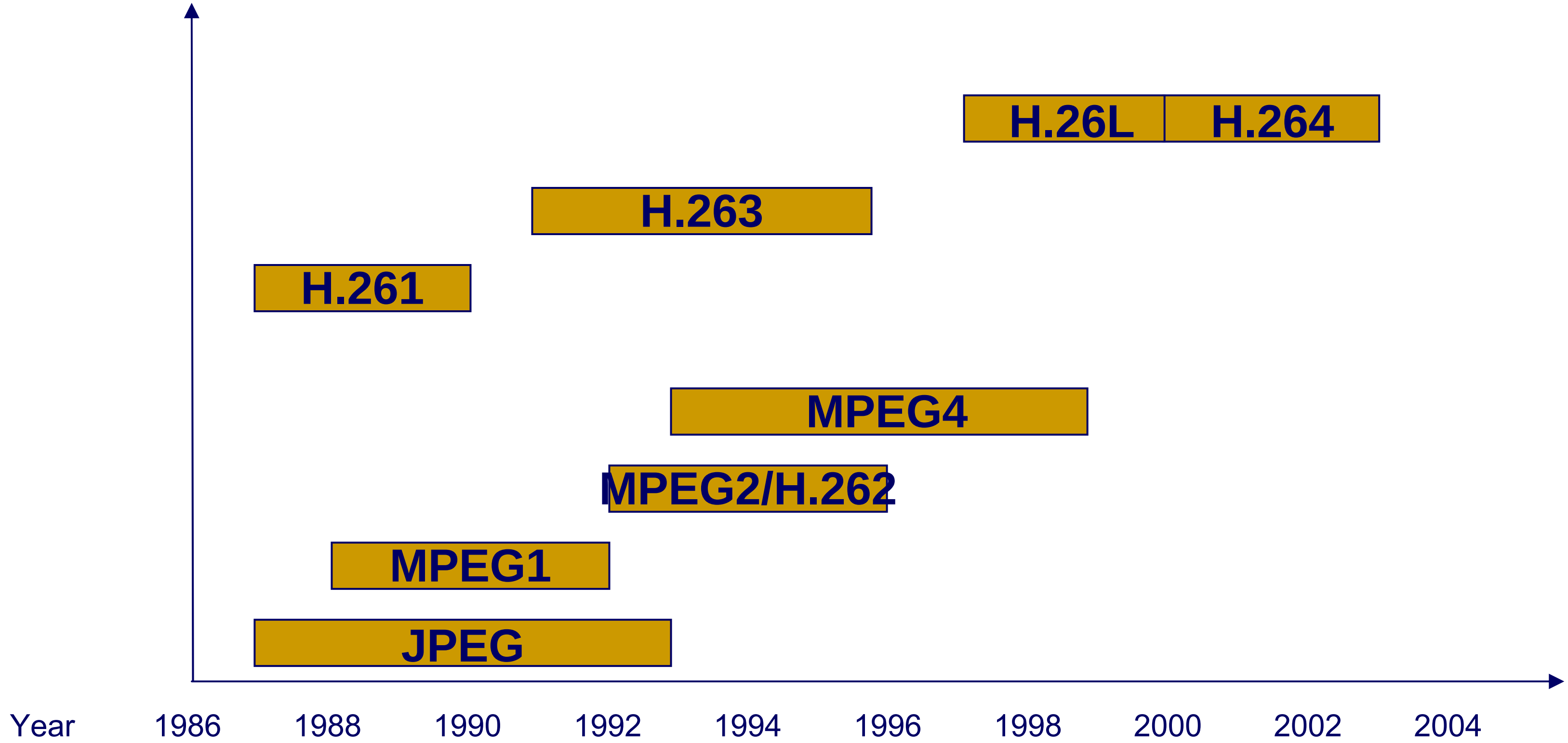
Video Coding Standards: Up to H.264

Jing Hu and J. D. Gibson

UCSB

Time table



UCSB

Joint(ITU+ISO) Photographic Experts Group (JPEG)

- JPEG targets:

