H.264 AVC (Advanced video coding)

references:
- Ostermann, J. ; Bormans, J. ; List, P. ; Marpe, D. more authors,
  Video coding with H.264/AVC: tools, performance, and complexity, Circuits and Systems Magazine, IEEE
- Iain E. G. Richardson, H.264 and MPEG-4 Video Compression Video Coding for Next-generation Multimedia, The Robert Gordon University, Aberdeen, UK
- http://twins.ee.nctu.edu.tw/courses/multimedia_c_05spring/handout/MMC13%20H264.pdf

Why the buzz about H.264? It's the bit rate!

- H.264 can encode video with approximately **3 times fewer bits** than comparable MPEG-2 encoders.
- H.264 is up to **twice** as efficient as MPEG-4 Part 2 (natural video) encoding

Figure.1A MPEG-2

Figure.1B H.264

In figure 1, you can see the difference of encoding via MPEG-2 and H.264 at 100 kpbs.

When and Where will You See H.264/AVC?

- H.264 is the most widely used codec on the planet, with significant penetration in optical disc, broadcast, and streaming video markets.
  - BBC HD, ITV HD, Sky HD
  - Youtube, Apple, Microsoft, Adobe, Hulu, etc.
  - Browser support:

IE	Edge *	Firefox	Chrome	Safari	Opera	iOS Safari *	Opera Mini *	Android Browser *	Chrome for Android
8								4.3	
9			45					4.4	
10	12	42	46			8.4		4.4.4	
11	13	43	47	9	34	9.2	8	46	47
	14	44	48		35				
		45	49		36				
		46	50						

The *Scope* of Picture and Video Coding Standardization

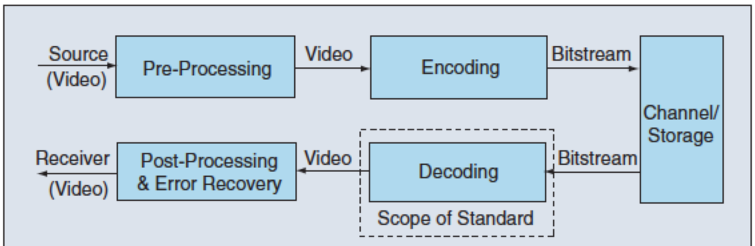


Figure 1. Scope of video coding standardization: Only the syntax and semantics of the bitstream and its decoding are defined.

- The H.264 standard defines the syntax and semantics of the bit stream as well as the processing that the decoder needs to perform.
- Manufactures of decoders can only compete in areas like cost and HW requirements.
- Optional post-processing is another area where different manufactures will provide competing tools to create a decoded video stream optimized for the targeted application.

MPEG Standards

- MPEG-1
  - ISO/IEC 11172
  - 1993 IS, design focus on non-interlaced SIF (352x240)@1.5Mbps
- MPEG-2
  - ISO/IEC 13818
  - 1994 IS, Optimized at "NTSC quality" CCIR601 video@6-10Mbps
- MPEG-4
  - officially ISO/IEC 14496
  - Part 2. video : 2001 IS, content based video coding
    - Roughly follows H.263 design and adds all prior features and various "trick modes" and (most important) shape coding
  - **Part 10. advance video coding (AVC) – ITU H.264**
    - 2004 IS, 50% bit rate reduction than other video standard
    - **Similar structure but vary in every details**

JVT/H.264/AVC

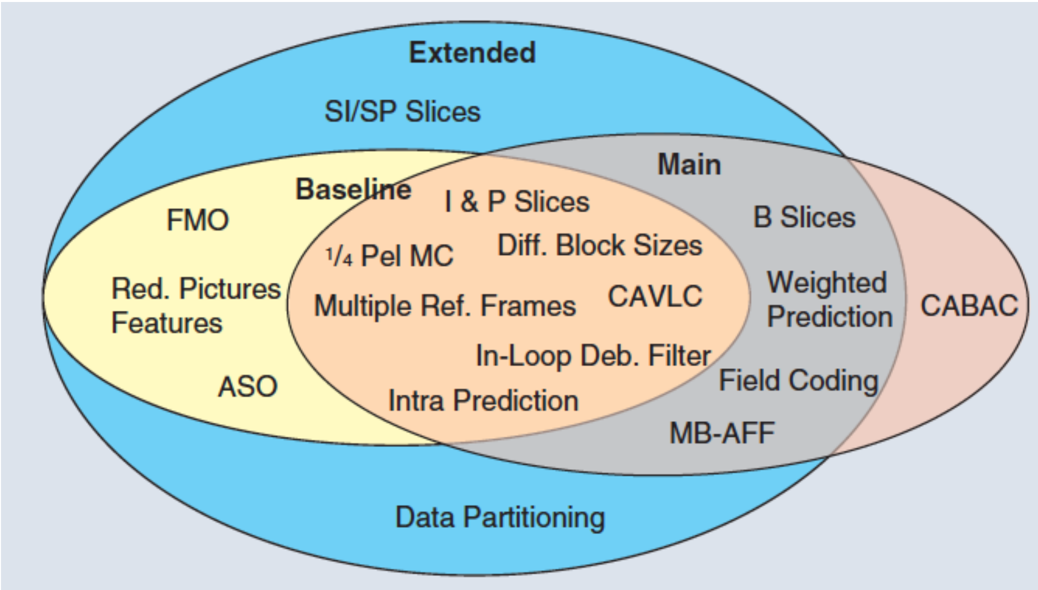
- Same design to be approved in both ITU-T and MPEG
- In ITU-T this will be a new & separated standard
  - ITU-T Recommendation H.264
  - ITU-T systems (H.32x) to be modified to support it
- In MPEG this will be a new "part" in the MPEG-4 suite
  - Separate codec design from prior MPEG-4 visual
  - New part 10 called "advanced video coding" (similar to "AAC" position in MPEG-2 as separate codec)
  - Not backward compatible with prior standards (incl. the prior MPEG-4 visual spec. – core technology is different)
  - MPEG-4 Systems / File Format modifying to support it

