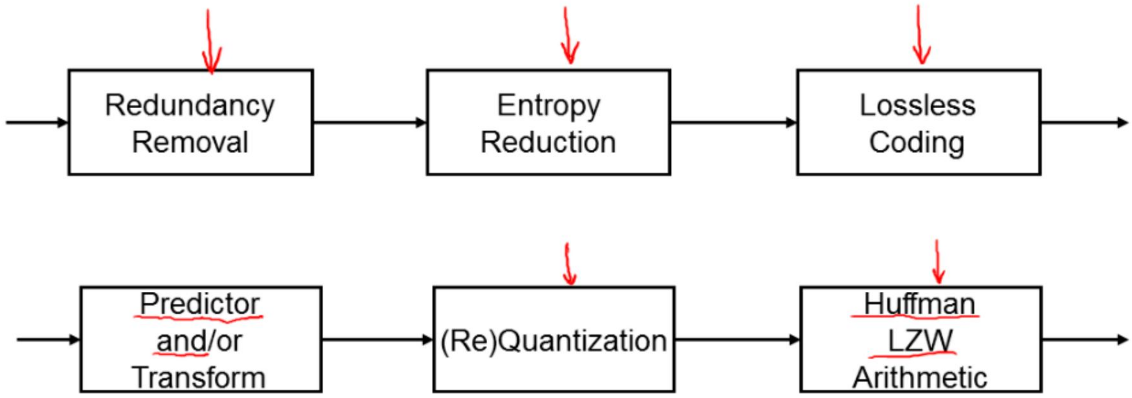
Data Compression System



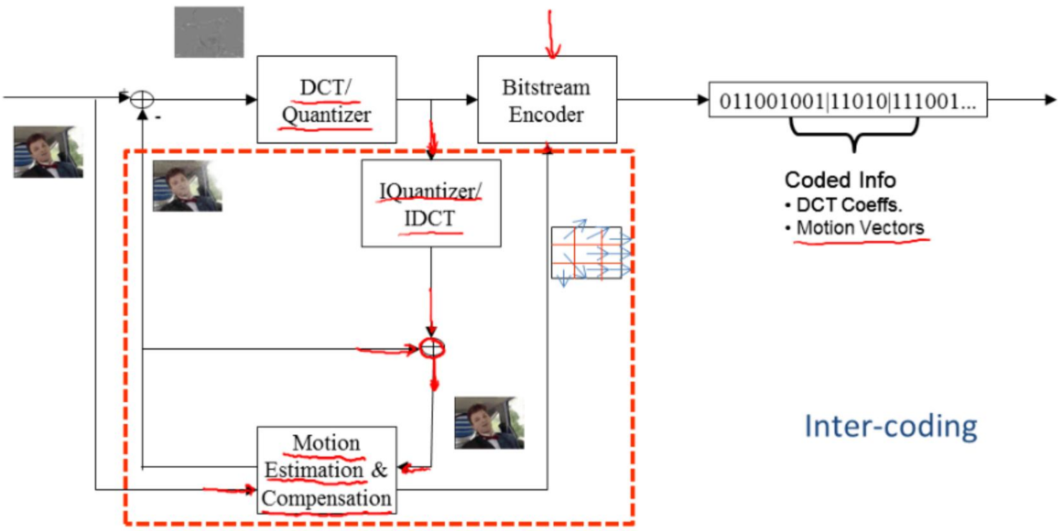
Outline

- Applications
- Hybrid Motion Compensated Video Coding
- Video Coding Standards (H.261, H.263, MPEG-1, MPEG-2, MPEG4, H.264, H.265)

Applications

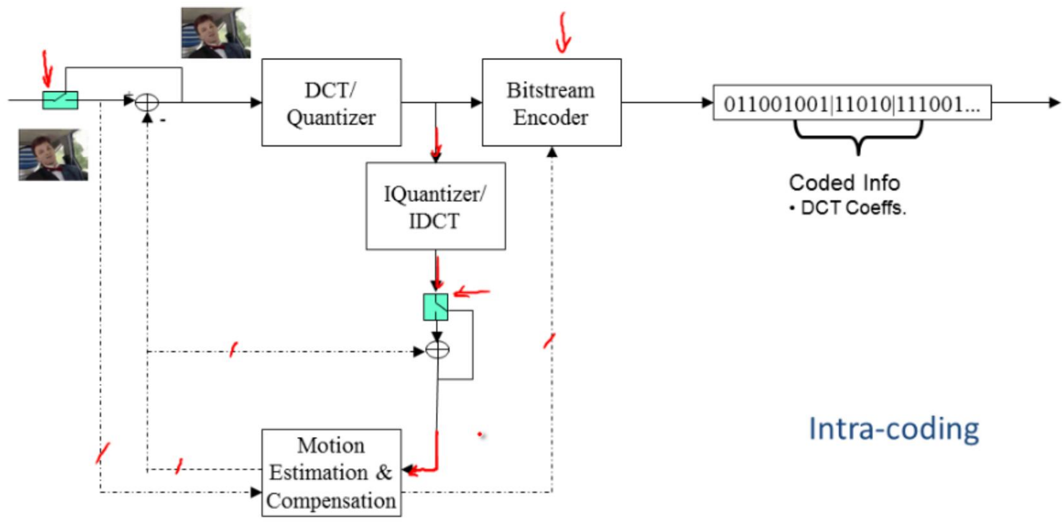
- Video content acquisition and editing systems
- DVDs, Blu-ray discs
- Broadcast of TV signals over satellite, cable, terrestrial transmission systems
- Mobile and internet network video
- Real-time conversation applications (video chat, video conferencing, telepresence systems)
- Security applications
- Stereo and multi-view capture, display, and transmission