8 bits/pixel monochrome images	→	0.083 bits/pixel as	"recognizable"
		0.25	"useful"
		0.75	"excellent"
		2.25	"indistinguishable"

- Color treatments:

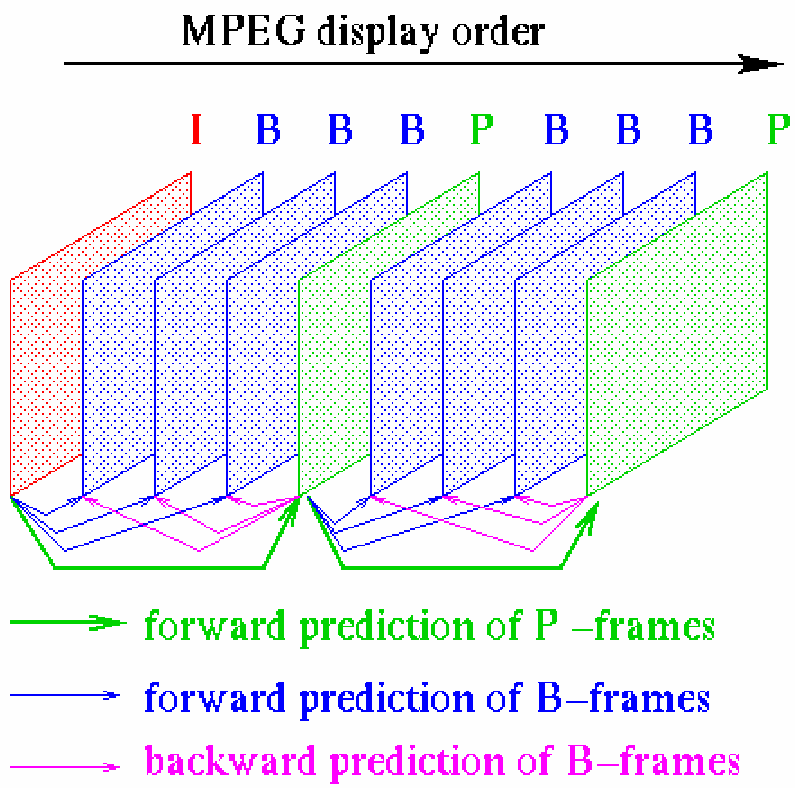
Red	→	Luminance (Y)	**Two modes:**
Green		Color difference B-Y (C_B)	4:2:2
Blue		Color difference R-Y (C_R)	4:2:0

- JPEG coder
  - 8×8 DCT (why DCT?)
  - Quantization (Two tables by Lohscheller 1984)
  - Zig-zag scanning and run-level description
  - Entropy coding (Huffman and arithmetic coding)
- Motion JPEG (Video coded as sequences of JPEG images)

UCSB

MPEG-1 = JPEG + Motion Prediction + Rate Control

- Early motivation: to encode motion video at 1.5Mbits/s for transport over T1 data circuits and for replay from CD-ROM
- Defines the decoder but not the encoder
- Frames (pictures)
  - Intra-coded using JPEG
  - Inter-coded using (interpolated) motion estimation & compensation and JPEG for the residules
    - Predicted and Bi-directional
- MacroBlocks (MBs)
  - 16×16 pixels block
- Rate control
  - buffer at each end
  - Test Model 5 (TM5)

