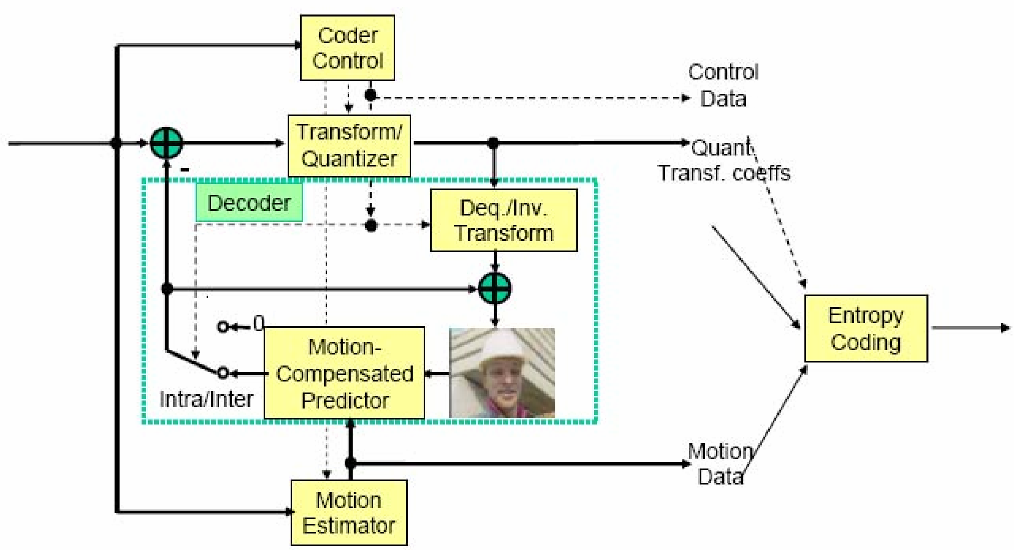
1. Video Standards and Systems[1]

Framework: Developments in broadcasting, personal computers, communication technologies and services, such as better data compression algorithms, fiber optic networks, distributed & network computing, DSP devices, and digital recording offer a variety of IT products in the near future. Driving the R&D in this e field are the consumer and commercial applications:

Digital television broadcasting	2 . . . 6 Mbps (10…20 Mbps for HD)	MPEG-2
DVD video	6 . . . 8 Mbps	MPEG-2
Internet video streaming	20 . . . 200 kbps	Proprietary, similar to H.263, MPEG-4, or H.264/JVT
Videoconferencing, videotelephony	20 . . . 320 kbps	H.261, H.263
Video over 3G wireless	20 . . . 100 kbps	H.263, MPEG-4

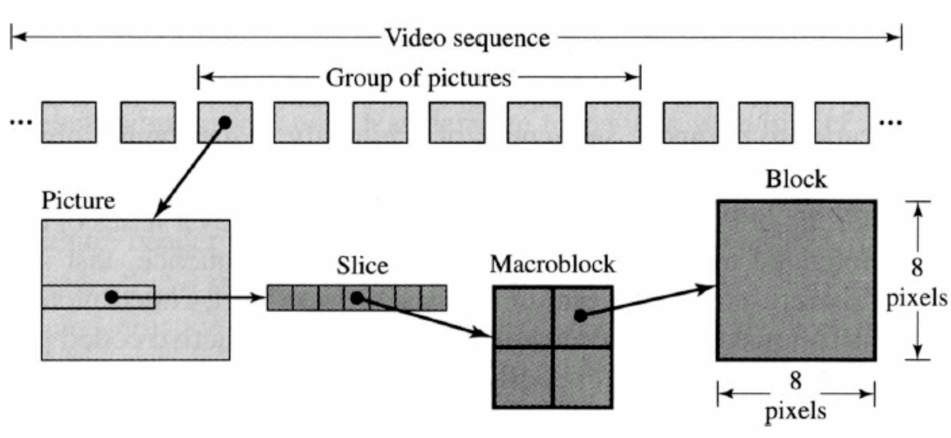
- Digital TV, including Advanced TV and HDTV @20 Mbps over 6.0 MHz taboo channels.
- Multimedia and desktop video @1.5 Mbps on CD-ROM or DVT (DAT) or HD storage and Video over IP (VOIP). (Do not mix with the term VoIP, which stands for "Voice over IP."
- Videoconferencing, H.261-H.264, MPEG4-MPEG21, and the new JVT Standard based on MPEG4.12, @ 384 Kbps and higher using *px64* ISDN channels and IP.
- Videophone & mobile image communication @ 10-33.4 Kbps on cellular & voice channels.
- Common tread to all of these systems is the principle of "Motion-Compensated Hybrid Coding" and the systems associated with it.



[1] The material in this chapter has been provided by Professor A. Murat Tekalp at Koc University, Istanbul, Turkey and VCDemo has been provided by Professor Inald Lagendijk of Delft University of Technology, The Netherlands.

Highlights of these systems:

1. Hierarchical Syntax



2. Syntax Hierarchy

SYNTAX HIERARCHY AS USED IN DIFFERENT VIDEO CODING STANDARDS*

Syntax layer	Functionality	Standard
Sequence (SC)	Definition of entire video sequence	H.261/3, MPEG-1/2
VOL (SC)	Definition of entire video object	MPEG-4
GOP (SC)	Enables random access in video stream	MPEG-1/2
GVOP (SC)	Enables random access in video stream	MPEG-4
Picture (SC)	Primary coding unit	H.261/3, MPEG-1/2
VOP (SC)	Primary coding unit	MPEG-4
GOB (SC)	Resynchronization, refresh, and error recovery in a picture	H.261/3
Slice (SC)	Resynchronization, refresh, and error recovery in a picture	MPEG-1/2
Video Packet (SC)	Resynchronization and error recovery in a picture	MPEG-4
MB	Motion compensation and shape coding unit	H.261/3, MPEG-1/2/4
Block	Transform and compensation unit	H.261/3, MPEG-1/2/4

*Each layer starts with a header. An SC in a syntax layer indicates that the header of that layer starts with a start code.

All-Digital TV

TV being the most commonly used image communication system used in the world is most appealing to developers and manufacturers of IT to be associated with. Digital TV has been in the agenda for long time, but it been recently taken into an active playground due to breakthrough in compression algorithms and the DSP technology with the efforts in the area of *High Definition TV* (HDTV).

Major Development Efforts for HDTV Research, Delivery and Products		
Japan	NHK	
Europe	EUREKA-95	
USA	Grand Alliance	ATT&, General Instruments, MIT, Philips N.A., David Sarnoff Research Center, Thompson Consumer Electronics, Zenith

NHK system being a hybrid system has had very limited success and eventual abandonment of the whole thing! All-digital TV is the big thing today. Although, the present TV standards are analog, digital TV signals find routine use in digital TV applications and it might even be a must for all-digital TV to be backward compatible with the good old TV sets! Here are the existing TV standards:

- **NTSC**
  1. 2:1 interlace with 4:3 aspect ratio.
  2. 525 lines/frame, 29.97 frames/s, (262.5 lines/field, 59.94 files/s.)
  3. Perceptually 340 lines/frame, 420 resolvable pels/line (Kell factor).
  4. Analog transmission over taboo 6.0 MHz channels.
  5. 68 channels in US: (54-88 MHz; Ch.2-6) + (174-216; Ch.7-13 MHz) + (470-806 MHz; Ch.14-69).
- **PAL**
  6. 2:1 interlace with 4:3 aspect ratio.
  7. 625 lines/frame, 50 fields/s.
  8. Analog transmission over 8.0 MHz channel.
- **SECAM** (Similar features as in PAL but different setup.)

US Grand Alliance:

After the initial efforts on ATV in Japan and IDTV and EQTV in Europe, FCC in US initiated a HDTV broadcast standard in 1987. After 6 years of competitive efforts of various groups to come up with a winner system, 4 candidates with all-digital architectures were taken into the test phase. All had flaws and nobody was declared winner.