Hybrid Motion Compensated Video Encoding



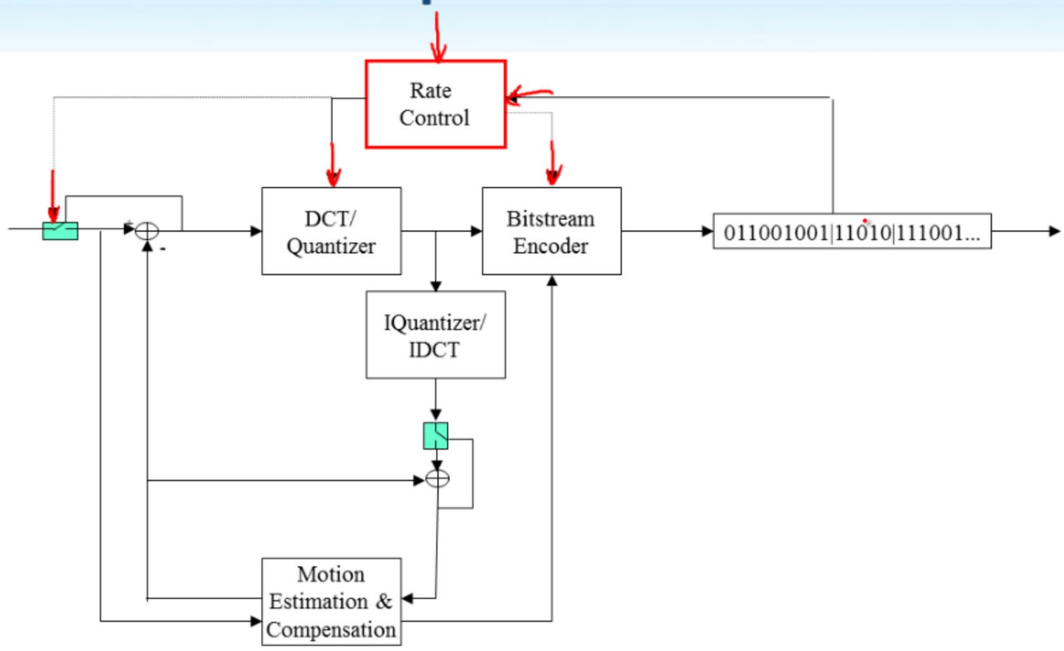
Inter-coding

Hybrid Motion Compensated Video Encoding

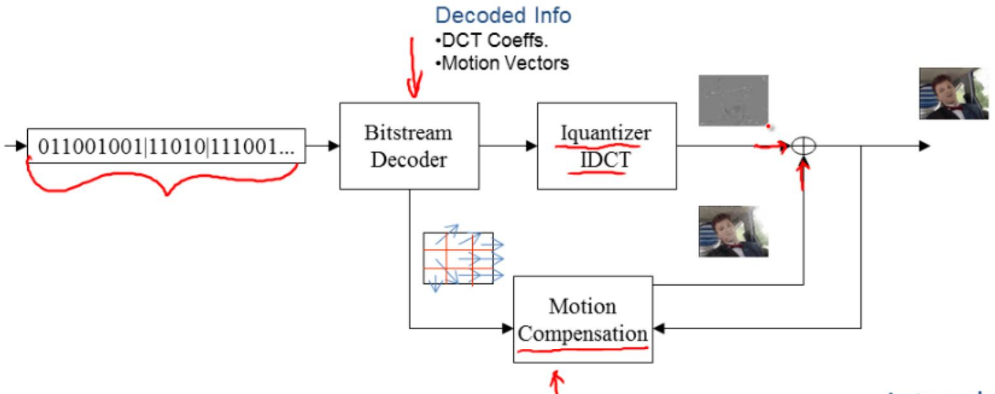


Intra-coding

Hybrid Motion Compensated Video Encoding

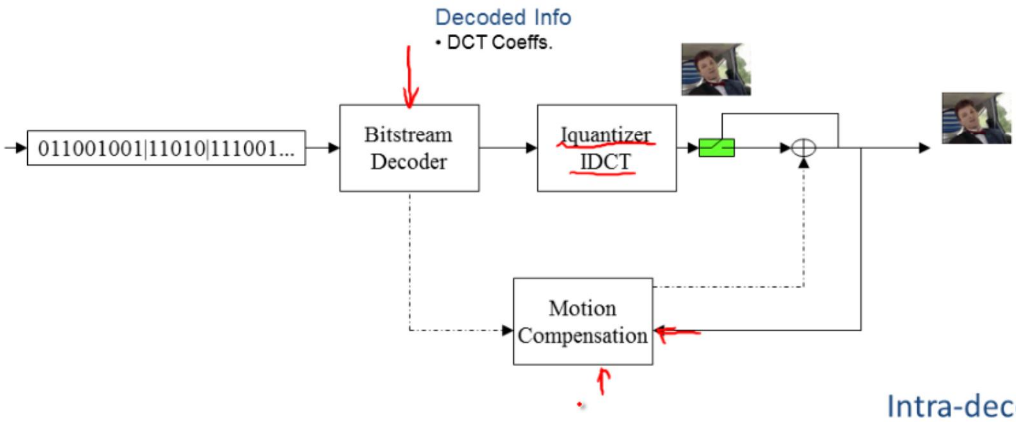


Hybrid Motion Compensated Video Decoding



Inter-decoding

Hybrid Motion Compensated Video Decoding



Intra-decoding

International Video Coding Standards

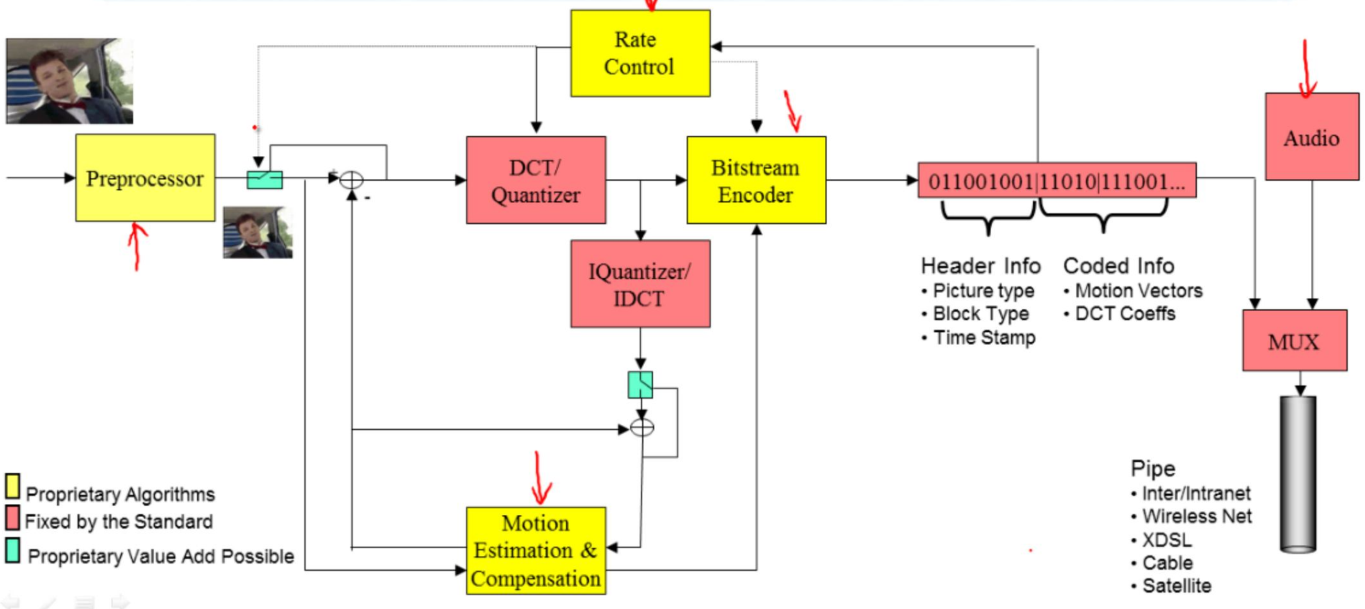
- ITU-T Video Coding Experts Group (VCEG):
  - H.261, H.263
- ISO/IEC Moving Pictures Experts Group (MPEG):
  - MPEG-1, MPEG-4 Visual
- JCT-VC (Joint Collaborative Team on Video Coding):
  - H.262/MPEG-2 Video
  - H.264/MPEG-4 AV
  - H.265/MPEG-H Part 2

Rule of Thumb: Same performance with previous standard at ½ the bit rate!

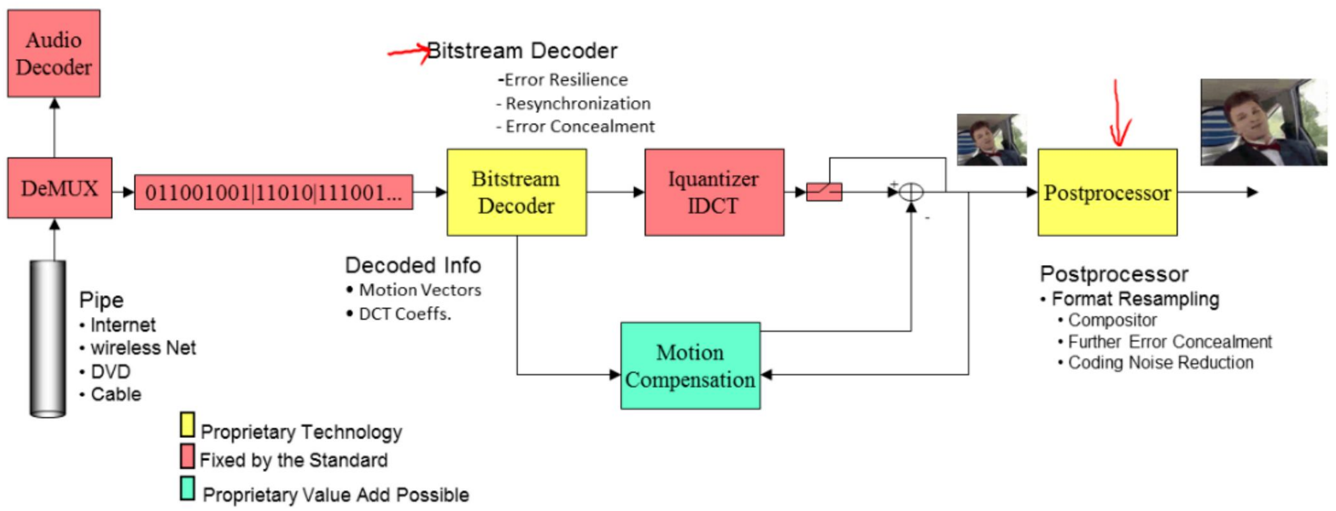
International Video Coding Standards

- They all use the Hybrid Motion Compensated Encoding Scheme
- All standards are *generic*. They do not specify the operations of encoder but the syntax and semantics of the coded bitstream and the decoding process
- Standards follow a toolkit approach: application need use only the features it needs

Hybrid Motion Compensated Video Encoding



Hybrid Motion Compensated Video Decoding

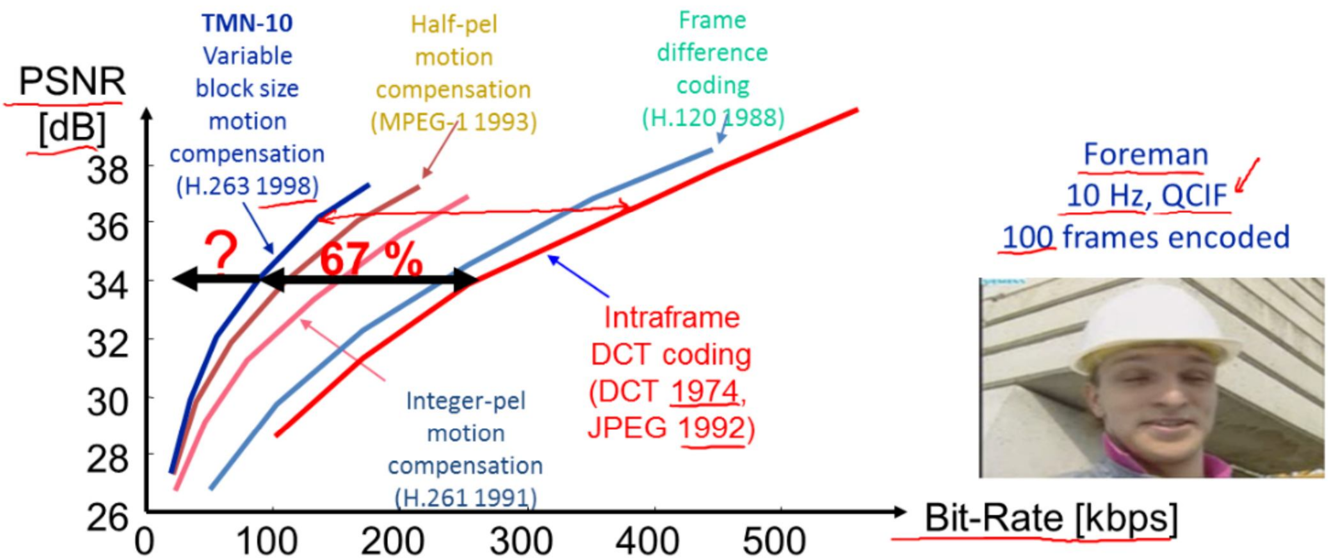


Early Standards Comparison

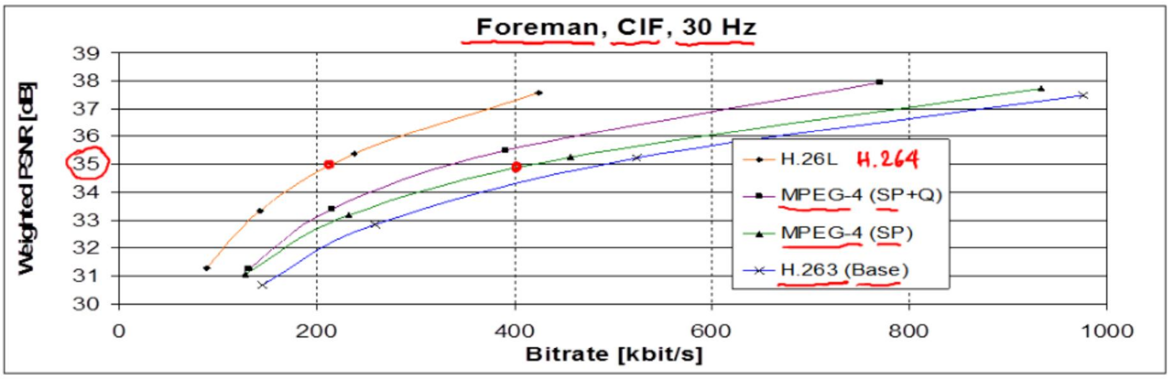
MPEG
Entertainment ⟶ ⟵ Communication
ITU

	MPEG-1	MPEG-2	MPEG-4	H.263	H.261
Dates of Standardization	11/92	11/94	1/99 Version 1 1/00 Version 2	5/96 Version 1 1/98 Version 2	12/90 Version 1 5/94 Revised
Primary Applications	Digital Storage Media	Broadcast/ DVD/ HDTV	Web Authoring, Multimedia Compression, Wireless Videophone	Desktop/ Wireless Video-conferencing	Wireline Video-conferencing
Typical Video Bitrates	1.5 Mbps	4-6 Mbps	20 Kbps - 6 Mbps	20-384 Kbps	128-384 Kbps
Typical Video Frame Size	352x240 (SIF)	720x480 (Rec. 601)	176x144 (QCIF) 352x288 (CIF) 720x480 (601)	176x144 (QCIF) 352x288 (CIF)	176x144 (QCIF) 352x288 (CIF)
Typical Associated Audio Quality	Stereo CD Quality	Surround Sound	Speech/Music/ Stereo CD/Surround Sound	Speech	Speech

Standards Comparison



Standards Comparison



Average Bitrate Savings

Coder	H.264/MPEG-4 AVC HP	MPEG-4 ASP	H.263 HLP	MPEG-2/ H.262 MP
UHD HEVC MP H.265	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%

QCIF / CIF

For equal PSNR for entertainment applications

http://www.streamingmedia.com

H.261 Video Compression Standard

- Part of H.320 family of recommendations
- Approved by ITU-T Dec. 1990
- Application: provision of video telephony and videoconferencing services over ISDN
- Rates: p x 64 kbps, p=1,…,30
- Picture Formats
  - Progressive (non-interlaced)
  - Fixed 4:3 aspect ratio
  - CIF (videoconferencing) Y=352x288, Cb=Cr=176x144, 30 fps
  - QCIF (video telephony) Y=176x144, Cb=Cr=88x72, 15 or 7.5 fps

H.263 Video Compression Standard

- Part of H.324 family of recommendations
- Approved by ITU-T 1995
- Application: video transmission over wireless and PSTNs
  - Video telephony, videoconferencing, security, surveillance, interactive games
- Rates: 28.8 – 56 kbps
- Picture Formats
  - Progressive (non-interlaced), 15 or 7.5 fps
  - Fixed 4:3 aspect ratio
  - QCIF Y=176x144, Cb=Cr=88x72
  - S-QCIF Y=128x96, Cb=Cr=64x68

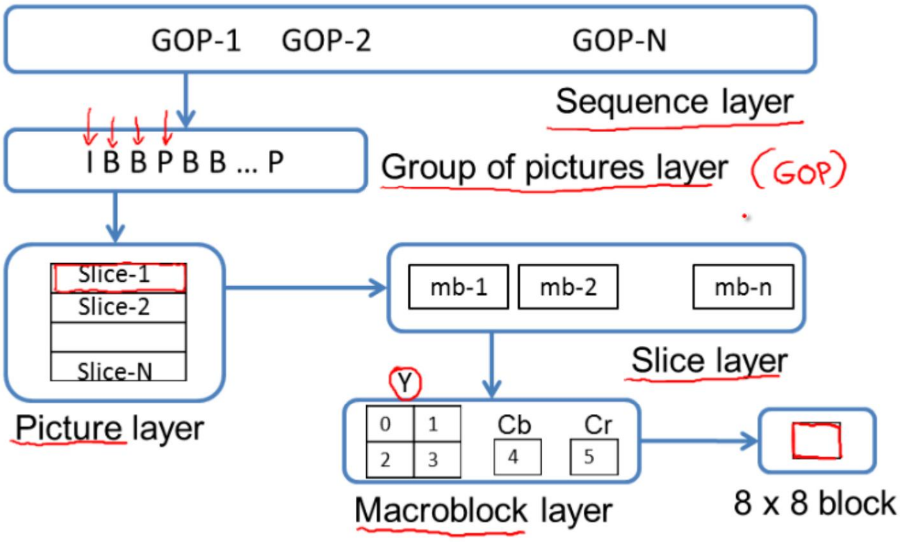
H.263 vs. H.261

- Bitrates < 64 kbps
- Unrestricted motion vectors
- Advance MC
  - ½ pel motion vectors
  - no loop filter
  - overlapping block MC
- I-, P-, B-, and PB-frames are used
- Same performance with H.261 at ½ the bit rate!