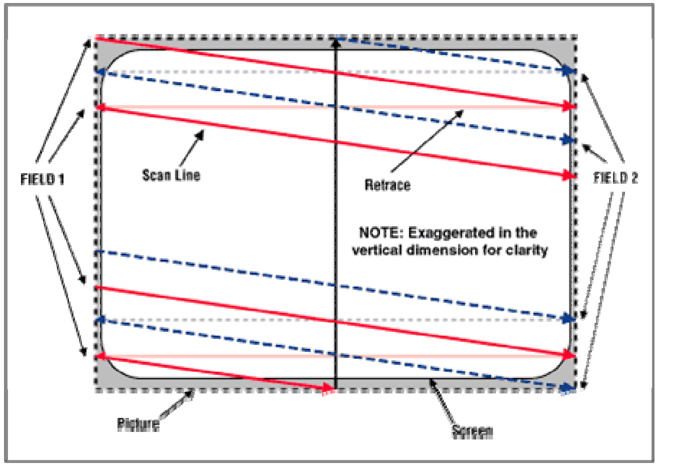


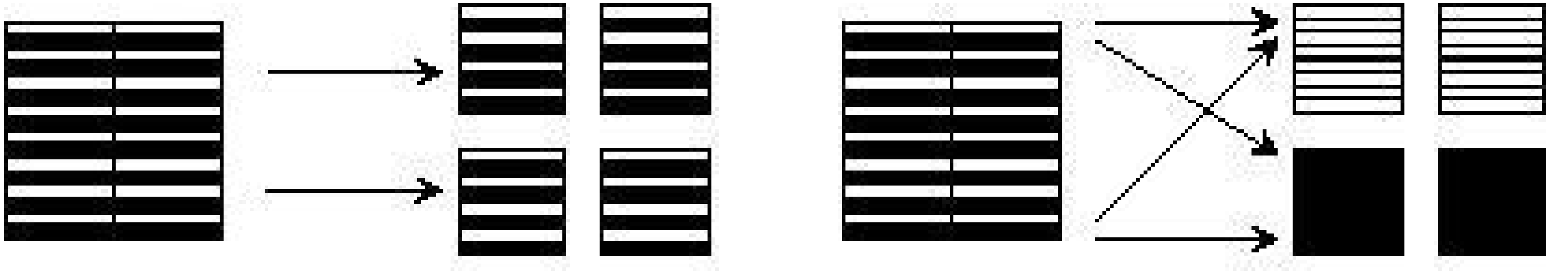
MPEG-2 = MPEG-1 +

- Improvments
  - Color space: could support 4:2:2 and 4:4:4 coding
  - Quantization: could have 9- or 10- bit precision for DC coefficients
  - Concealment motion vectors: used when an intra-MB is lost
  - Pan and Scan: supports display of different aspect ratios, e.g., 16:9
- Profiles and levels
  - Profiles: define the tools or syntactical elements
  - Levels: define the permissible ranges of parameters
- Interlace tools
- Scalable coding profiles
- System layer: define two bit stream constructs
  - Program stream (PS): modeled on MPEG-1 (backward compatibility)
  - Transport stream (TS): more robust, does not need a common time base, designed for use in error-prone environment.

MPEG-2 – Interlace Tools

- Interlaced Scanning: Image flicker is less apparent because the image is painted twice as many times as what is in non-interlaced scanning.
- Frame Pictures and Field Pictures
  - two fields are processed sequentially or not
- Frame DCT and Field DCT
  - Field pictures usually use field DCT
  - Frame pictures use field DCT when there is obvious vertical motion
- Frame Prediction and Field Prediction



