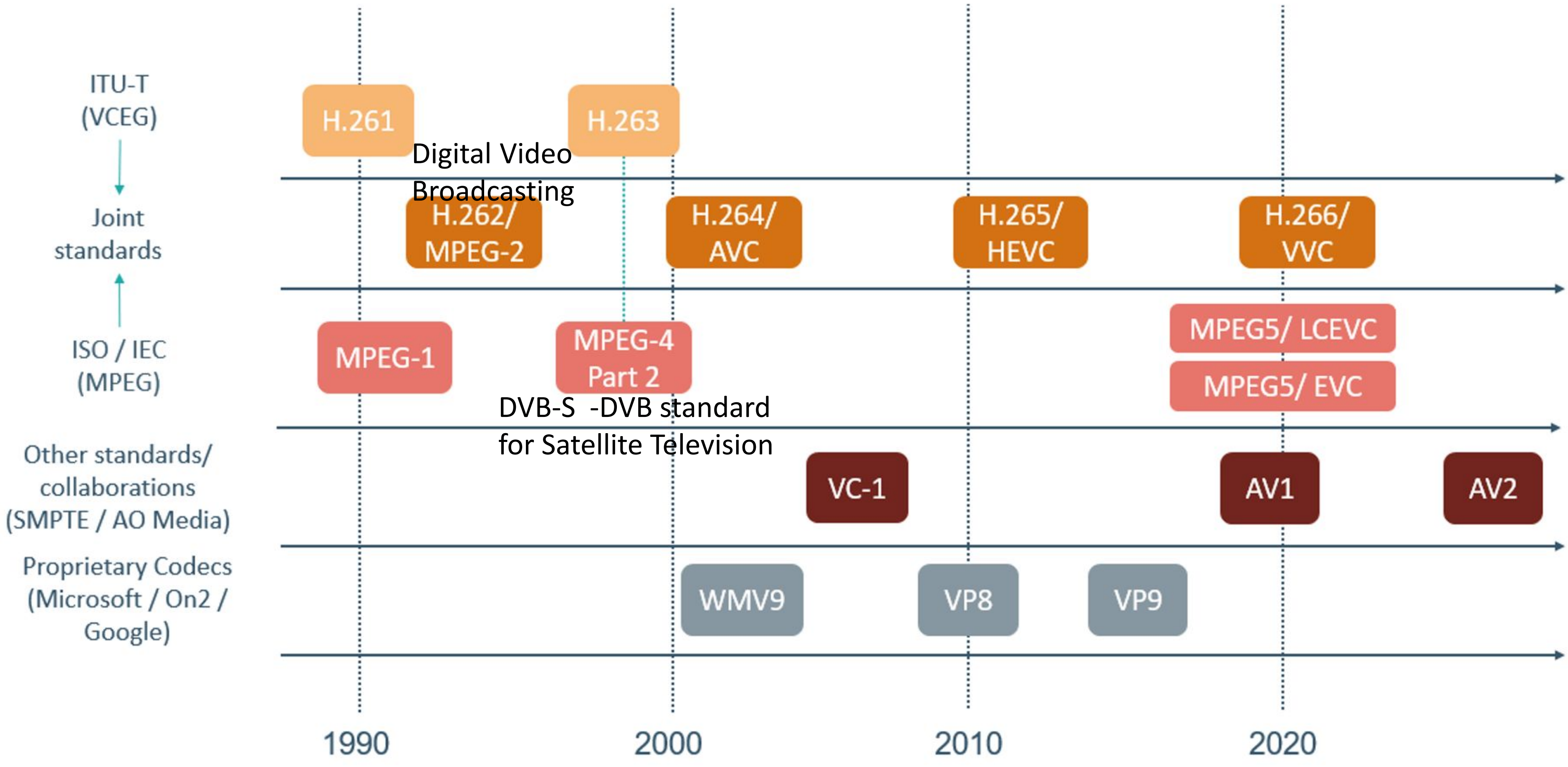
Spracovanie obrazu, grafika a multimédiá

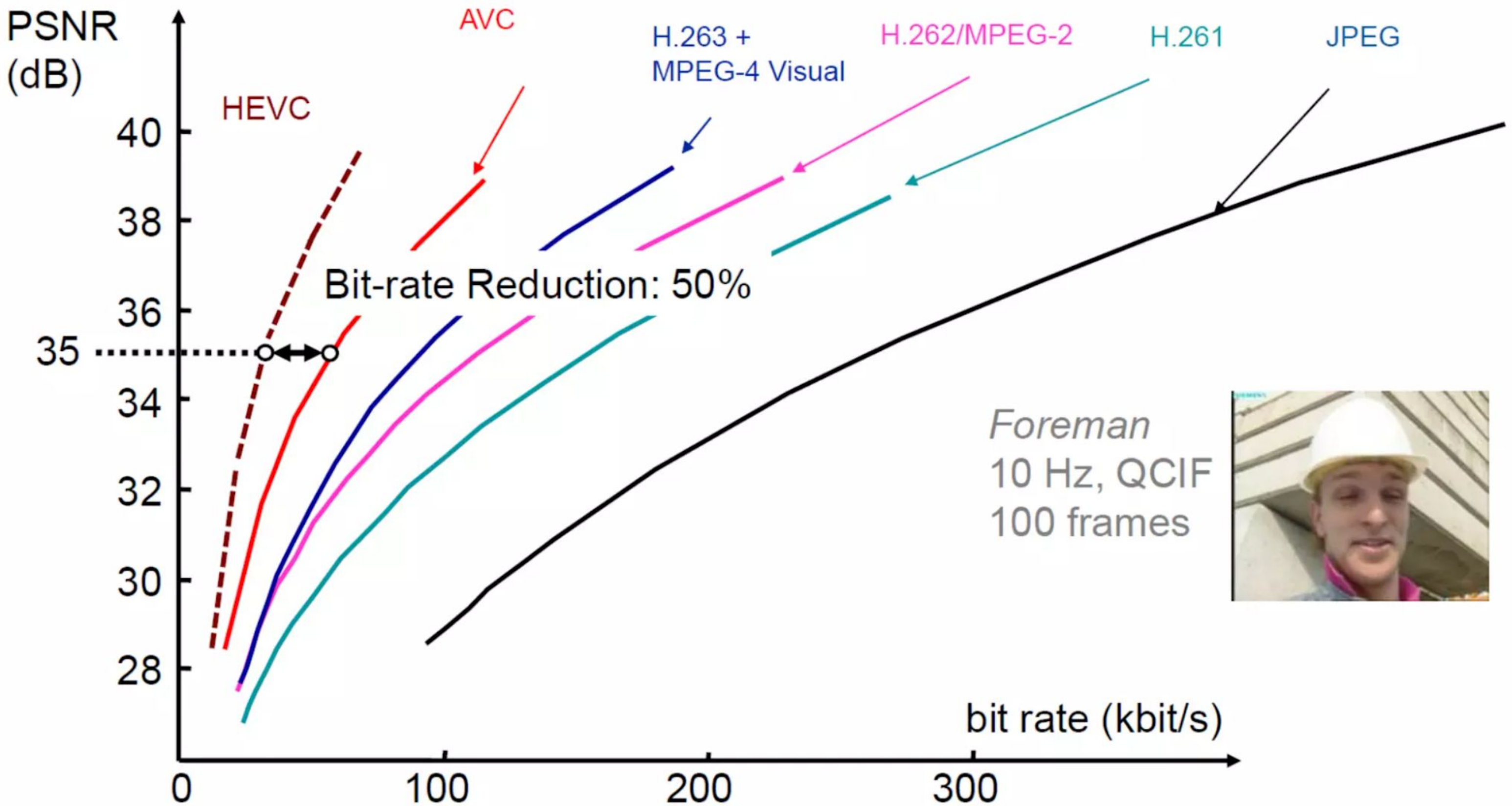
Video compression

Video Compression Standards

Timeline of MPEG/VCEG standards and examples of alternative codecs



History of video coding standardization (1985 ~ 2013)



3

The Moving Picture Experts Group (MPEG)

- established by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC).

-The MPEG group is responsible for

the development of a range of standards and technical reports relating to video, audio and multimedia content.

Video compression basic principles

Video compression

-coding techniques to reduce redundancy in video data

combination of:

- spatial image compression
- temporal motion compensation

Main idea:

Video compression typically operates on square-shaped groups of neighbouring pixels - macroblocks.

These blocks of pixels are compared from one frame to the next and the video compression codec sends only the differences within those blocks.

Inter-frame / Intra-frame compression

- inter-frame compression uses one or more earlier or later frames in a sequence to compress the current frame

- intra-frame compression uses only the current frame (still images compression).

Inter-frame compression

Inter-frame compression

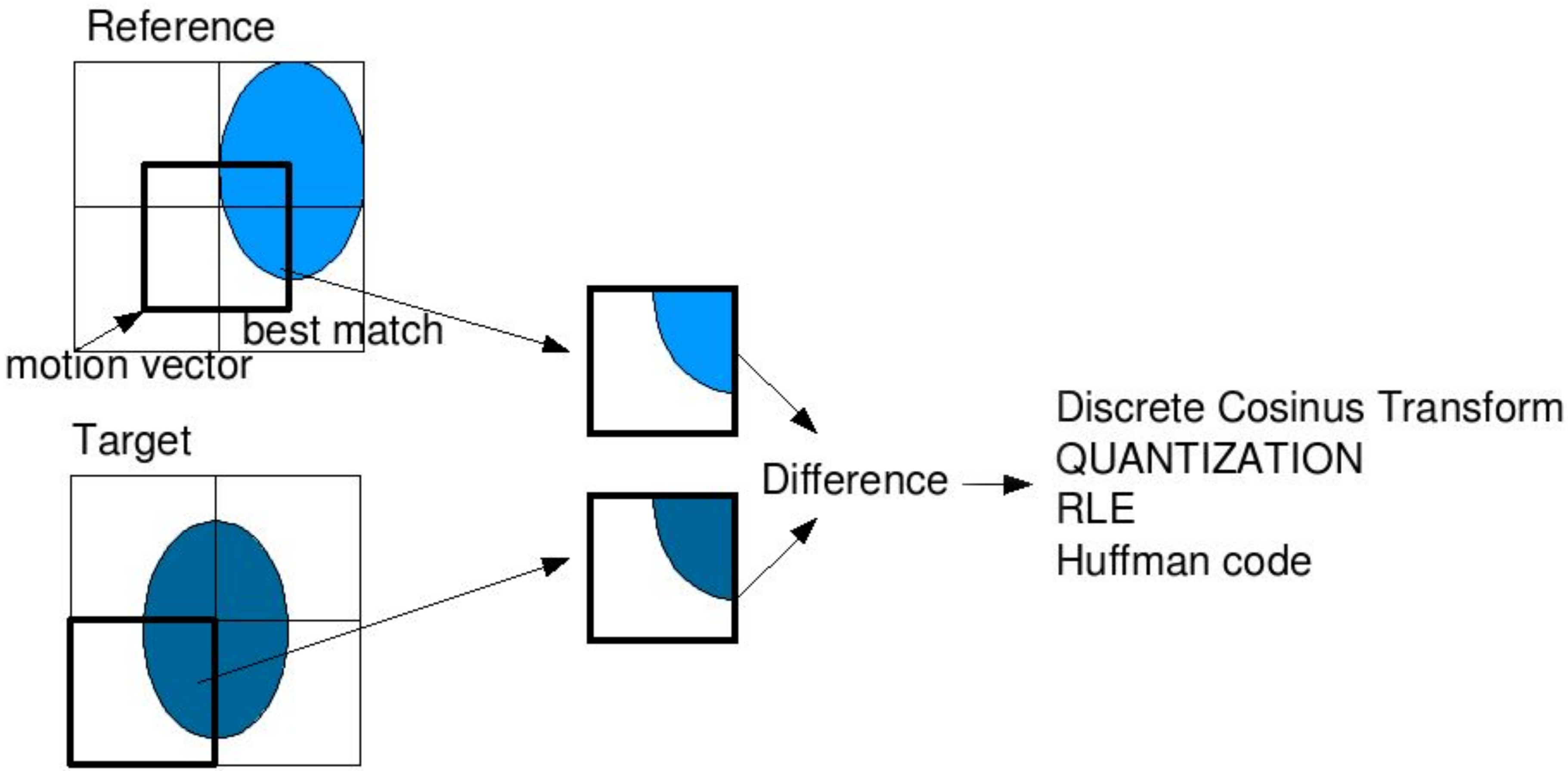
This kind of prediction tries to take advantage from

temporal redundancy between neighbouring frames allowing to achieve higher compression rates.

Inter-frame compression using macroblocks

- an inter-coded frame is divided into blocks known as macroblocks.
- instead of directly encoding the raw pixel values for each block, the encoder will try to find a block similar to the one it is encoding on the reference frame(previously encoded frame) - this process is done by a block matching algorithm.
- if the encoder succeeds on its search, the block could be encoded by a motion vector, which points to the position of the matching block at the reference frame.
- the process of motion vector determination is called motion estimation.

Inter-frame compression using macroblocks



Intra-frame compression

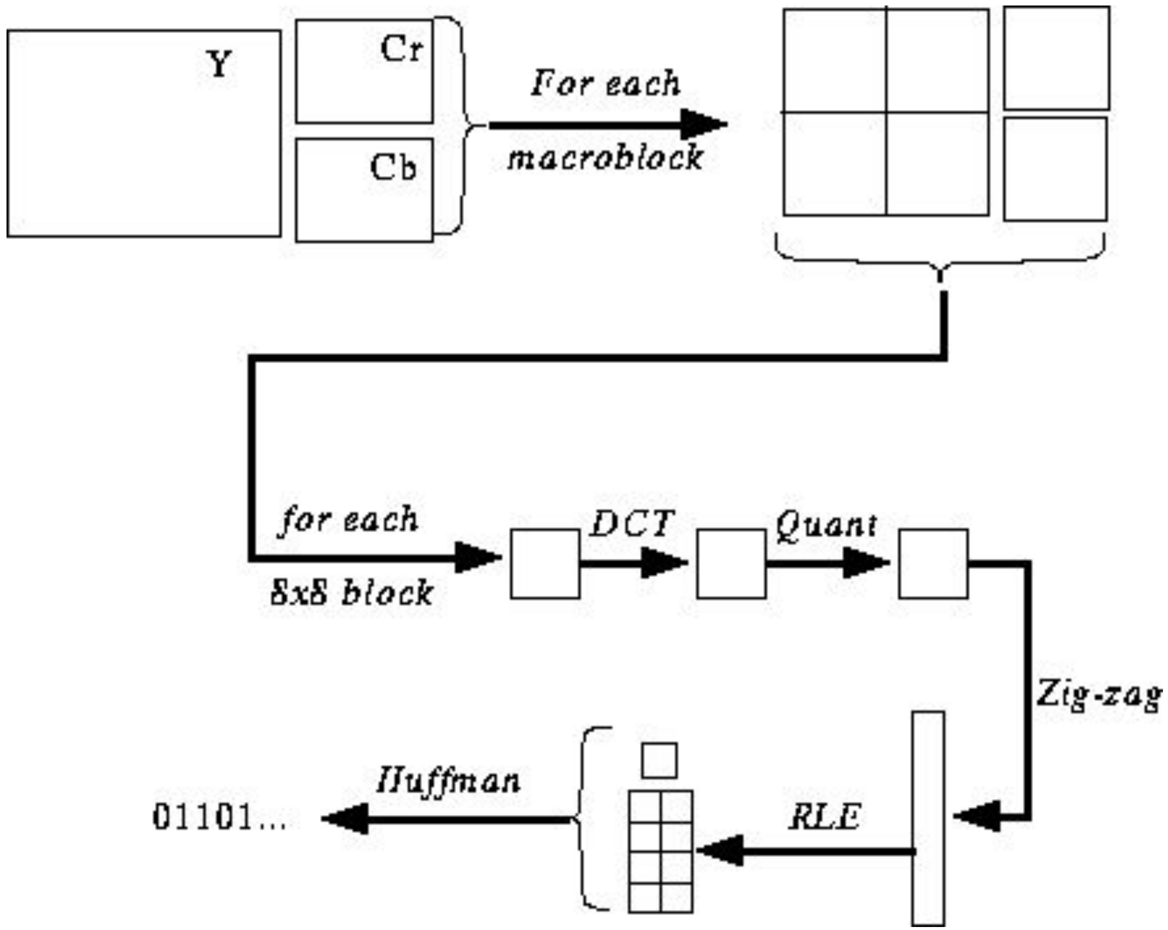
- Image compression applied to each frame.
- Can therefore be lossless or lossy,
- Data in the YCbCr data format
- The coding process varies greatly depending on which type of encoder is used (e.g.JPEG , H.264)

- The most common steps usually include:
  - partitioning into macroblocks,
  - transformation (e.g., using a DCT or wavelet),
  - quantization
  - entropy encoding.