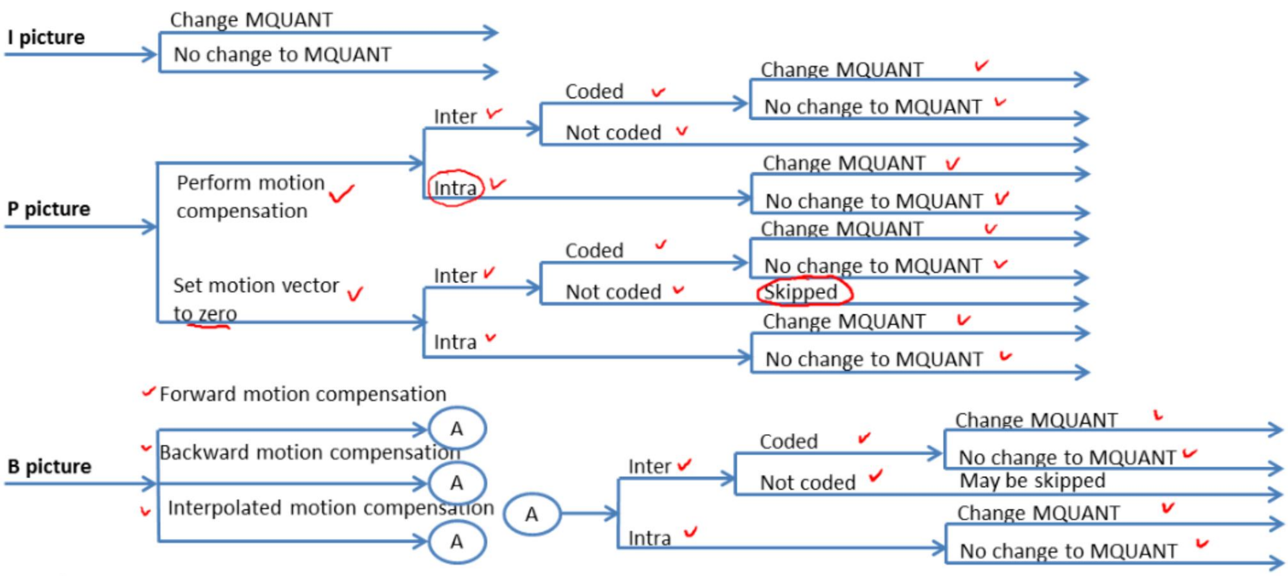
MPEG Video

- MPEG-1 (1992): ISO standard 11172 "*Coding of moving pictures and associated audio for storage media at up to about 1.5 Mbps*"
- MPEG-2 (1994): ISO standard 13818 "*Generic coding of moving pictures and associated audio*"
- MPEG-4 (1999), MPEG-7, MPEG-21, H.264, H.265

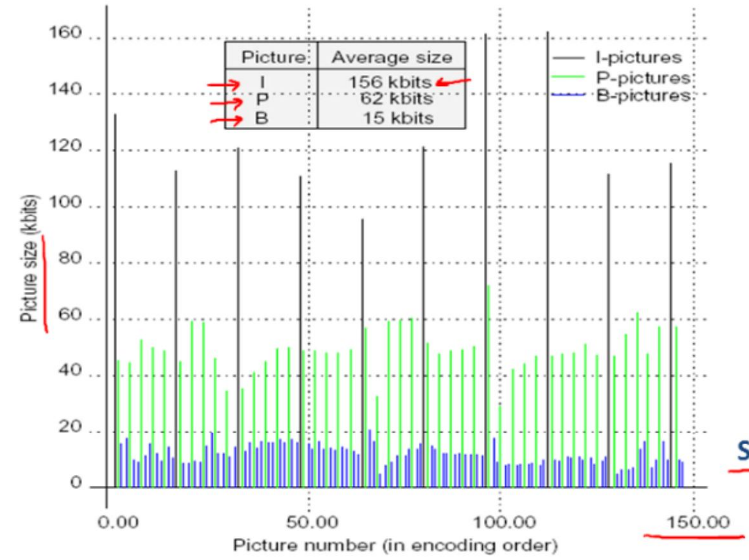
MPEG-1 Video Bitstream



Macroblock Coding



MPEG-1: Example of Bit Distribution

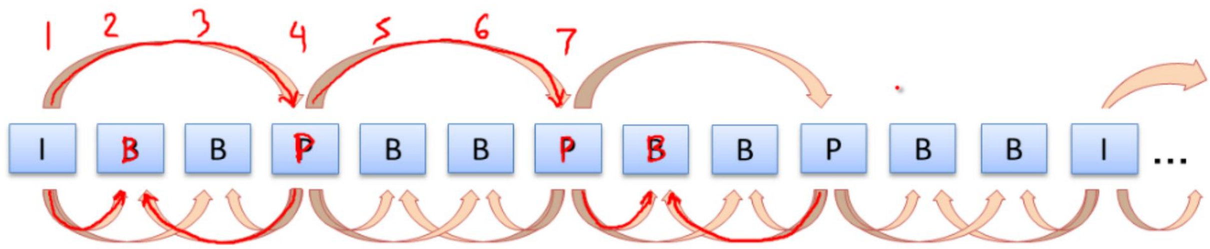


GOP
16 pictures
I BBPBB...

SIF with 1.15 Mbps

V. Bhaskaran and K. Kostantinides, Image and Video Compression Standards, Kluwer, 2nd edition, 1997.

MPEG-1 Example Frame Sequence



1 4 2 3 7 5 6 ...

1 2 3 4 5 6 7

Example of macroblock distribution

Picture Type	Macroblock Type				
	I	P	B	Zero MV	Skipped
I	3,300				
P	897	8,587		5,128	568
B	60	7,356	22,845		429

x6

Example of the number of the 8×8 blocks coded

1 MB=6blocks

Picture Type	Macroblock Type			
	I	P	B	Zero MV
I	19,800			
P	5,382	30,730		18,146
B	360	8,176	18,853	

Example of the distribution of the computational load (MPEG1 decoding)

Decoding Function	Load (%)
Bit stream header parsing	0.44
Huffman decoding and inverse quantization	19.00
Inverse 8 x 8 DCT	22.10
Motion compensation	38.64
Color transformation and display	19.82

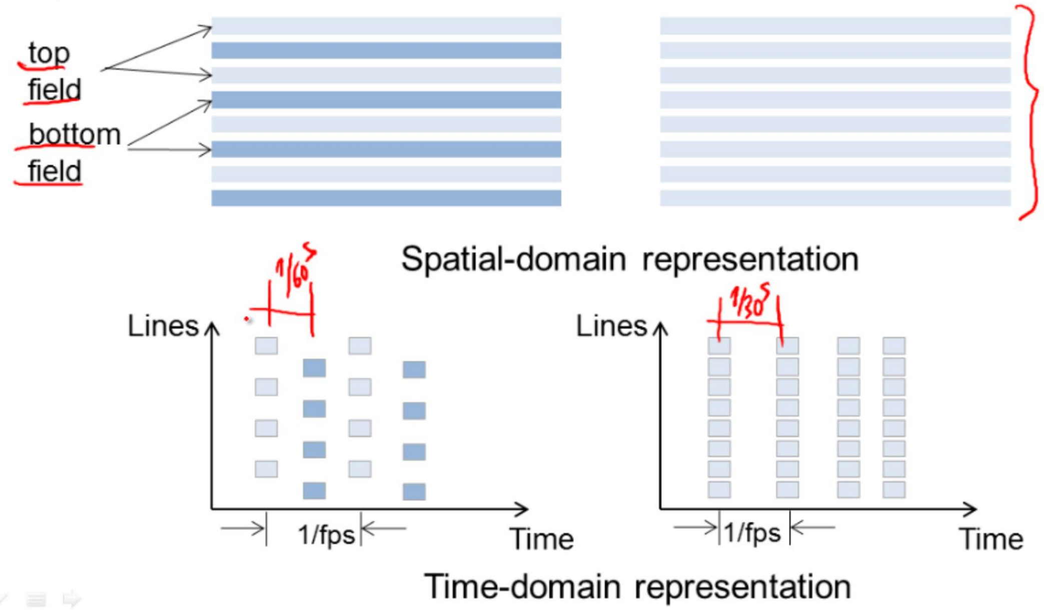
Why MPEG-2?

- MPEG-2 concept is similar to MPEG-1 but includes extensions to cover a wider range of applications
- Primary application during the definition process: all-digital transmission of broadcast TV quality video at coded bitrates between 4 and 9 Mbps
- MPEG-2 syntax was found to be efficient for other applications, such as HDTV

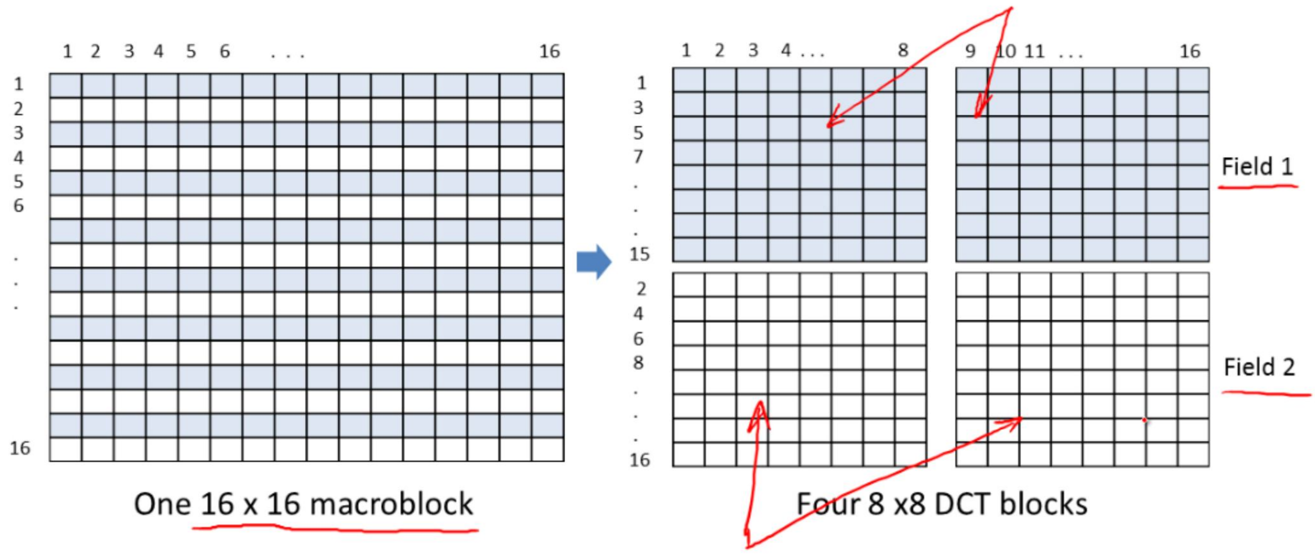
Characteristics

- MPEG-2 is backwards compatible with MPEG-1
- Like MPEG-1 all input pictures can be coded as I-, P- or B-pictures
- Most significant enhancement over MPEG-1 is the addition of syntax for efficient coding of interlaced video
- MPEG-2 adopted a toolkit-like approach (divided into *profile* and *levels*)
- Scalable bitstreams