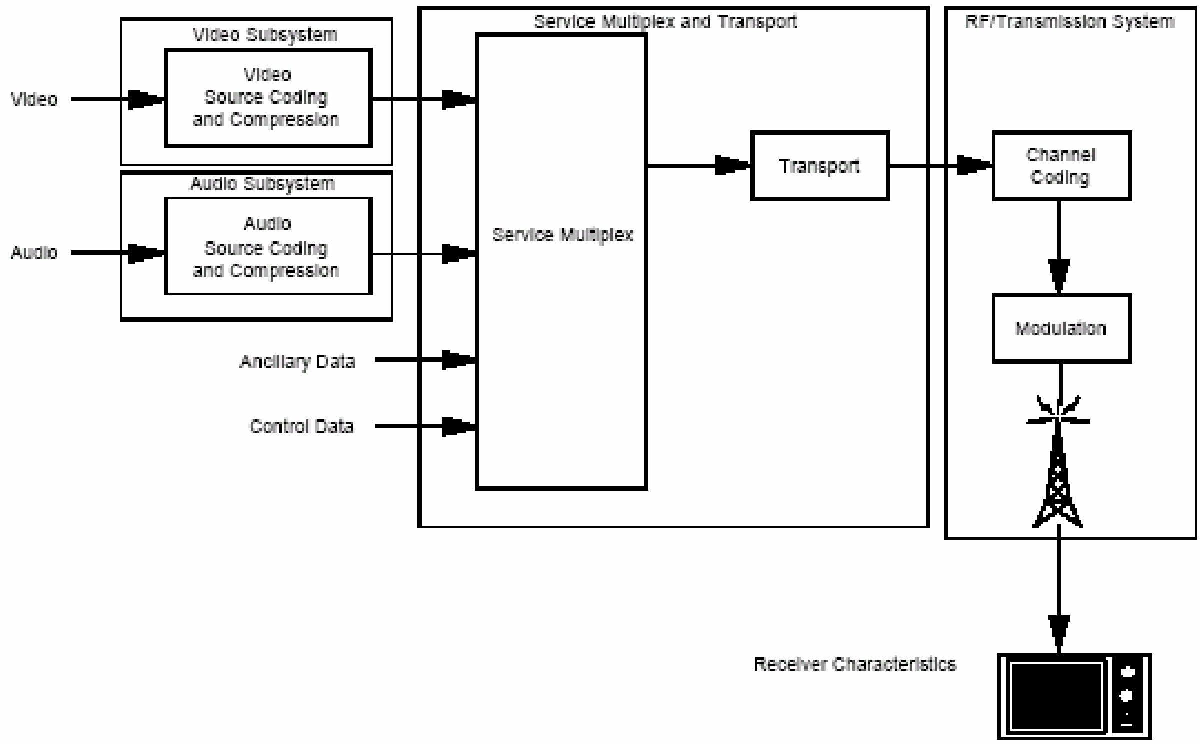
In May 1993, 4 groups consisting of (a) AT&T and Zenith Electronics Corporation; (b) General Instrument Corporation & MIT; (3) Philips Consumer Electronics & Thomson Consumer Electronics, and (4) the David Sarnoff Research Center came together to form the "***Digital HDTV Grand Alliance***." Their combined proposal became the HDTV Standard. Key points:

1. Video compression algorithm will be MPEG-2, main profile, high-level.
2. MPEG-2 transport mechanism will be used.
3. Dolby AC-3 audio system will be used, and
4. Three modulation techniques: 4-level VSB, 6-level VSB, and 32 QAM will be tested to complete specification.

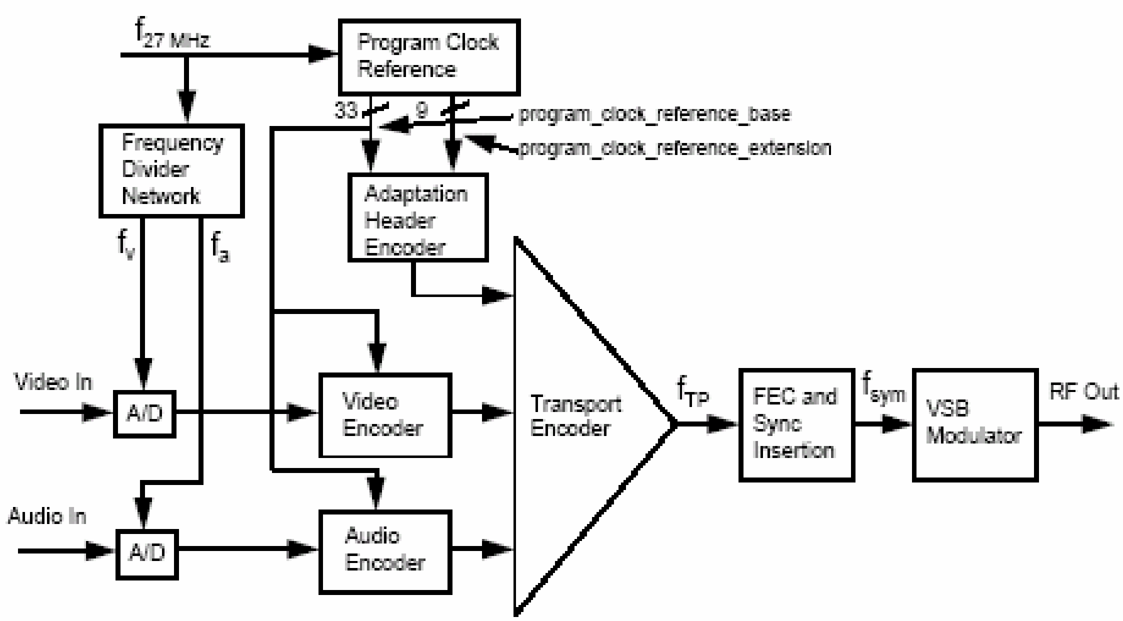
System Block Diagram:

A basic block diagram representation of the ATSC standard is shown below, which is based on a model adopted by the International Telecommunication Union, Radio-communication Sector (ITU-R), Task Group 11/3 (Digital Terrestrial Television Broadcasting). According to this model, the digital television system can be seen to consist of three subsystems.3

- Source coding and compression
- Service multiplex and
- RF/transmission



ITU-R digital terrestrial television broadcasting model.

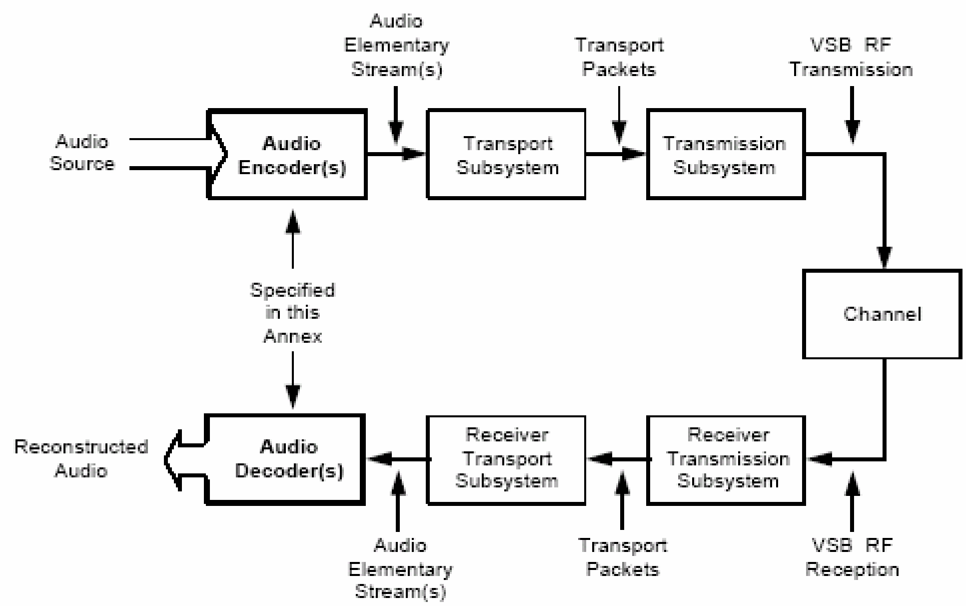


High level view of encoding equipment.

Possible Video Input Signals

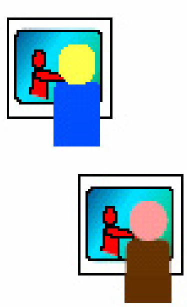
While not required by this standard, there are certain television production standards, shown in the table below that define video formats that relate to compression formats specified by the HDTV standard.

Video Standard	Active Lines	Active Samples/ Line
SMPTE 274M	1080	1920
SMPTE 296M	720	1280
ITU-R BT.601-4	483	720



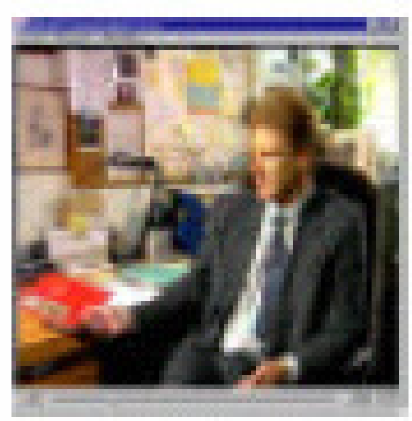
Audio subsystem in the digital television system.

Broadcast Applications:



TV:

- Traditional TV
- Narrowcast TV
- Video-on-Demand (VoD)
- Interactive TV
- 3-D TV
- ...