Intra-frame compression

Intra-frame compression coding

- similar to JPEG still image video encoder

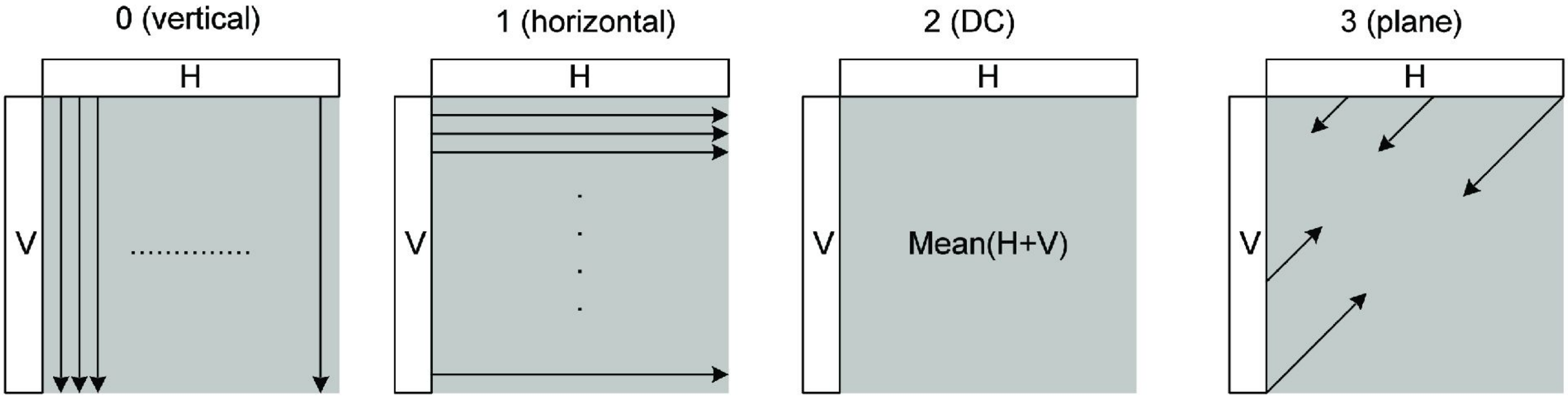


Intra-prediction

Predict the similarity between the neighbouring pixels in one frame in advance and exploit differential coding transform coding to remove the redundancy.

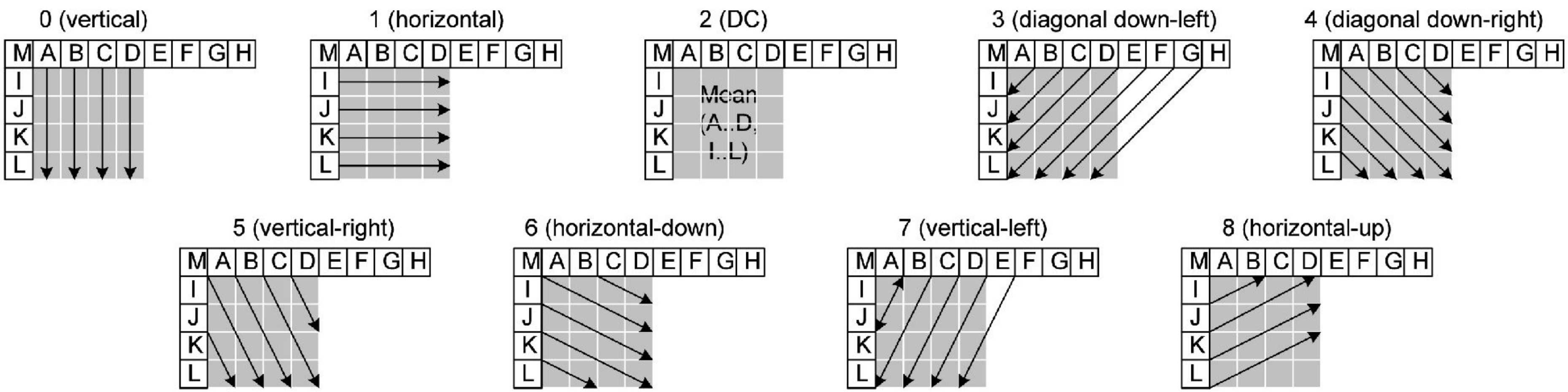
Intra-prediction

The first two modes extrapolate pixels from the vertical and horizontal neighboring pixels, respectively. The third mode estimates the DC coefficient of the block. A DC coefficient is the average of all pixels in a block. The last mode (called Plane mode) performs a planar prediction, by generating a gradient estimated from the neighbouring pixels.



Intra-prediction
Modes for 4x4 sub-blocks.

Mode 0 (vertical): extrapolation from upper samples (H),
Mode 1 (horizontal): extrapolation from left samples (V),
Mode 2 (DC): mean of upper and left-hand samples (H+V),
Mode 3 (Plane): a linear "plane" function is fitted to the upper and left-hand samples H....



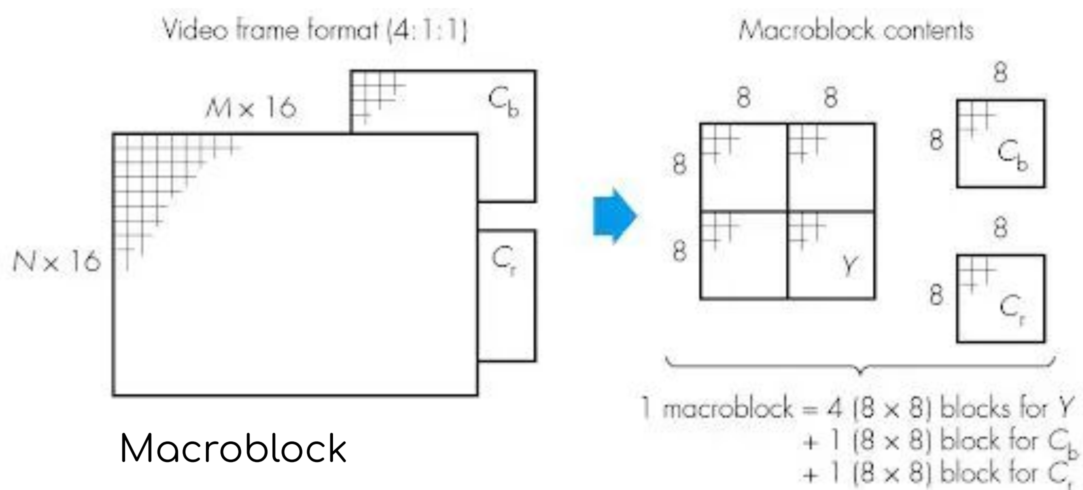
Macroblocs

Macroblock

MPEG-1 operates on video in a series of 8x8 blocks for quantization.

The set of chroma + luma blocks, with a resolution of 16x16, is called a macroblock.

- A macroblock is the smallest independent unit
- Motion vectors operate solely at the macroblock level.



Macroblock

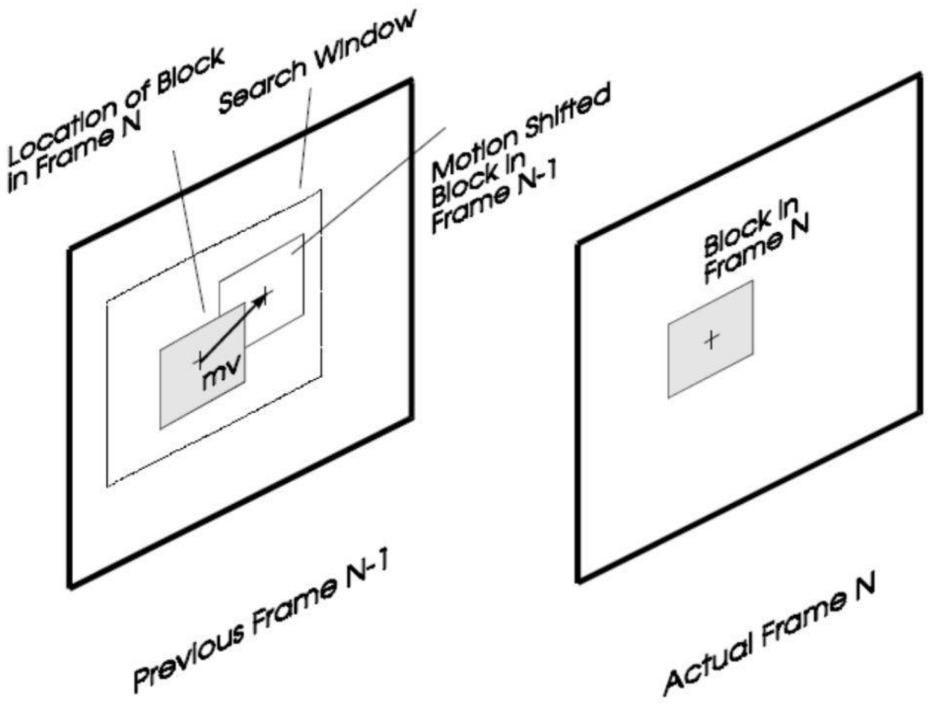
Video coding MPEG1, MPEG2

Motion Compensated Prediction

Block matching approach for motion compensation:

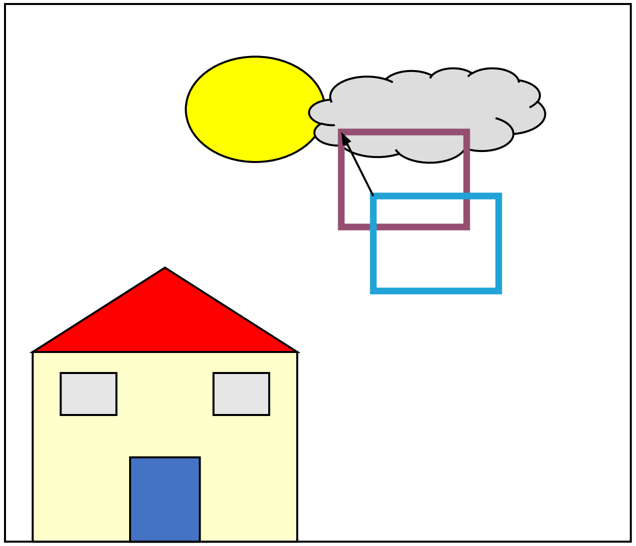
One motion vector (mv) is estimated for each block in the actual frame N to be coded.

The motion vector points to a reference block of same size in a previously coded frame N-1.

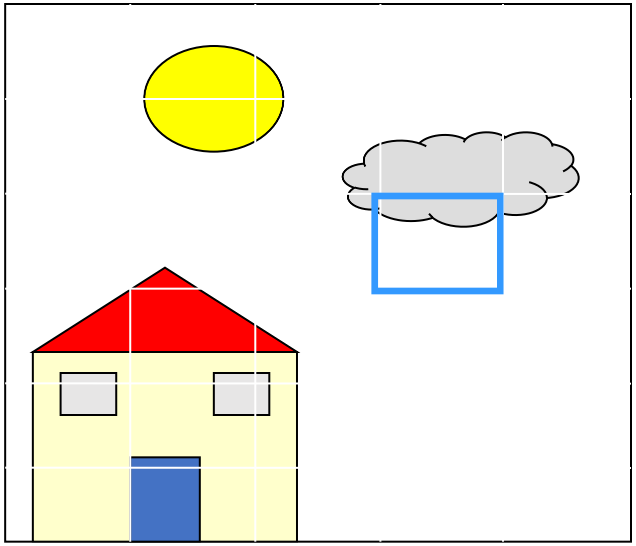


Motion Vectors

A motion vector (MV) describes the offset between the location of the block being coded (in the current frame) and the location of the best-match block in the reference frame



T=1 (reference)



T=2 (current)

Motion vectors

Only blocks that change will be updated

Motion estimation

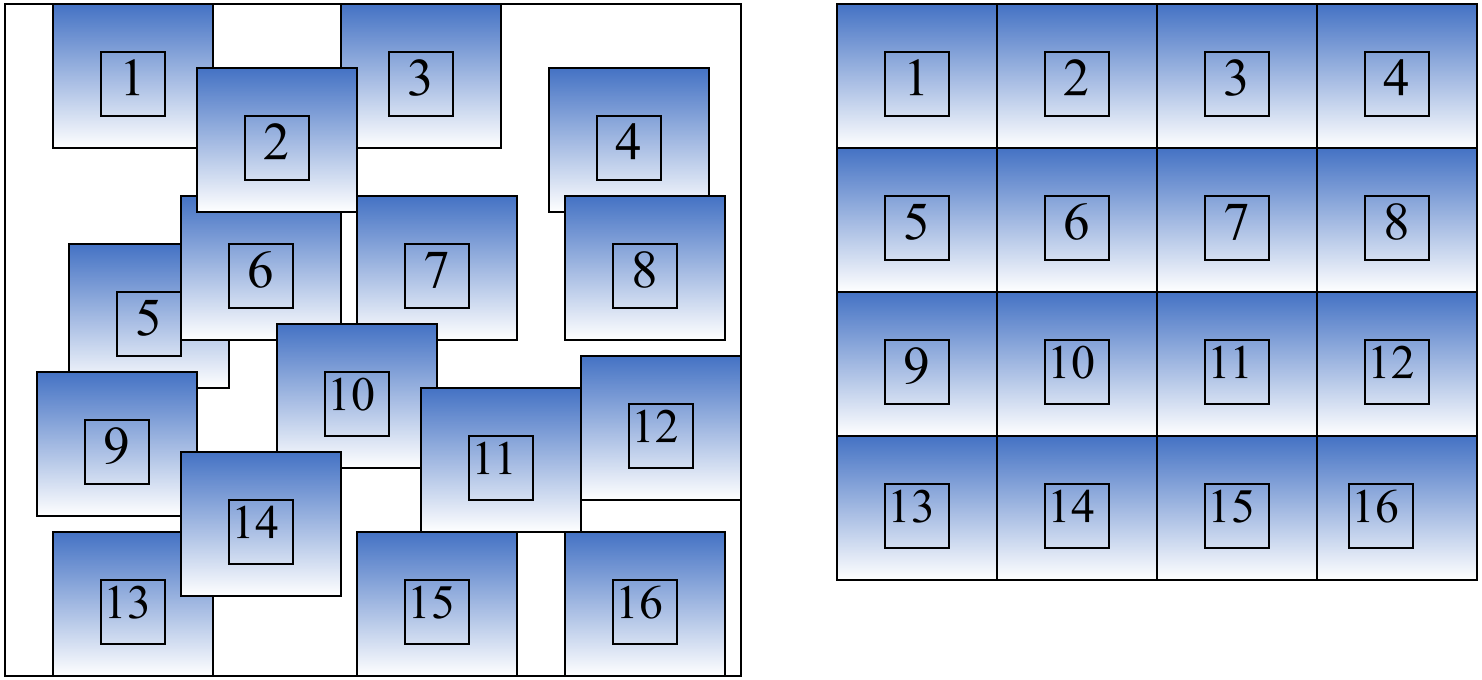
Motion vectors record the distance between two areas on screen based on the number of pixels called pels (pel – abbreviation of pixel).