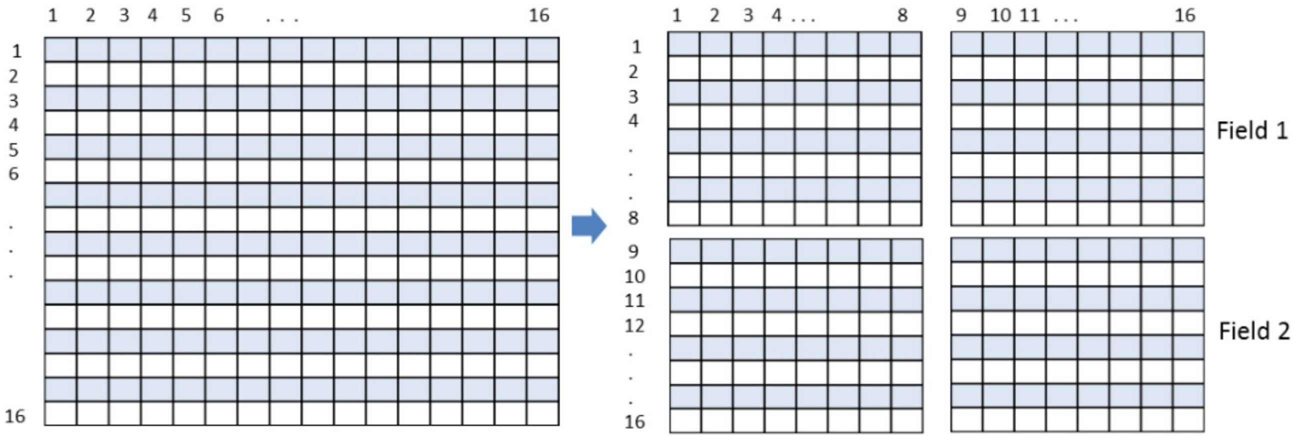
Interlaced vs. Progressive Scans



MPEG-2 DCT Block Derivation with I-frames: field mode



MPEG-2 DCT Block Derivation with I-frames: frame mode

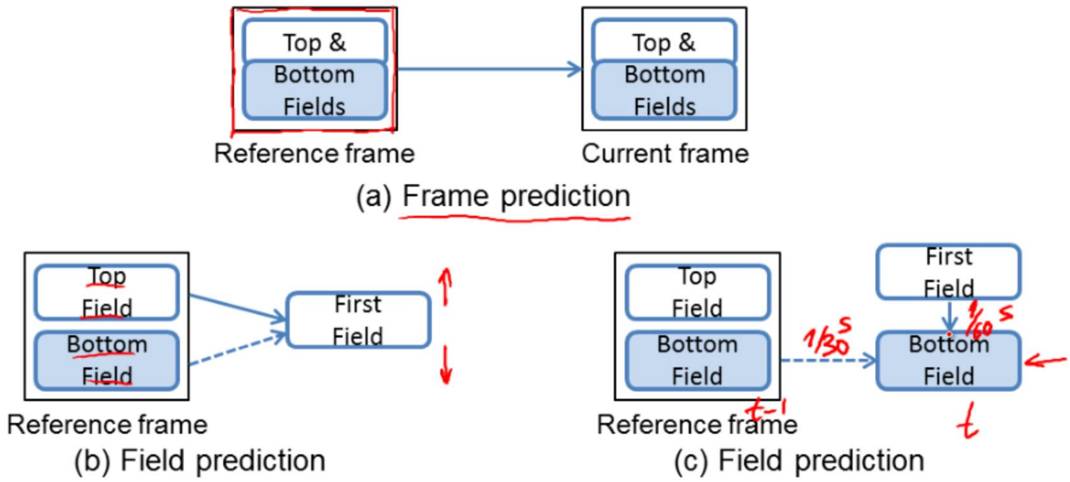


One 16 x 16 macroblock

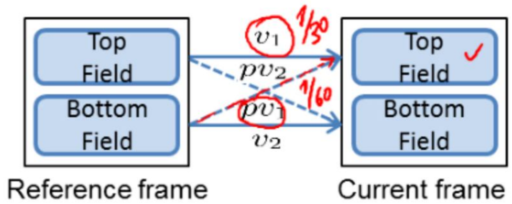
Four 8 x8 DCT blocks

Prediction Models and Motion Compensation

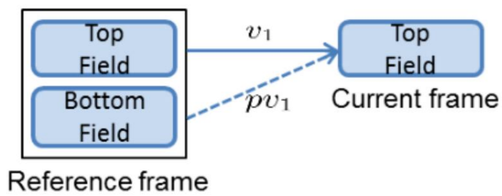
Frame and field prediction is supported as in example below



Dual-prime Motion Compensation



Field prediction in frame picture



Field prediction in field picture

Applicable only to P-pictures and to GOPs that have no B-pictures between these P-Pictures and the reference frame picture

$$pv_1(x) = \frac{v_1(x)}{2} + \delta(x) \qquad pv_1(y) = \frac{v_1(y)}{2} - 1 + \delta(y)$$

$$pv_2(x) = \frac{3v_2(x)}{2} + \delta(x) \qquad pv_2(y) = \frac{3v_2(y)}{2} + 1 + \delta(y)$$

$$\delta(x), \delta(y) \in \{-1, 0, 1\}$$

Only δ is included in the bitstream (not pv_1, pv_2)

MPEG-2: Profiles and levels

Levels		Profiles: Nonscalable: Simple 4:2:0	Profiles: Nonscalable: Main 4:2:0	Profiles: Scalable: Main+ 4:2:0	Profiles: Scalable: Next 4:2:2
High	Max resolution/ rate (Hz)	N/A	1920 x 1152/60	N/A	1920 x 1152/60
	Min. resolution/ rate (Hz)	N/A	N/A	N/A	960 x 576/30
	Bitrate (Mbits/s)	N/A	80	N/A	100 (all layers) 80 (base+mid) 25 (base layer)
High-1440	Max resolution/ rate (Hz)	N/A	1440 x 1152/60	1440 x 1152/60	1440 x 1152/60
	Min. resolution/ rate (Hz)	N/A	N/A	720 x 576/30	720 x 576/30
	Bitrate (Mbits/s)	N/A	60	60 (all layers) 40 (base+mid) 15 (base layer)	80 (all layers) 60 (base+mid) 20 (base layer)
Main	Max resolution/ rate (Hz)	720 x 576/30	720 x 576/30	720 x 576/30	720 x 576/30
	Min. resolution/ rate (Hz)	N/A	N/A	N/A	352 x 288/30
	Bitrate (Mbits/s)	15	15	15 (all layers) 10 (base layer)	20 (all layers) 15 (base+mid) 4 (base layer)
Low	Max resolution/ rate (Hz)	N/A	352 x 288/30	352 x 288/30	N/A
	Min. resolution/ rate (Hz)	N/A	N/A	N/A	N/A
	Bitrate (Mbits/s)	N/A	4	4 (all layers) 3 (base layer)	N/A

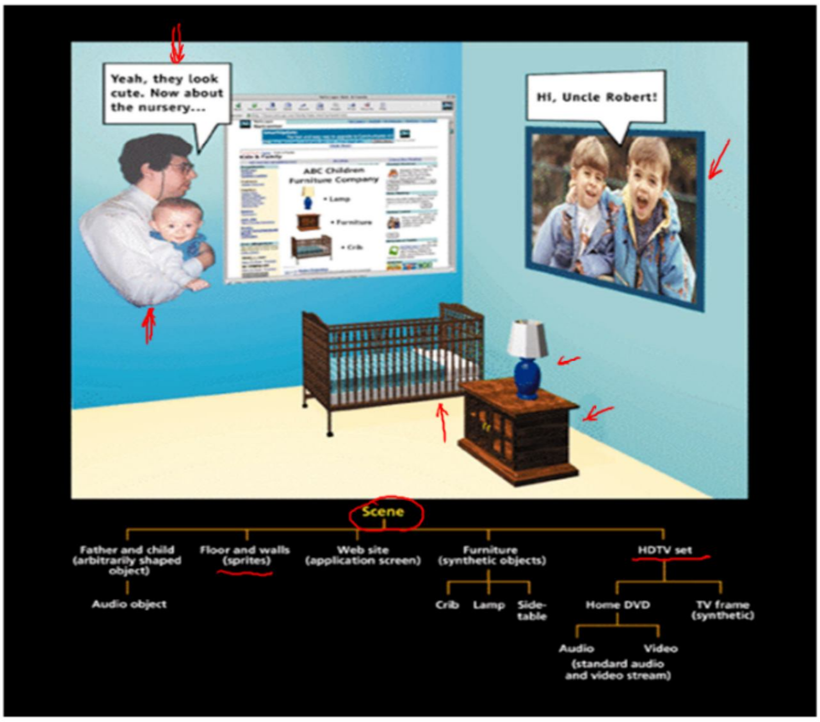
HDTV

- Advanced Television (ATV) – North America (Grand Alliance)
  - 16/9 aspect ratio
  - 1280x 720
  - Video: MPEG-2 Main Profile-High Level (MP@HL)
  - Audio: Dolby AC-3
- Digital Video Broadcast (DVB) – Europe
  - 4/3 aspect ratio
  - 1440 samples per line, 1152 (1080 visible) lines per frame
  - Video: MPEG-2 SSP@H1440 -spatially scalable profile at high 1440
  - Audio: MPEG Audio Layer 2
- Multiple sub-Nyquist Sampling Encoding (MUSE) – Japan
  - 16/9 aspect ratio
  - 1920 samples per line, 1035 lines per frame
  - Video: similar to MP@HL

What is New with MPEG-4?

- Ability to access and manipulate individual elements that make up each scene
  - cartoon: reposition, delete, alter movements of individual characters within the scene

MPEG-4 Scene: Object-oriented composition



Rob Koenen, KPN Research

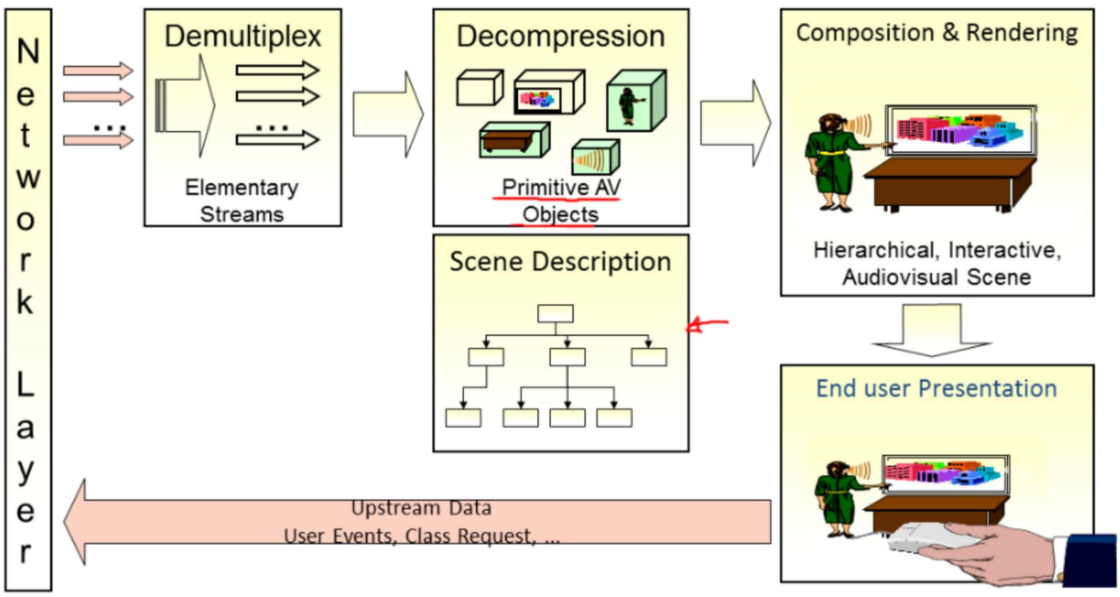
What is MPEG-4?