Source: The Emerging H.264/AVC Video Coding Standard, by T. McMahon 9

Profiles and Levels

- Many standards contain different configurations of capabilities – often based in “profiles” & “levels”
  - ***A profile*** is usu. a set of **algorithmic features**
  - ***A level*** is usu. a degree of **capability** (e.g., pixel resolution or speed of decoding, # of MB per second)
- H.264/AVC currently has four Profiles
  - **Baseline** (good for most applications up through D-Cinema)
  - **Main** (adds interlace, B-Slices and CABAC efficiency gains)
  - **Extended** (the so-called streaming profile)
  - **High** (adds ABT, 4:2:2, 4:4:4, for digital cinema)
    - 4:2:0/8 bit: High
    - 4:2:0/10: High 10
    - 4:2:2/10: High 4:2:2
    - 4:4:4/12: High 4:4:4
- H.264/AVC has many (11) Levels
  - Built to match popular international production and formats
  - From QCIF to D-Cinema

Profiles and Levels of H.264



Profiles and Levels of H.264 (1)

- Three profiles now: Baseline, Main, and Extended
- Baseline (Progressive, Videoconferencing & Wireless)
  - I and P slices (not B)
  - In-loop deblocking filter
  - 1/4-sample motion compensation
  - Tree-structured motion segmentation down to 4x4 block size
  - VLC-based entropy coding
  - Some enhanced error resilience features
    - Flexible macroblock ordering/arbitrary slice ordering
    - Redundant slices

Profiles and Levels of H.264 (2)

- Main Profile (esp. Broadcast)
  - All Baseline features except enhanced error resilience features
  - B pictures
  - CABAC
  - MB-level frame/field switching
  - Adaptive weighting for B and P picture prediction
  - Interlace
  - Note. Main profile is not a superset of baseline profile
- Extended Profile
  - All Baseline features
  - B pictures
  - More error resilience: Data partitioning
  - SP/SI switching pictures
  - Note. Extended profile is a superset of baseline profile

How to support video everywhere?

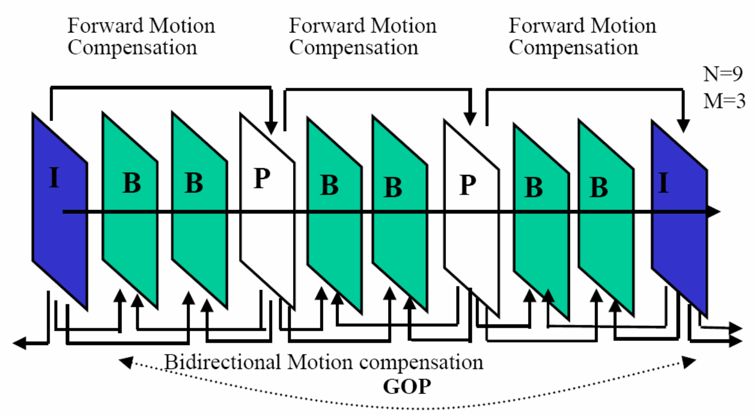
Video Resolution and Format	Video Frame Rate (fps)	Bit-rate (kbps)	H.264 Profile
192x108p	6.25	31	Baseline
192x108p	25	86	Baseline
256x144p	25	156	Baseline
384x216p	25	281	Baseline
448x252p	25	437	Baseline
512x288p	25	437	Main
640x360p	25	688	Baseline
704x396p	25	827	Main
544x576p	25	929	High
704X396p	50	1570	Main
704x576i	25	1570	High
896x504p	25	1374	High
960x540p	50	2812	High
1280x720p	50	5070	High
1920x1080i	25	8000	High

VIDEOFACTORY

The Video Factory Project is a BBC in-house programme designed to encode, package and deliver all iPlayer video content.

Technical Summary of MPEG-1/2

- Compression based on Motion-compensated interframe DCT
- "Group of pictures (GOP)" basic video unit for compression
- Motion compensation with I, P, B frame structure
  - Intra (I) frames (intraframe DCT)
  - Predicted (P) frame (DCT with forward MC)
  - Bidirectional(B) frame (DCT with forward/backward MC)



Two-Layer Design of AVC

AVC employs a two layer design:

- The video coding layer (VCL) defines the efficient representation of the video.
- The network adaptation layer (NAL) converts the VCL representation into a format suitable for specific transport layers for storage media.
  - For circuit-switched transport like MPEG-2, the NAL delivers the coded video as an ordered stream of bytes containing start codes such that these transport layers and the decoder can robustly and simply identify the structure of the bit stream.

Byte Stream Format

  - For packet switched networks like RTP/IP or TCP/IP, the NAL delivers the coded video in packets without these start codes.