MPEG-1 video uses a motion vector (MV) precision of one half of one pixel, or half-pel.

The finer the precision of the MVs, the more accurate the match is likely to be, and the more efficient the compression.

Motion Compensation

The blocks being predicted are on a grid



The blocks used for prediction are NOT

Motion Vector Search Metrics - Error measure

1. Mean squared error

$$\Sigma(b(B_{ref})-b(B_{curr}))^2$$

2. Mean abs. error

$$\Sigma|b(B_{ref})-b(B_{curr})|$$

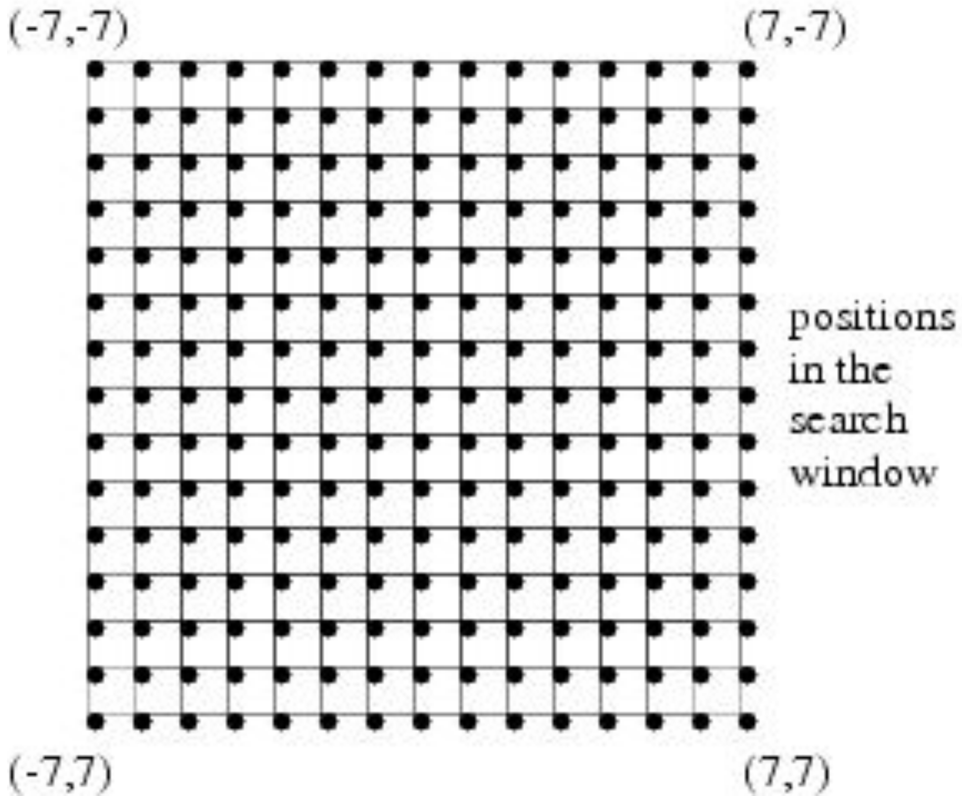
Motion Vector Search

Given error measure, how to efficiently determine best-match block in search window?

- Full search: best results, most computation
- Logarithmic search – heuristic, faster
- Hierarchical motion estimation

Motion Vector Search

Full search: Evaluate every position in the search window



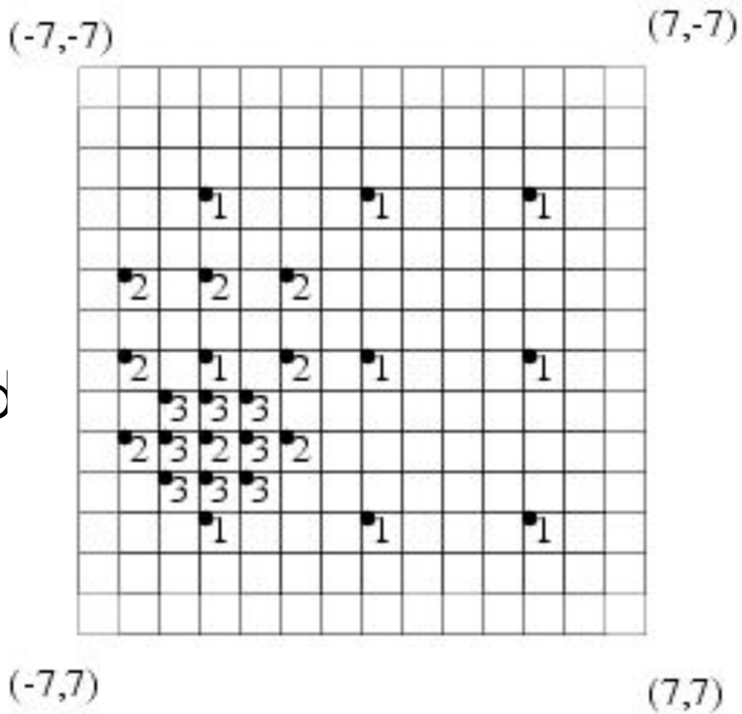
Motion Vector Search
Logarithmic Search

- First examine positions marked 1.

- Choose best of these (lowest error measure) and examine positions marked 2 surrounding it

- Choose the best of these, and examine the positions marked 3

- Final result = best of these



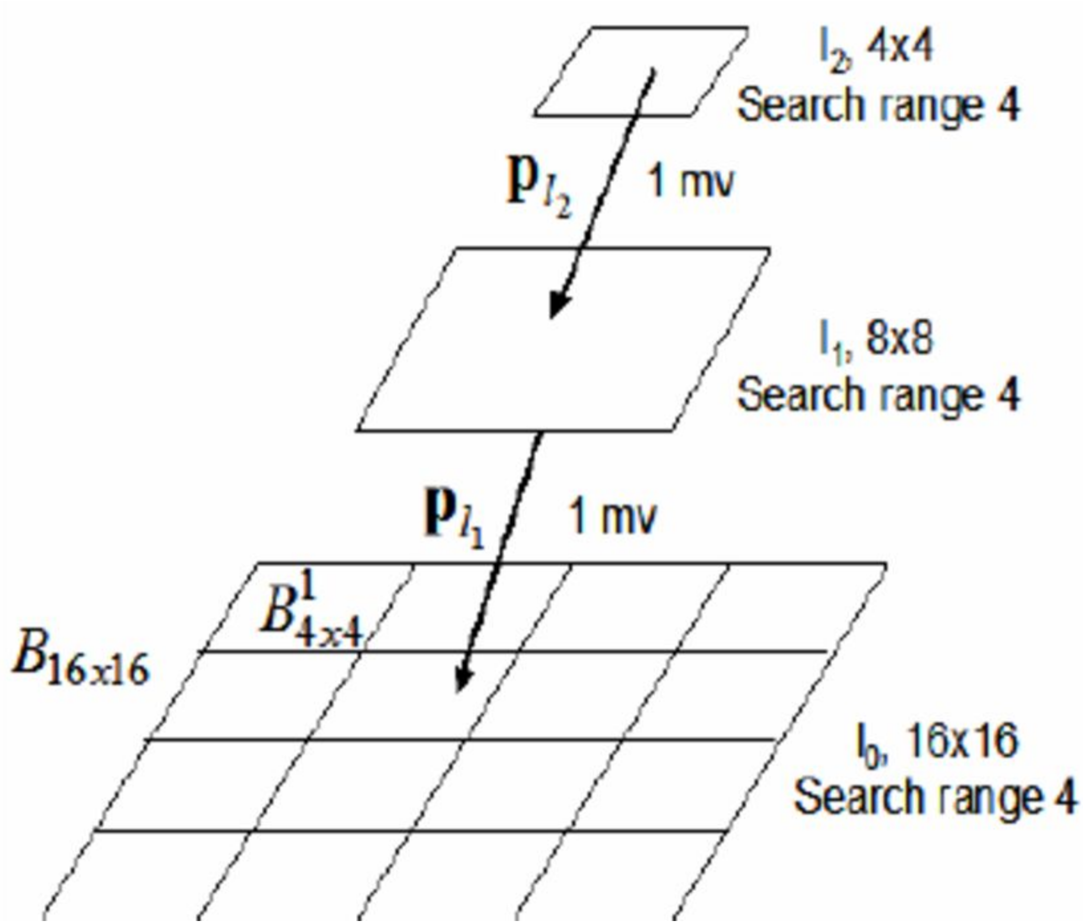
Hierarchical Motion Estimation (Pyramid Approach)

Use an averaging filter on the image, then downsample by a factor of 2

Conduct a search on the downsampled image (only ¼ of the size)

Given the results of the search on the downsampled image, return to the full resolution image and refine the search there

Hierarchical Motion Estimation



Motion Compensation

The standards do not specify HOW the encoder will find the motion vectors (MVs)

The encoder can use exhaustive/fast search, MSE /MAE/other error metric, etc.

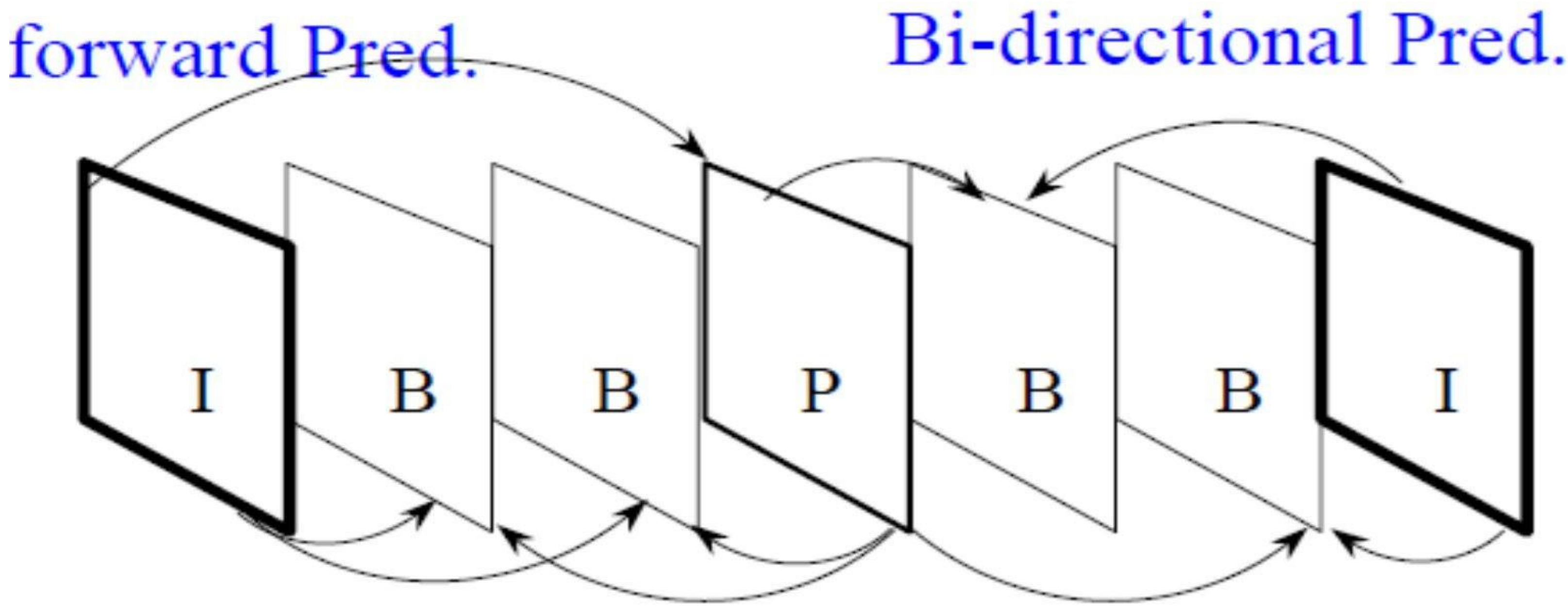
The standard DOES specify

- The allowable syntax for specifying the MVs
- What the decoder will do with them

What the decoder does is to grab the indicated block from reference frame, and *glue it in place*

MPEG-1: Basic Compression Modes

A typical assignment of compression modes for the frames of an image sequence



Intra-frame
I-frame

Intra-frame - they can be decoded independently of any other frames.

I-frames can be considered effectively similar to baseline JPEG images.

Predicted-frame
P-frame

Predicted-frame.

They may also be called forward-predicted frames, or inter-frames

Motion vectors and prediction errors are coded.

P-frames store the difference in image from the frame (either an I-frame or P-frame) immediately preceding it.

B-frame
Bidirectional-frame

Bidirectional-frame.

(bi-directional predictive compression)

They may also be known as backwards-predicted frames or B-

pictures.

