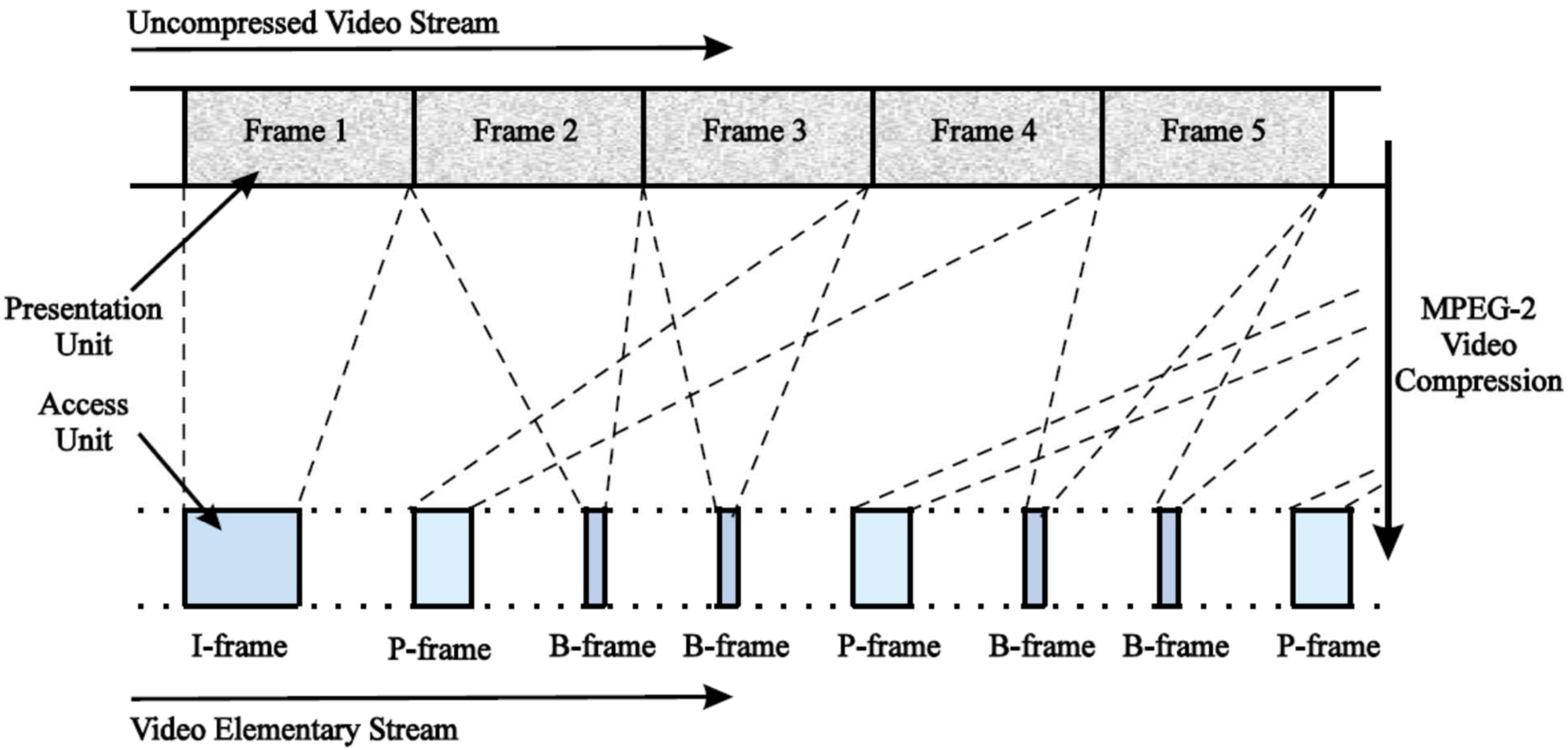
Glossary and essential components

- Program:
  - a single broadcast service or channel.
- Elementary Stream:
  - a program comprises one or more elementary streams. An elementary stream is a single MPEG-compressed component of a program (i.e. coded video or audio).
- The output of an MPEG-2 multiplexer:
  - a contiguous stream of 8-bit-wide data bytes. The multiplex may be of fixed or variable data rate and may contain fixed or variable data rate elementary streams.

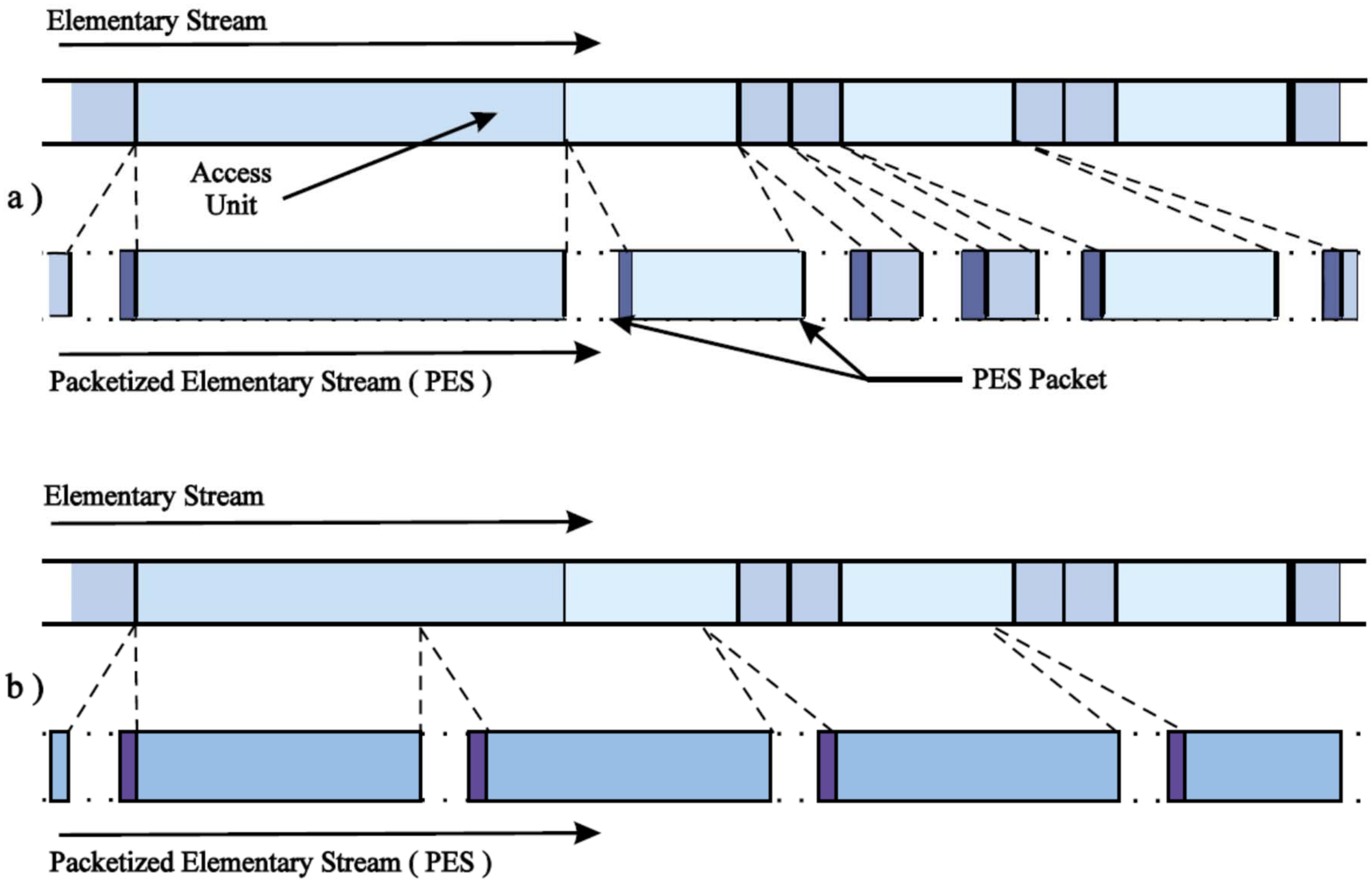
Transport Stream and Program Stream

- Transport Stream:
  - a multiplex devised for multi-programme applications so that a single transport stream can accommodate many independent programmes. It comprises a succession of 188-byte-long packets called transport packets.
- Program stream:
  - it can accommodate a single programme only, for storage and retrieval of programme material from digital storage media. Intended for use in error-free environments.