AVC Layered Structure

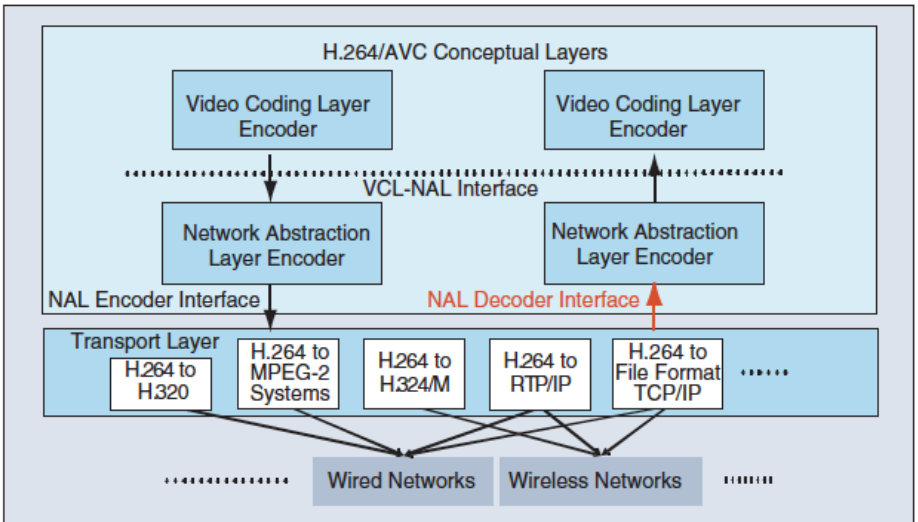
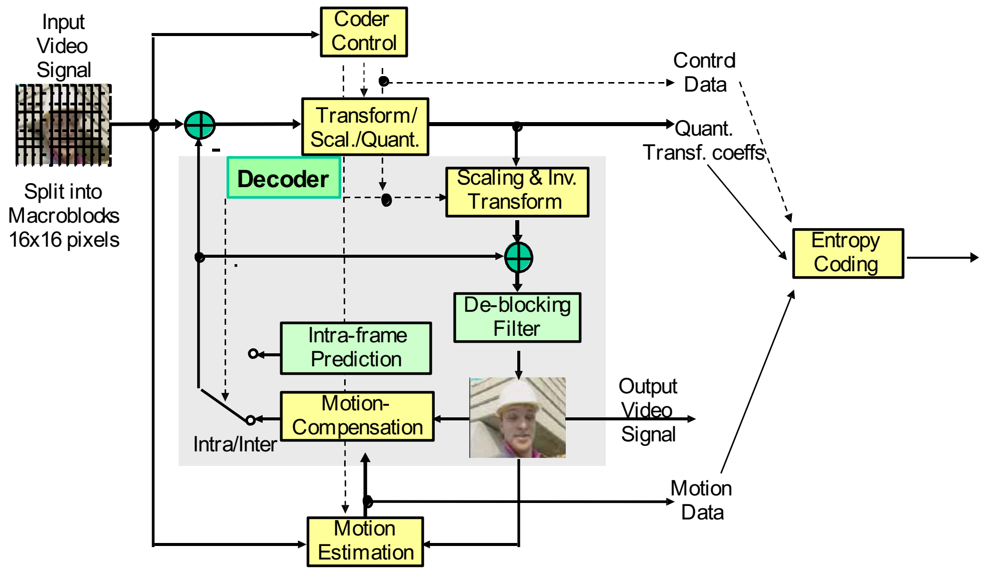


Figure 2. H.264/AVC in a transport environment: The network abstraction layer interface enables a seamless integration with stream and packet-oriented transport layers (from [7]).

AVC Video Features

- Predictive coding
  - 13-mode luma, 4-mode chroma intra prediction
  - Gazillion modes for inter prediction
- 16-bit integer combined transform/quantization
  - Exact forward-inverse transform pair is used
  - Transform block size is 4x4
- Two different entropy coding methods
  - Universal VLC and Context Adaptive VLC
  - Context Adaptive Binary Arithmetic Coding
- In-loop filter

Basic Structure for Video Standard

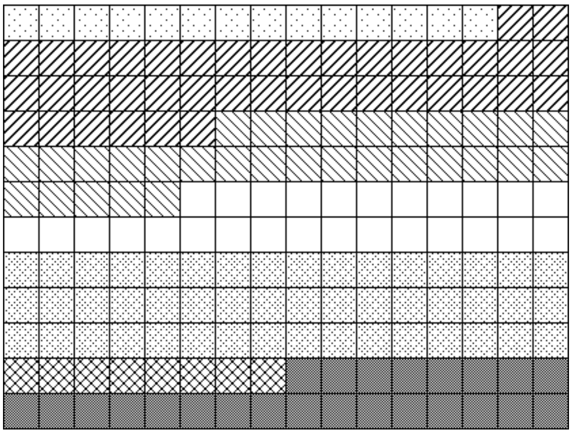


Notion of slices

Added notion of *slice* for synchronization after loss/corrupt data.

- Each slice is independently decodable.
- Errors do no propagate across the boundaries of slice.

Example: picture with 7 slices:



Slices Types

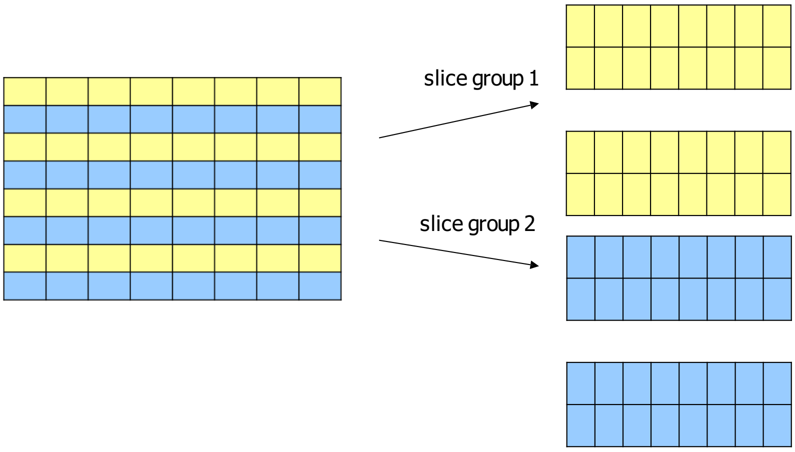
- Each slice type limits the MB partition types and prediction modes.

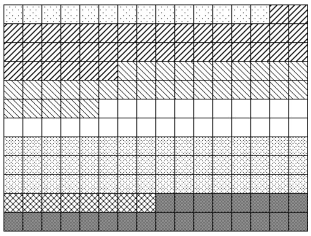
Slice type	Description	Profile(s)
I (Intra)	Contains only I macroblocks (each block or macroblock is predicted from previously coded data within the same slice).	All
P (Predicted)	Contains P macroblocks (each macroblock or macroblock partition is predicted from one list 0 reference picture) and/or I macroblocks.	All
B (Bi-predictive)	Contains B macroblocks (each macroblock or macroblock partition is predicted from list 0 and/or list 1 reference pictures) and/or I macroblocks.	Extended and Main
SP (Switching P)	Facilitates switching between coded streams; contains P and/or I macroblocks.	Extended
SI (Switching I)	Facilitates switching between coded streams; contains SI macroblocks (a special type of intra coded macroblock).	Extended

list0: previous frames
list1: forward frames

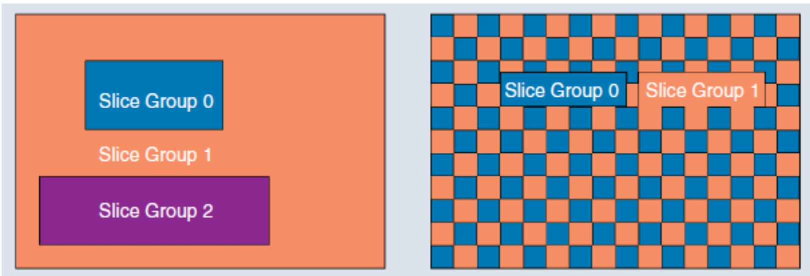
Flexible Macroblock Ordering

- Macroblocks in a frame can be assigned to slice groups
- Each slice group is further divided into slices