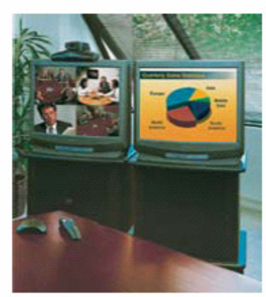


Further details on ATSC Standard can be found in: ATSC Standard: Digital Television Standard (A/53), Revision C at he URL: http://www.atsc.org

Videoconferencing: It refers to interactive distance conferencing with multi-media features over limited bandwidth channels (56 kb through DSL) with clear sound and sharp full-motion video. Two major product levels: Personal & Group

- Personal
  - Desktop, primarily for one-to-one communication
  - PC-based software, with simple video camera
    - PC handles audio and video compression
    - Requires Pentium-III or better PC
  - Either video capture card or direct via-USB
  - NetMeeting or similar software

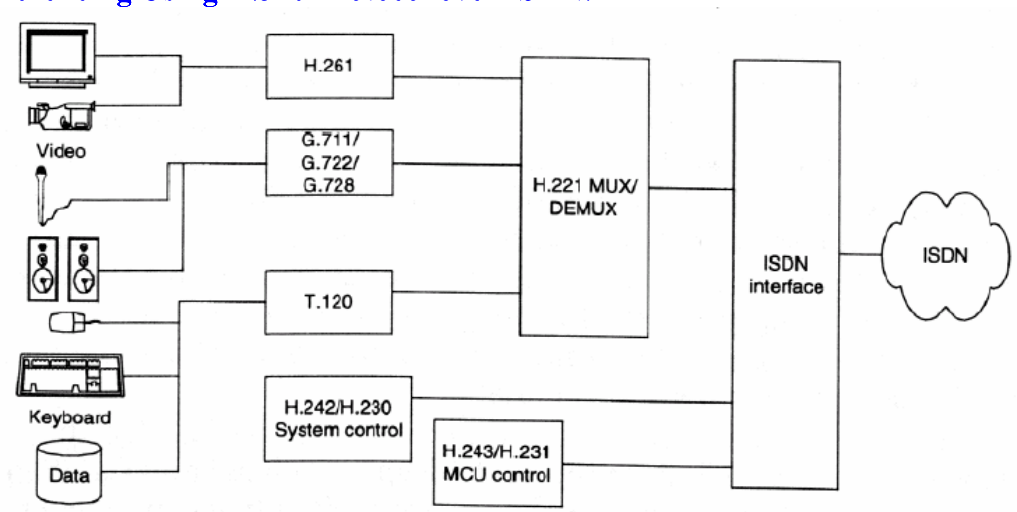
- Group
  - Meeting room, primarily for group-to-group communication
  - Dedicated system based around PC
    - Bulk of processing & software based on PC platform
    - Dedicated hardware
      - audio DSP echo cancellation,
      - sound detection
      - video compression/processing
      - camera motor control
      - remote keyboards/control



Videoconferencing Standards:

	H.320	H.321	H.322	H.323	H.324
Approved Date	1990	1995	1995	1996	1996
Network	Narrowband switched digital ISDN	Broadband ISDN, ATM	Guaranteed bandwidth packet-switched networks	Packet-switched networks (LAN/ WAN, ATM)	PSTN or POTS - the analog phone system
Video	H.261 H.263	H.261 H.263	H.261 H.263	H.261 H.263	H.261 H.263
Audio	G.711 G.722 G.728	G.711 G.722 G.728	G.711 G.722 G.728	G.711 G.722 G.723 G.728 G.729	G.723
Multiplexing	H.221	H.221	H.221	H.225.0	H.223
Control	H.230 H.242	H.242	H.242 H.230	H.245	H.245
Multipoint	H.231 H.243	H.231 H.243	H.231 H.243		
Data	T.120	T.120	T.120	T.120	T.120
Comm. Interface	I.400	AAL I.363 AJM I.361 PHY I.400	I.400 & TCP/IP	TCP/IP AAL	V.34 modem

Videoconferencing Using H.320 Protocol over ISDN:

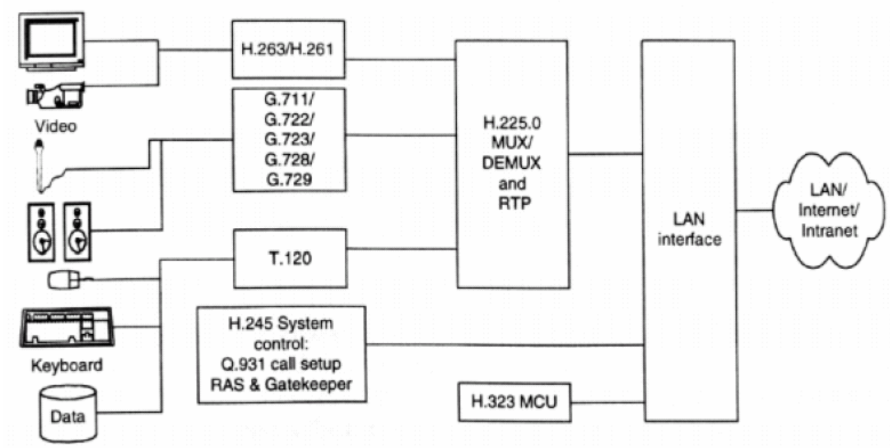


H.320 Protocol Highlights:

1. Supports QCIF and CIF resolutions
2. It uses H.261 Video Codec as a default codec.
3. Frame rates of 7.5, 10, 15 or 30 frames/second
   - Class 1: QCIF at 7.5 frames/second, no encoding, no motion compensation
   - Class 2: QCIF/CIF at 17 frames/second max, limited motion compensation
   - Class 3: CIF at 30 frames/second max, full encoding; motion compensation

4. Three possible Audio Codecs
   G.711 for Class 1
   G.722 and G.728 for Classes 2 &3.

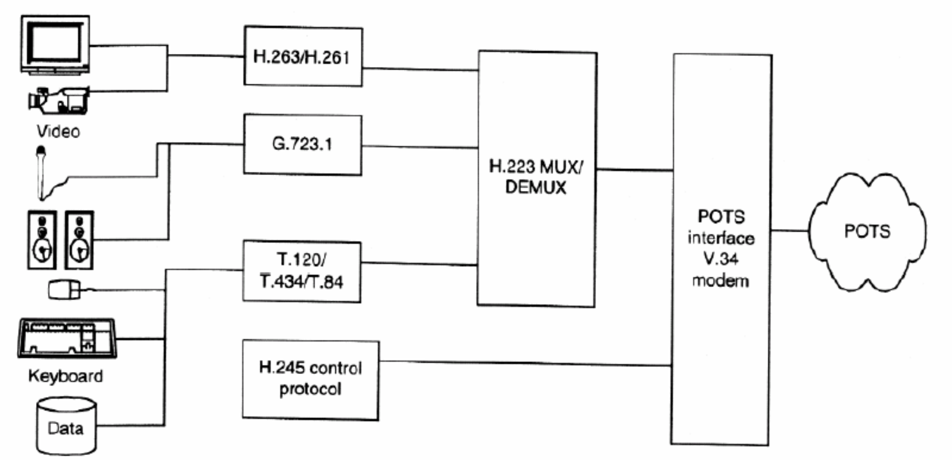
H.323 Protocol for LAN/ATM



H.323 Protocol Highlights

– Extends H.320 to LANs (Intranet and Internet) with no guaranteed QoS support Also specifies video conferencing over ATM including ATM QoS. Includes standards for system control of clients, multipoint servers, gateways and gatekeepers
– Defines the Real-time Transport Protocol (RTP)
– H.261 and H.263 Video Codecs and G.729 Audio Codec

H.324 Protocol for Terrestrial Public Telephone Service (POTS):



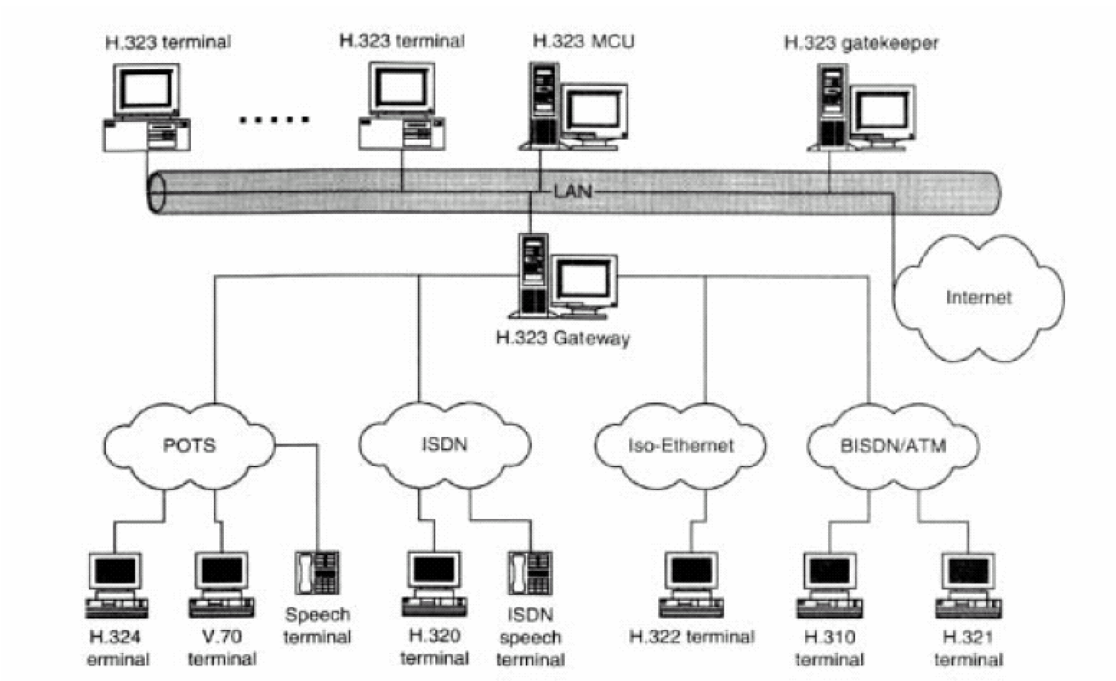
• Designed for Plain Old Telephone System (POTS) connections
  – 56Kbps modems
  – 28.8 and 33.3Kbps modems (V.34) can also be used

• Can use H.261 but most implementations will use H.263 as it can support Sub-QCIF (SQCIF) 128x96 frame size
  • G.723 Audio Codec
  • Codec algorithms are better than in H.320, but the overall quality is worse due to the more limited bandwidth available

Audio Codecs for Videoconferencing Standards and Systems:

1. G.711: Pulse Coded Modulation PCM encoding of 3.1KHz audio to 64kbps digital using either A-law (Europe) or $\mu - law$ mu-law encoding (Japan & North America).
2. G.721: Adaptive DPCM (ADPCM) of 3.1KHz audio to 32Kbps digital.
3. G.722: Defines how 7.5KHz audio is encoded to a 64Kbps using ADPCM.
4. G.723: Dual rate speech coder for transmitting 5.3 and 6.4Kbps.
5. G.728: Compression using Low Delay Code Excited Linear Prediction (LDCELP) to achieve 3.4KHz audio signal to a digitized signal of 16Kbps.