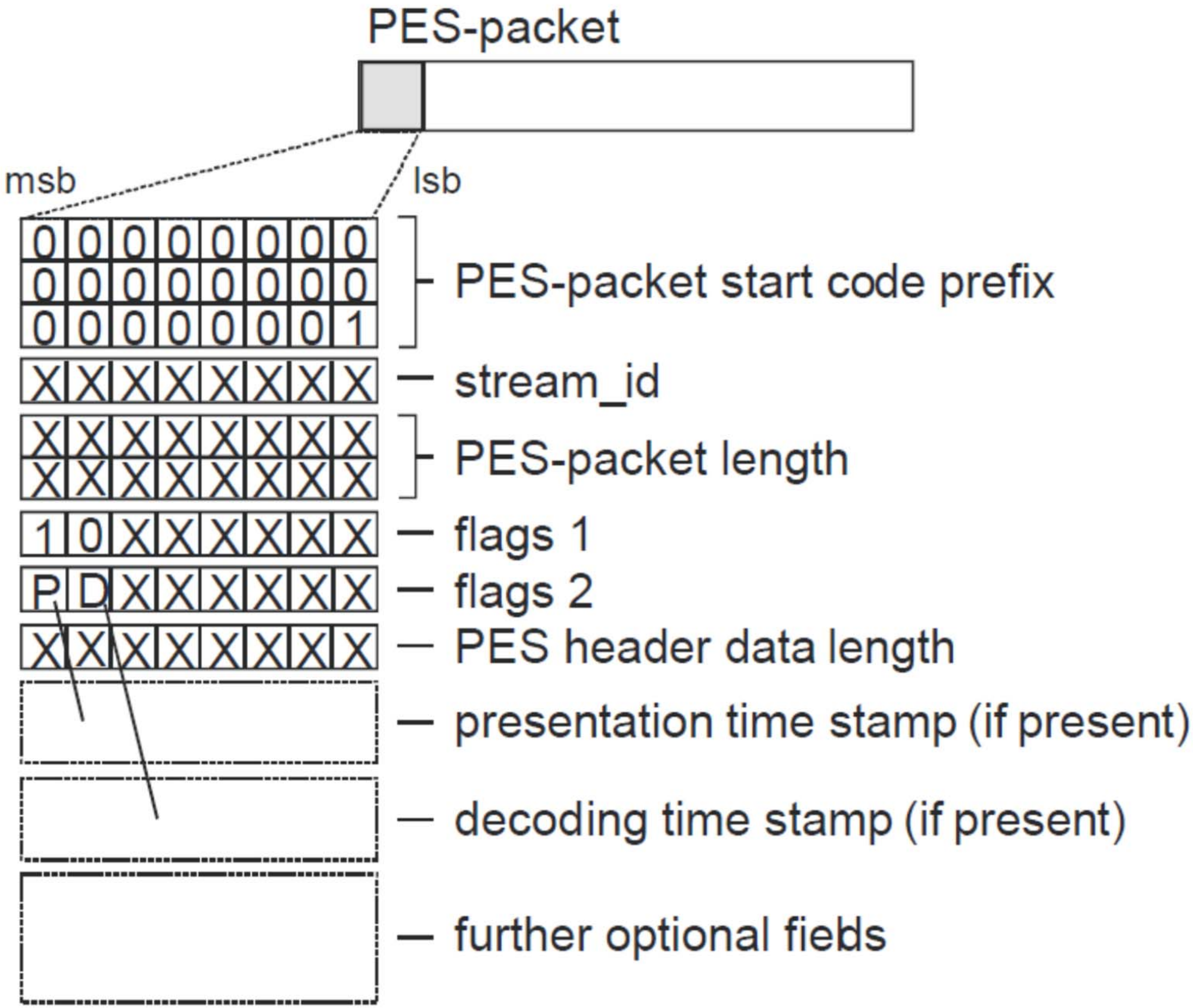
From Presentation Units to Access Units



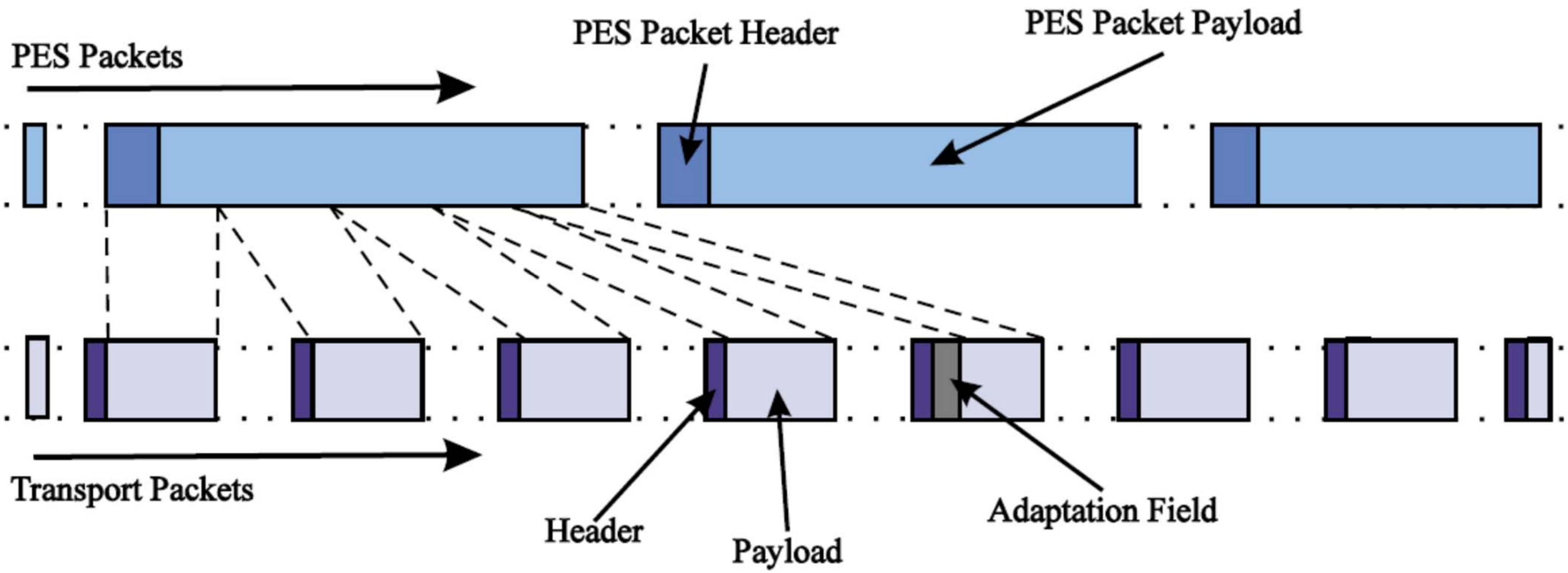
From Elementary Streams to PES