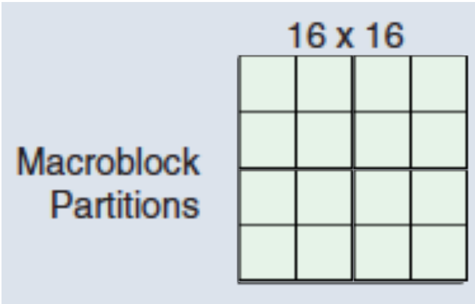


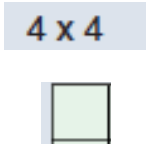
Flexible Macroblock Ordering



Macroblocks might be transmitted out of raster scan order in a flexible and efficient way. Specific macroblock allocation maps enable the efficient application of features such as slice interleaving, dispersed macroblock allocation using checkerboard-like patterns, one or several foreground slice groups and one left-over background slice groups, or sub-pictures within a picture to support, e.g., isolated regions.

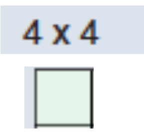
H.264 Coding Scheme (Intra prediction 1)

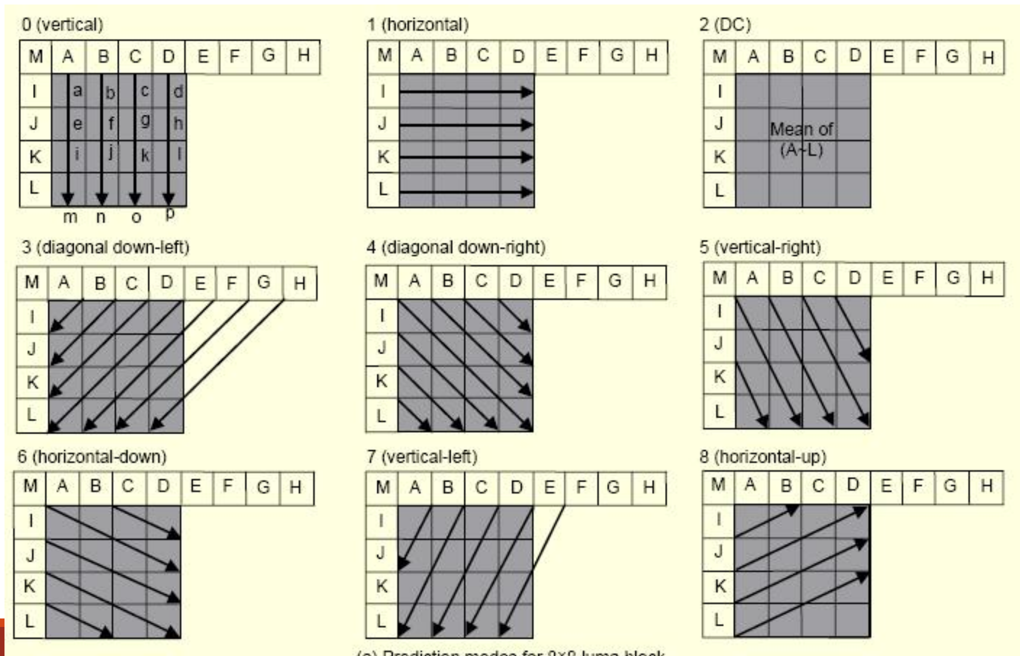




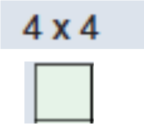
- In H.264/AVC, two different types of intra prediction are possible for the prediction of the luminance component Y.
- The first type is called INTRA_4x4 and the second one INTRA_16x16.

H.264 Coding Scheme (Intra prediction 2)



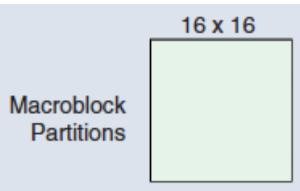


H.264 Coding Scheme (Intra prediction 2)

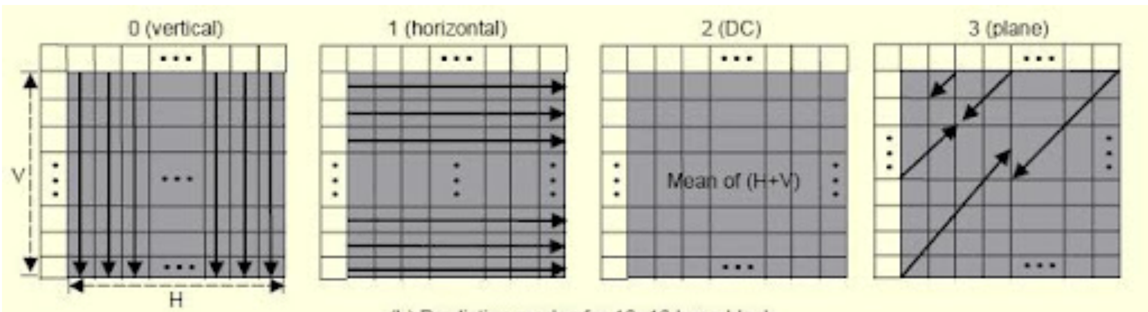


Mode 0 (Vertical)	The upper samples A, B, C, D are extrapolated vertically.
Mode 1 (Horizontal)	The left samples I, J, K, L are extrapolated horizontally.
Mode 2 (DC)	All samples in P are predicted by the mean of samples A . . . D and I . . . L.
Mode 3 (Diagonal Down-Left)	The samples are interpolated at a 45° angle between lower-left and upper-right.
Mode 4 (Diagonal Down-Right)	The samples are extrapolated at a 45° angle down and to the right.
Mode 5 (Vertical-Right)	Extrapolation at an angle of approximately 26.6° to the left of vertical (width/height = 1/2).
Mode 6 (Horizontal-Down)	Extrapolation at an angle of approximately 26.6° below horizontal.
Mode 7 (Vertical-Left)	Extrapolation (or interpolation) at an angle of approximately 26.6° to the right of vertical.
Mode 8 (Horizontal-Up)	Interpolation at an angle of approximately 26.6° above horizontal.