- Coding of Audio Visual Objects
- The Next Generation Multimedia Communication Standard
  - New Video and Audio Coding Tools
  - Merging of **Natural** and **Synthetic** Data
  - Robust Bitstream Syntax
  - Flexible Systems Layer for Interactivity

Application Space of MPEG-4

- Streaming Media for the Internet
- Entertainment/Interactive Games/DVD
- Content Based Storage/Retrieval
- Real Time Communications
- Surveillance
- Mobile Multimedia
- Broadcast and Studio Post-Production
- Interactive TV

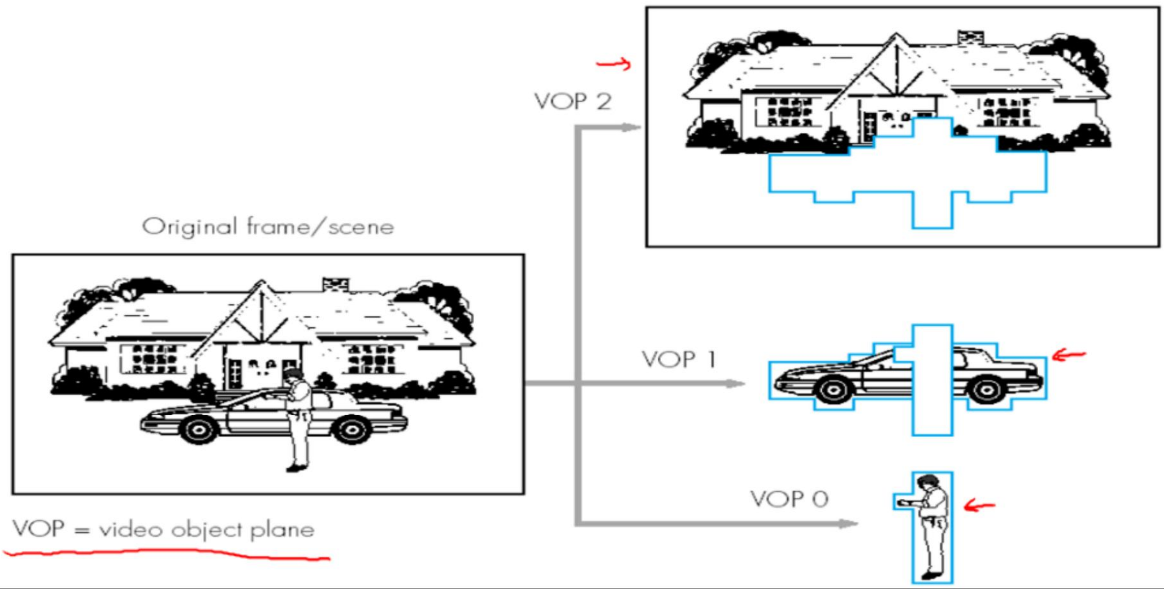
MPEG-4 Composition & Interaction



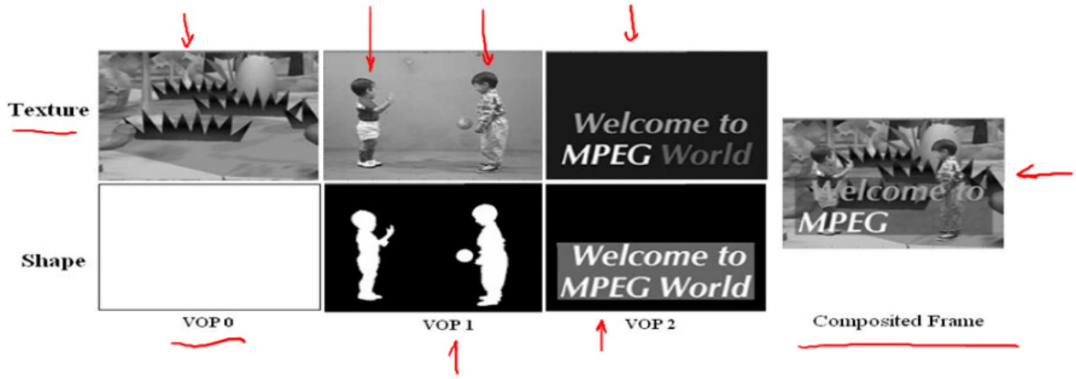
Coding of Visual Objects

- Natural Textures, Images and Video
  - Frames (Progressive and Interlaced)
  - Arbitrarily Shaped Objects
  - Scalability, Still Image Texture Maps
  - Sprites
  - Error Resilience
- Synthetic Objects
  - Facial Animation
  - Animated Meshes
  - Body Animation (Version 2)

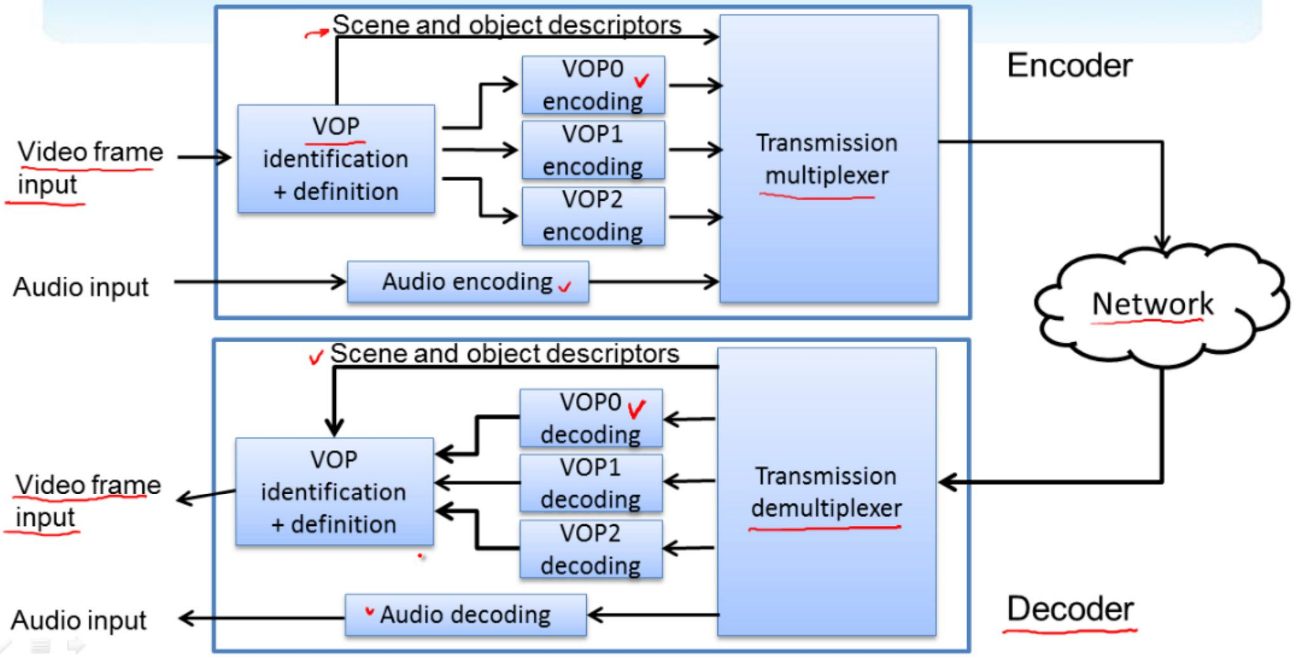
Content-based Video Coding Principles



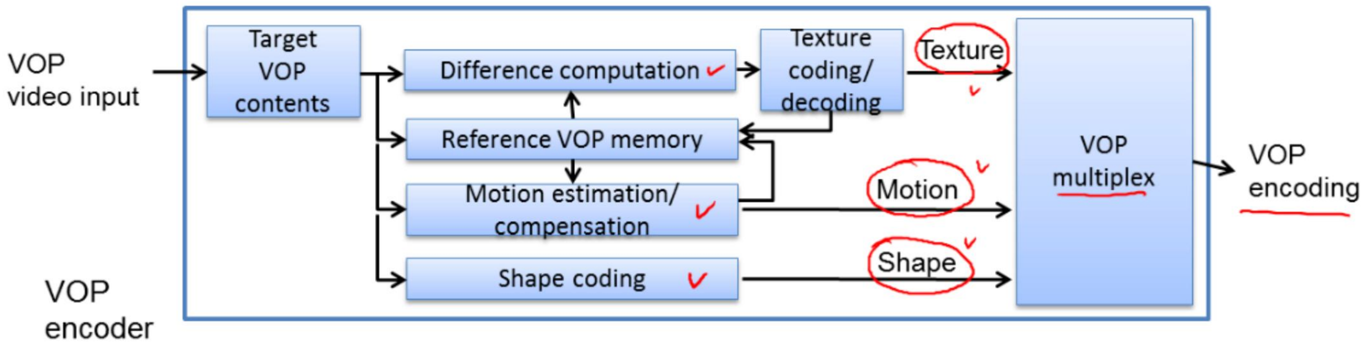
Texture-Shape Decomposition



MPEG-4 Coder/Decoder

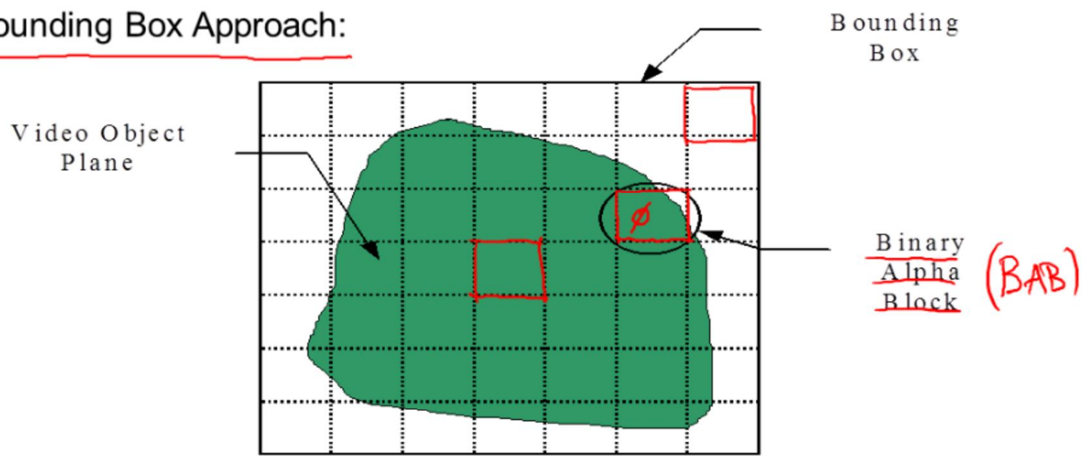


VOP Encoder Schematic



Shape Coding

- Bounding Box Approach:

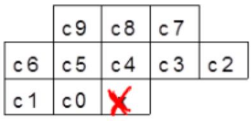


- Texture is Padded Block DCT Coded

Shape Coding

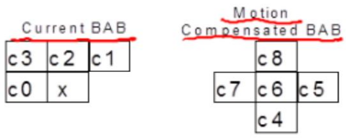
- Context-Based Arithmetic Encoding (CAE)
  - Binary Shape Pixels in Binary Alpha Blocks (BAB)