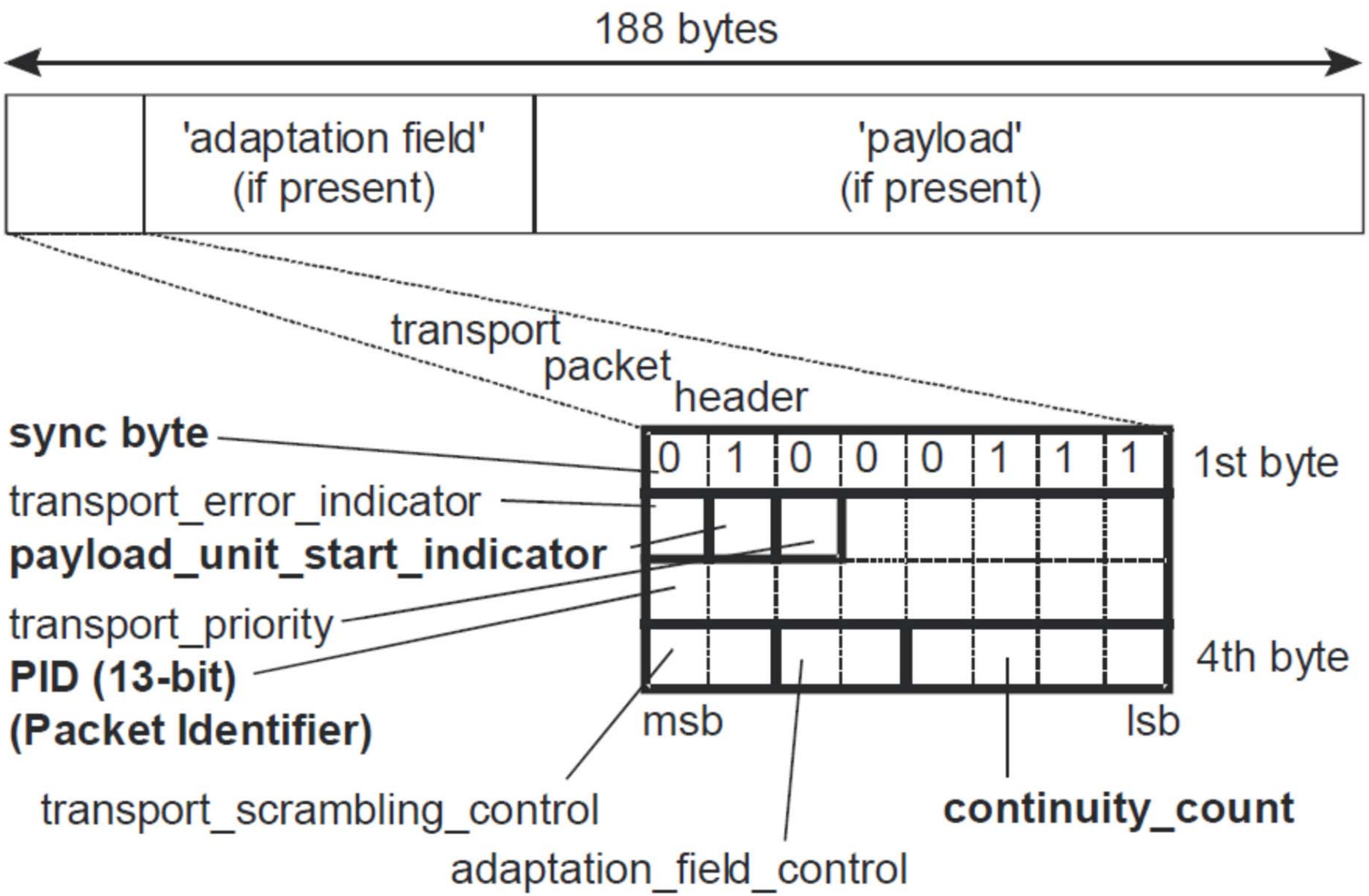
PES Header Information



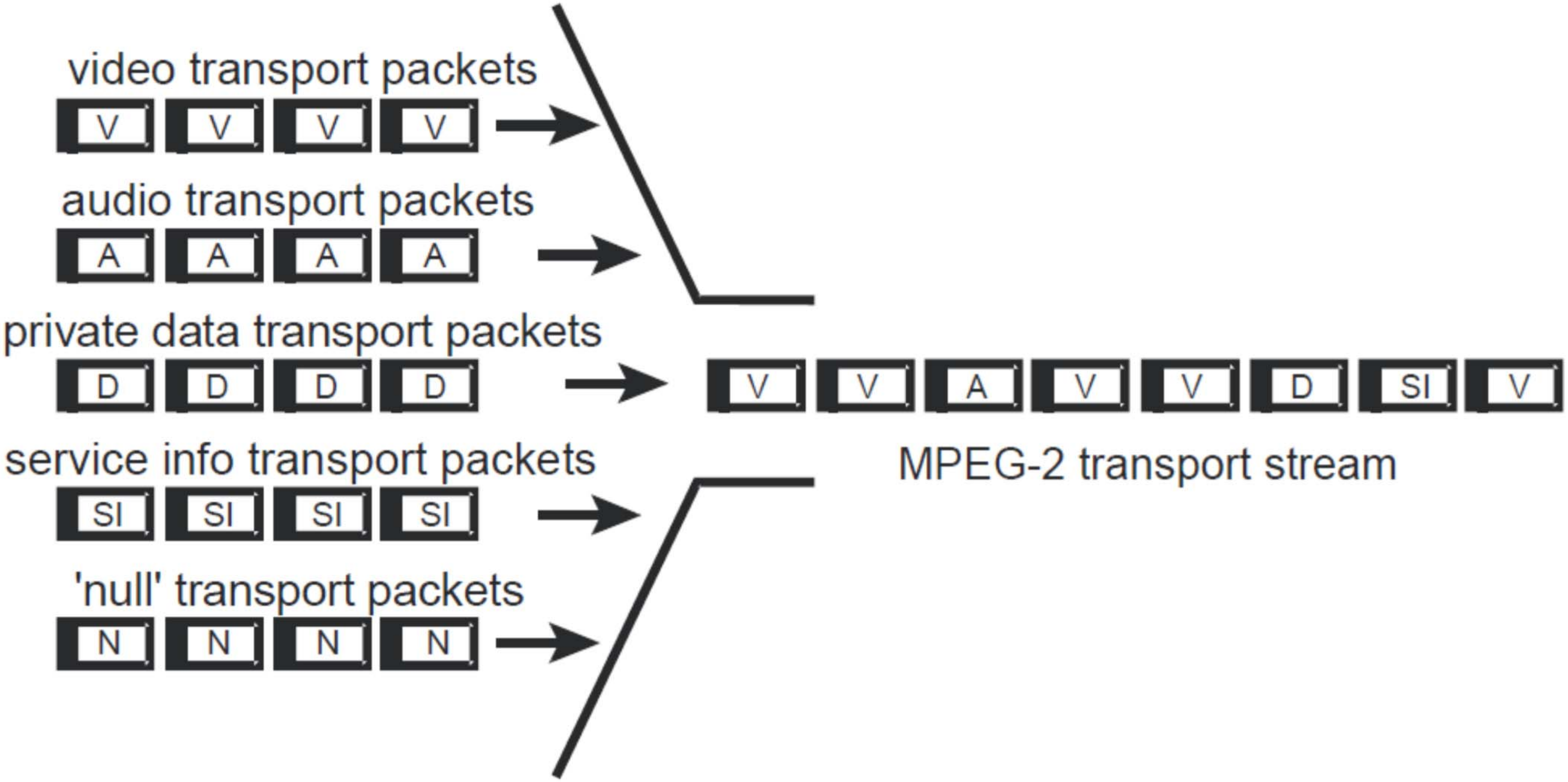
From PES to Transport Stream Packets (188 or 204 Bytes)