H.264 Coding Scheme (Intra prediction 3)



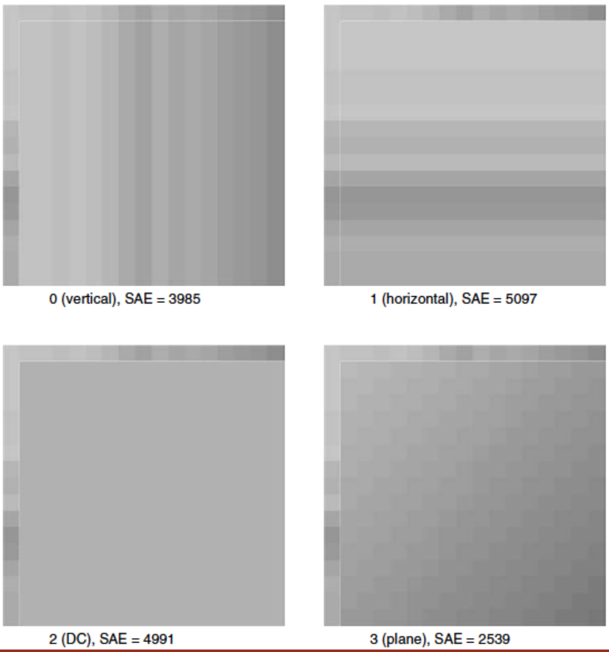
Mode 0 (vertical)	Extrapolation from upper samples (H)
Mode 1 (horizontal)	Extrapolation from left samples (V)
Mode 2 (DC)	Mean of upper and left-hand samples (H + V).
Mode 4 (Plane)	A linear 'plane' function is fitted to the upper and left-hand samples H and V. This works well in areas of smoothly-varying luminance.



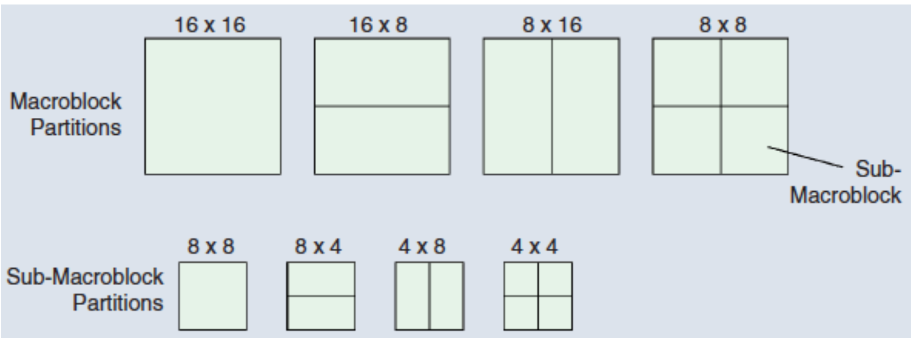
(b) Prediction modes for 16×16 luma block

Very high efficiency if signal is smooth within the entire macroblock.

H.264 Coding Scheme (Intra prediction 3)

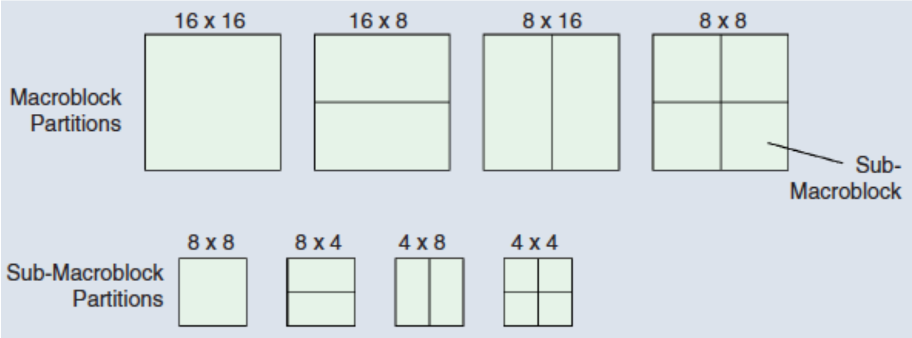


H.264 Coding Scheme (Inter prediction 1)



- In case of motion compensated prediction macroblocks are predicted from the image signal of already transmitted reference images.
- Partitions with luminance block sizes of 16x16, 16x8, 8x16, and 8x8 samples are supported.
- In case of an 8x8 sub-macroblock in a P-slice, one additional syntax element specifies if the corresponding 8x8 sub-macroblock is further divided into partitions with block sizes of 8x4, 4x8 or 4x4.

H.264 Coding Scheme (Inter prediction 2)



- larger partition sizes would lead to fewer bits for encoding the motion vector, but could lead to more bits spent in encoding the residual (actual partition – motion predicted partition) for partitions with high detail.
- On the other hand, choosing smaller partition sizes leads to more bits spent in encoding the motion vector, but potentially fewer bits in encoding the residual.