Intra Shape



$2^{10} = 1024$

Inter Shape

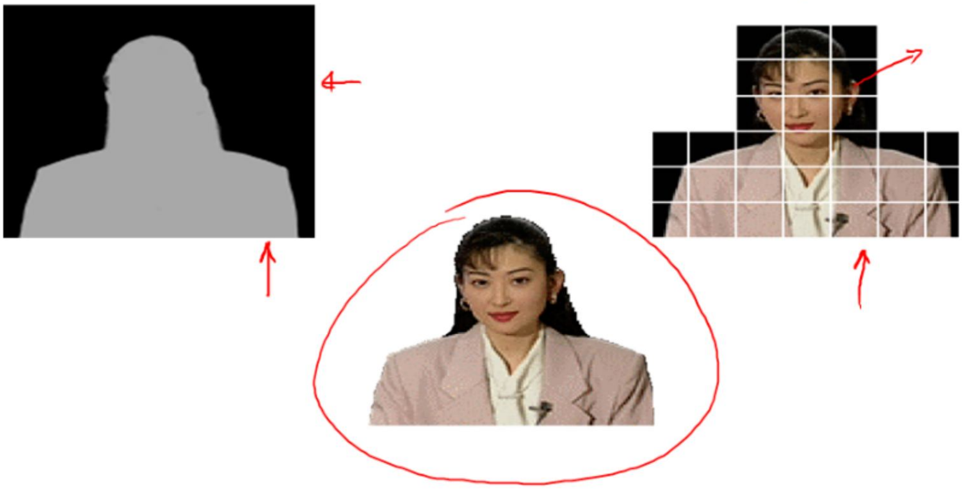


$2^9 = 512$

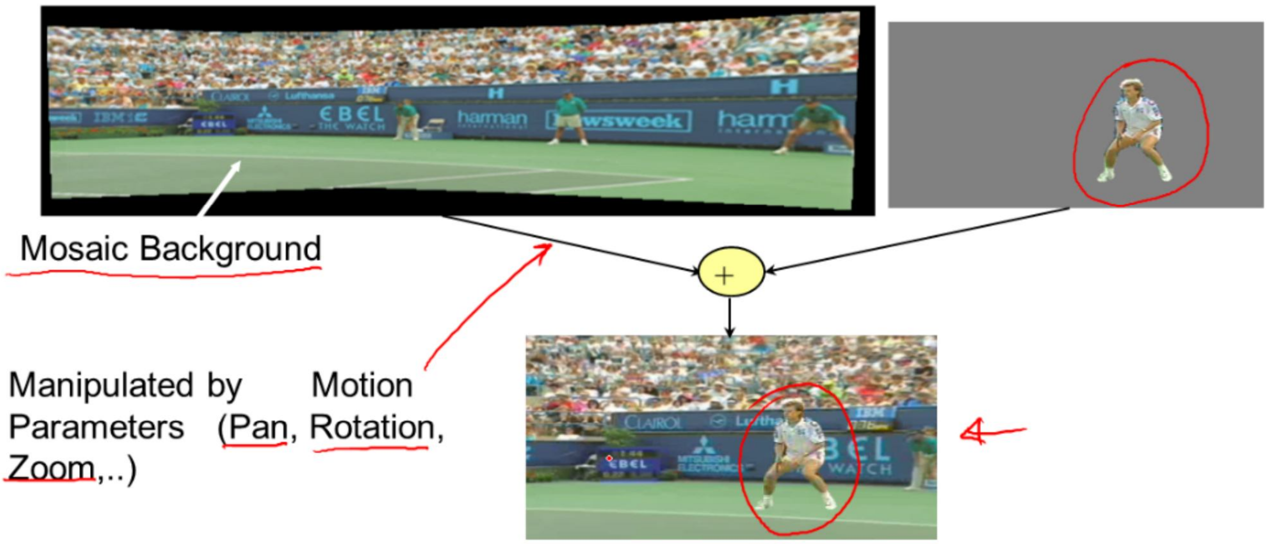
Experimental Results

Texture Coding

- Pad Macroblocks - Extract Shape



Sprites



Face Object

- Shape, Texture and Expressions
  - parameters in the incoming bitstream
  - remote and local control of parameters
- Three Sets of Data Used
  - *Facial Definition Parameters (FDPs)*
  - *Facial Animation Parameters (FAPs)*
  - *Facial Interpolation Transform (FIT)*
- Defines a Specific Face Via
  - 3D feature points
  - 3D mesh/scene graph
  - Face Texture
  - Face Animation Table (FAT)

What is H.264?

- The latest (and greatest) offered by the JVT
- Initial goal was "twice the compression efficiency as anything else" (c. 1998)

  - o Is "H.264" the same as "MPEG4 Part 10"?
  - o Is "H.264" the same as "H.264/AVC"?
  - o Is "H.264" the same as "H.26L"?

New features in H.264

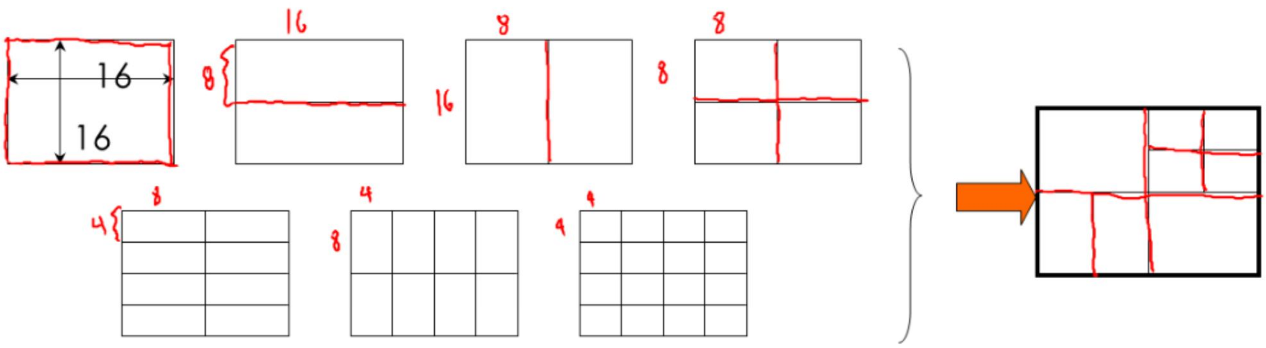
- Motion compensation and intra-prediction
- Image transform
- Deblocking filters
- Entropy coding
- Frames and slices

Variable Block-Size MC

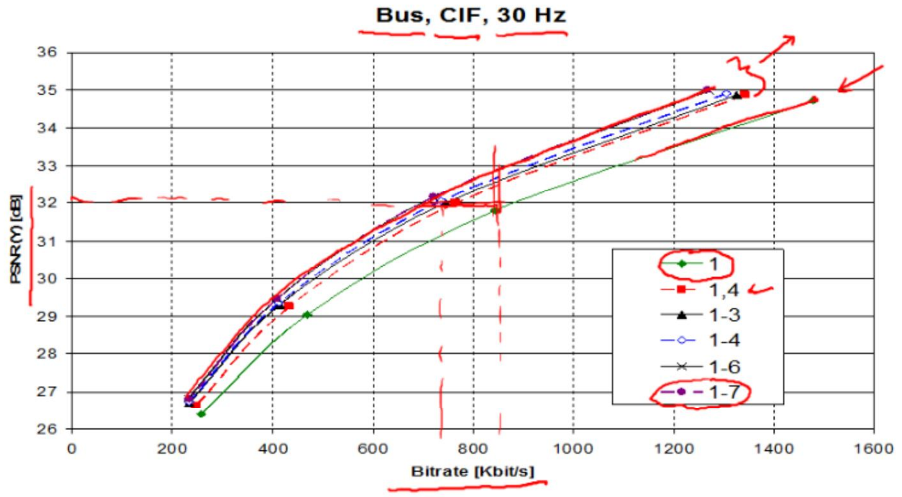
- Motivation: size of moving/stationary objects is variable
  - Many small blocks may take too many bits to encode
  - Few large blocks give lousy prediction
- In H.264, each 16x16 macroblock may be:
  - Kept whole
  - Divided horizontally (vertically) into two sub-blocks of size 16x8 (8x16)
  - Divided into 4 sub-blocks
  - In the last case, the 4 sub-blocks may be divided once more into 2 or 4 smaller blocks.