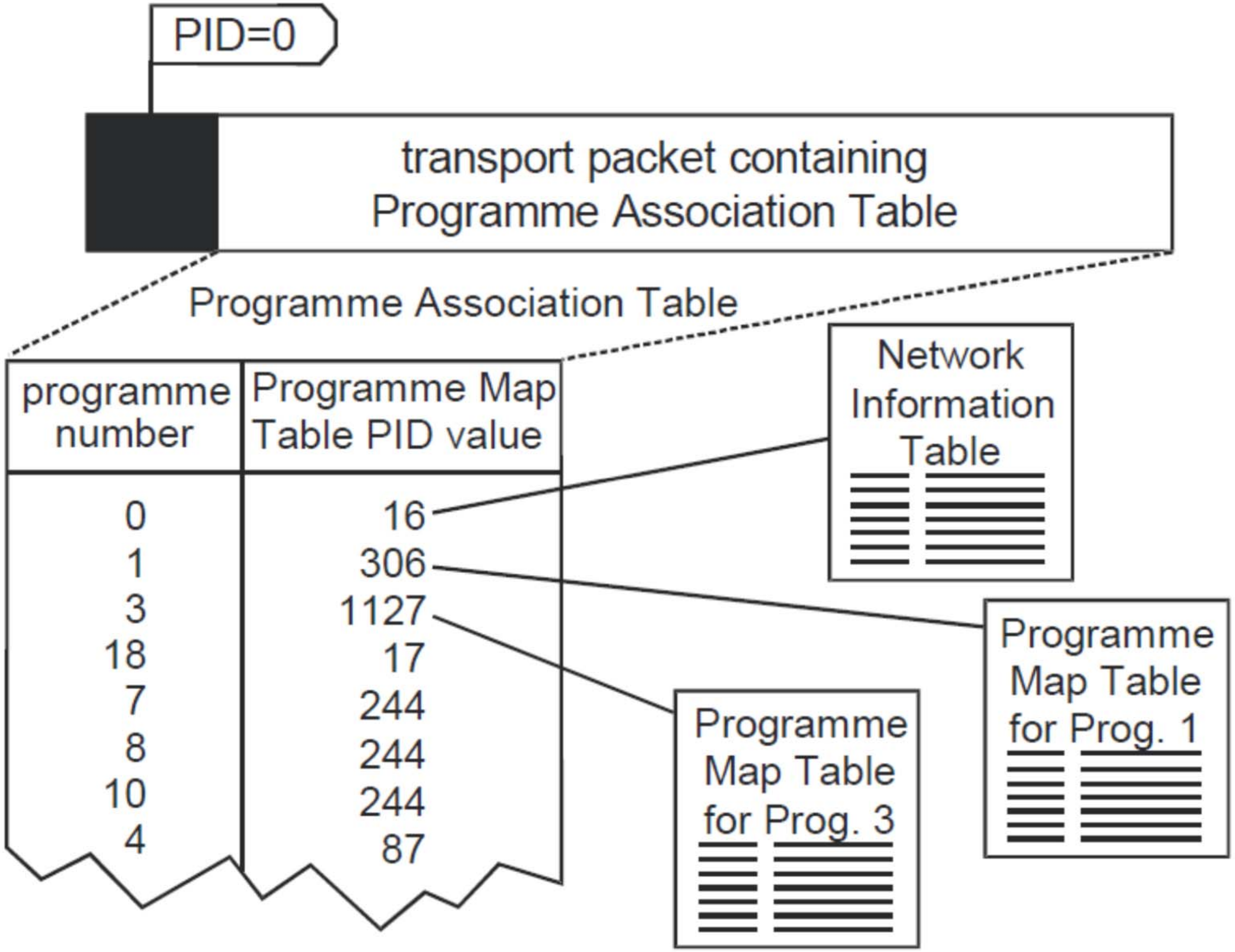


Transport stream Packets Header Information



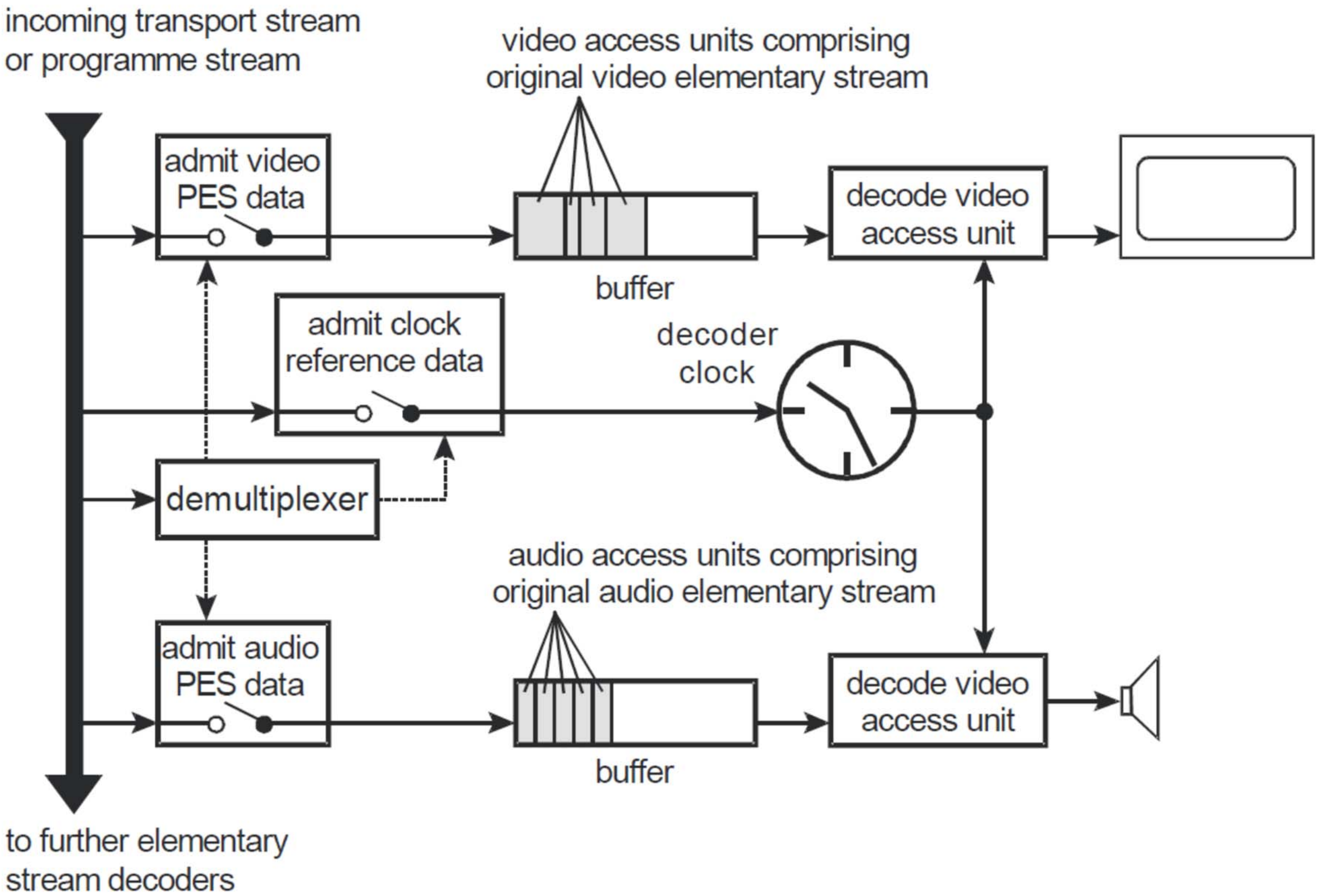
MPEG-2 Transport Stream Multiplexer



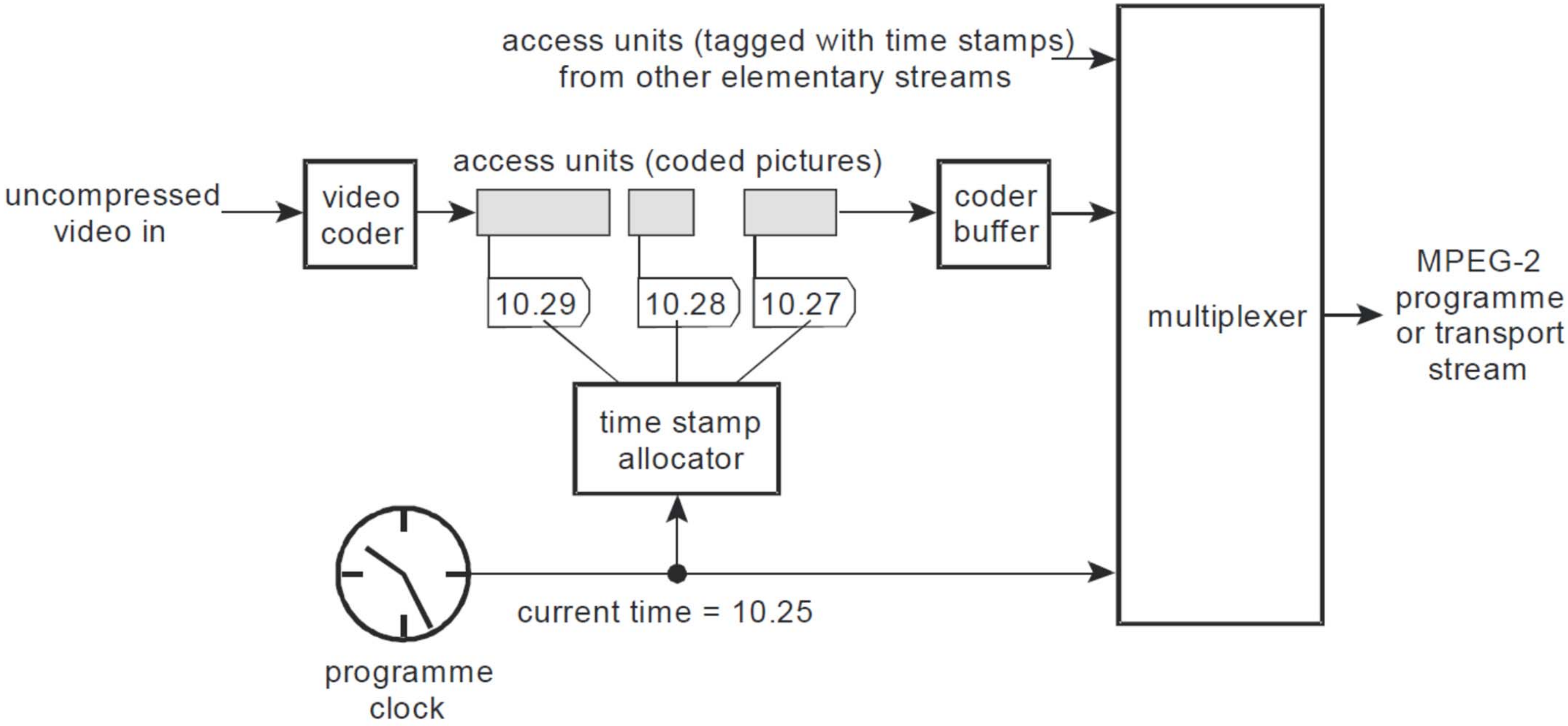


PID=0
transport packet containing
Programme Association Table
Programme Association Table
programme number
Programme Map Table PID value
0 16
1 306
3 1127
18 17
7 244
8 244
10 244
4 87
Network Information Table
Programme Map Table for Prog. 1
Programme Map Table for Prog. 3