H.264 Coding Scheme (Inter prediction 2)

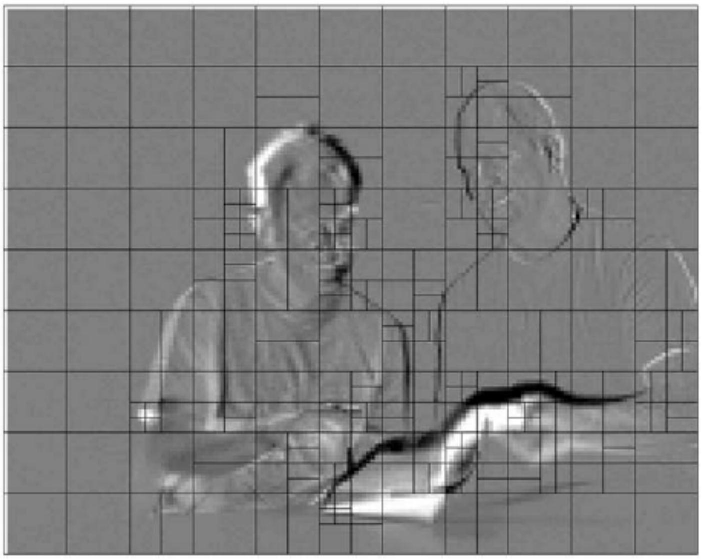
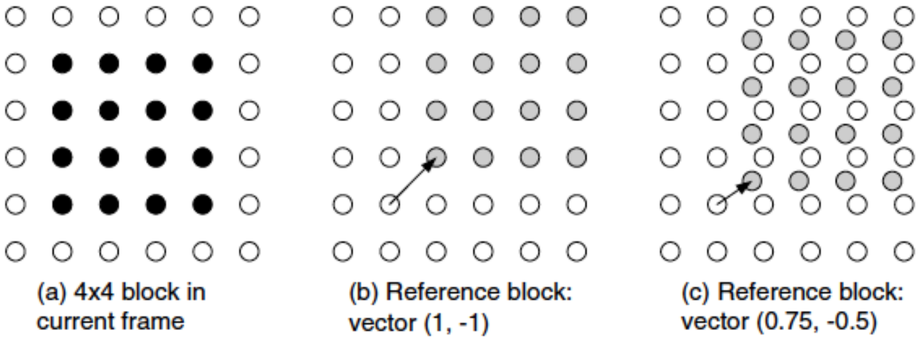


Figure 6.12 Residual (without MC) showing choice of block sizes

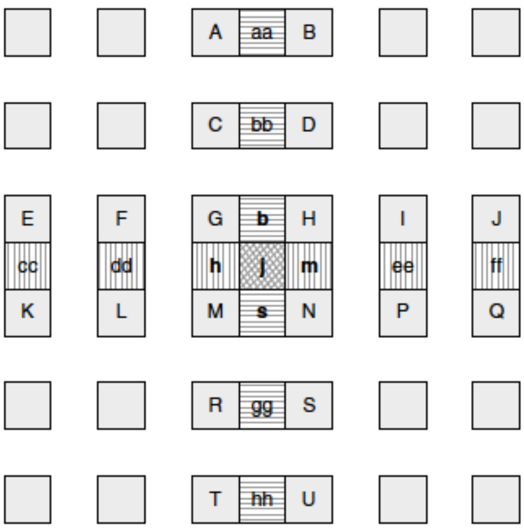
- The H.264 reference encoder selects the 'best' partition size for each part of the frame.
- In this case the partition size that minimises the amount of information to be sent, and the chosen partitions are shown superimposed on the residual frame.
- In areas where there is little change between the frames (residual appears grey), a 16x16 partition is chosen.
- In areas of detailed motion (residual appears black or white), smaller partitions are more efficient.

H.264 Coding Scheme (Inter prediction 3 - MV)



- Each partition or sub-macroblock partition in an inter-coded macroblock is predicted from an area of the same size in a reference picture.
- The offset between the two areas (the motion vector) has quarter-sample resolution for the luma component (and one-eighth-sample resolution for the chroma components).
- The samples at sub-sample positions do not exist in the reference picture and so it is necessary to create them using interpolation from nearby coded samples.

H.264 Coding Scheme (Inter prediction 3 - MV)



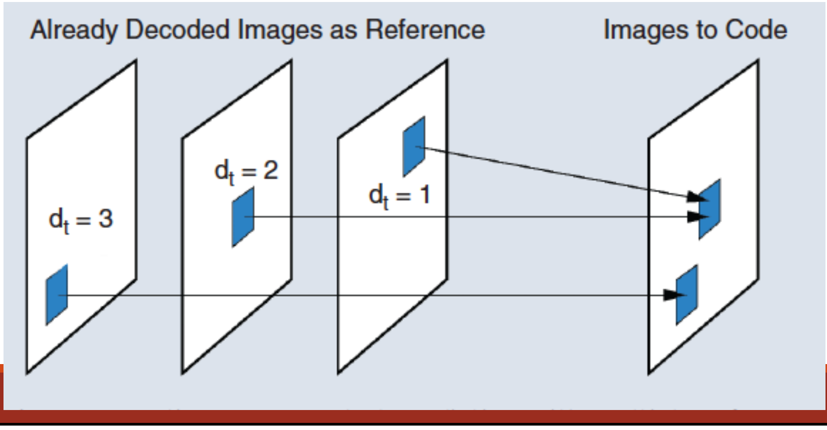
- The samples half-way between integer-position samples ('half-pel samples') in the luma component of the reference picture are generated first.
- Each half-pel sample that is adjacent to two integer samples (e.g. b, h, m, s) is interpolated from integer-position samples using a **six tap Finite Impulse Response (FIR) filter with weights** (1/32,−5/32, 5/8, 5/8,−5/32, 1/32).
- For example, half-pel sample b is calculated from the six horizontal integer samples E, F, G, H, I and J:

$$\mathbf{b} = \mathrm{round}((E - 5F + 20G + 20H - 5I + J)/32)$$

- Once all the half-pel samples are available, the samples at quarter-step ('quarter-pel') positions are produced by **linear interpolation**.

H.264 Coding Scheme (Inter prediction 4)

- In former MPEG standards this reference image is the most recent preceding image.
- In H.264/AVC it is possible to refer to several preceding images. For this purpose, an additional picture reference parameter has to be transmitted together with the motion vector. This technique is denoted as motion-compensated prediction with multiple reference frame.



Transform and quantization of h.264

- Transformation
  - Previous video coding standards relied on Discrete Cosine Transform (DCT)
    - complexity
  - H.264/MPEG-4 part 10, uses an integer transform with a similar coding gain as a 4x4 DCT.
    - considered as an approximation of 4x4 DCT
    - integer or fixed-point arithmetic

Transform and quantization of h.264

- Quantisation
  - A total of 52 values of Qstep are supported by the standard, indexed by a Quantisation Parameter (QP).
  - Qstep doubles in size for every increment of 6 in QP.
  - The wide range of quantiser step sizes makes it possible for an encoder to control the tradeoff accurately and flexibly between bit rate and quality.