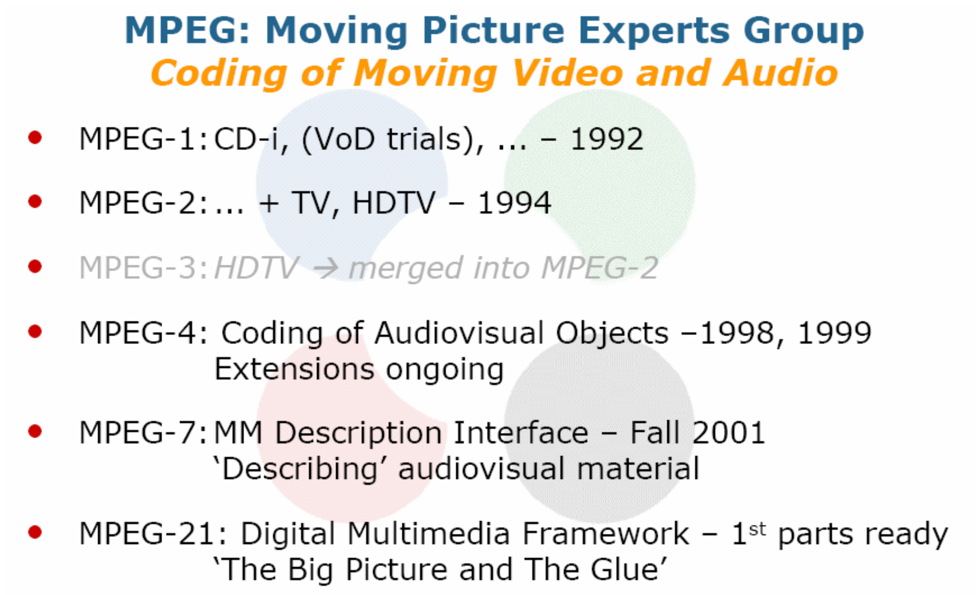
Videoconferencing Network Layout and the System Block Diagram:



The ISO MPEG-4 and Beyond

The MPEG Ad-Hoc Group of 1993 was commissioned to develop a fundamentally generic video coding standard at rates below 64 Kbps. Later, this mission has been modified to "provide an audio-visual coding standard allowing for

1. Interactivity,
2. Compression based on Discrete Wavelet Transform instead of DCT
3. Universal accessibility with high degree of flexibility, and
4. Scalability and Extensibility.
5. Very similar to XML.



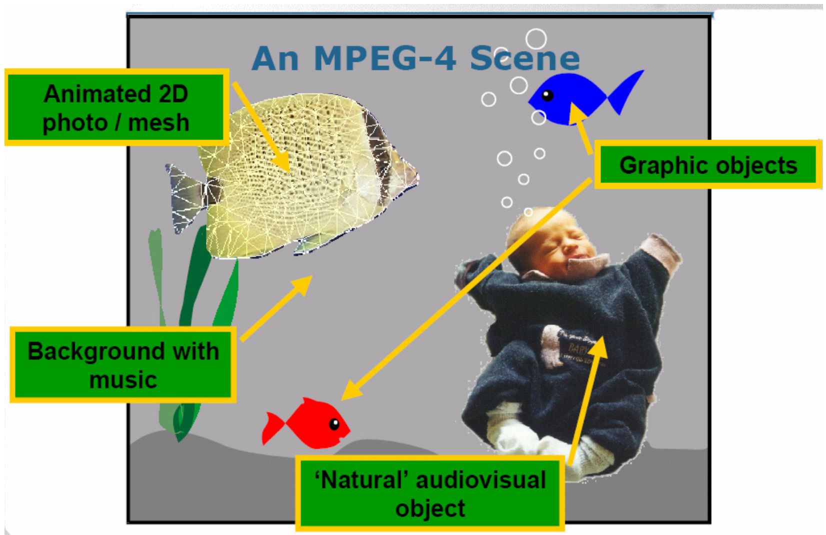
- **As of June 2003 MPEG-4 part 10 --(JVT Codec, Audio extensions and some Systems refinements—is the latest standard (AVC H.264).**

MPEG-4, 7, 21 levels include eight functionalities that are not supported by the existing platforms:

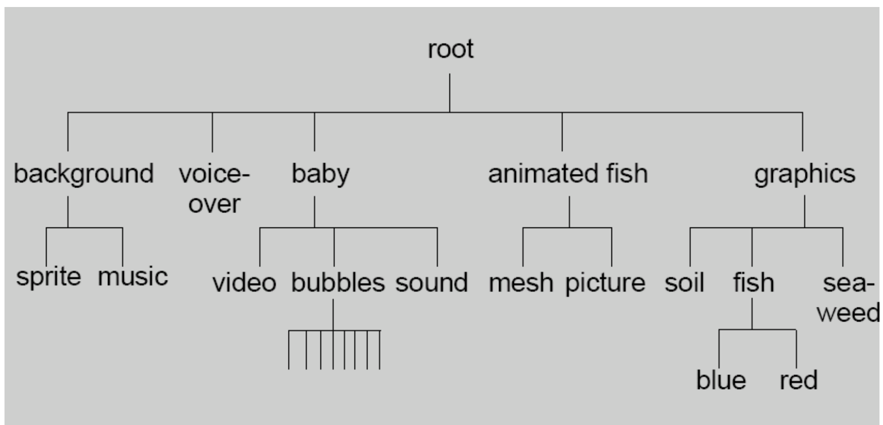
1. Content-based manipulation & bit stream editing.
2. Content-based multimedia data access tools.
3. Content-based scalability.
4. Coding of multiple concurrent data streams.
5. Hybrid natural and synthetic data coding.
6. Improved coding efficiency.
7. Improved temporal access at very low bit rates.
8. Robustness in error-prone environments.

Possible Application Areas:

- Architecture, real estate, and interior design (searching for ideas)
- Broadcast media selection (radio channel, TV channel)
- Cultural services (history, museums, art galleries, etc.) and digital libraries
- E-Commerce (personalized advertising, on Commerce (personalized advertising, on-line catalogues, various directories.
- Education (repositories of multimedia courses, search for support material)
- Home entertainment
- Investigation (human/human characteristics recognition, forensics)
- Journalism
- Directory services (yellow pages, Tourism, Geographical information systems)
- Multimedia editing
- Remote sensing
- Social (even dating services)
- Surveillance