Clock Reference Data for Synchronization of ES



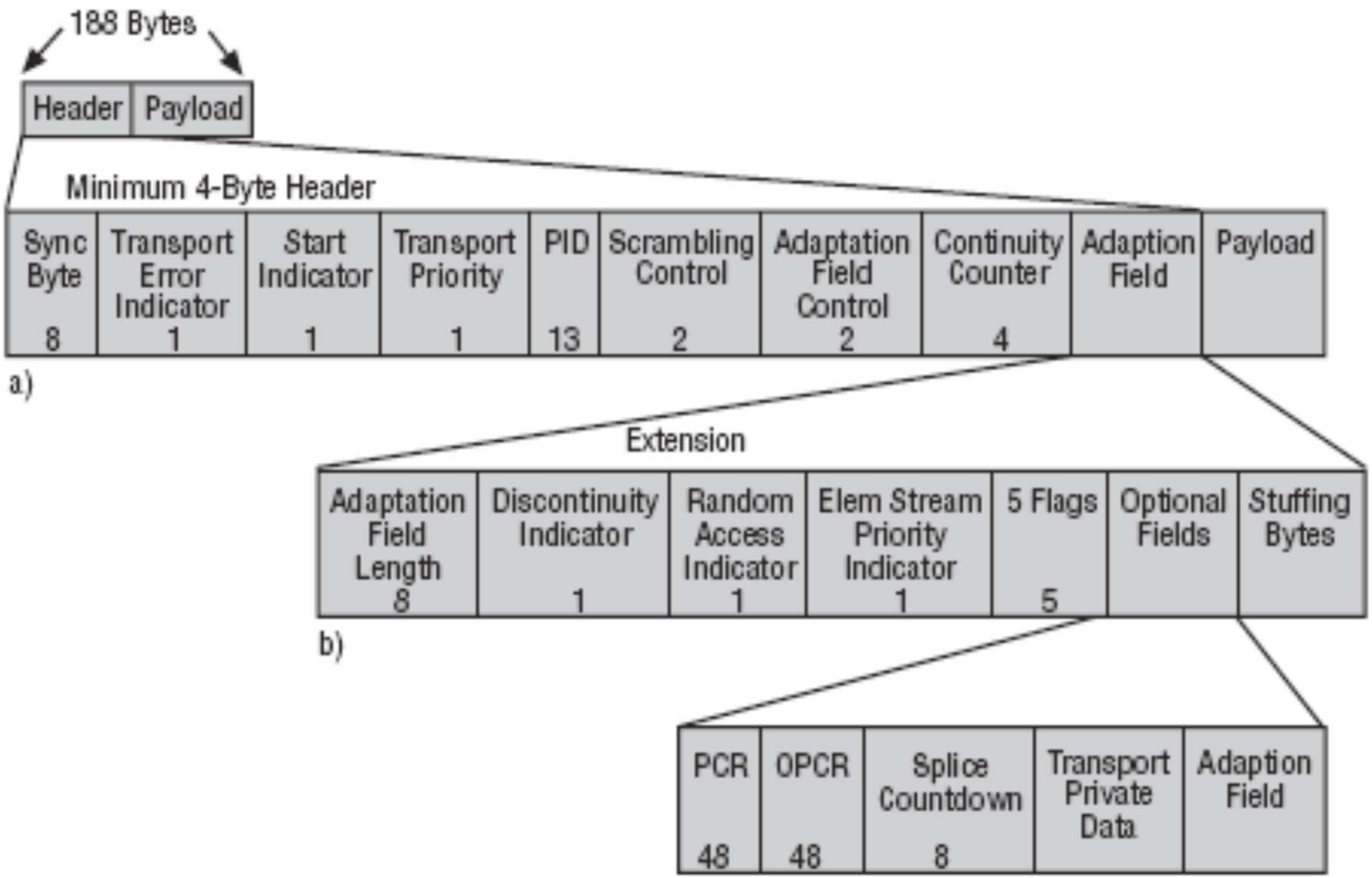
Clock Reference Insertion



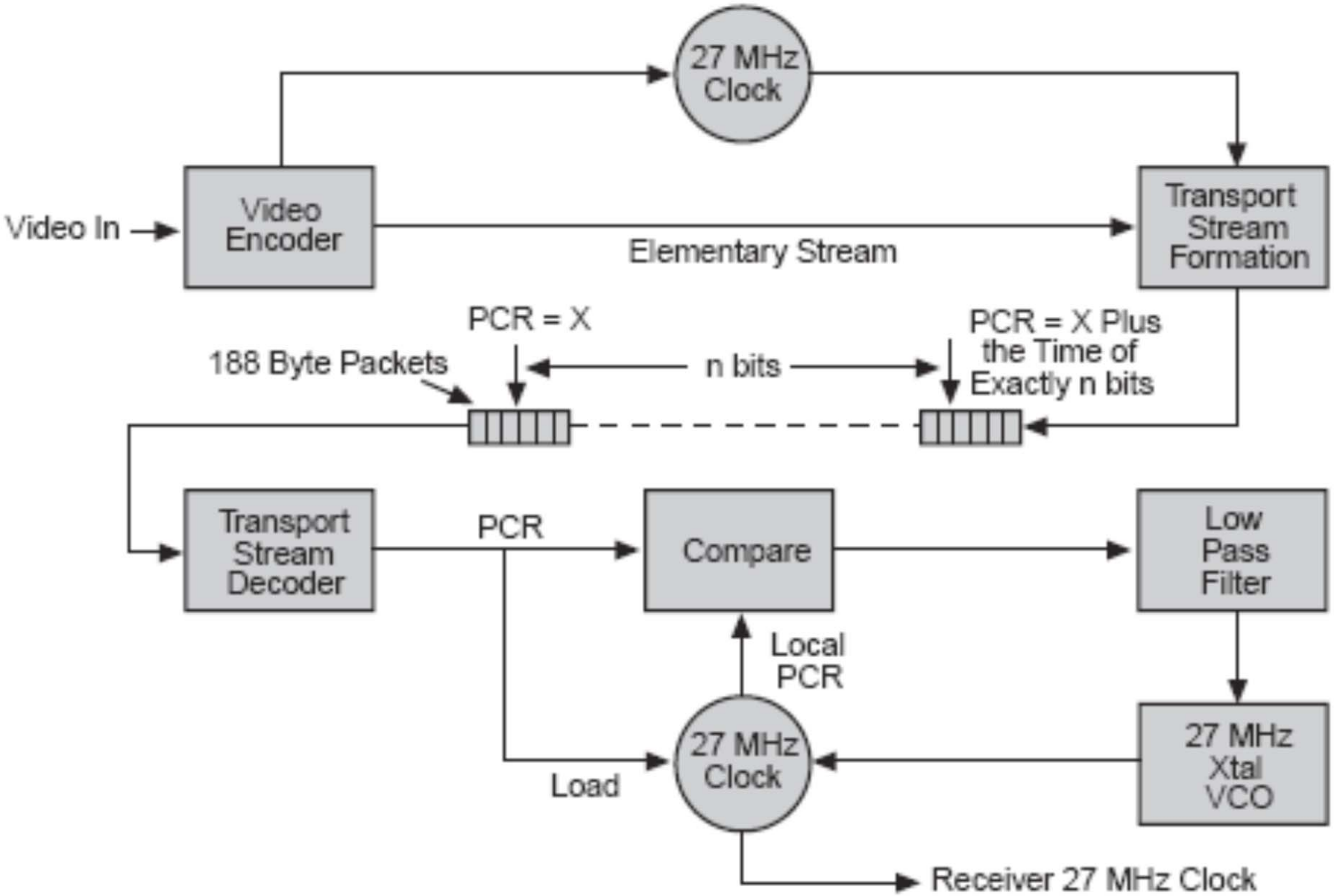
TS Functionality

- MPEG-2 Transport Stream
  - Flexible Multiplexer for multichannel transmission and storage
  - Configurable for different application requirements
  - Several functionality supported:
    - Stream synchronization
    - User information on top of video-audio ES
    - User information streams
  - Stable and well established specification (emulation code are avoided)