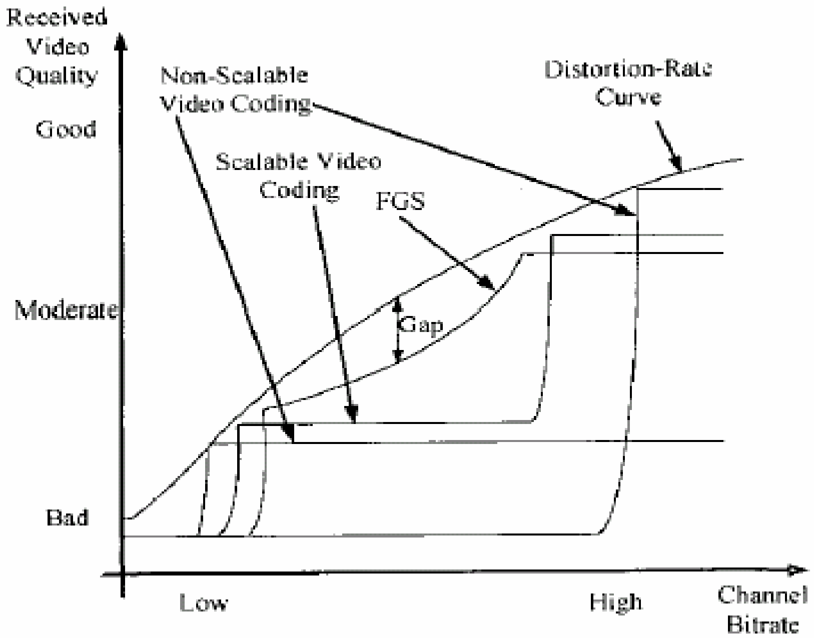


Frame DCT

Field DCT

MPEG – Scalable Coding (SC)

- Non-scalable coding
  - To optimize video quality at a given bit rate.
- Base and enhancement layer SC
  - To optimize video quality at two given bit rates.
  - SNR SC (different quantization accuracy)
  - Temporal SC (different frame rates)
  - Spatial SC (different spatial resolution)



- Fine granularity scalability (FGS)
  - To optimize the video quality over a given bit rate range
  - Also has base layer and enhancement layer
  - Enhancement layer uses bit-plane coding
    - Bit-plane coding considers each quantized DCT coefficient as a binary integer of several bits instead of a decimal integer of a certain value
    - Frequency weighting and selective enhancement

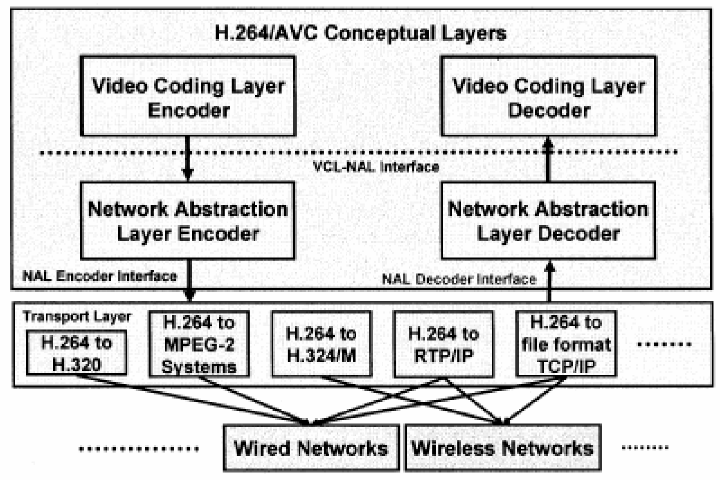
UCSB

MPEG-4 = MPEG-2+Objects+Other Enhancements

- Objects (optional)
  - Video (texture+shape), image, audio, speech, text, etc.
  - Encoded using different techniques
  - Transmitted independently
  - Composited at the decoder using BInary Format for Scenes (BIFS)
- Improvements in MPEG-4 version2
  - Global motion compensation (GMC)
  - Quarter pixel motion compensation
  - Shape-adaptive DCT
- Why is MPEG-4 not a success as MPEG-2?
  - Not substantially better than MPEG-2
  - Suffers from its sheer size and flexibility
  - Issue of licensing

Advanced Video Coding/ ITU-T Recommendation H.264/ ISO/IEC MPEG-4 (Part 10)

- H.264 structure
  - Video coding layer (VCL)
  - Network abstraction layer (NAL)
- Possible applications of H.264
  - Conversational services operated below 1Mbps with low latency.
    - ISDN-based H.320
    - H.324/M in circuit-switched channels
    - H.323
  - Entertainment services operated between 1-8+ Mbps with moderate latency such as 0.5-2s in modified MPEG-2/H.222.0 systems.
    - Broadcast via satellite, cable, terrestrial or DSL
    - DVD for standard and high-definition video
    - Video-on-demand via various channels
  - Streaming services operated at 50-1500kbps with 2s or more of latency.