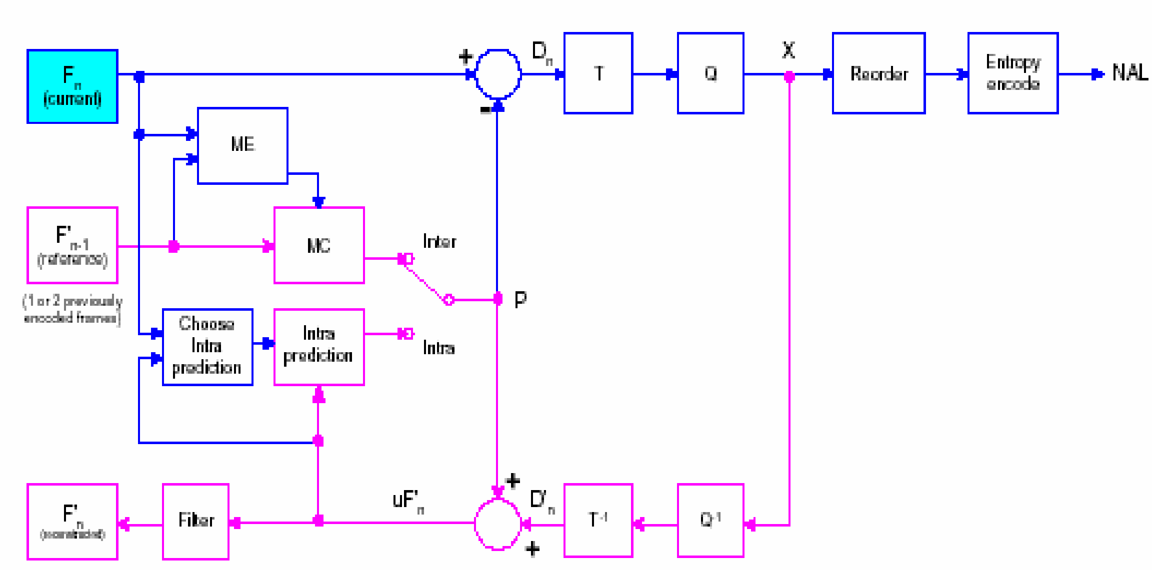


Scene Tree

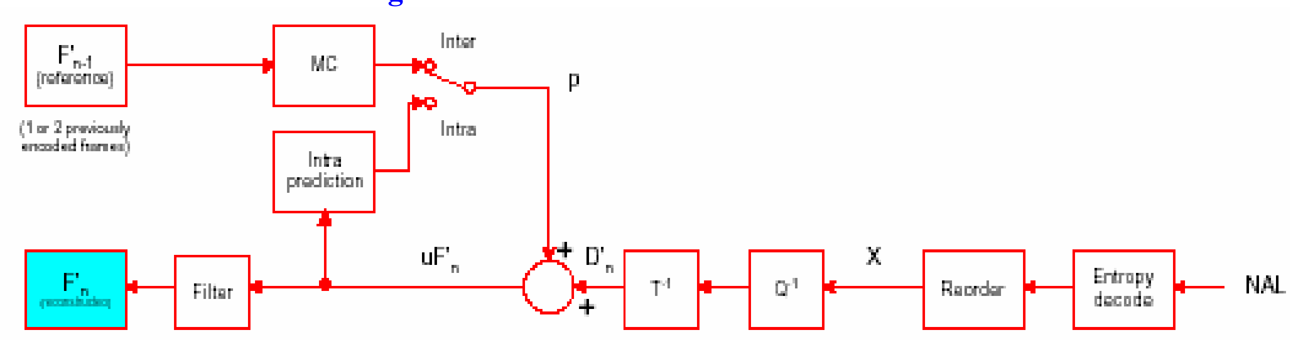


Products:

MPEG-4 Encoder Block Diagram:



MPEG-4 Decoder Block Diagram:



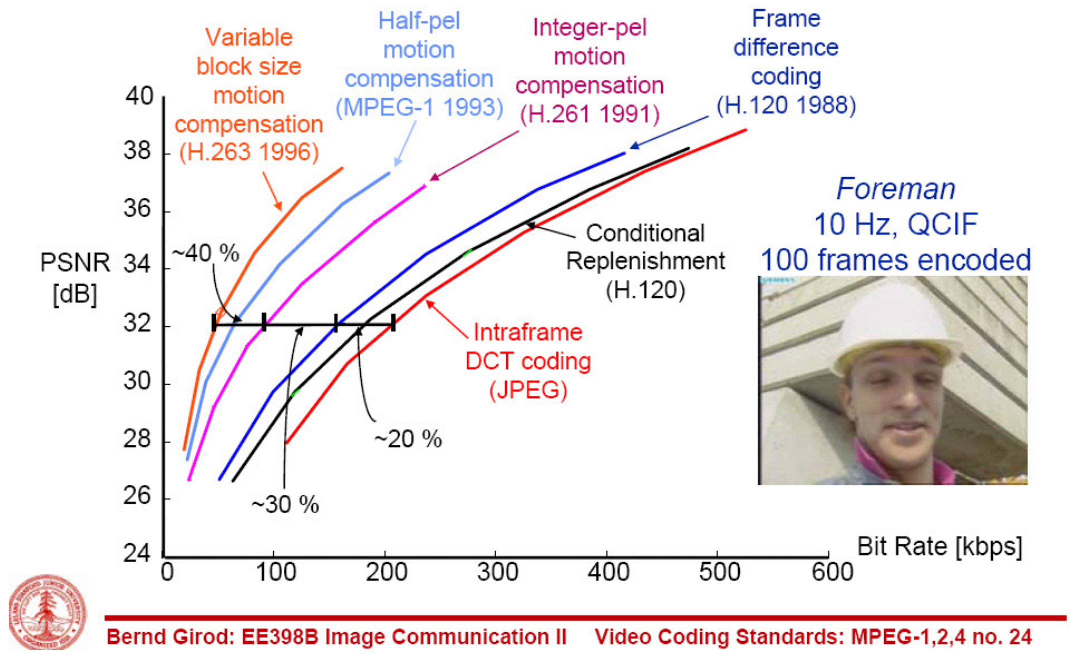
Emerging H.26L Video Compression Standards[2]

- H.26L family codecs (ITU-T Q.6/SG16 (VCEG) standardization activity for video coding especially aimed at 3G mobile networks and broadcast
- Possible formation of a joint video team with MPEG
- Goal for H.26L: 50% bit-rate reduction for same fidelity wrt earlier standards
- 1999: 3 proposals for definition of a first test model
  - HHI: Warping/OBMC motion model, wavelet- and context-based adaptive coding (CABAC)
  - Nokia: Affine motion model, multiple block transforms
  - Telenor: Block-matching with variable block-sizes, 4x4-DCT
- Current Status: Finalized on June 2003 as the 8th test model (TML-8) based on Telenor proposal.

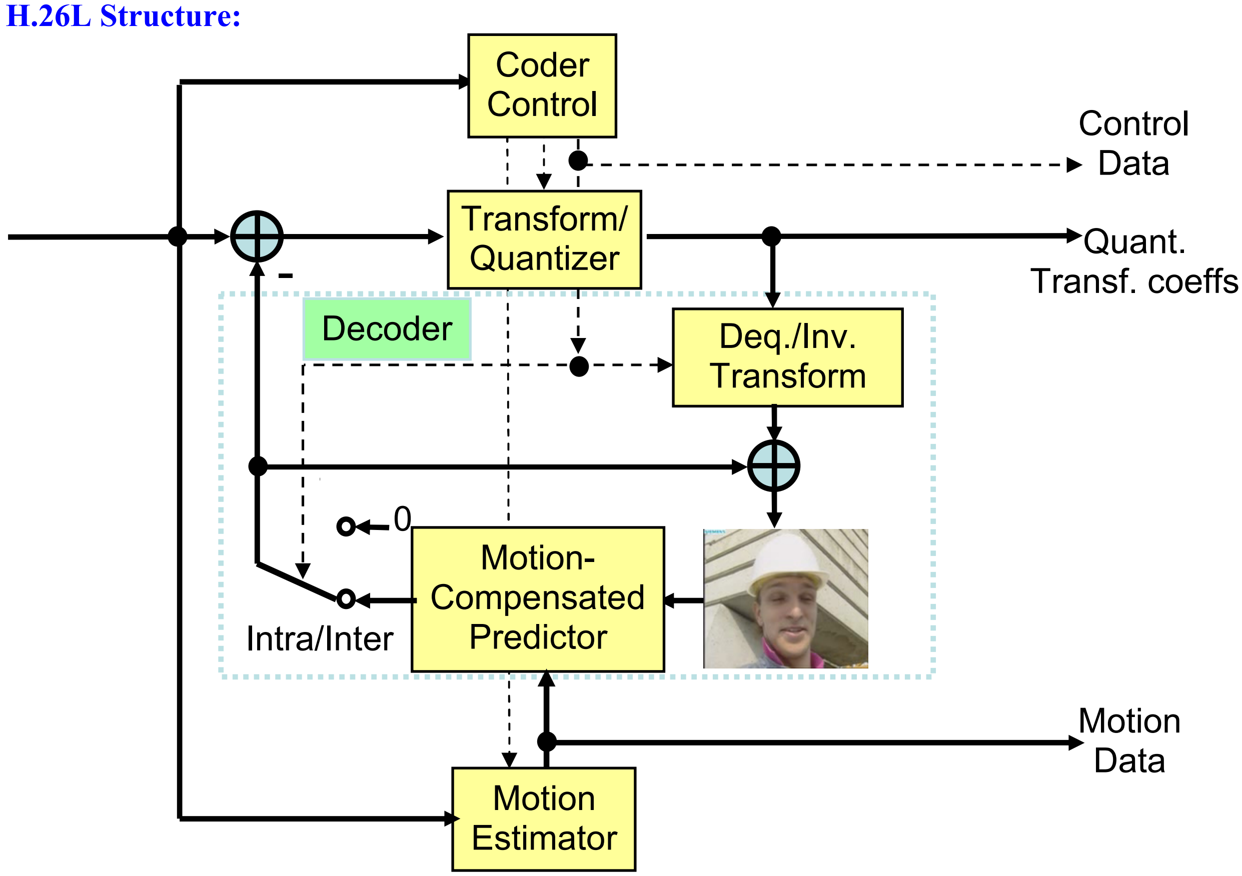
---

[2] The material on H.26L are provided by Thomas Wiegand of Heinrich Hertz Institute Berlin, Germany and Gary Sullivan of Microsoft. This topic also appears as the AVC Standard prepared by the Joint Video Team (JVT) formed by ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG). formed a Joint Video Team (JVT)

Video Compression Performance Progress: (From Berndt Girod's Video Coding Standards Course lecture notes at Stanford.)