MPEG-2 Transport Stream Packet



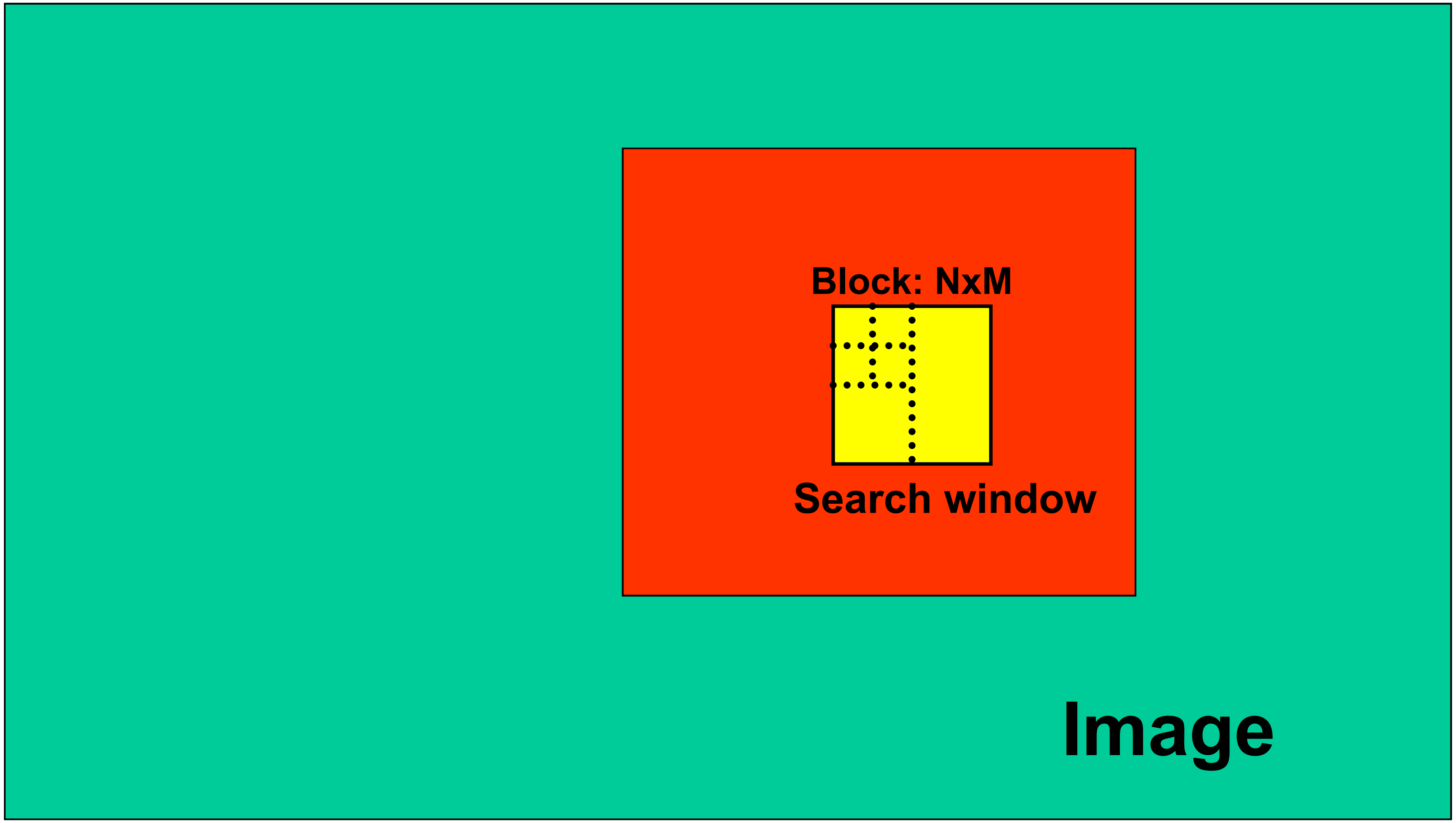
Program Clock Reference



Motion estimation in Video Sequences

- Three main families:
  - Gradient based:
    - Based on iterative minimization of DFD
    - One specific vector for each pixel (dense vector field)
  - Frequency based:
    - Fourier transform
    - Correlation of phase difference
  - Block correspondence (Block Matching):
    - Minimization of block error
    - Sum of Absolute Difference (SAD) in a search window

$$SAD(d) = \sum_{x \in B} \left| g_t(x) - g_{(t-\tau)}(x + d_{t,\tau}(x)) \right|$$



Gradient Based

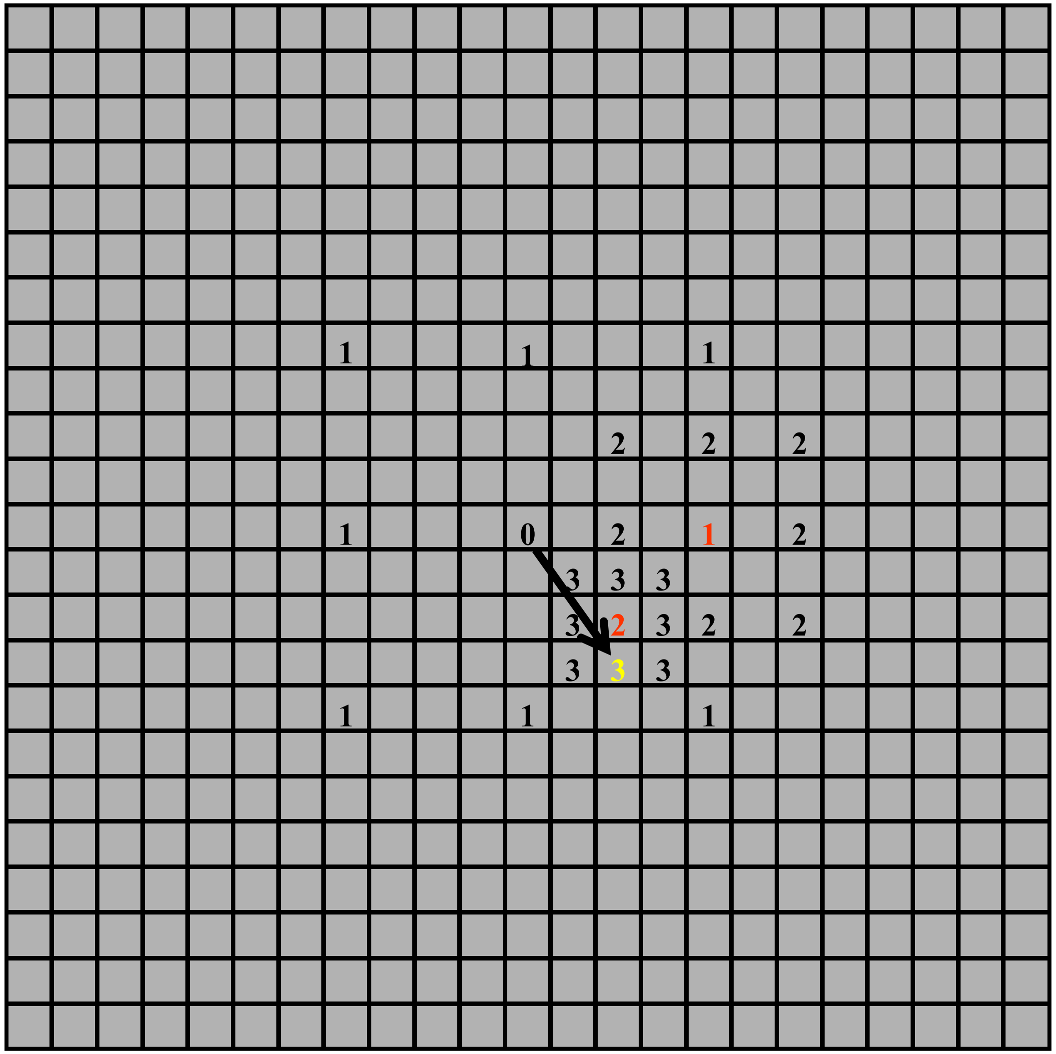
$$DFD(x,d) = I_{n-1}(x+d) - I_n(x)$$

- Minimisation of the Displaced Frame Difference:
  - Taylor expansion:
    - No exact solution
    - Iterative estimation that minimize of DFD
  - Drawbacks
    - Sensible to noise, non-linearities (occlusions)
    - Cannot detect large displacements (local minima)
    - Larger neighbourhoods (matrix inversions)
    - Multiresolutions can improve results