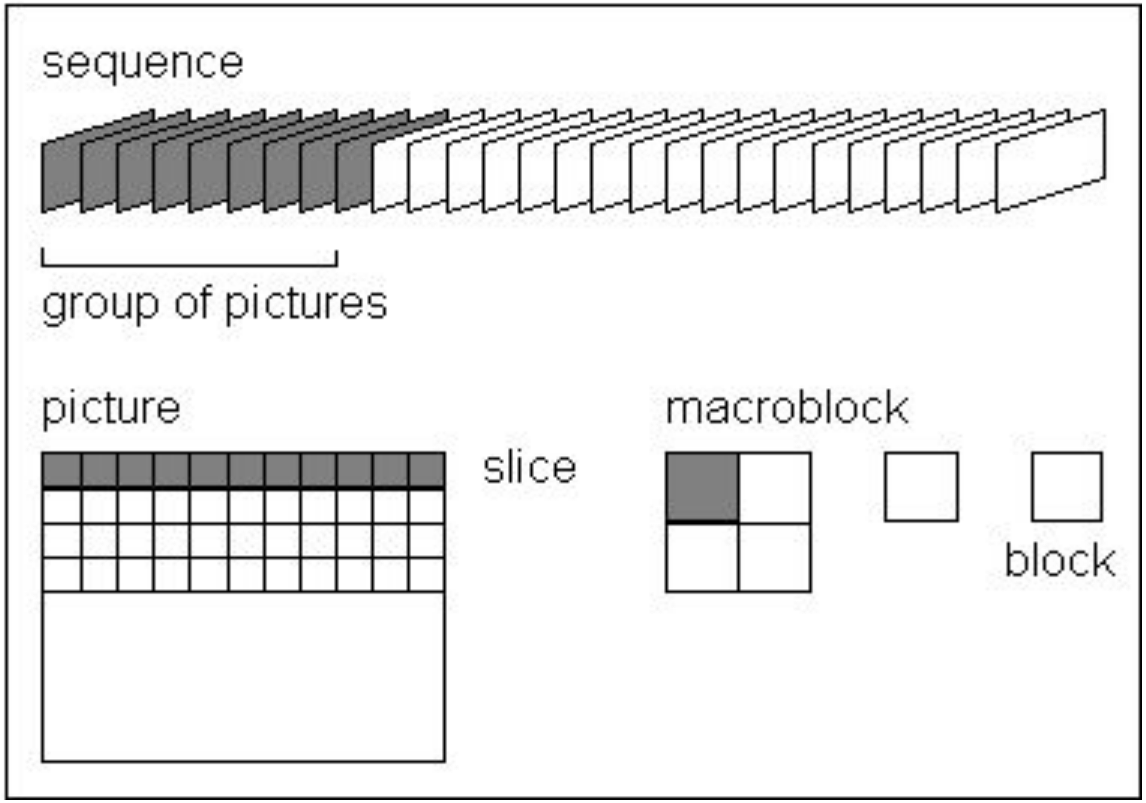
B-frames are quite similar to P-frames, except they can make

predictions using both the previous and future frames .

Temporal picture structure

Temporal picture structure

Example of temporal picture structure.



Temporal picture structure
MPEG Video Sequence in Layers

- Sequence – arbitrary length; a video clip, a complete program item, or a concatenation of programs

- Group of Pictures (GOP) – determine the random access to the sequence

- Picture – three types (I, P, B)

- Slice contains any number of sequential macro blocks encoded without reference to any other slice for intra frame addressing and (re)synchronization (e.g., error recovery)

- Macro block
- Block

GOP - Group of Pictures

Group of successive pictures within a coded video stream.

Repeating sequence of I-, P- and B-pictures.

Always begins with an I-picture.

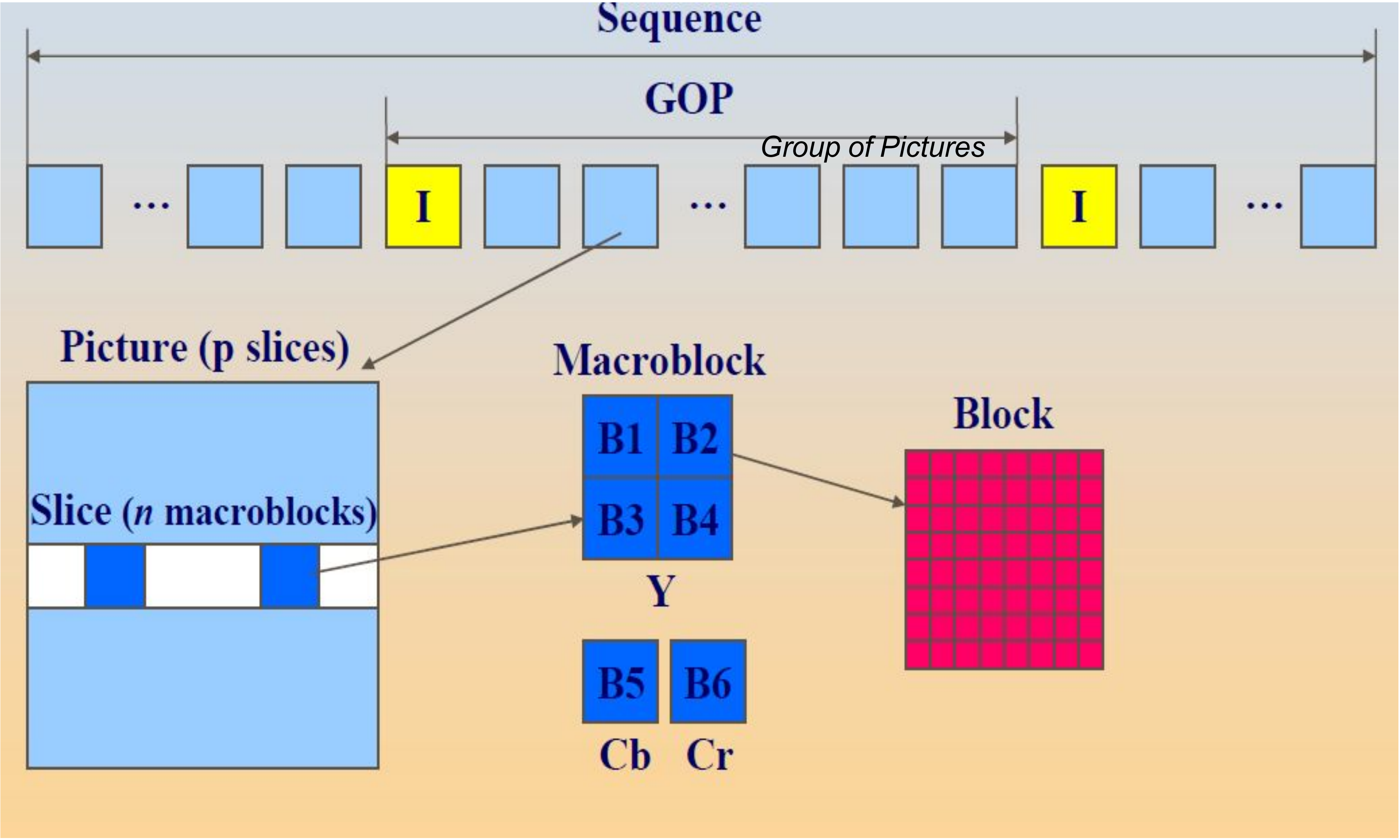
Each coded video stream consists of successive GOPs.

The GOP structure is often referred by two numbers, for example, M=3, N=12.

The first number tells the distance between two anchor frames (I or P).

The second one tells the distance between two full images (I-frames): it is the GOP size.

MPEG Video Sequence in Layers



Motion Estimation/Compensation

Performed on luminance macroblock (16×16)

Supporting half-pixel (half-pel) motion compensation for improved performance (MPEG1)

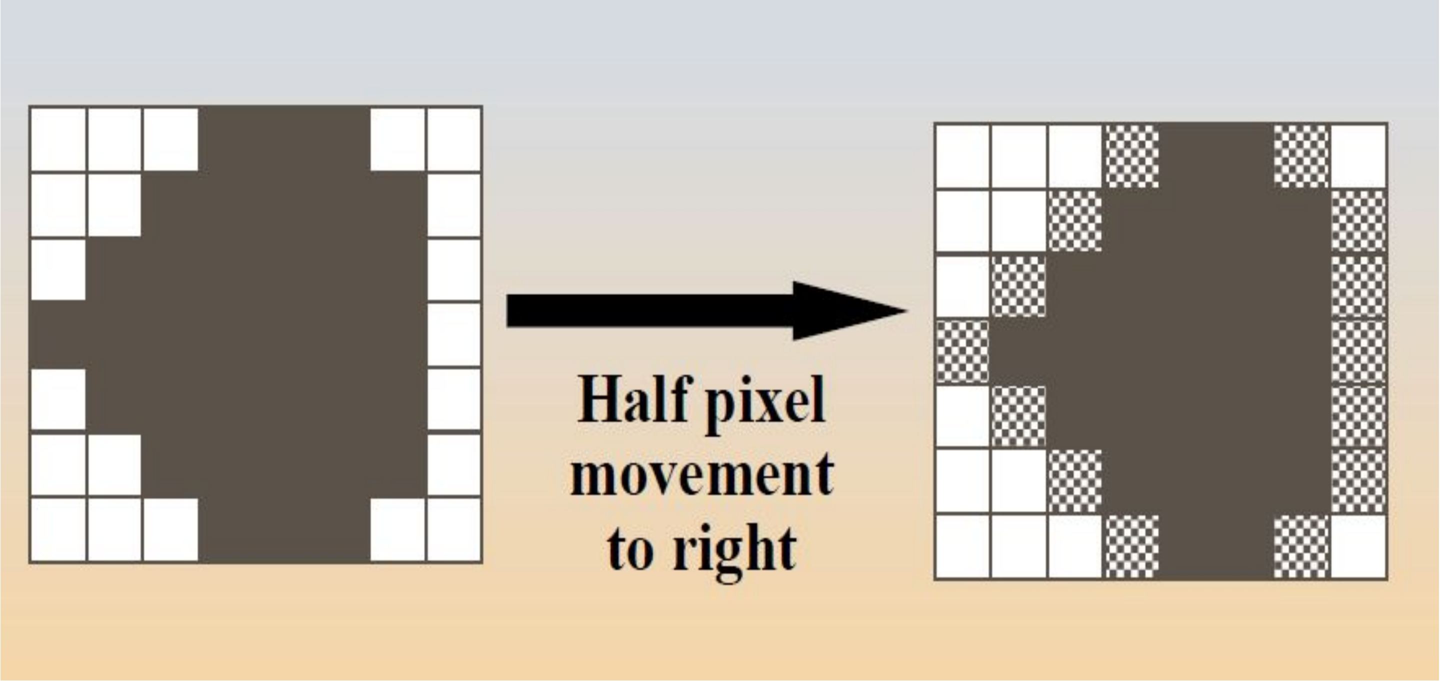
Maximum motion vector range

-512 to + 511.5 for half-pixel motion vector

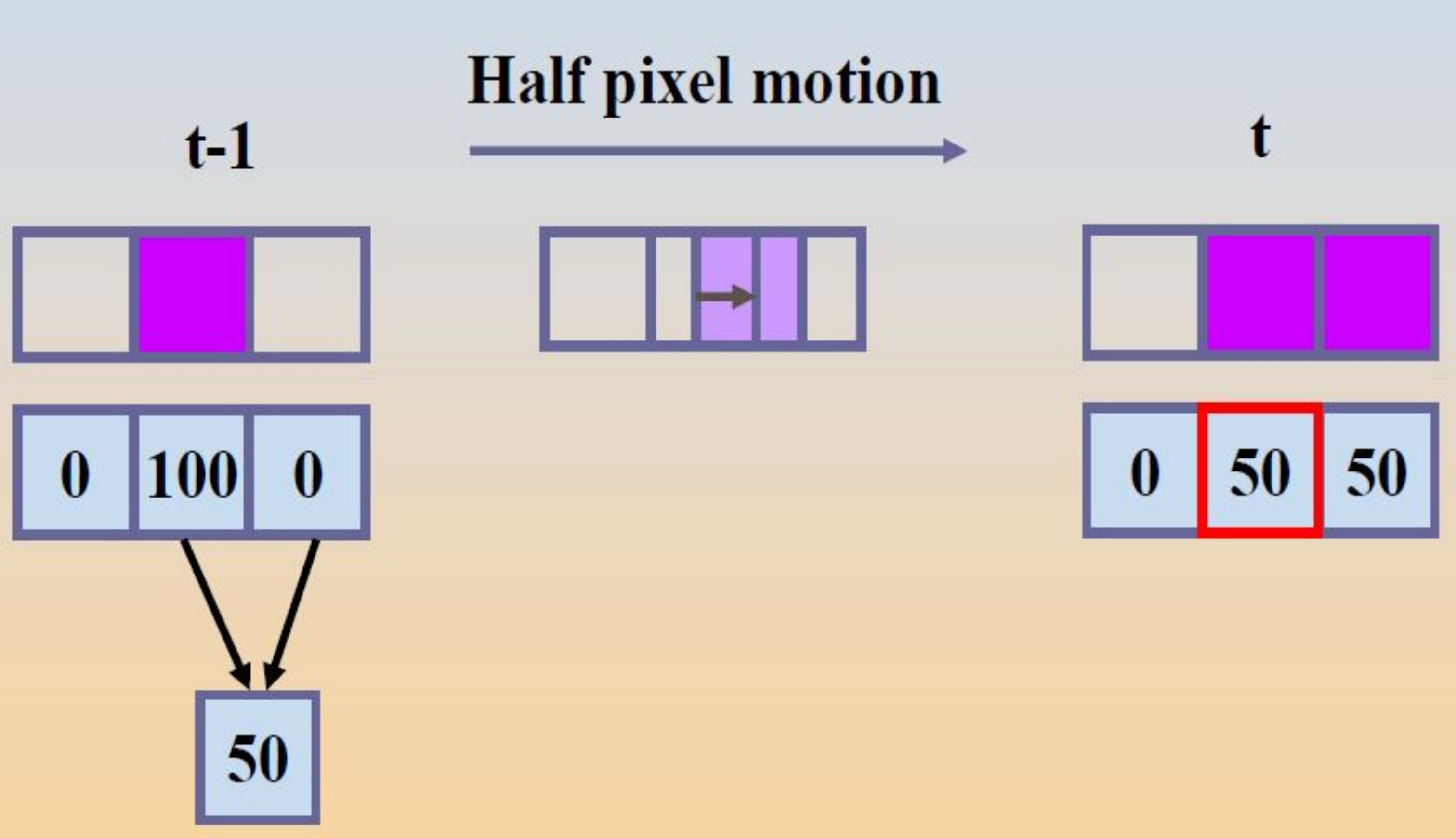
-1024 to 1023 for full pixel motion vector