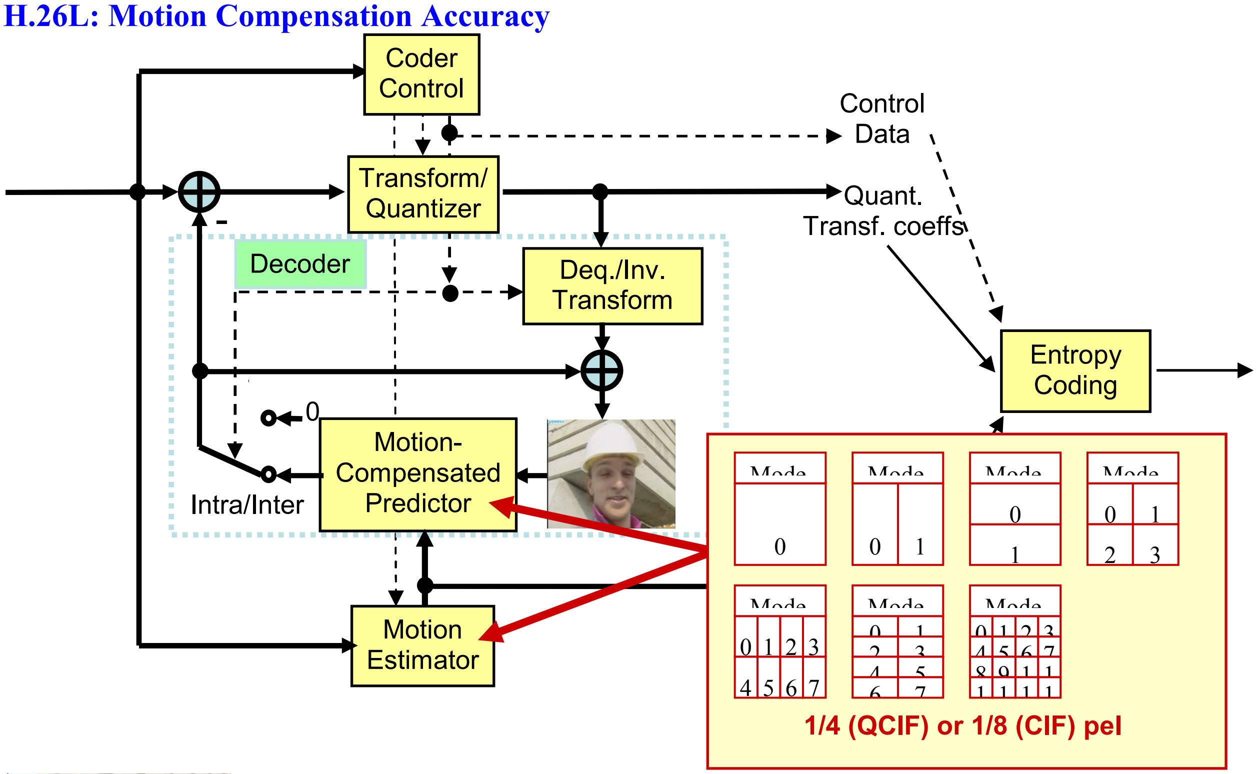
H.26L Structure:



The H.26L TML-8 Design (Part 1):

- Still using a hybrid of DPCM and transform coding as in prior standards.
- Common elements with other standards include:
  - 16x16 macroblocks
  - Conventional sampling of chrominance and association of luminance and chrominance data
  - Block motion displacement
  - Motion vectors over picture boundaries
  - Variable block-size motion
  - Block transforms (not wavelets or fractals)
  - Scalar quantization

H.26L: Motion Compensation Accuracy

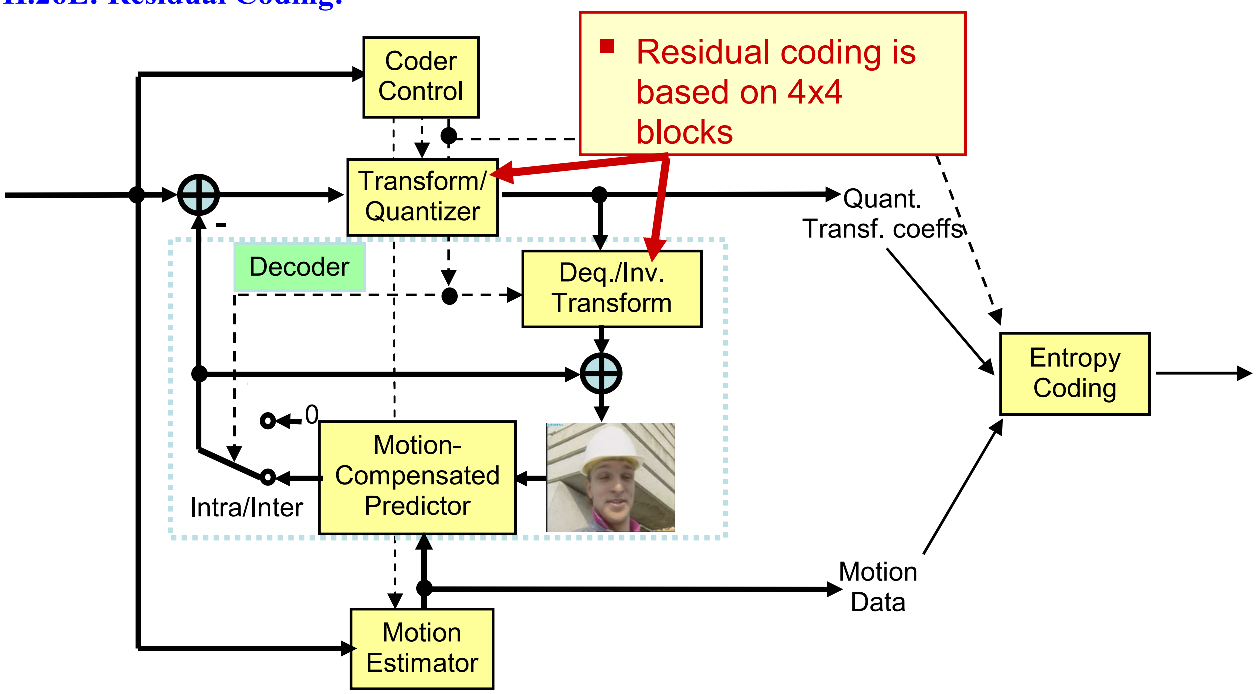


Multiple Reference Frames are used for Motion Compensation
The H.26L TML-8 Design (Part 2): Motion Compensation

- Various block sizes and shapes for motion compensation (7 segmentations of the macroblock: 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4)
- Multiple reference pictures (per H.263++ Annex U)
- Temporally-reversed motion
- B picture prediction weighting
- New "SP" transition pictures for sequence switching
- 1/4 sample (sort of per MPEG-4) and 1/8 sample accuracy motion:

1. 6x6 tap filtering to 1/2 sample accuracy, bilinear filtering to 1/4 sample accuracy, special position with heavier filtering
2. 8x8 tap filtering applied repeatedly for 1/8 pel motion

H.26L: Residual Coding:



The H.26L TML-8 Design (Part 3):

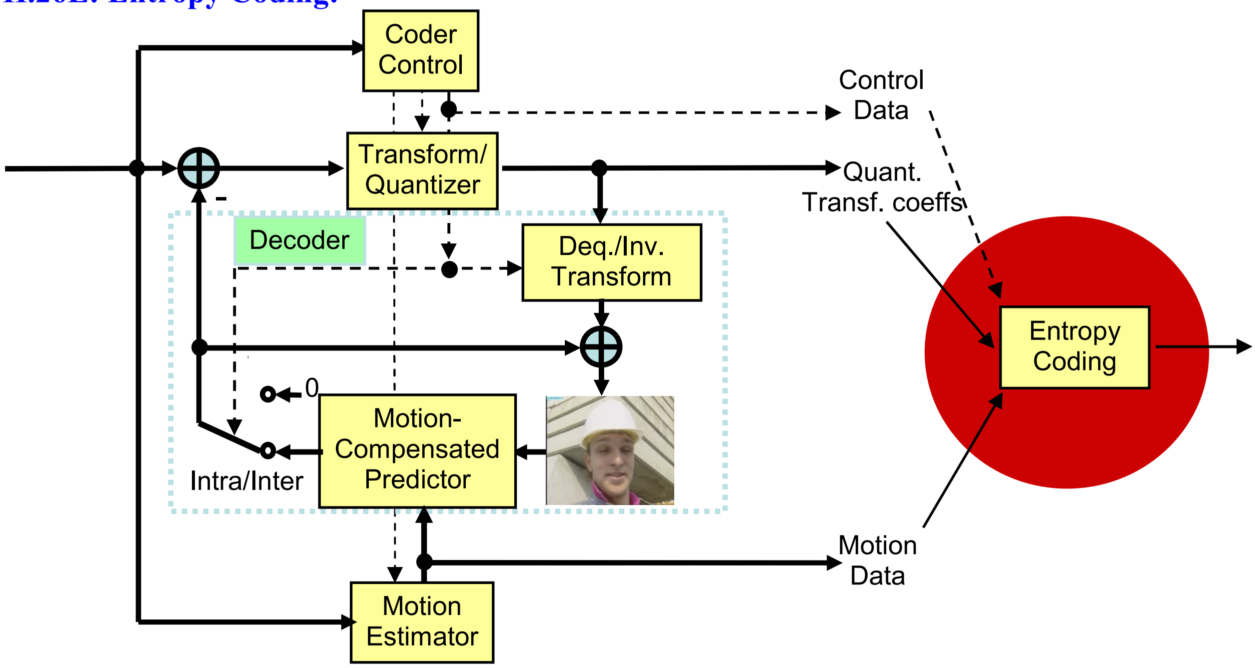
- Transform
  - Integer transform approximating a DCT
  - Based primarily on 4x4 transform size (all prior standards used 8x8)
  - Expanded to 8x8 for chroma by 2x2 DC transform
- Intra Coding Structure
  - Directional spatial prediction (6 types luma, 1 chroma)
  - Expanded to 16x16 for luma intra by 4x4 DC Xfm