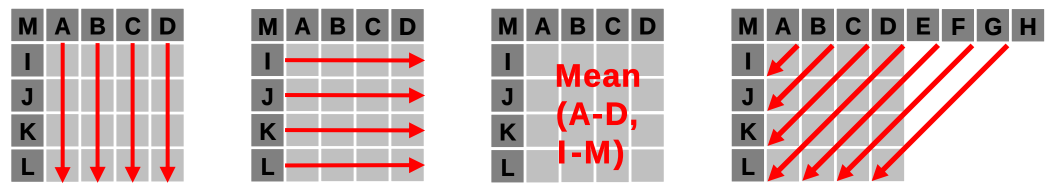
UCSB

Intra-Coded Macroblocks

	H.264	MPEG-1/2/4, H.261/3
Prediction in space domain	■ Spatial prediction ■ Encode the prediction modes (Use predictive coding if 4x4 modes are used)	■ No spatial prediction
Transform	■ Integer transform of residue	■ 8x8 Discrete Cosine Transform (DCT) for pixel values
Quantization	■ Quantization including scaling	■ Quantization
Prediction in frequency domain	■ No coefficient prediction	■ Coefficient prediction (for DC values in MPEG-2 and AC values in the first row and column in MPEG-4)

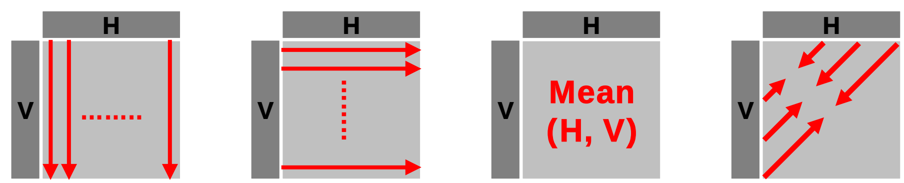
Spatial Prediction for Intra-Coded MBs

- luma
  - 4x4: 9 modes

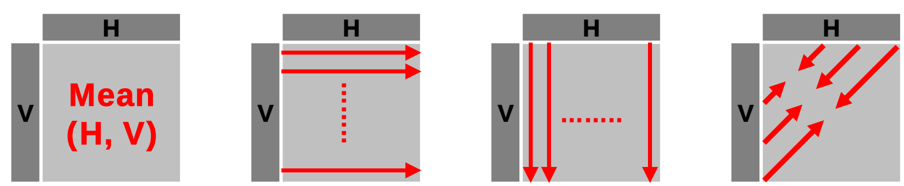


…

  - 16x16: 4 modes



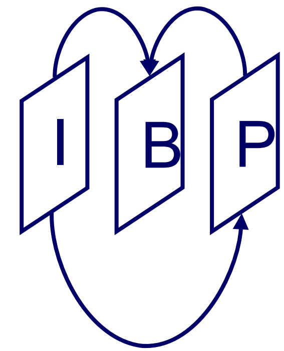
- chroma
  - 8x8: 4modes



  - The same prediction mode is always applied to both chroma blocks

UCSB

Inter-Coded Macroblocks

	H.264	MPEG-1/2/4, H.261/3
References	▪ Permits up to 15 (2 mostly used) reference pictures ▪ Bi-predictive B-slices ▪ A P-slice may reference a picture that has B-slices ▪ Supports explicit weighting coefficients and (a+b)/2 type	▪ A P-slice references only one I-picture ▪ Bi-directional B-slices  ▪ Only permit (a+b)/2 type prediction weighting
Block Sizes	▪ Tree-structured (16x16 → 16x8, 8x16, 8x8 → 8x4, 4x8, 4x4)	▪ Either 16x16 or 8x8
Motion Estimation	▪ half or ¼-pixel accuracy ▪ 6-point interpolation for half-pixel and 2-point linear interpolation for ¼-pixel	▪ MPEG2 permits half-pixel accuracy and MPEG4 permits ¼-pixel accuracy ▪ 2-point linear interpolation

Transform and Quantization – Type 3 (2)

- 52 quantization stepsizes (Qstep) indexed by quantization parameters (QP)