Chrominance (Cb, Cr) motion vectors are half of luminance MB's

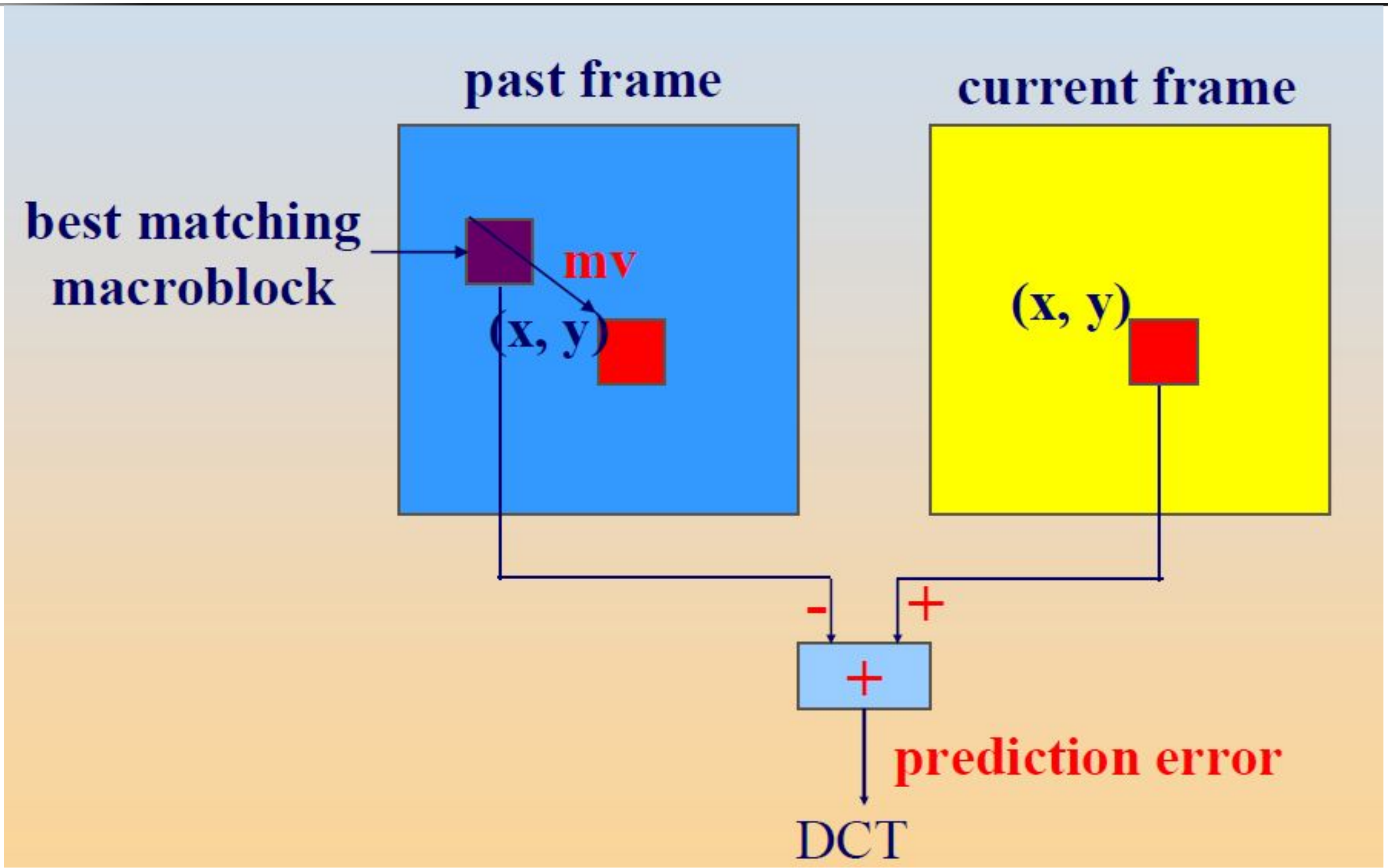
Half-Pixel Motion Compensation



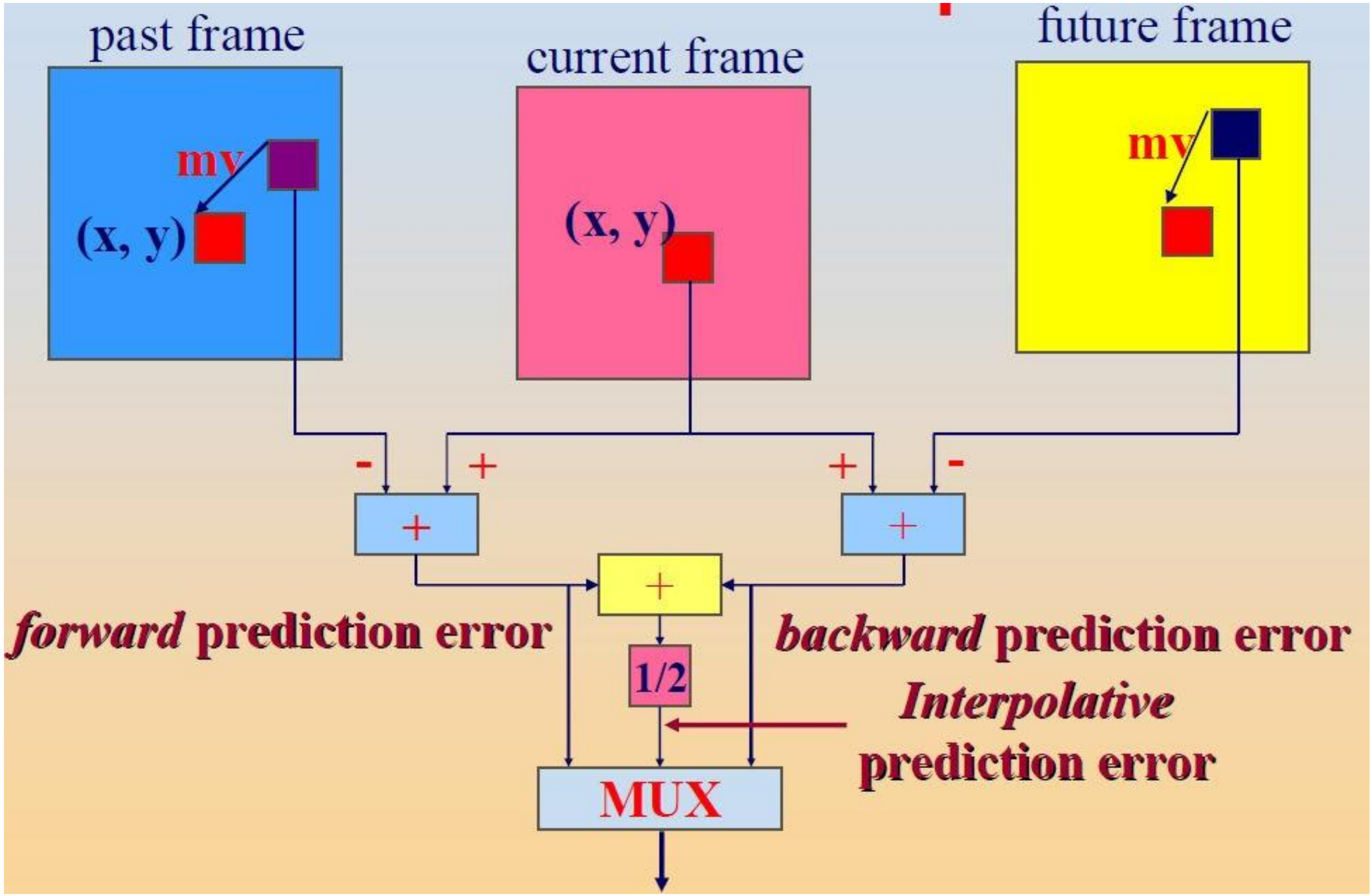
Half-Pixel Motion Compensation



Forward Motion Compensation



Bidirectional Motion Compensation



DCT

Picture Types and MB Types

I-frame

- All I-macroblocks

P-frame

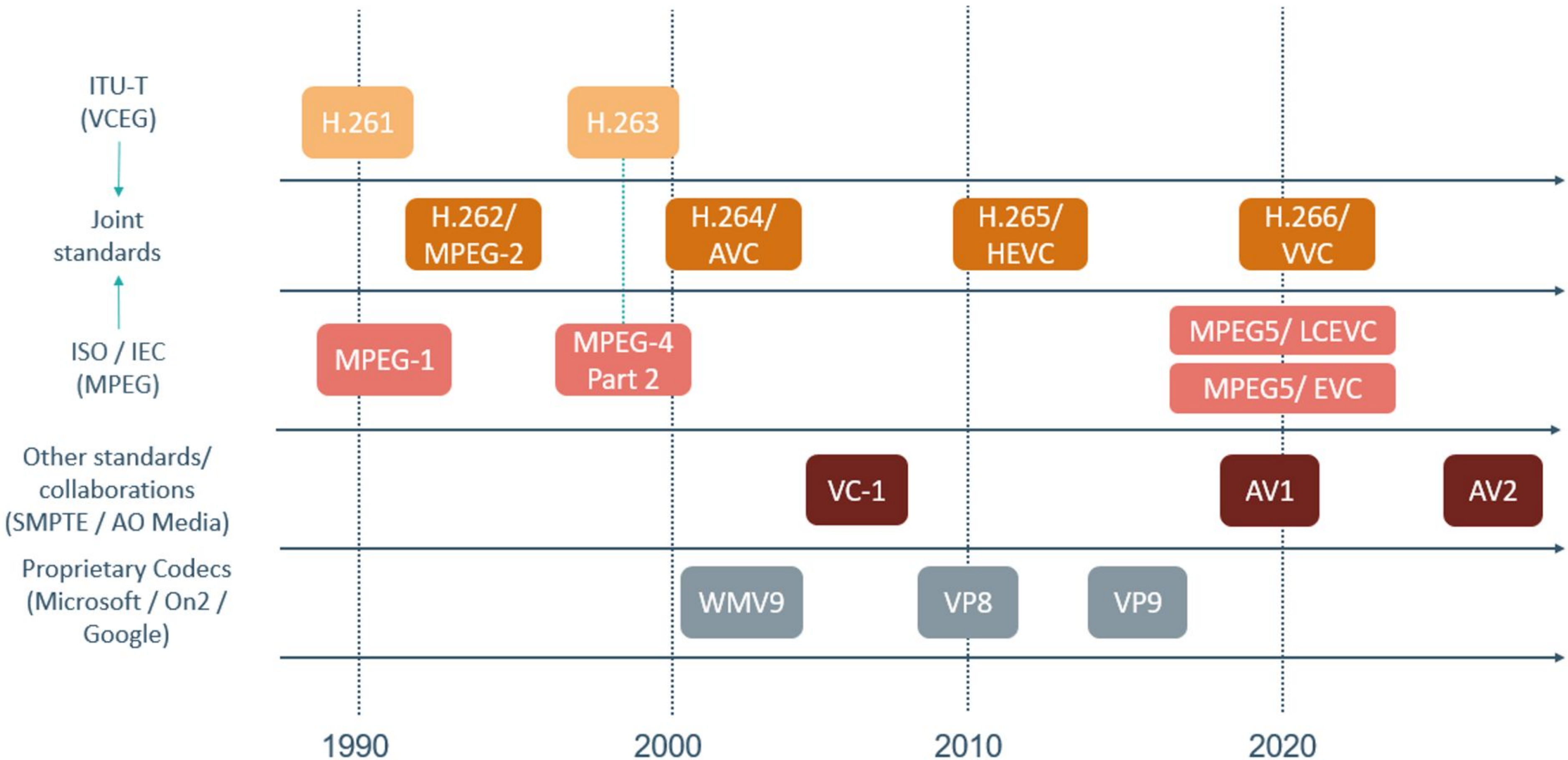
- P-macroblocks
- I-macroblocks
- Skip

B-frame

- B-macroblocks
- Forward prediction
- Backward prediction
- Bi-directional prediction
- I-macroblocks
- Skip

H.264

Timeline of MPEG/VCEG standards and examples of alternative codecs



H.264 Inter-frame prediction improvements

H.264 is Part 10 ISO/IEC FCD 14496-10 of MPEG 4 'Advanced Video Coding' (also to be published as ITU-T H.264/AVC)

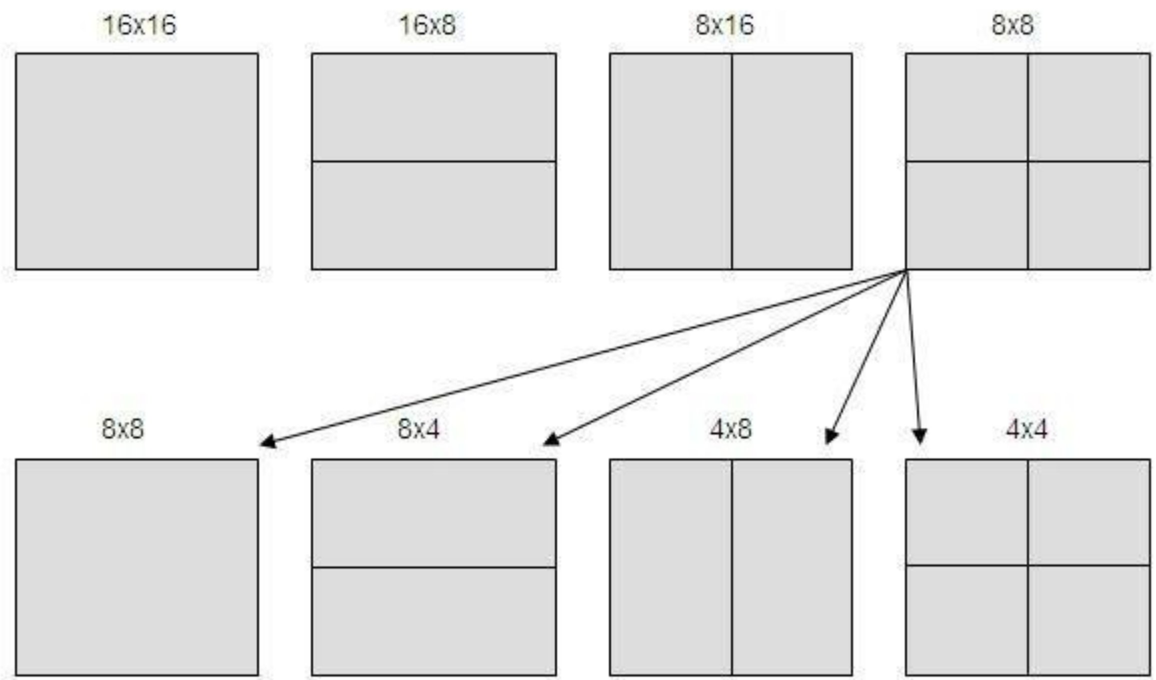
The most important improvements of this technique in regard to previous standard are MPEG-2:

- More flexible block partition
- Resolution of up to ¼ pixel motion compensation
- Multiple references
- Enhanced Direct/Skip Macroblock

More flexible block partition

Luminance block partition of 16 x 16 (MPEG-2), 16 x 8, 8 x 16, 8 x 8.

Last case allows divide the block in new blocks of 4 x 8, 8 x 4, 4 x 4.