The H.26L TML-8 Design (Part 4):

- Quantization
  - Two inverse scan patterns
  - Logarithmic step size control
  - Smaller step size for chroma (per H.263 Annex T)
- Deblocking Filter (in loop)

- Distinct Network Adaptation Layer (NAL) design for network transport
    - Slice-structure coding
    - Data partitioning

H.26L: Entropy Coding:



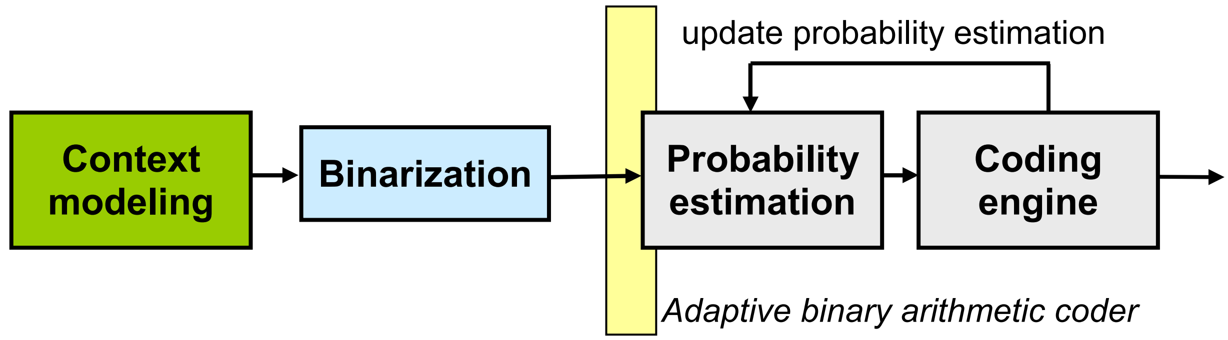
1. Entropy Coding in H.26L Based on Universal Variable Length Code (UVLC): Relatively simple design with some disadvantages:

- Probability distribution is static
- Correlations between symbols are ignored, i.e. no conditional probabilities are used
- Code words must have integer number of bits (Low coding efficiency for highly peaked probability densities, pdfs)

2. Context-based adaptive binary arithmetic codes (CABAC):

- Usage of adaptive probability models
- Exploiting symbol correlations by using contexts
- Non-integer number of bits per symbol by using arithmetic codes
- Restriction to binary arithmetic coding
    - Simple and fast adaptation mechanism
    - But: Binarization is needed for non-binary symbols
    - Binarization enables partitioning of state space

Technical Overview of CABAC:



Chooses a model conditioned on past observations

Maps non-binary symbols to a binary sequence

Uses the provided model for the actual encoding and updates the model

Bit Assignment (Binarization):

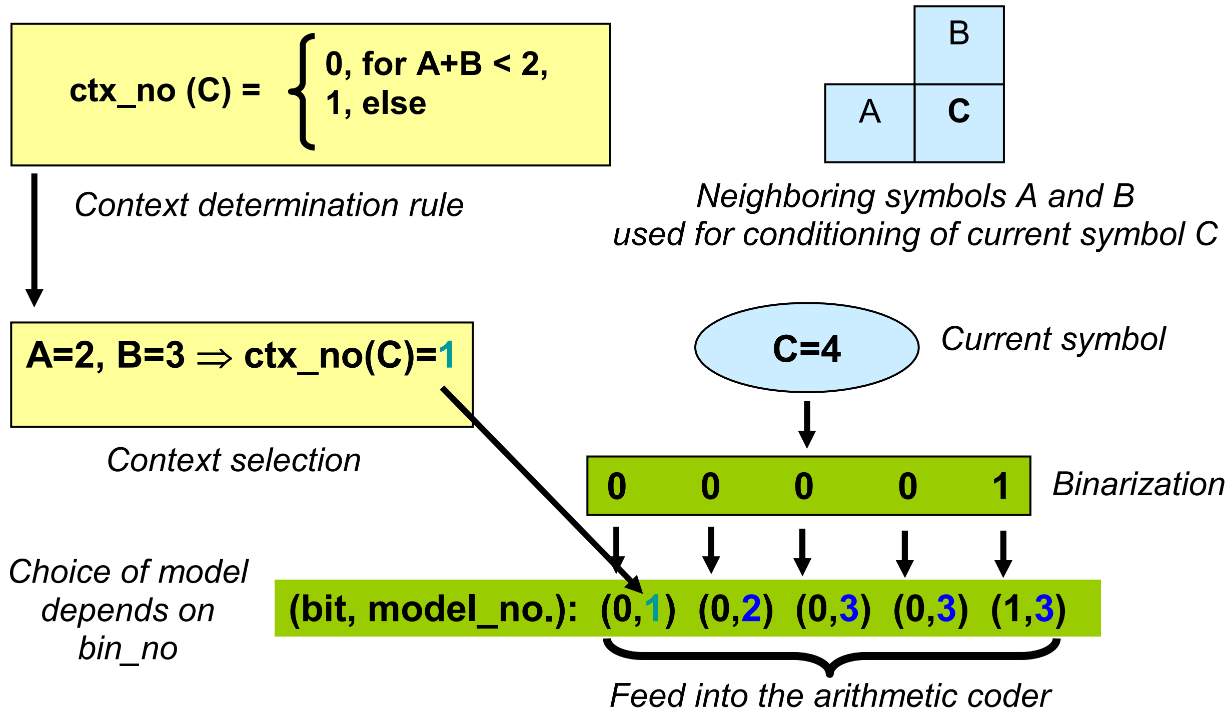
Symbol	Binarization
0	1
1	0 1
2	0 0 1
3	0 0 0 1
4	0 0 0 0 1
5	0 0 0 0 0 1
6	0 0 0 0 0 0 1
.	...
Bin_no	1 2 3 4 5 6 7 ...

Mapping to a binary sequence using the unary code tree:

- Applies to all non-binary syntax elements except for macroblock type
- Ease of implementation
- Optimal codes for a geometric pdf: $p(x) = 2^{-(x+1)}$
- Discriminate between binary decisions (bins) by their position in the binary sequence

Uses different models for different bin_numbers in the arithmetic coder

Example: Context Modeling and Binarization



Probability Estimation and Adaptation:

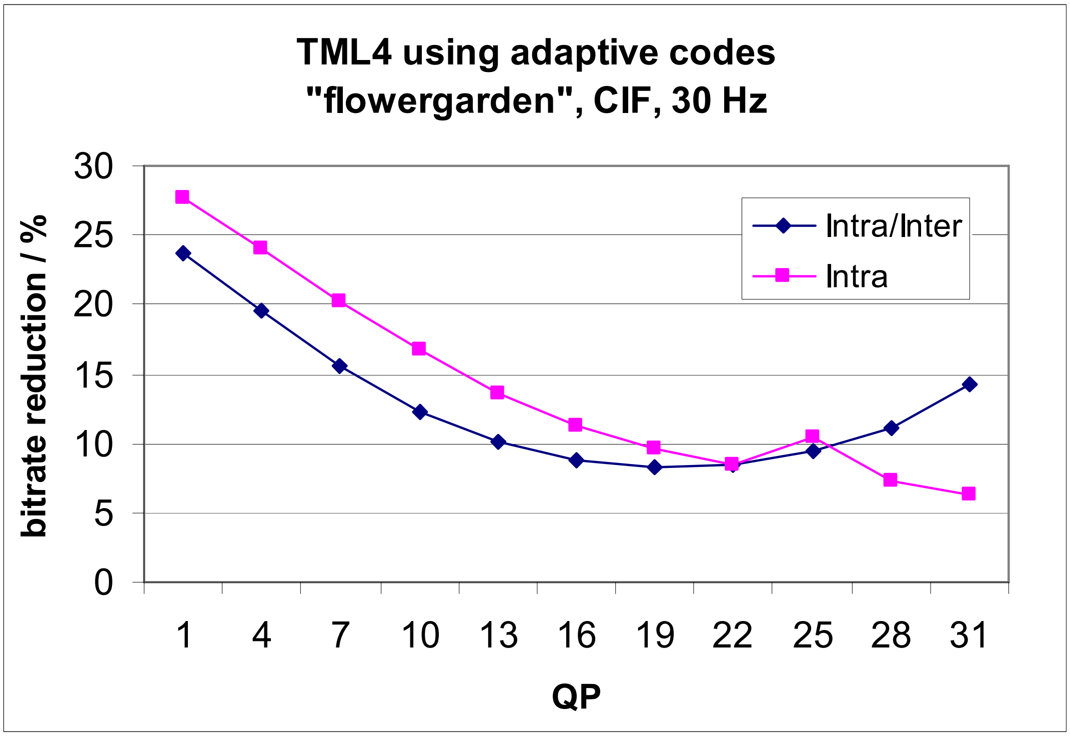
- Each model only consists of two counters: counts(„0“), counts(„1“)
- Coding with multiple contexts (models) is easy to obtain because of a clean interface between model and coder
- Model (probability estimate) is updated after each symbol is encoded

Therefore, we have an adaptive model.

Binary Arithmetic Coding:

- Standard implementations use integer arithmetic
- Fast, multiplication-free variants of binary arithmetic coder exists: e.g. MQ-coder used in JBIG-2, JPEG-LS, JPEG-2000
- *Estimation*: Increase in computational complexity lower than 10% (MQ) and 20% (Standard-AC) of the total decoder execution time at medium bitrate

Results: Bit-Rate Reduction:
Test Sequence: Flower Garden with CIF Format and 30 frames/second.