Variable Block-Size MC

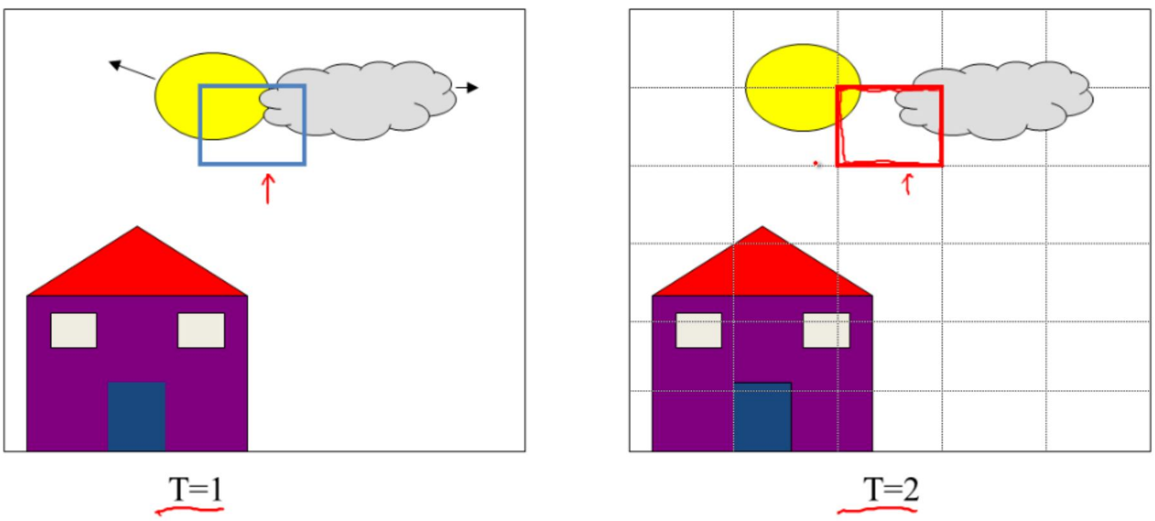
- 7 motion compensation block sizes



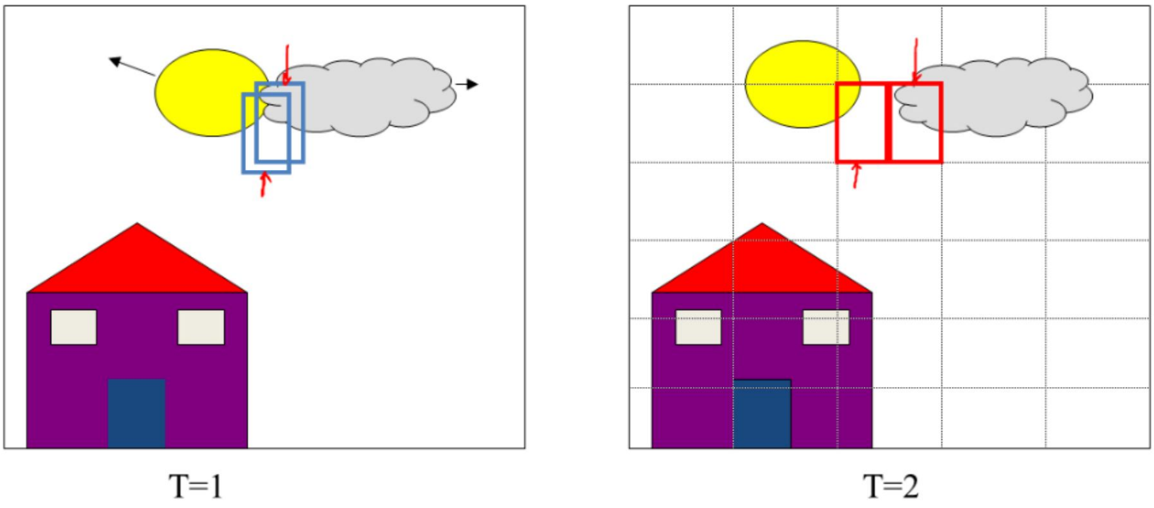
Example



Motion Scale Example



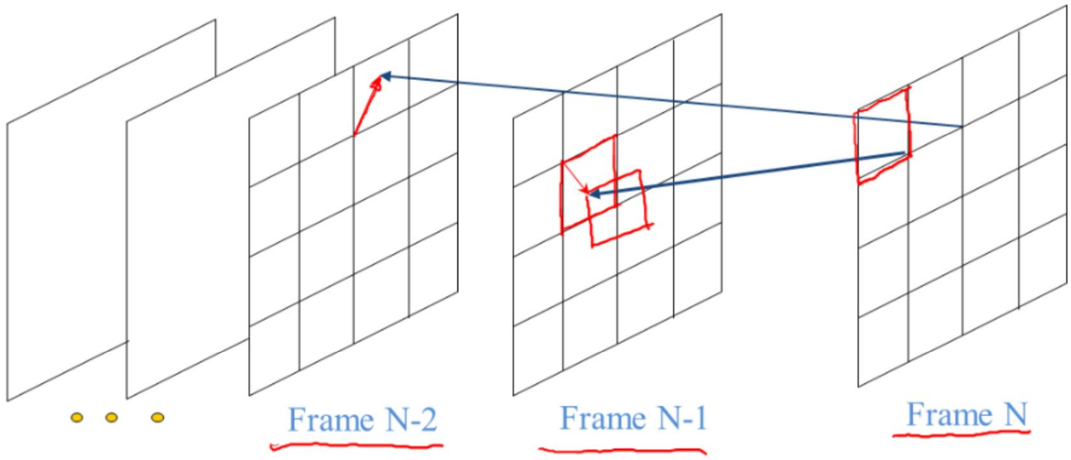
H.264 VBS Example



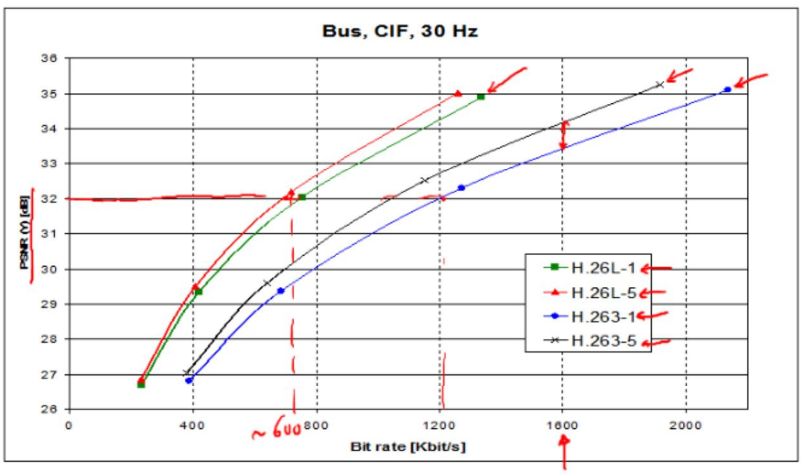
Arbitrary Reference Frames

- In H.263, the reference frame for prediction is always the previous frame
- In MPEG and H.26L, some frames are predicted from both the previous and the next frames (bi-prediction)
- In H.264, *any* one frame may be used as reference:
  - Encoder and decoder maintain synchronized buffers of available frames (previously decoded)
  - Reference frame is specified as index into this buffer
- In bi-predictive mode, each macroblock may be:
  - Predicted from one of the two references
  - Predicted from both, using weighted mean of predictors

Multi-Reference Frame MCP



Multi-Frame MCP Performance

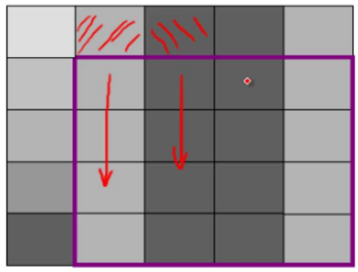


Intra Prediction

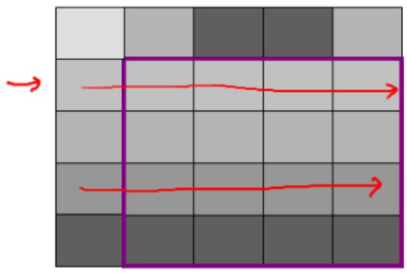
- Motivation: intra-frames are natural images, so they exhibit strong spatial correlation
  - Implemented to some extent in H.263++ and MPEG-4, but in transform domain
- Macroblocks in intra-coded frames are predicted based on previously-coded ones
  - Above and/or to the left of the current block
  - The macroblock may be divided into 16 4x4 sub-blocks which are predicted in cascading fashion
- An encoded parameter specifies which neighbors should be used to predict, and how

Intra-Prediction Example

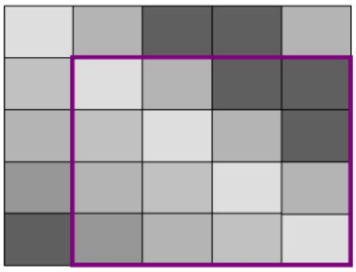
Intra-Prediction: Vertical



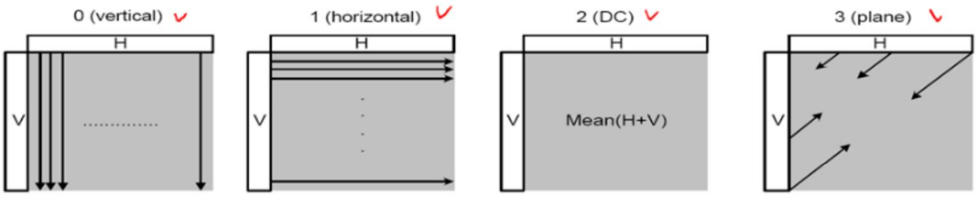
Intra-Prediction: Horizontal



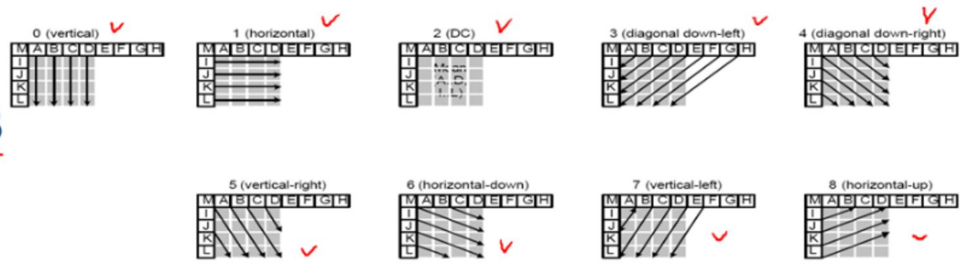
Intra-Prediction:Main Diagonal



- Baseline Profile supports I16MB and I4MB intra prediction modes
  - I16M



  - I4MB



New features in H.264

V
- Motion compensation and intra-prediction
- Image transform
- Deblocking filters
- Entropy coding
- Frames and slices
- Relevance to our work

H.264 Image Transform

- Motivation:
  - DCT requires real-number operations, which may cause inaccuracies in inversion
  - Better motion compensation means less spatial correlation – no need for 8x8 transform
- H.264 uses a *very* simple integer 4x4 transform
  - A (pretty crude) approximation to 4x4 DCT
  - Transform matrix contains only +/-1 and +/-2
    - Can be computed with only additions, subtractions, and shifts
- Results show negligible loss in quality (~0.02dB)

Deblocking Filters

- Motivation: block-based MC and transforms generate blocking artifacts
  - Very visible to human eye at low bit-rates
- Previous standards applied simple filters to “smudge” edges between blocks
- H.264 adaptively chooses for *each* edge which one of 5 deblocking filters to apply.
  - For instance, if both blocks have the same motion vector, less filtering is needed.
- Improves *objective* quality as well: about 7-9% reduction in bit-rate for same PSNR.

Deblocking Filter

No Deblocking Filter
PSNR = 35.07 dB

With Deblocking Filter
PSNR = 35.43 dB

New features in H.264

- Motion compensation and intra-prediction
- Image transform
- Deblocking filters
- Entropy coding
- Frames and slices

v

Entropy Coding

- Motivation: traditional coders use fixed, variable-length codes
  - Essentially Huffman-style codes
  - Non-adaptive
  - Can't encode symbols with probability > 0.5 efficiently, since at least one bit required
- H.263 Annex E defines an arithmetic coder
  - Still non-adaptive
  - Uses multiple non-binary alphabets, which results in high computational complexity

Entropy Coding: CABAC

- Arithmetic coding framework designed specifically for H.264
- Binarization: all syntax symbols are translated to bit-strings
- 399 predefined context models, used in groups
  - E.g. models 14-20 used to code macroblock type for inter-frames
  - The model to use next is selected based on previously coded information (the context)

Applications

- Demands strong processor
- Cellular products
  - ✓ Usually baseline, need to simplify handset side (both encoder and decoder)
  - ✓ Many variants decode-only
- Set-top boxes
  - ✓ Broadcasting – usually in conjunction with transcoders
  - ✓ Personal media – camcorder storage
- Usually overkill for surveillance, but sometimes used in high-end systems

H.264 The Magnificent

- Does H.264 live up to expectations?
  - ✓ *Very* competitive compression
  - ✓ Application-friendly features embedded in standard (e.g. flexible slicing, FMO, 16-bit transform)
  - ✓ Large selection of optional tools in higher profiles (e.g. FRExt, film grain analysis)

H.264 The Cumbersome

- Is H.264 approximately at the same complexity level as other standards (H.263+, MPEG2 etc)?
  - More complicated base features
    - ✓ 4x4 transform
    - ✓ smaller inter pred. blocks
  - More features to run
    - ✓ mult. ref. frames
    - ✓ quarter-pel ME/MC
    - ✓ loop deblocking
    - ✓ CAVLC/CABAC

Summary

- Some very strong results:
  - ¼-pixel prediction
  - CABAC
  - Deblocking filters
- Very exciting! First major improvement since mid 90's
  - Two-fold improvement in performance!

What is HEVC?

- The latest video compression standard by JCT-VC (Joint Collaborative Team on Video Coding)
- Initial goal: 50% bit rate reduction at the same quality compared to H.264 High Profile (HP)
  - ITU: H.265/H.NGVC (Next Generation Video Coding)
  - ISO/IEC: MPEG-H Part 2

Comparison with H.264

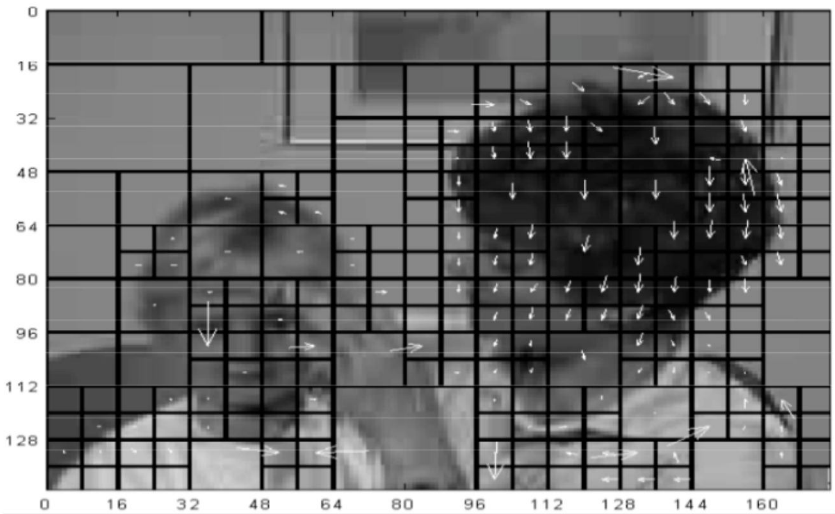
- Better coding efficiency
  - More than 35% bit rate reduction compared with H.264 HP
- Support of high-resolution video
  - HEVC: up to 8K UHD (8192*4320)
  - H.264: up to 4K UHD (4096*2160)
- Support of stereo and multi-view
- Increased use of parallel processing