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

- Quantization

$$Z_{ij} = \text{round}\left(W_{ij} \cdot \frac{PF}{Qstep} \right) \;\Rightarrow\; Z_{ij} = \text{round}\left(W_{ij} \cdot \frac{MF}{2^{qbits}} \right)$$

$$\text{where } \frac{MF}{2^{qbits}} = \frac{PF}{Qstep} \text{ and } qbits = 15 + \text{floor}(QP/6)$$

- Integer arithmetic

$$|Z_{ij}| = (|W_{ij}|.MF + f) >> qbits$$
$$\text{sign}(Z_{ij}) = \text{sign}(W_{ij})$$

where $f=2^{qbits}/3$ for intra MBs and $2^{qbits}/6$ for inter MBs to control the quantization width near the origin (the “dead zone”)

- The advantages of the new transform and quantization scheme:
  - Integer transform avoids the inverse-transform mismatch.
  - Smaller blocksize (4*4) leads to a significant reduction in ringing artifacts.
  - No multiplication involved. Requires only 16-bit arithmetic.

Entropy Coding

Parameters to be coded	entropy_coding_mode=0	entropy_coding_mode=1
Macroblock type (Intra/Inter)	Exponential Golomb codes (Exp_Golomb) Variable Length Coding (VLC)	Context-based Adaptive Binary Arithmetic Coding (CABAC)
Coded block pattern		
Quantizer parameter		
Reference frame index		
Motion vector		
Residual data	Context-adaptive variable length coding (CAVLC)	

UCSB

Deblocking Filters

- A boundary-strength (BS) parameter is assigned to every 4×4 block

Block modes and conditions	Boundary-Strength parameter (BS)
One of the blocks is intra-coded and the edge is a MB edge	4
One of the blocks is intra-coded	3
One of the blocks has coded residuals	2
Difference of block motion ≥ one luma sample distance	1
Motion compensation from different reference frames	1
Else	0

- BS = 0 → No filtering
  BS = 1-3 → Slight filtering
  BS = 4 → Strong filtering
- Filters only when
  - $|P_0-Q_0|< \alpha$
  - $|P_1-P_0|< \beta$
  - $|Q_1-Q_0|< \beta$

P3	P2	P1	P0	Q0	Q1	Q2	Q3

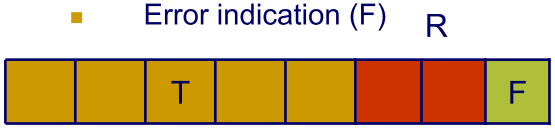
- <u>Thresholds α and β</u> depend on the <u>average quantization parameter (QP)</u>
- The deblocking filtering accounts for 1/3 of the computational complexity of a decoder.

Contributions of the VCL Tools

Spatial Prediction for Intra-coded Macroblocks	Saves 6-9% bits
Temporal Prediction	Saves around 50% bits
Transforms	PSNR less than 0.02dB
Logarithmic Quantization	A change in step size by 12% also saves 12% bits
CAVLC	Saves 5-8% bits
CABAC	Saves 5-15% bits over CAVLC
Picture-adaptive frame/field (PAFF) coding	Saves 16%-20% bits
MB-adaptive frame/field (MBAFF) coding	Saves 14-16% bits over PAFF
Deblocking Filter	Saves 5-10% bits

H.264 Over IP

- Network Abstraction Layer Unit (NALU)
  - A byte stream of variable length
  - 1-byte header
    - NALU type (T)
    - NALU importance (R)
    - Error indication (F)

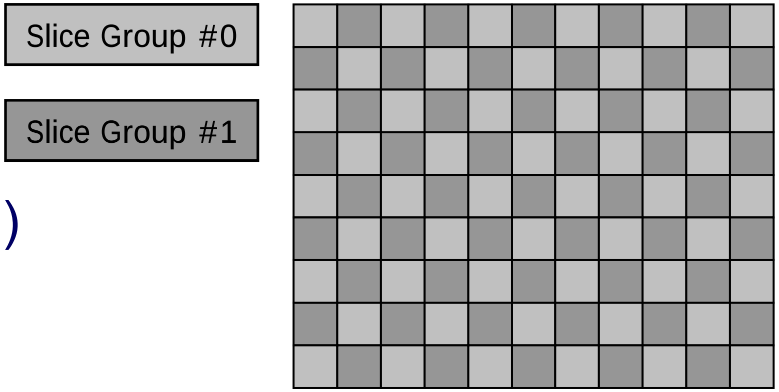


- RTP packetization
  - Simple packetization
    - One NALU in one RTP packet
    - NALU header as RTP header
  - NALU fragmentation
  - NALU aggregation