Quantisation step sizes in H.264 CODEC

QP	0	1	2	3	4	5	6	7	8	9	10	11	12	…
QStep	0.625	0.6875	0.8125	0.875	1	1.125	1.25	1.375	1.625	1.75	2	2.25	2.5	…
QP	…	18	…	24	…	30	…	36	…	42	…	48	…	51
QStep		5		10		20		40		80		160		224

Transform and quantization of h.264

- Quantisation

Original frame (violin frame 2)

Reconstructed, QP = 32 (no filter)

Reconstructed, QP = 36 (no filter)

Transform and quantization of h.264

- de-blocking filter
  - A filter is applied to each decoded macroblock to reduce blocking distortion.
  - The deblocking filter is applied after the inverse transform in the encoder (before reconstructing and storing the macroblock for future predictions) and in the decoder (before reconstructing and displaying the macroblock).
  - The filter smooths block edges, improving the appearance of decoded frames.
  - The filtered image is used for motion-compensated prediction of future frames and this can improve compression performance because the filtered image is often a more faithful reproduction of the original frame than a blocky, unfiltered image.

Transform and quantization of h.264

Reconstructed, QP = 36 (no filter)

Reconstructed, QP = 36 (with filter)

Coding

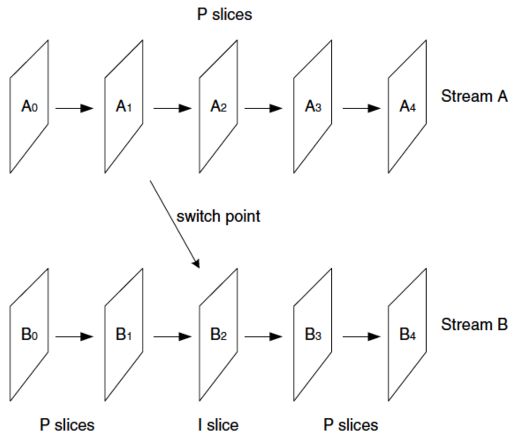
In those earlier methods, specifically tailored but fixed variable length codes (VLCs) such as Huffmann were used. it was implicitly assumed that the underlying statistics are stationary. However, residual data in a motion-compensated predictive coder shows a highly non-stationary statistical behavior.

H.264/AVC specifies two alternative methods of entropy coding:

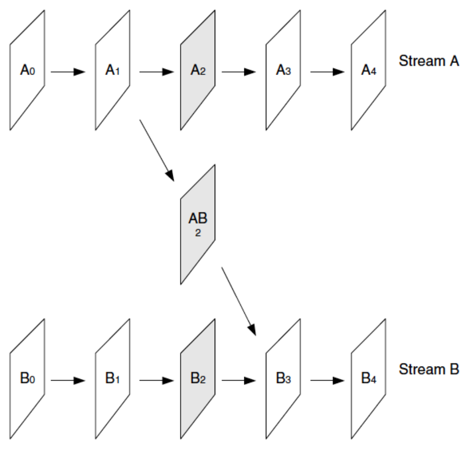
- CAVLC: a low-complexity technique based on the usage of context-adaptively switched sets of variable length codes.
- CABAC: the computationally more demanding algorithm of context-based adaptive binary arithmetic coding (CABAC).
- Both methods represent major improvements in terms of coding efficiency.

Error Robustness and Network Friendliness

1. The video coding layer (VCL) and the network adaptation layer (NAL)
2. Intra-frames, slicing, Flexible Macroblock Ordering
3. Switching between different bit streams encoded at different bit rates (version switching). SP and SI slices are specially-coded slices that enable efficient switching.

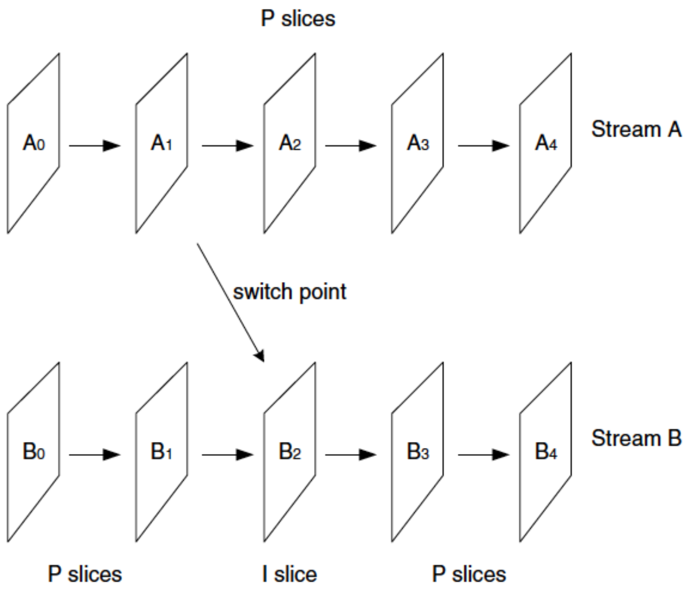


Switching streams using I-slices



Switching streams using SP-slices

Error Robustness and Network Friendliness

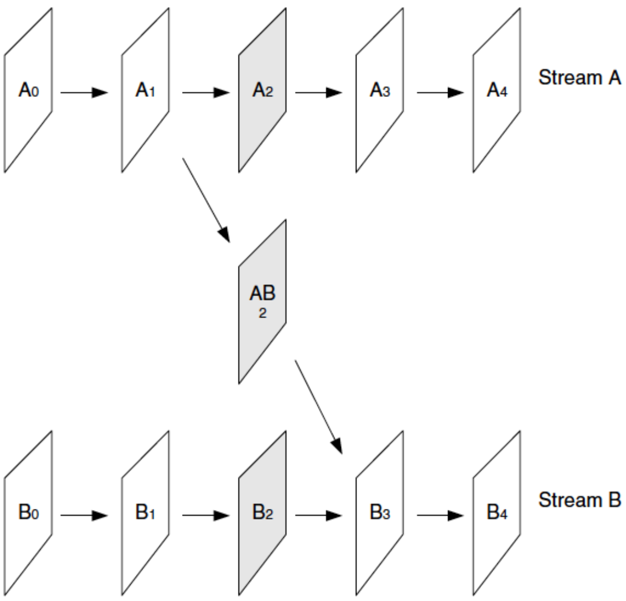


Switching streams using I-slices

Challenge: switching from Stream A to Stream B:

- One solution is to code frame B2 as an I-slice.
- Because it is coded without prediction from any other frame, it can be decoded independently of preceding frames in stream B.
- However, an I-slice is likely to contain much more coded data than a P-slice and the result is an undesirable peak in the coded bitrate at each switching point.