CTU and CTB

- Coding Tree Units (CTU) and Coding Tree Blocks (CTB)
  - CTU of sizes 16x16, 32x32, 64x64
  - Larger CTUs more suitable for HD content, better coding efficiency

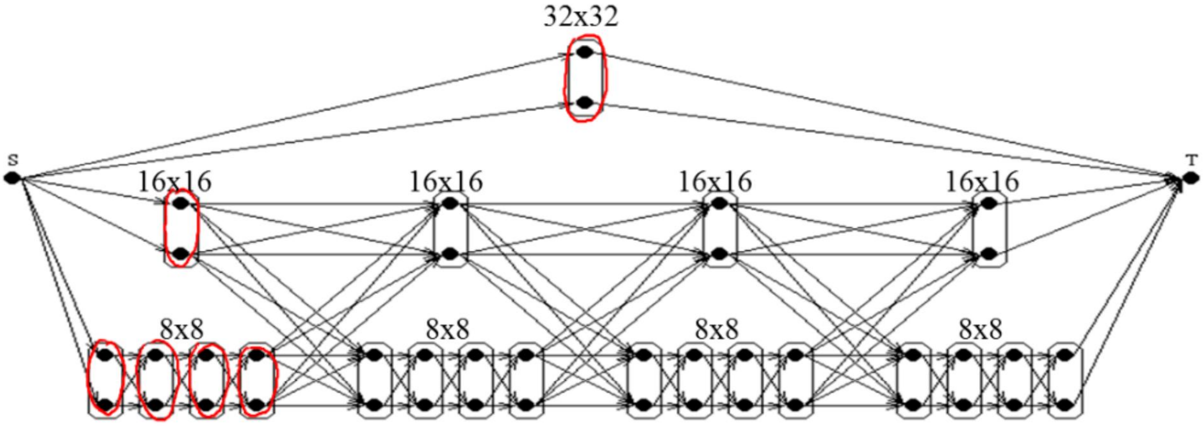
	Maximum CTU size used in video encoding		
	64x64	32x32	16x16
Bit rate	0%	+2.2%	+11%
Encoding time	100% ✓	82%	58%
Decoding time	100% ✓	111%	160%

QT-based Encoding



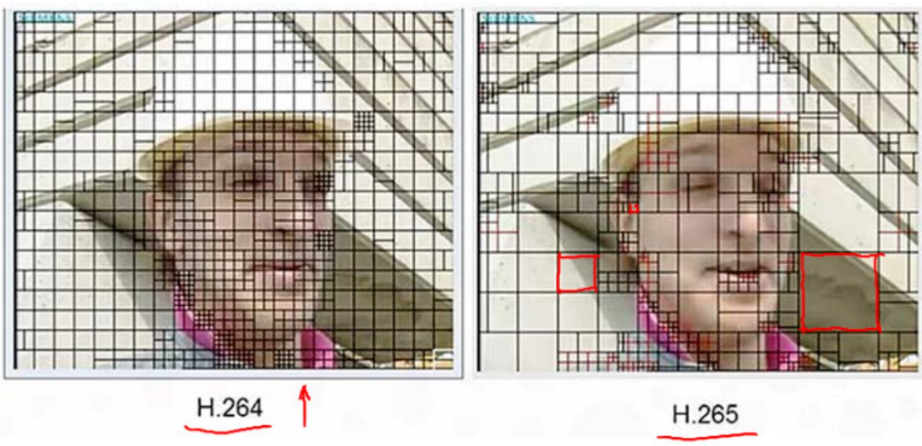
G. M. Schuster and A. K. Katsaggelos, *Rate-Distortion Based Video Compression: Optimal Video Frame Compression and Object Boundary Encoding*, Kluwer Academic Publishers, 1997.

Multilevel Trellis



Experimental Result

CTUs



Images from HEVC webinar by Elemental Technologies.

Improved Motion Compensation

- Quarter-sample precision (same as H.264)
- 7-tap or 8-tap filters for fractional sample interpolation
- H.264 uses 6-tap filter for half-sample interpolation and linear interpolation for quarter samples

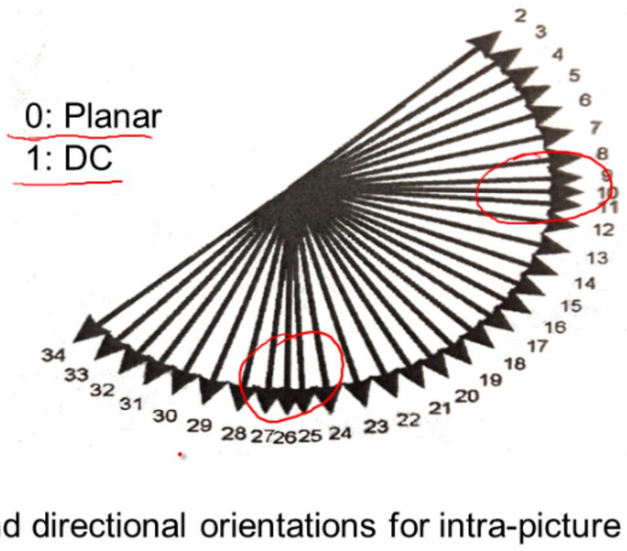
Motion Vector Signaling

- Advanced Motion Vector Prediction (AMVP) is used to derive candidate for MV
- Merge mode is used to inherit MV from temporal or spatial neighbors
- Improved “skip” or “direct” motion

Intra-Prediction

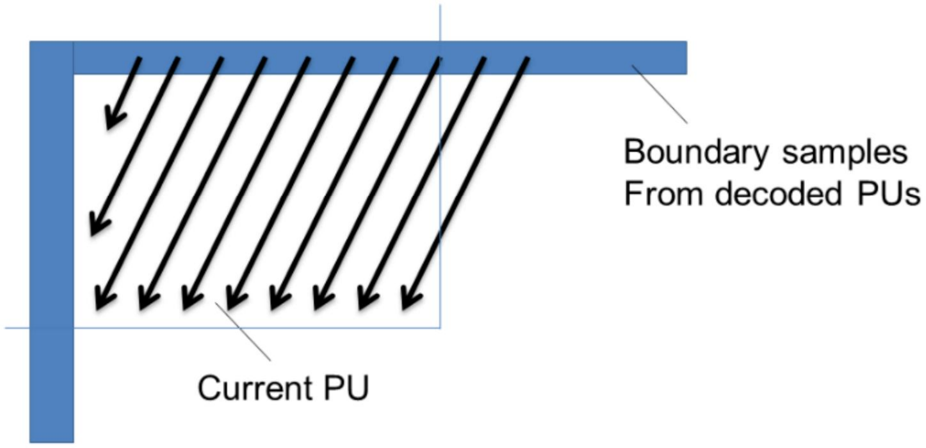
- 33 directional modes v.s. 8 in H.264
- Planar mode: surface fitting
- DC/flat mode

Intra-Prediction



Modes and directional orientations for intra-picture prediction

Intra-Prediction



Directional Mode 29

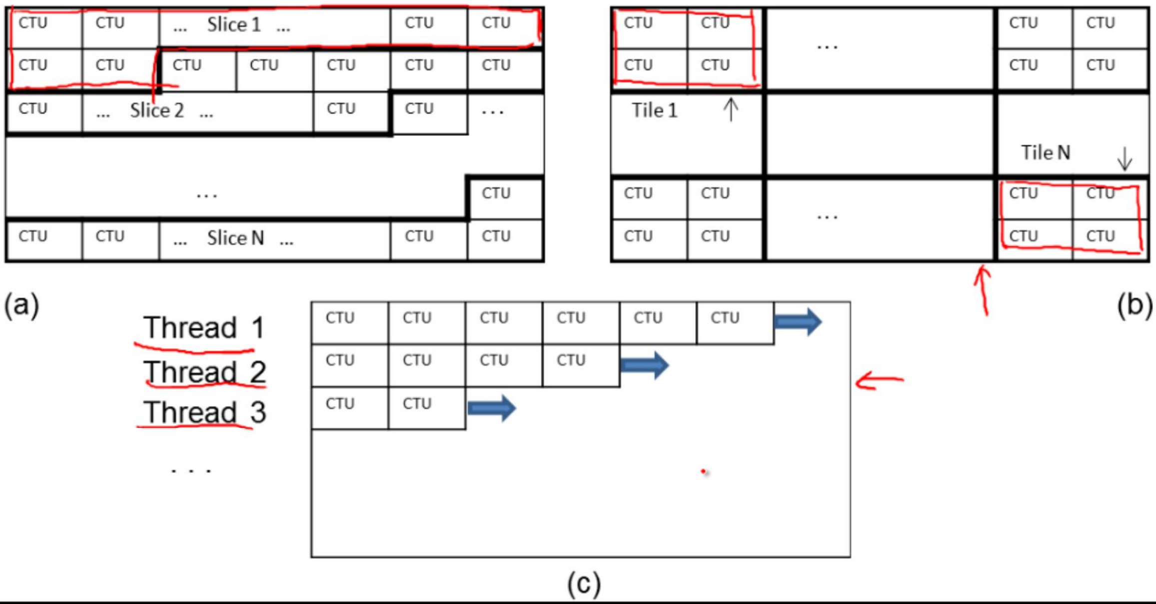
Entropy Coding

- CABAC with improved throughput
- No VLC (as opposed to both arithmetic and VL coding in H.264)

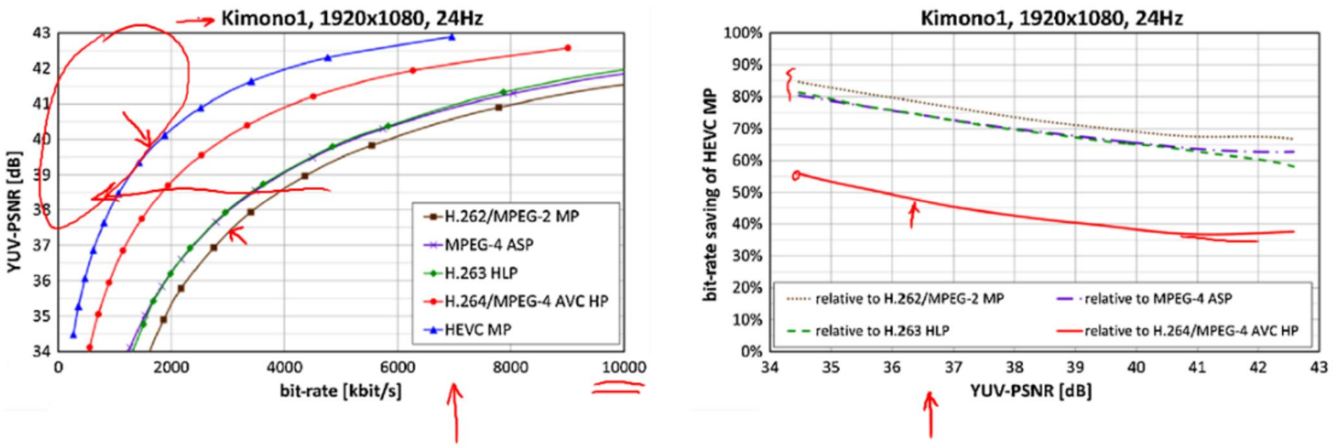
In-loop Filters

- Used in inter-picture prediction loop
- Deblocking filter: similar to but simpler than that in H.264; better support of parallel processing
- Sample adaptive offset (SAO) filter: applied after deblocking filter; better reconstruction of original signal amplitude by using offset from a look-up table

Slices and Tiles



Experimental Result



G.J. Sullivan, H. Schwarz, T. K. Tan, T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards – Including High Efficiency Video Coding (HEVC)", *IEEE Transactions on Circuits d Systems for Video Technology*, vol. 22, no. 12, 2012