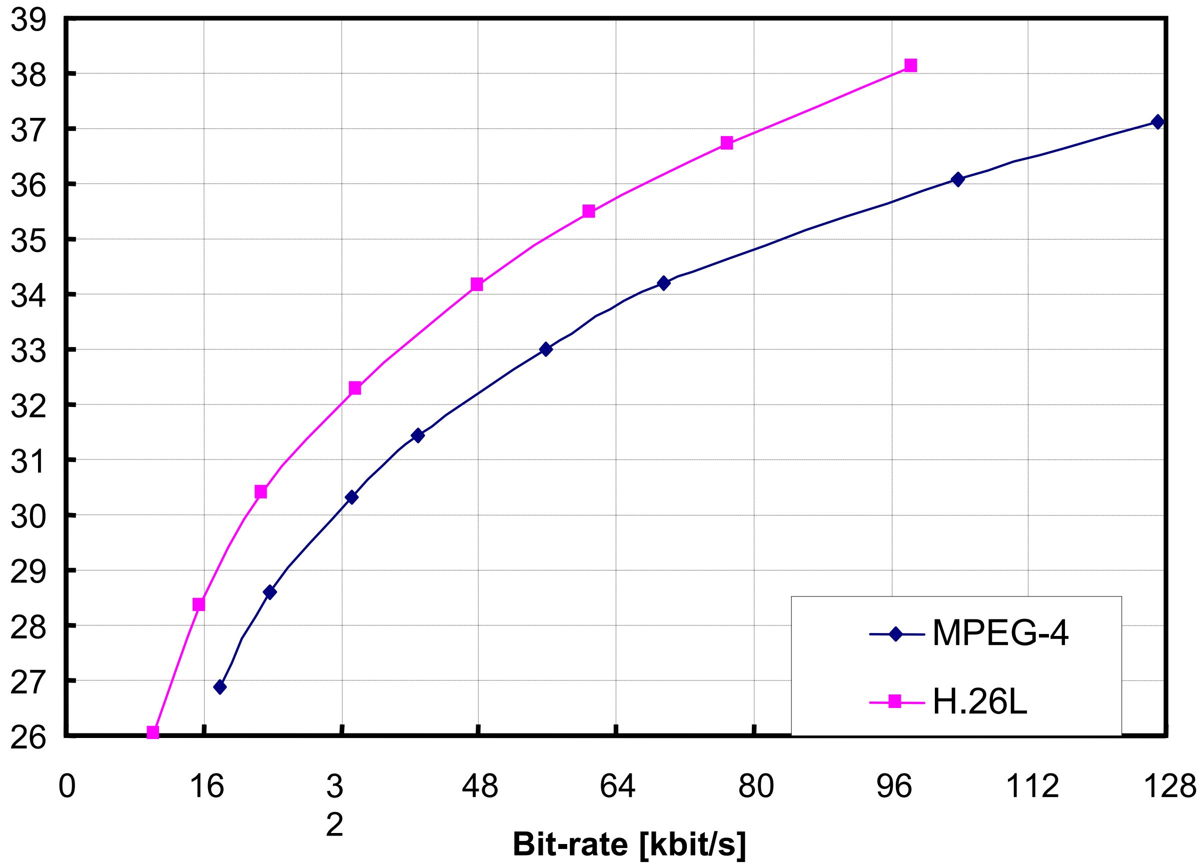


Comparison of H.26L to MPEG-4:

- MPEG-4: Advanced Simple Profile (ASP)
  - Motion Compensation: 1/4 pel
  - Global Motion Compensation
- H.26L:
  - Motion Compensation: 1/4 pel (QCIF), 1/8 pel (CIF)
  - Using CABAC entropy coding
  - 5 reference frames in 7 of 8 cases (News: 17 / 25)
- Both:
  - Sequence structure *IBBPBBP...*
  - QPB=QPP+2 (step size: +25%)
  - Search range: 32x32 around 16x16 predictor

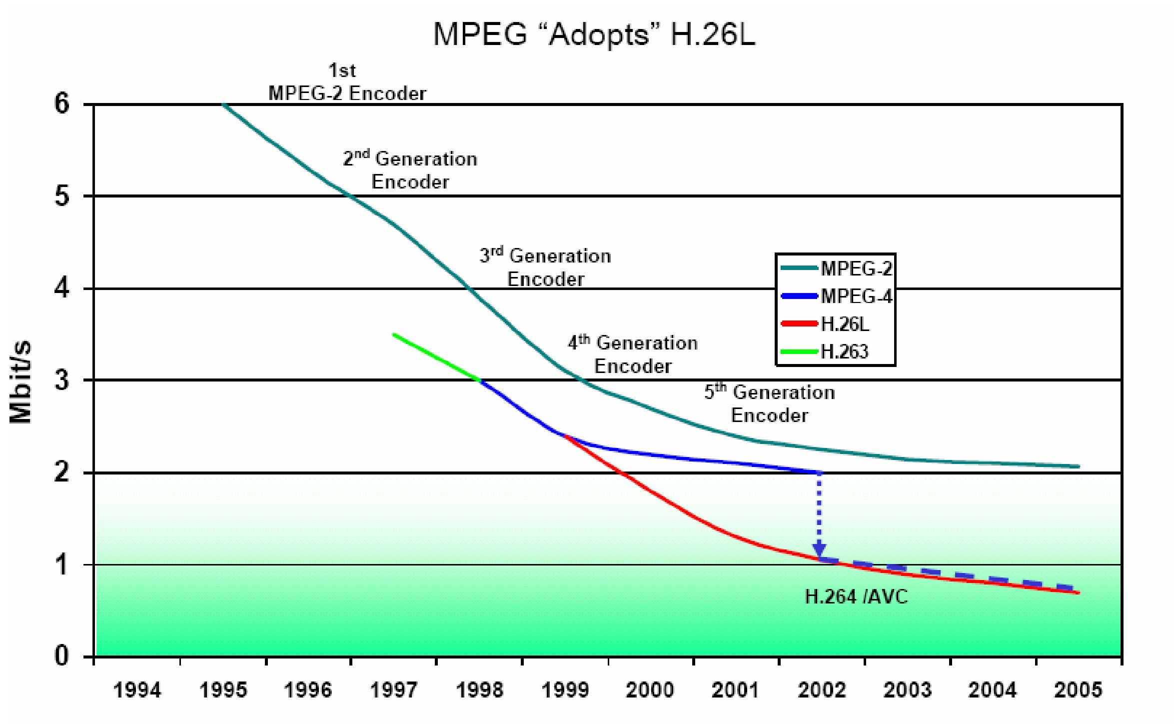
RD Curves:

Test Sequence: Foreman with QCIF format at 10 frames/second
Peak Signal-to-Noise Ratio **(PSNR**) in dB versus Bit-Rate in kbits/second



PSNR Results: H.26L vs. MPEG-4 Average PSNR for Luminance at Three Different Bit Rates
Test Sequences: Foreman, News Anchor, Container Ship and Tempete

Average A,B,C = 2.0 dB Average A,B,C,D,E,F = 2.1 dB			Average PSNR of Luminance		
			H.26L TML-8	MPEG-4 ASP	Gain
Case A (Avg. 1.9)	32 kbit/s QCIF, 10 fps	Foreman	31.9 dB	30.1 dB	+1.8 dB
		News	36.0 dB	33.3 dB	+2.8 dB
		Container Ship	38.5 dB	36.8 dB	+1.8 dB
		Tempete	29.3 dB	27.8 dB	+1.5 dB
Case B (Avg. 2.3)	64 kbit/s QCIF, 15 fps	Foreman	34.7 dB	32.8 dB	+1.9 dB
		News	39.4 dB	35.8 dB	***+3.7 dB***
		Container Ship	40.5 dB	38.6 dB	+1.9 dB
		Tempete	31.3 dB	29.4 dB	+1.9 dB
Case C (Avg. 1.9)	128 kbit/s CIF, 15 fps	Foreman	33.3 dB	31.3 dB	+2.0 dB
		News	38.7 dB	35.9 dB	+2.9 dB
		Container Ship	36.7 dB	35.4 dB	+1.3 dB
		Tempete	28.8 dB	27.6 dB	+1.3 dB



Conclusions

- H.26L Standard design is based on hybrid video coding
- Similar in spirit to other standards but with important differences
- Entropy coding can be conducted using
  - One VLC
  - Context-based adaptive arithmetic coding
- Context-based adaptive arithmetic coding provides improvements of 5-15 %
- H.26L delivers significant performance gain over existing standards including H.261-MPEG-4
- Bit-rate savings up to 50 % against MPEG-4.

1. More Information can be found in the H.26L standardization Documents and the associated websites:
http://www.itu.int/ITU-T/studygroups/com16/jvt/JVTToR.pdf

2. Draft Document can be downloaded from the ftp site:
ftp://ftp.imtc-files.org/jvt-experts/2002_12_Awaji/JVT-F100d2ncm.zip

3. JVT FTP Site: ftp://ftp.imtc-files.org/jvt-experts/
4. Wiegand's website: http://bs.hhi.de/
5. VQEG website: http://www-ext.crc.ca/vqeg/frames.html