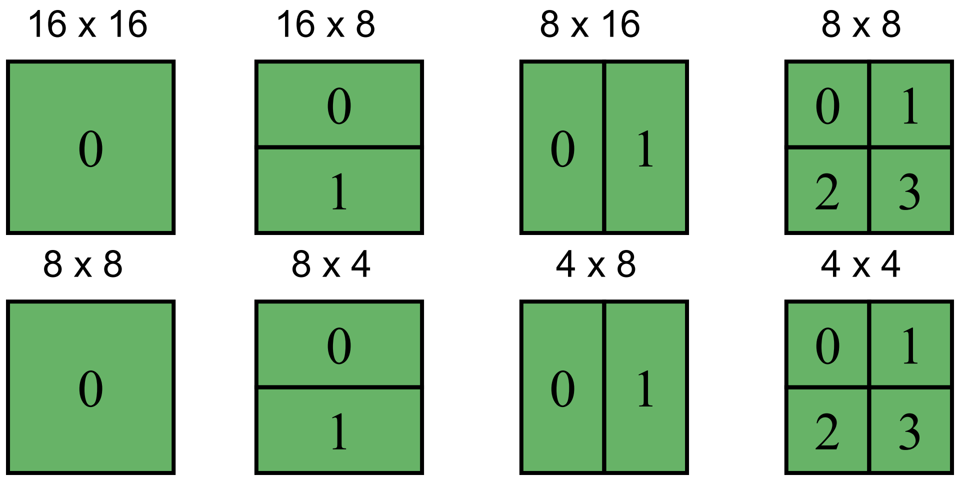


Variable Block Size

The fixed block size may not be suitable for all motion objects

- Improve the flexibility of comparison
- Reduce the error of comparison

7 types of blocks for selection



Resolution of up to ¼ pixel motion compensation

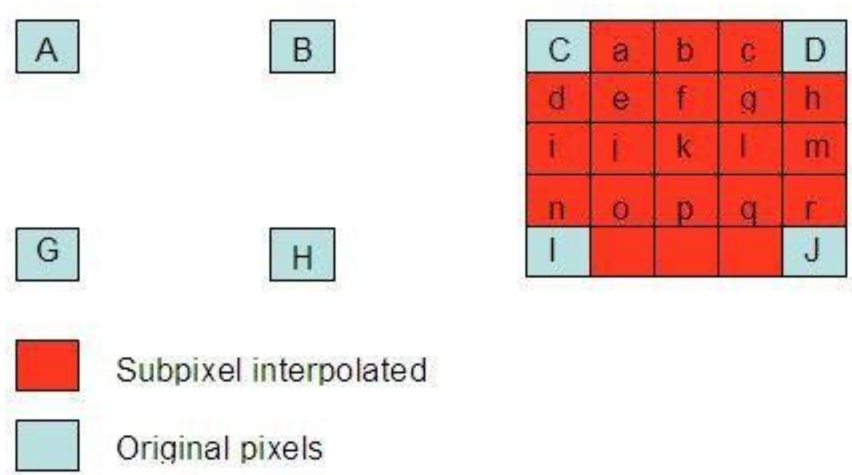
Pixels at half-pixel position are obtained by applying a filter of length 6.

H=[1 -5 20 20 -5 1]

For example:

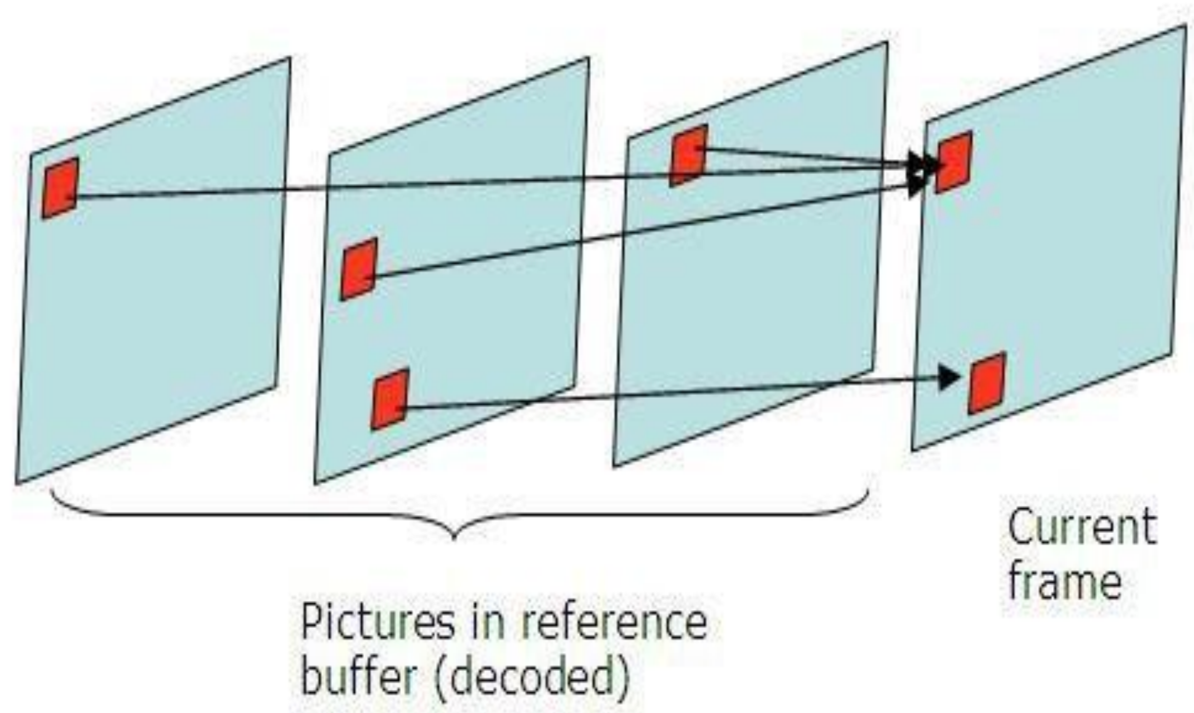
b=A - 5B + 20C + 20D - 5E + F

Pixels at quarter-pixel position are obtained by bilinear interpolation. (MPEG-2 allowed a ½ pixel resolution)



Multiple references

Multiple references to motion estimation allows finding the best reference in 2 possible buffers



H.265

High Efficiency Video Coding (HEVC) H.265

- provides superior video quality and up to twice the data compression as the previous standard (H.264/MPEG-4 AVC)

- HEVC can support Ultra High Definition video

- This standard was developed by the ISO/IEC Moving Picture Experts Group (MPEG) and ITU-T Video Coding Experts Group (VCEG), through their Joint Collaborative Team on Video Coding (JCT-VC)
- HEVC is also known as ISO/IEC 23008-2 MPEG-H Part 2 and ITU-T H.265

High Efficiency Video Coding (HEVC) H.265

Two resolutions are defined as UHDTV:

4K UHDTV is 3840 pixels wide by 2160 pixels tall), which is four times as many pixels as 1920x1080 (2.07 megapixels).

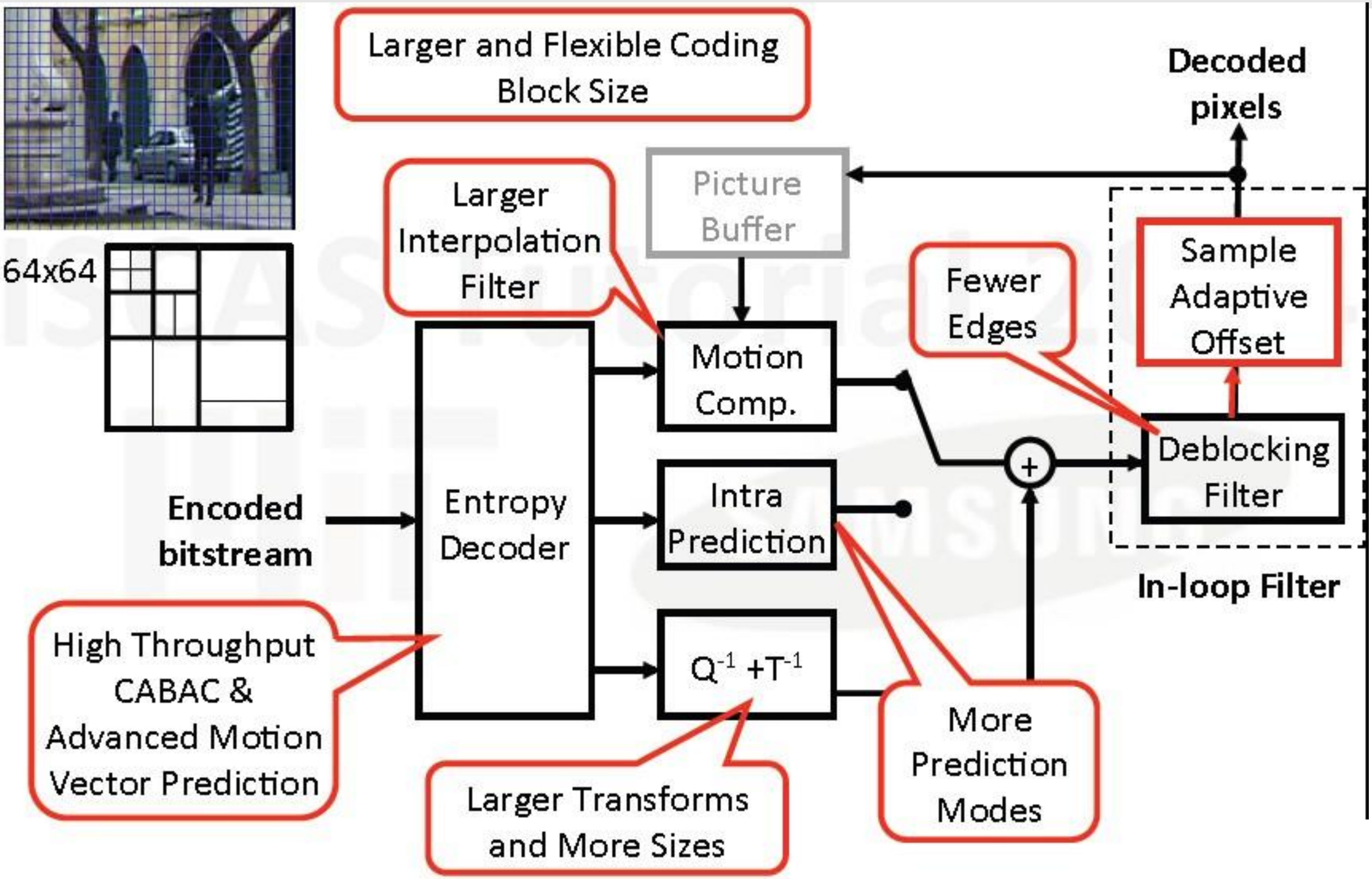
8K UHDTV is 7680 pixels wide by 4320 pixels tall (33.18 megapixels), which is sixteen times as many pixels as current HDTV. The p in 2160p and 4320p stands for progressive scan or non-interlaced.

High Efficiency Video Coding (HEVC) H.265

Major contributors to the higher compression performance of HEVC:

larger block structures with flexible sub-partitioning mechanisms.

H.265/HEVC vs. H.264/AVC Decoder



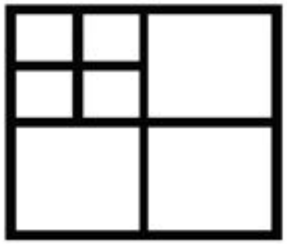
HEVC main improvements

- Larger Coding Blocks :Where H.264/AVC defines macroblocks up to 16×16 pixels, HEVC can describe a much larger range of block sizes, up to 64 x 64 pixels.
- Flexible Coding Block Structure :HEVC allows predicted blocks to be coded in different block sizes.
- Motion vectors with much greater precision : There are 35 Intra-prediction directions, compared with only 9 for H.264/AVC.
- Adaptive Motion Vector Prediction, a new method to improve inter-prediction.
- An improved deblocking filter
- Sample Adaptive Offset – an additional filter that reduces artifacts at block edges

Large Transforms

- HEVC supports 4x4, 8x8, 16x16, 32x32 integer transforms
  - Two types of 4x4 transforms (IDST-based for Intra, IDCT-based for Inter); IDCT-based transform for 8x8, 16x16, 32x32 block sizes
  - Integer transform avoids encoder-decoder mismatch and drift caused by slightly different floating point representations.
  - Parallel friendly matrix multiplication/partial butterfly implementation
  - Transform size signaled using Residual Quad Tree
- Achieves 5 to 10% increase in coding efficiency
- Increased complexity compared to H.264/AVC
  - 8x more computations per coefficient
  - 16x larger transpose memory

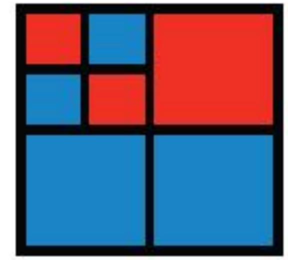
Represent residual of CU with TU quad tree