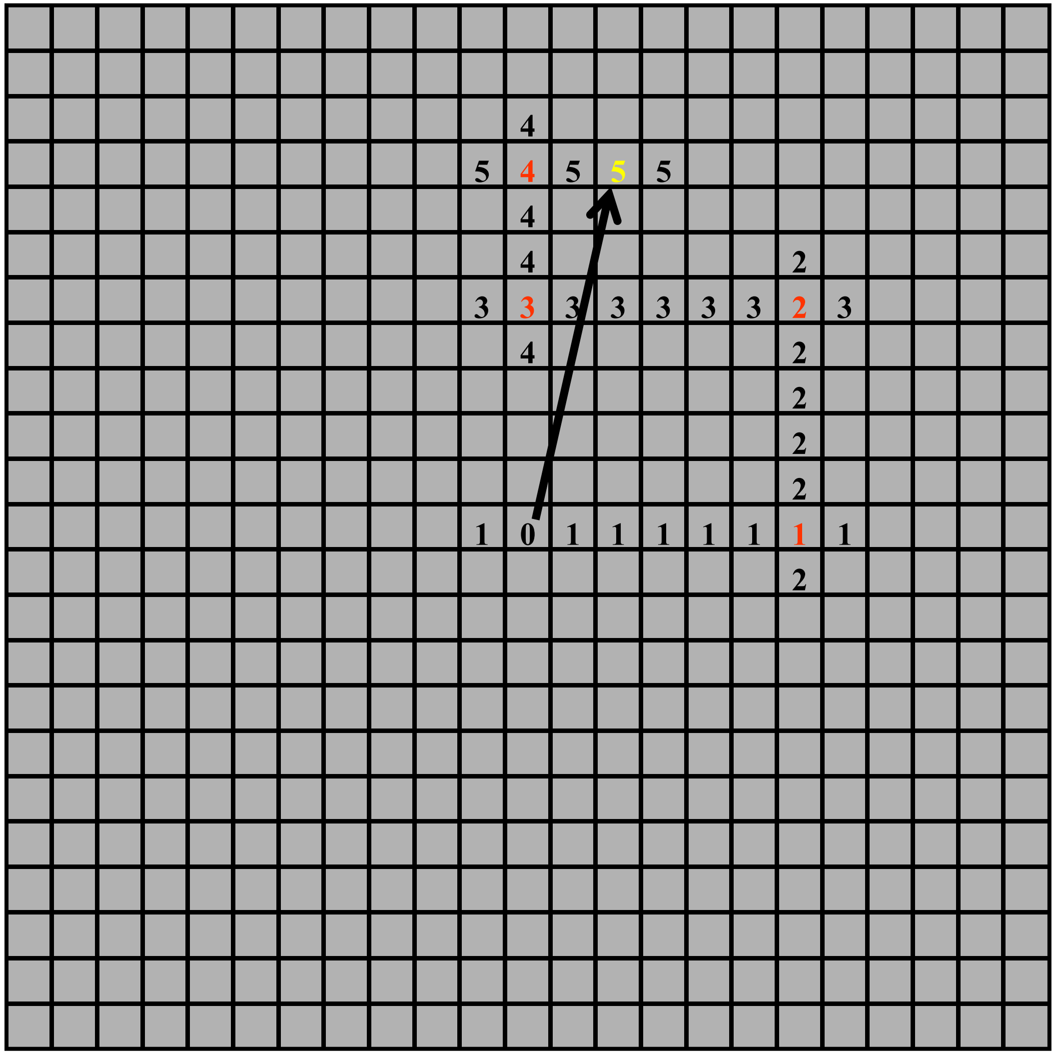
Frequency Based

- Find the peaks of Phase Correlation:
  - Peaks corresponds to displacements:
    - Robust to noise
    - Accurate displacements
    - Pixels that correspond to displacement are not known
  - Drawbacks
    - Good only to detect vector candidates
    - Region partitions and correlations need to be applied to find correspondences
    - Very complex processing

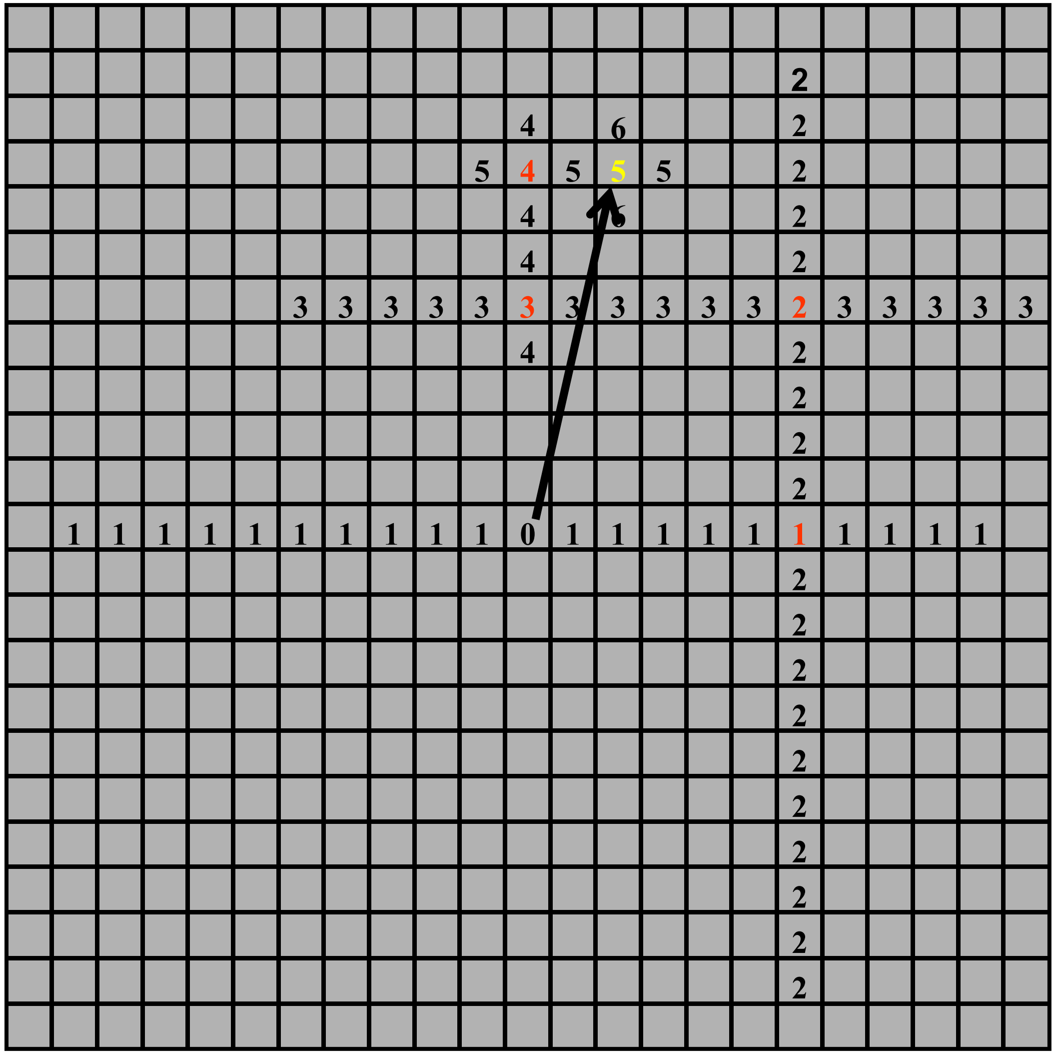
Block Matching: N-Step Search



Block Matching: Gradient Search

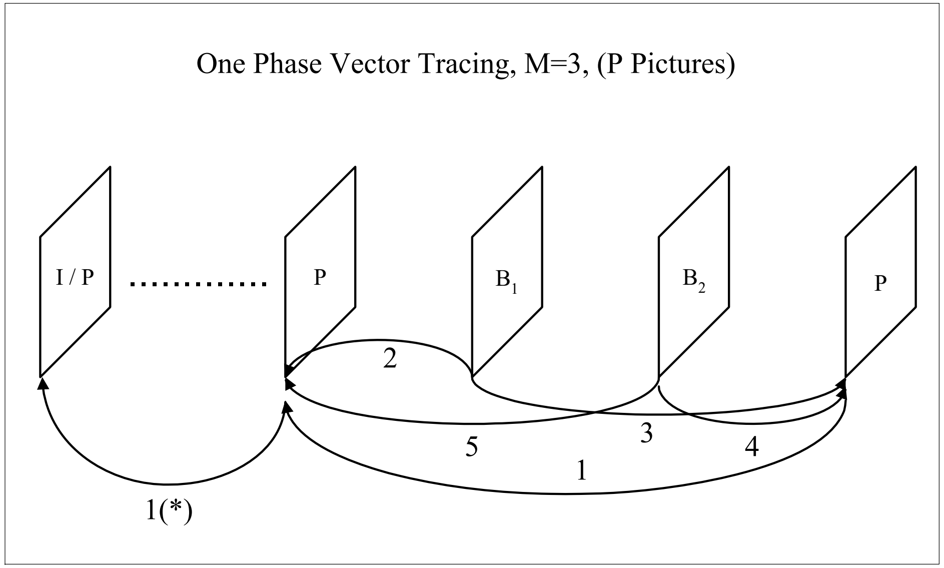


Block Matching: Gradient Search Variants



Vector tracing schemes (I)