OSI/RM	**Protocols and specifi-cations for H.264**
Application Layer	RTP (Real-Time Transport Protocol)
Presentation Layer	Header size: IP/UDP/RTP = 20+8+12=40 bytes
Session Layer	Media-Unaware RTP payload specifications to reduce the loss rates observed by the decoder. Packet duplication/Packet based FEC/Audio redundancy coding Control protocols: H.245, SIP (Session Initiation Protocol), SDP (Session Description Protocol), RTSP (Real-Time Streaming Protocol)
Transport Layer	UDP (User Datagram Protocol)
Network Layer	IP: best effort service
Data Link Layer	
Physical Layer	

UCSB

Error-Resilience Tools



- Parameter sets
  - Sequence parameter set
  - Picture parameter set
- Flexible macroblock ordering (FMO)
  - Allows to assign MBs to slices in an order other than scan order
- Arbitrary slice ordering (ASO)
  - Improved end-to-end delay in real-time applications
- Redundant slices (RS)
  - Redundant representations are coded using different coding parameters
- Data partitioning with Unequal Error Protection (UEP)
- Feedback from decoder to encoder
  - Acknowledging correctly received slices (ACK)
  - Not acknowledging message (NAK)

Comparison of all the coding standards

	Applications	Bitrate	Coding efficiency	Input format	complexity
MPEG-1	VCD	1.5 mbits/sec (1.15 mbits/sec for video data)		maximum frame size 4095x4095, maximum frame rate 60 frames/sec	
MPEG-2/H.262	Digital TV standard by ASTC& DVB, DVD	5-10Mbits/s At first, the main focus of MPEG-4 was the encoding of video and audio at very low rates. In fact, the standard was explicitly optimized for three bit rate ranges: Below 64 kbits/sec, 64 to 384 kbits/sec, 384 kbits/s to 4 mbits/s.			For the same sample rate, and MPEG-2 encoder is about 50 percent more complex than MPEG-1 encoder.
MPEG-4	multimedia and Web compression				
H.261/2/3/3++	Videoconferencing and videotelephony	typically 384 kbit/s for videoconferencing and less than 128 kbits/s for videophone defined for ISDN bit rates of p x 64 bits/s where p = 1-30		CIF(360/352x288):progressive format, frame rate 20, 15, 10 or 7 Hz. QCIF: half the resolution of CIF.	
H.264		Video bit rate ranges 64kbps – 240mbps Compression gain of 1-3dB over MPEG-4, 1-5dB over H.263, 3-6dB over MPEG-2	Twice of MPEG-2	QCIF to 4kx2k	Decoder processing rate 250k-250m pixels.s

References

- Peter Symes, Digital Video Compression, McGraw-Hill, 2003.
- Joan L. Mitchell, MPEG Video Compression Standard, Chapman & Hall : International Thomson, 1996.
- William B. Pennebaker and Joan L. Mitchell, JPEG Still Image Data Compression Standard, Van Nostrand Reinhold, 1993.
- H.264/MPEG-4 Part 10 White Paper, www.vcodex.com
- IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on the H.264/AVC Video Coding Standard, Vol. 13, No. 7, July 2003.
- B. Yan and K.W. Ng, A Survey on the Techniques for the Transport of MPEG-4 Video Over Wireless Networks, IEEE Transactions on Consumer Electronics, Vol. 48, No. 4, November 2002.
- MPEG Video Compression Technique, http://rnvs.informatik.tu-chemnitz.de/~jan/MPEG/HTML/mpeg_tech.html
- P. Hoyingcharoen and C. Schmidt, Overview of H.264/AVC, Signal Compression Lab, UCSB.