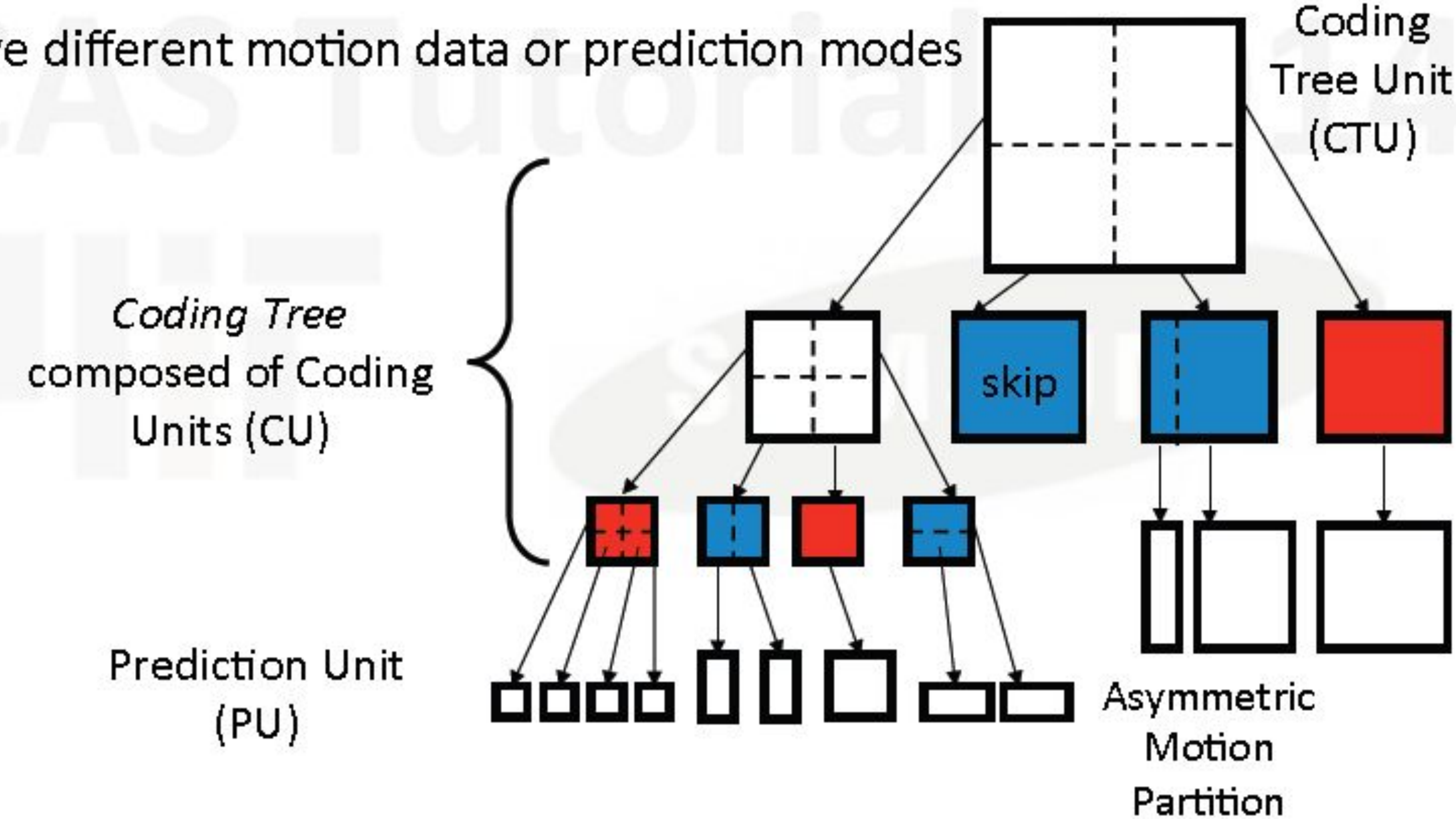


M. Budagavi et al., "Core Transform Design in the High Efficiency Video Coding (HEVC) Standard," IEEE JSTSP, 2013

Flexible Coding Block Structure

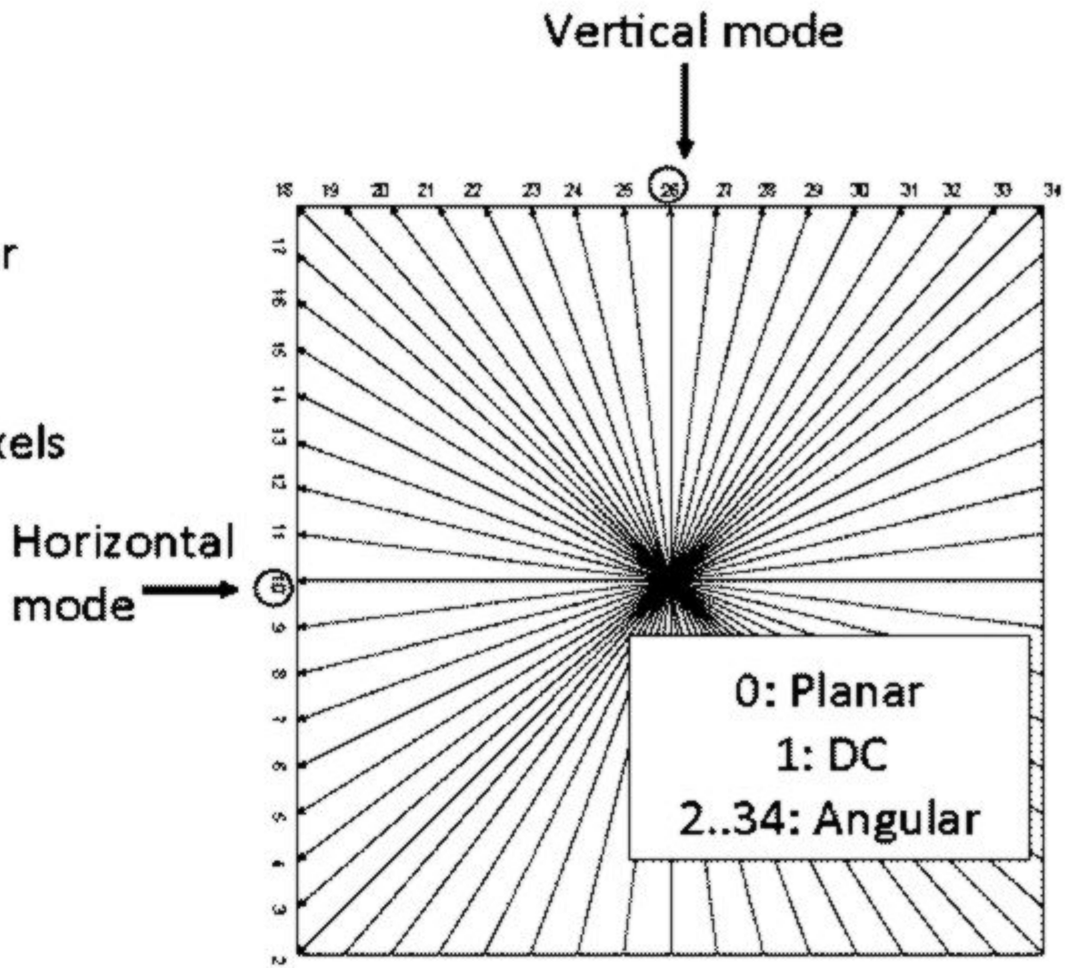
- Better adaptation to different video content
- CTU divided into Coding Units (CU) with Quad tree
- Coding units divided into prediction units (PU)
- PU have different motion data or prediction modes





Intra-prediction directions

- H.264/AVC has 10 modes
  - angular (8 modes), DC, planar
- HEVC has 35 modes
  - angular (33 modes), DC, planar
- Angular prediction
  - Interpolate from reference pixels at locations based on angle
- DC
  - Constant value which is an average of neighboring pixels (reference samples)
- Planar
  - Average of horizontal and vertical prediction



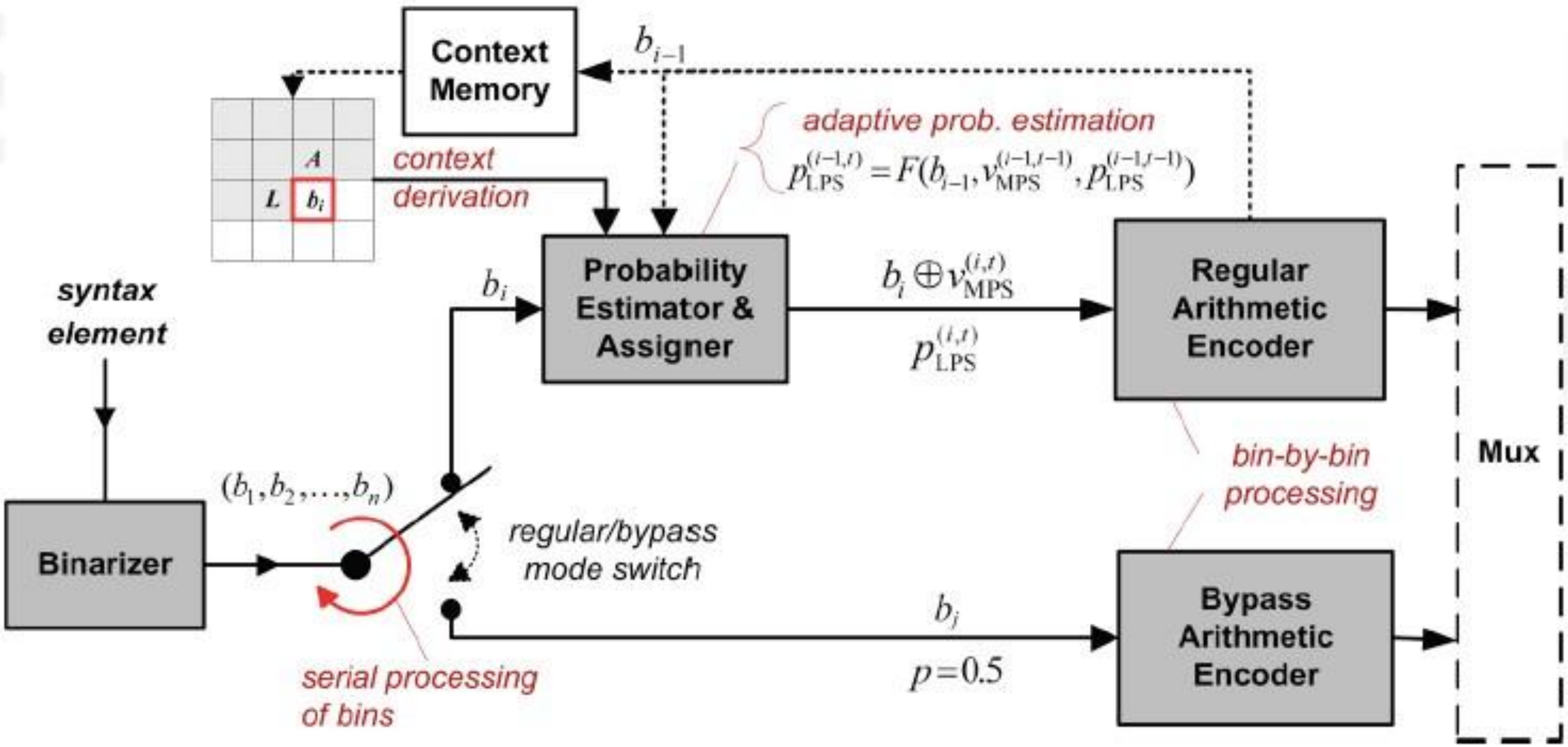
Remove Perceptual Redundancy

In-loop deblocking filtering

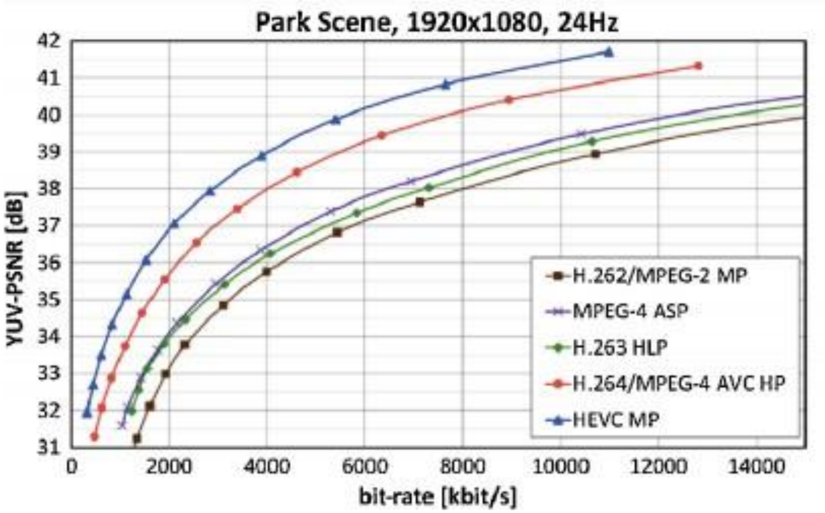
- Remove blocking artifact
  - Result from block based motion compensation
  - Result from block based transform coding

Entropy Coding

- Lossless compression of syntax elements
- HEVC uses Context Adaptive Binary Arithmetic Coding (CABAC)
  - 10 to 15% higher coding efficiency compared to CAVLC



Coding Efficiency of HEVC



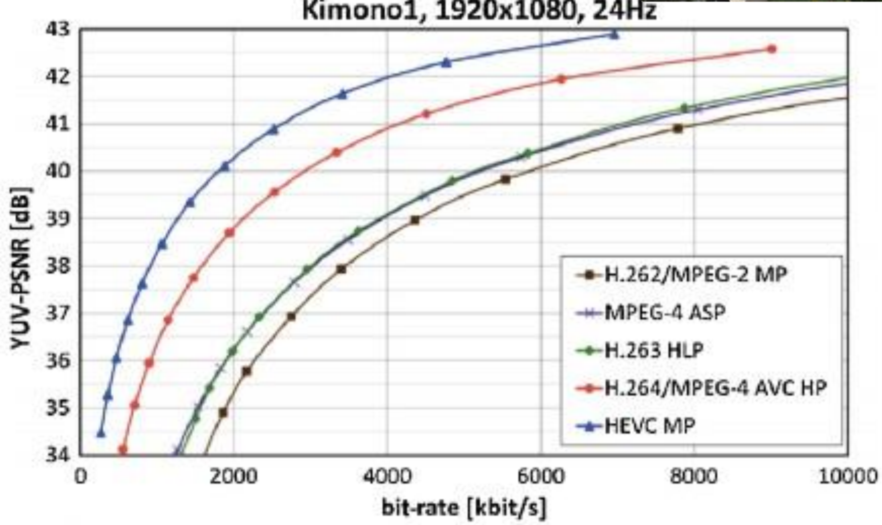


TABLE VI

AVERAGE BIT-RATE SAVINGS FOR EQUAL PSNR FOR ENTERTAINMENT APPLICATIONS

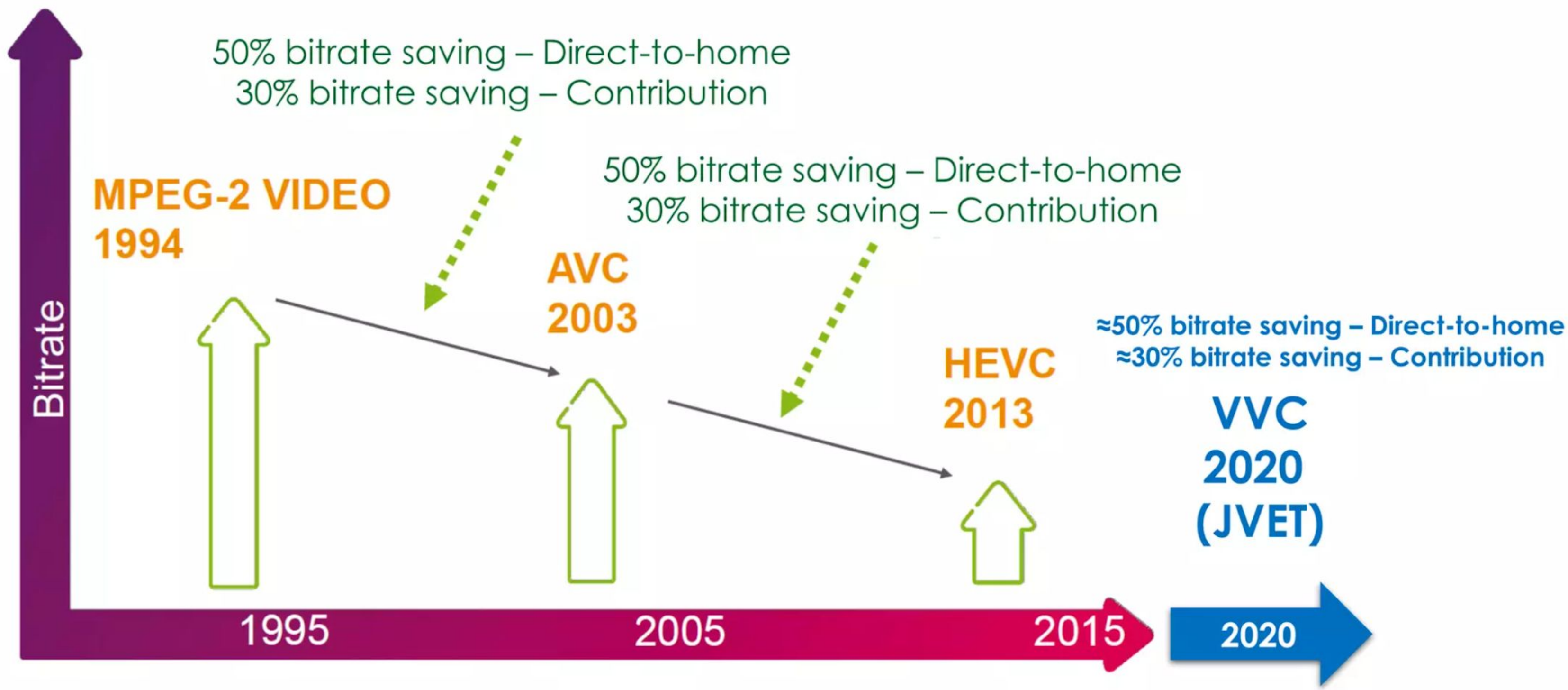
Encoding	Bit-Rate Savings Relative to			
	H.264/MPEG-4 AVC HP	MPEG-4 ASP	H.263 HLP	MPEG-2/ H.262 MP
HEVC MP	35.4%	63.7%	65.1%	70.8%
H.264/MPEG-4 AVC HP	–	44.5%	46.6%	55.4%
MPEG-4 ASP	–	–	3.9%	19.7%
H.263 HLP	–	–	–	16.2%