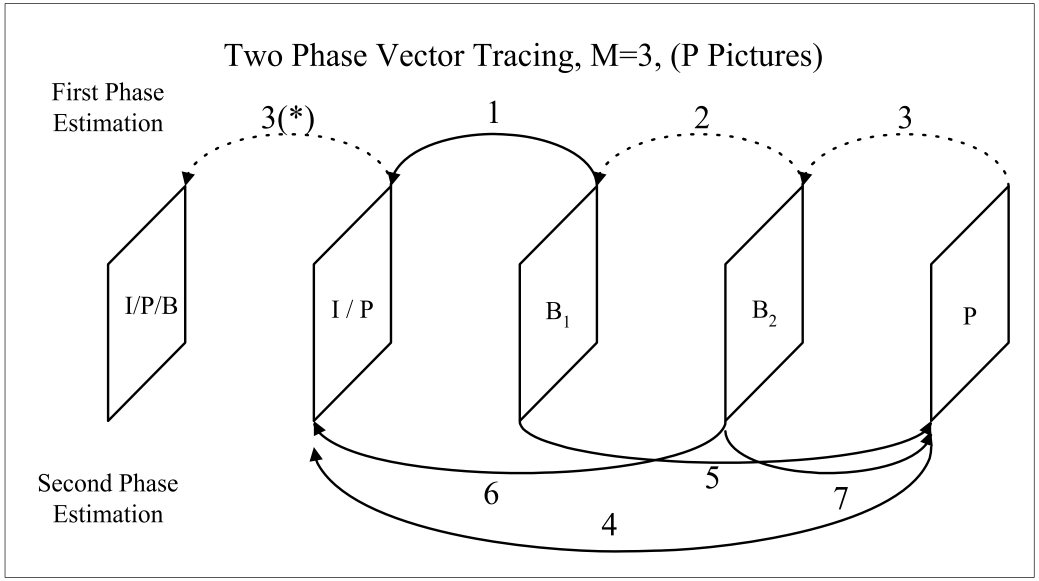
$$\left\{V_{1tr}\left(\bar{x},\bar{y},\bar{t}\right)\right\}=\left(\bigcup_{x=\bar{x}-\alpha}^{\bar{x}+\alpha}\bigcup_{y=\bar{y}-\beta}^{\bar{y}+\beta}\frac{j\left(\bar{t}\right)}{M}\left(V_{1tr}\left(x,y,\bar{t}+T\right)\right)\right)\bigcup\left(\bigcup_{\gamma,\delta\in E}V_{1tr}\left(\bar{x}-\gamma,\bar{y}-\delta,\bar{t}\right)\right)$$



One-phase vector tracing

Vector tracing schemes (II)

$$\left\{V_{2tr}(\bar{x},\bar{y},\bar{t})\right\} = \left(\bigcup_{x=\bar{x}-\alpha}^{\bar{x}+\alpha}\ \bigcup_{y=\bar{y}-\beta}^{\bar{y}+\beta} g_{2tr}\right) \bigcup \left(\bigcup_{\gamma,\delta\in E} V_{2tr}(\bar{x}-\gamma,\bar{y}-\delta,\bar{t})\right)$$



Two-phase vector tracing

VT integration in a genetic algorithm

- The genetic structure is a suitable framework to efficiently exploit the tracing information: best estimates from previous frames are inserted directly in the first population, best estimates from neighbor macrblocks are used to bias the random generation of the rest of the first population
- Exploitation of spatial & temporal correlation of motion vector fields
- Example: 4 “Generations”, “Population” of 20 MVs, 10 “Sons” per generation Þ 80 Matchings per MB

Further algorithm effects:

- delivers smooth motion vector fields
- good resistance against noise
- able to track fast motion

⇙ Delivers very high quality with a very low number of matchings/MB

An optimal MGS algorithm

One-phase

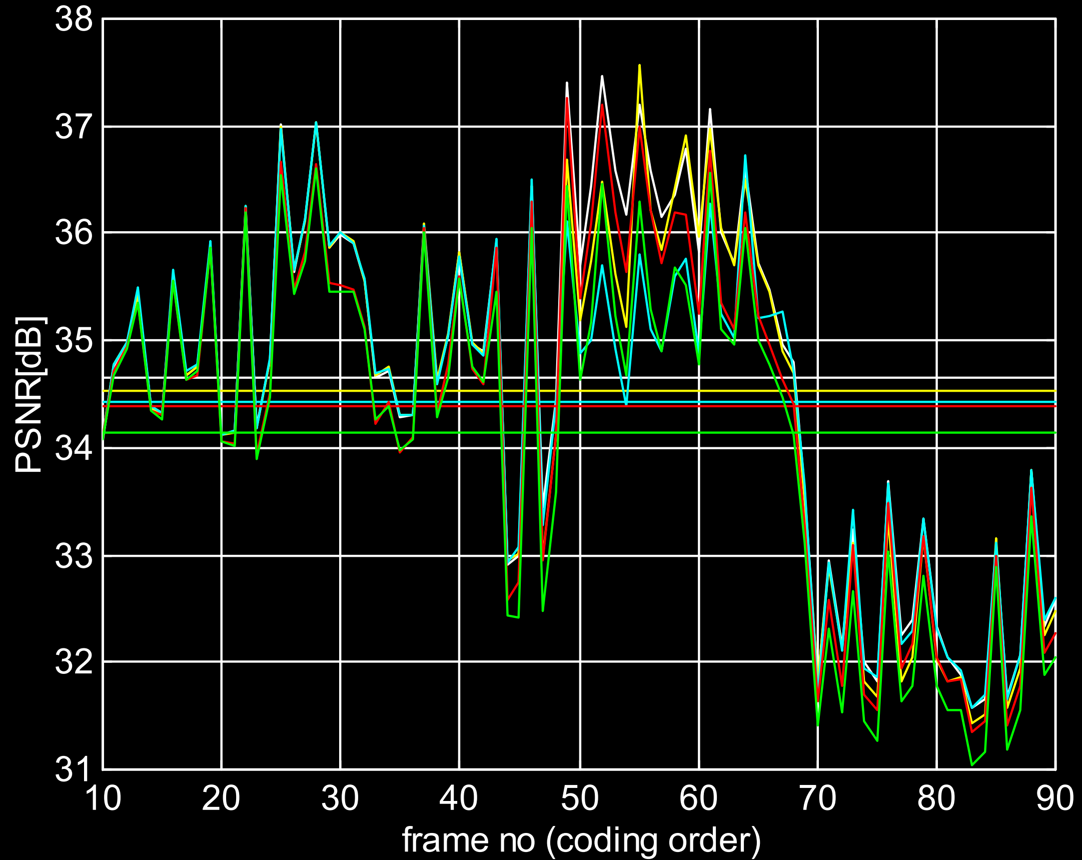
- 3 generations of 27 elements
- 9 vectors in the first population come from previous frames
- 18 vectors are generated through biased random generation around neighbors
- In successive populations: 20 vectors are the mean of the two parents, 7 vectors are random around the partial best

Two-phase

- The total amount of matchings per GOP is the same -> same number of operations
- The memory bandwidth is increased, since more estimation phases are necessary (60 % for M=3)

Algorithm Performance

■ Reference simulation: full search on ± 48 H ± 32 V
mean PSNR: 34.65 dB



FS ± 16 H ± 16 V
complex telescopic
mean PSNR = 34.54 dB

■ **FS ± 32 H ± 24 V**
mean PSNR = 34.42 dB

■ **2-phase HVTMGS**
on ± 48 H ± 32 V :
mean PSNR = 34.38 dB

■ **1-phase HVTMGS**
on ± 48 H ± 32 V :
mean PSNR = 34.14 dB

Simulation settings:

- **basketball**
  frame 10-90
  frame picture coding
- **704*576 pixel**
- **9 Mbit/s**
- **N=12, M=3**
- **frame&field prediction**
- **no SW offset**
- **All with same half pel**

Complexity reduction factor

$$r = \frac{H * V}{p \times G}$$

- H*V is the search window size (97*65)
- p is the population size (27)
- G is the number of generations (3)

$$r \cong 78$$