Error Robustness and Network Friendliness



Switching streams using SP-slices

Challenge: switching from Stream A to B:

- SP-slices are designed to support switching between similar coded sequences.
- At the switching point (frame 2 in each sequence), there are three SP-slices, each coded using motion compensated prediction (making them more efficient than I-slices).
- The key to the switching process is SP-slice AB2, created in such a way that it can be decoded using motion-compensated reference picture A1, to produce decoded frame B2.

Error Robustness and Network Friendliness

4. Data partitioning

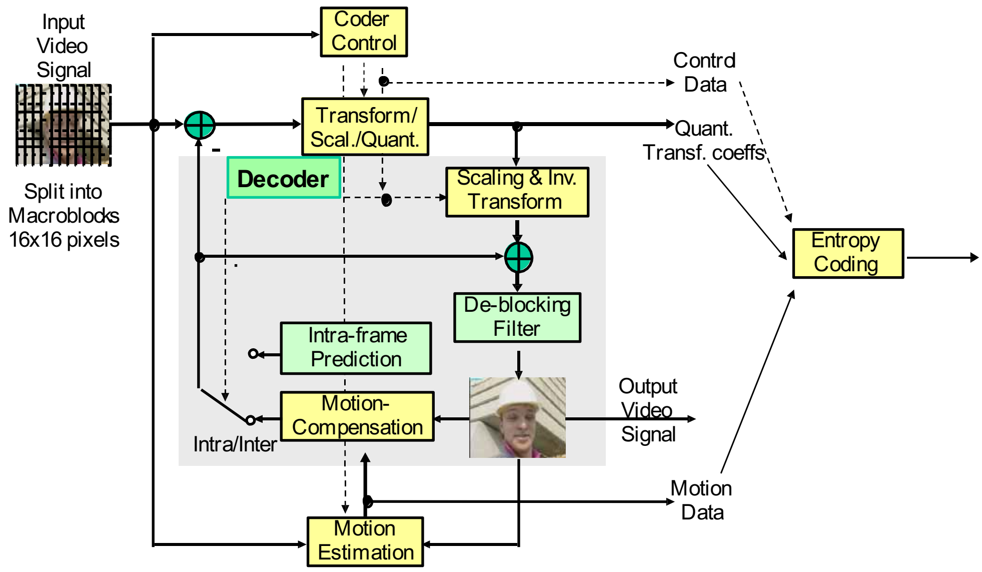
The coded data that makes up a slice is placed in three separate Data Partitions (A, B and C),each containing a subset of the coded slice.

- Partition A contains the slice header and header data for each macroblock in the slice
- Partition B contains coded residual data for Intra and SI slice macroblocks
- Partition C contains coded residual data for inter coded macroblocks (forward and bi-directional).

Each Partition can be placed in a separate NAL unit and may therefore be transported separately. (Why?)

- If Partition A data is lost, it is likely to be difficult or impossible to reconstruct the slice, hence Partition A is highly sensitive to transmission errors.
- Partitions B and C can (with careful choice of coding parameters) be made to be independently decodable and so a decoder may (for example) decode A and B only, or A and C only, lending flexibility in an error-prone environment.

Basic Structure for Video Standard



Coding efficiency

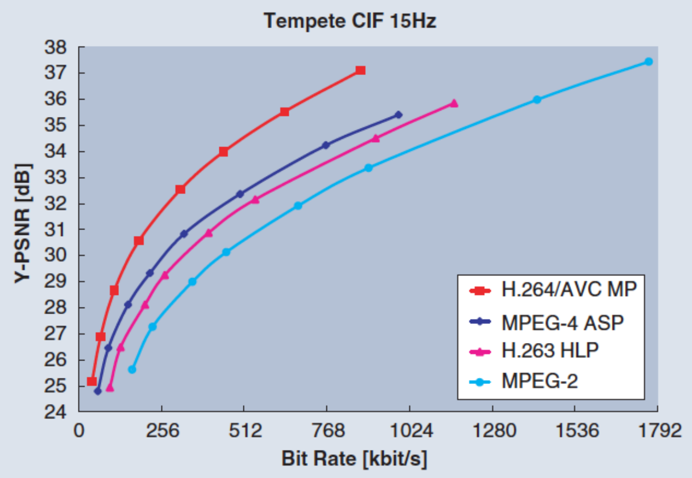


Figure 21. Luminance PSNR versus average bit rate for different coding standards, measured for the test sequence *Tempete* for video streaming applications (from [36]).

Video streaming application:

- H.264/AVC MP (Main Profile),
- MPEG-4 Visual ASP (Advanced Simple Profile),
- H.263 HLP (High Latency Profile),
- MPEG-2 Video ML@MP (Main Level at Main Profile)

Table 1.
Average bit rate savings for video streaming applications (from [10]).

Coder	Average Bit Rate Savings Relative To: MPEG-4 ASP	H.263 HLP	MPEG-2
H.264/AVC MP	37.44%	47.58%	63.57%
MPEG-4 ASP	–	16.65%	42.95%
H.263 HLP	–	–	30.61%

Conclusion

H.264/AVC provides an improved coding efficiency and a significant improvement in flexibility for effective use over a wide range of networks.

While H.264/AVC still uses the concept of block-based motion compensation, it provides some significant changes:

- Enhanced motion compensation capability using high precision and multiple reference frames
- Use of an integer DCT-like transform instead of the DCT
- Enhanced adaptive entropy coding including arithmetic coding
- Adaptive in-loop de-blocking filter