$$PSNR = 10\log_{10}\frac{(2^{bitdepth}-1)^2 * W * H}{\sum_i \{O_i - D_i\}^2}$$

J. R. Ohm et al., "Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, 2012

Versatile Video Coding (VVC) H266

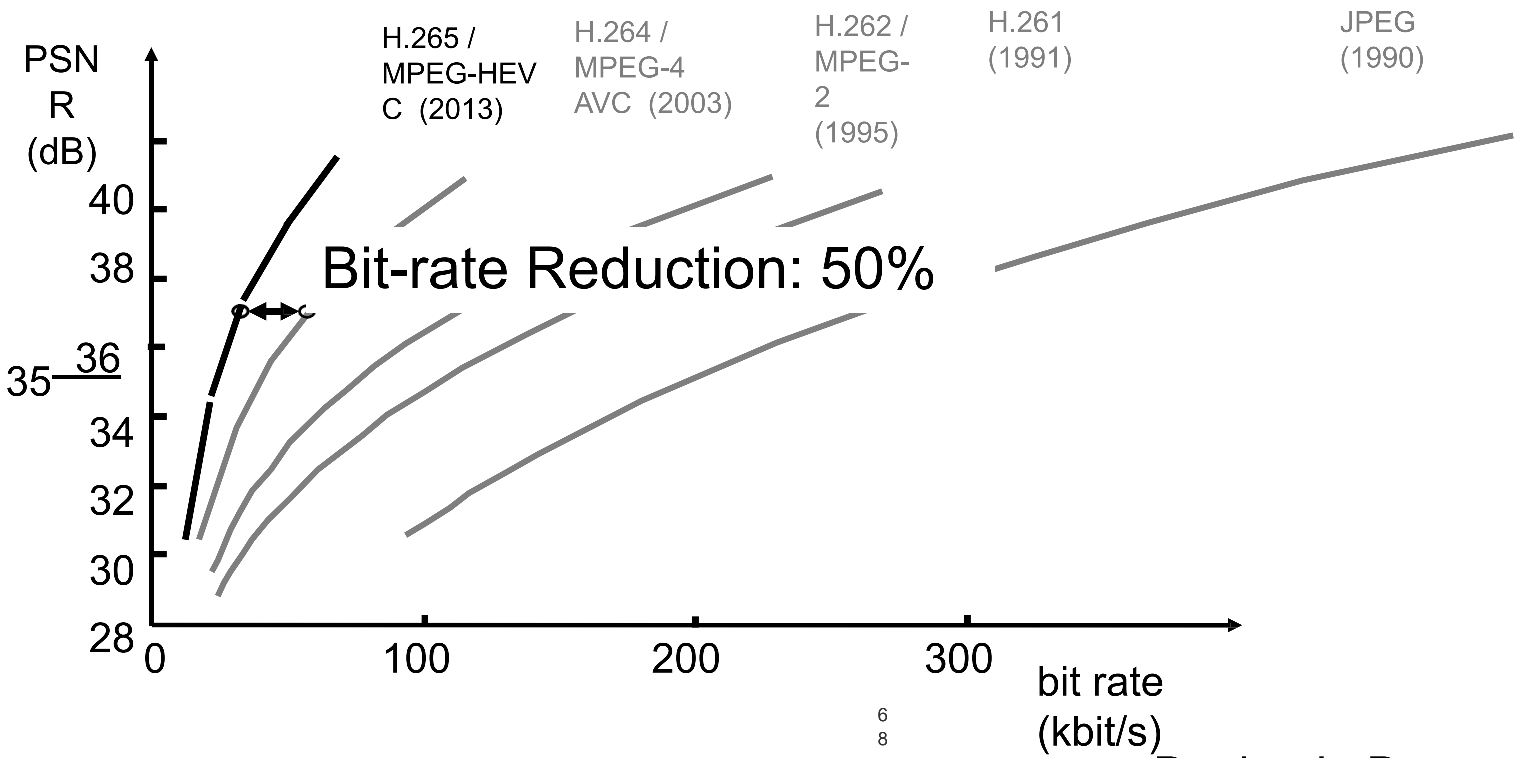
Versatile Video Coding (VVC) for Contribution and Distribution



46

VVC – Coding Efficiency

History of Video Coding Standards



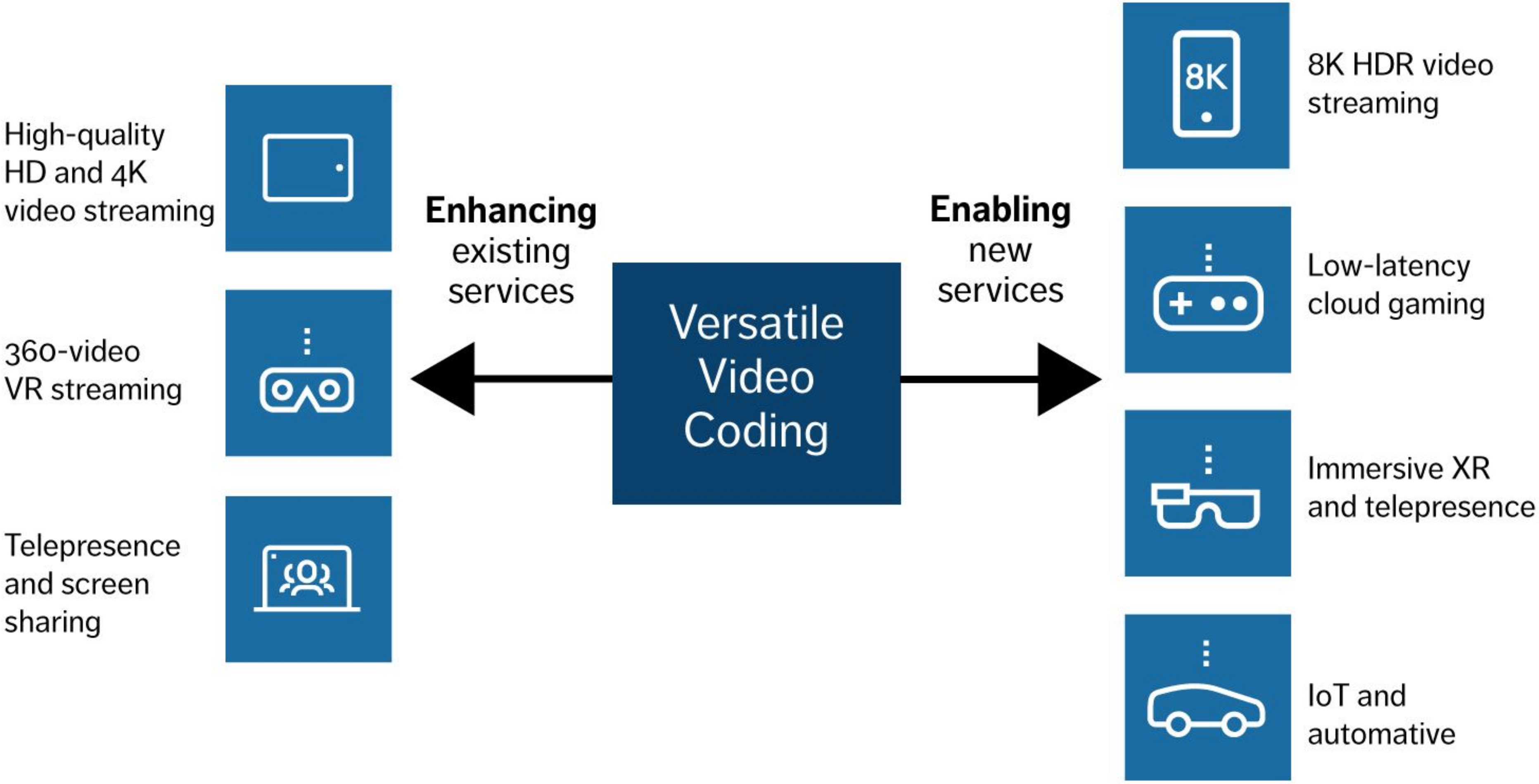
Benjamin Bross

Fraunhofer Heinrich Hertz Institute, Berlin

Versatile Video Coding (VVC) H266

- latest and most sophisticated video codec to date
- offers the highest compression efficiency of all video codecs
- it is particularly appropriate for higher resolution video streams due to its coding tools, which can operate on block sizes of up to 128x128 pixels and with 64x64 sample size transforms.
- can achieve a reduction in bitrate of **around 40 percent** for existing HD and 4K services deployed with HEVC, at the same visual quality.

Versatile Video Coding



Versatile Video Coding (VVC)

- versatility
  by efficient coding of a wide range of video content and applications.
- VVC includes tools specialized for different kinds of content:
  - video beyond standard- and high-definition including high resolution (up to 8K or even larger), high dynamic range (HDR), and wide color gamut;

Versatile Video Coding (VVC)

Joint ITU-T (VCEG) and ISO/IEC (MPEG) project

Coding Efficiency

50% over H.265/HEVC

HD / UHD / 8K resolutions

10bit / HDR

Versatility

Screen content

Adaptive resolution change

Independent sub-pictures

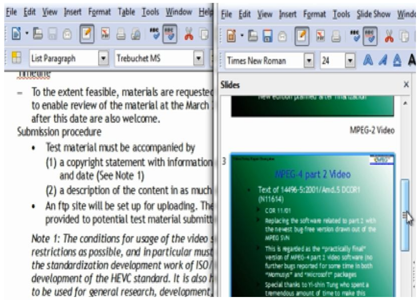
Many incremental improvements of classic hybrid video coding design

- Flexible Block Partitioning with Multi-type Tree (MTT)
- Separate Tree for Luma and Chroma (CST)
- Dependent Quantization (DQ)
- Bi-prediction with CU weights (BCW)
- Decoder-side motion vector refinement (DMVR)
- Symmetric motion vector difference (SMVD)
- Multiple Transform Set (MTS)
- Combined intra/inter prediction (CIIP)
- Low frequency non-separable transform (LFNST)
- Multi-reference line intra prediction (MRL)
- Adaptive Loop Filter (ALF)
- Intra block copy mode (IBC)
- Joint coding of chrominance residuals (JCCR)
- Sub-block transform (SBT)
- Affine Motion Compensation
- Intra sub-partitioning (ISP)
- Subblock-based Temporal Merging Candidates
- Matrix based intra prediction (MIP)
- Adaptive motion vector resolution (AMVR)
- Cross-component Linear Model (CCLM)
- Triangular partition mode (TPM)
- Luma mapping with chroma scaling (LMCS)
- Bi-directional optical flow (BDOF)
- Transform Skip Residual Coding (TSRC)
- Merge with MVD (MMVD)
- Quantized residua1l1DPCM

VVC – Versatility
Screen content coding (SCC)

- Application: new emerging content
  - Gaming
  - Screen sharing / remote desktop
  - …

Arena of Valor (1080p60)



Slide Editing (720p30)

- Problem: Video codecs typically optimized for natural video
- Solution: Special screen content coding tools

  VVC supports SCC already in v1

VVC – Versatility
Reference picture resampling (RPR)

Application: Adaptive streaming with resolution switching

- Problem: Pictures with different resolutions cannot reference each other in inter-picture prediction -> reduces coding efficiency
- Solution: Resample reference picture in case of different resolutions VVC supports reference picture resampling

More efficient resampling filters currently under investigation

RPR as enabler for spatial scalability in VVC v1 (exact design under investigation)

VVC – Versatility
Independent sub-pictures

- Application: Tiled streaming of 360-degree videos

- Problem: Managing a decoder pixel budget dynamically post-encoding

  -> throwing 24K video (parts) at a 4K decoder
- Solution: More efficient coding of independent sub-pictures

  (in-picture padding) Flexible block addressing for easier

  extraction and merging of sub-pictures

  HLS design to avoid slice header rewriting

Immersive video - 360-degree video

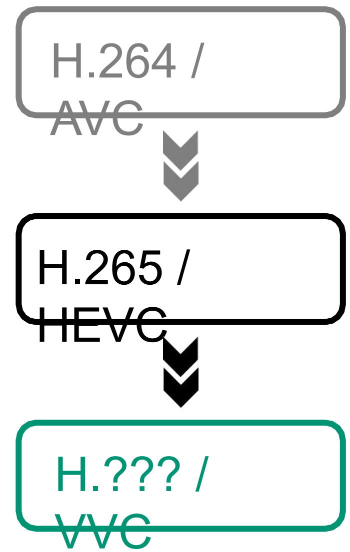
- immersive video is video in which every angle is recorded.

- During playback, which at present typically occurs on a virtual reality (VR) headset, the user is not constrained to a particular view but can look around freely. When the viewer does not change position, this is called three degrees of freedom (3DoF) immersive video, or more commonly 360-degree video.

- the use of subpictures in VVC significantly reduces the complexity of application systems.

Summary VVC

- Coding Efficiency – VVC Test Model 6.1 over HEVC (HM)
  - 38% PSNR-based bitrate reduction for HD and UHD
  - 8.9x encoder and 1.6x decoder runtime
- Versatility – enabled by:
  - Screen content coding tools (gaming, screen sharing,…)
  - Reference picture resampling (adaptive streaming)
  - Potential spatial scalability using RPR filters
  - Independent sub-pictures (360 video, ROI)
- Final Standard by July 2020



Benjamin Bross, Fraunhofer Heinrich Hertz Institute, Berlin