Lecture 6
VIDEO COMPRESSION – Part 2

- Standards

MPEG-1 (ISO/IEC-11172) (1/2)

- MPEG, part of ISO IEC/JTC1/SC29/WG11, started 1988
- VHS quality video at 1-1.5 Mbit/s for storage on CD-ROM
- Oct. 1989: Competitive tests for video coding
- Sept. 1990: video part becomes Committee Draft; IS in May 1993
- Standard specifies bitstream syntax and decoder
- MPEG-1 standard consists of
  - ISO/IEC 11172-1: MPEG-1 Systems
  - ISO/IEC 11172-2: MPEG-1 Video
  - ISO/IEC 11172-3: MPEG-1 Audio
  - ISO/IEC 11172-4: MPEG-1 Conformance
  - ISO/IEC 11172-5: MPEG-1 Software

MPEG-1 (2/2)

⇦ Only the decoder is standardized along with the bitstream syntax

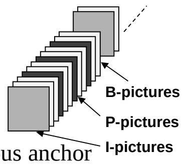
⇦ The video sequence is split into intra, predicted and interpolated frames

⇦ The video sequence is divided into group of pictures starting with intra frames

⇦ Motion vectors are obtained at half pel accuracy and sent to the decoder using DPCM at macroblock basis

⇦ Handles CCIR 601 and CIF formats

⇦ Covers bitrates of about up to 1.5 Mb/s

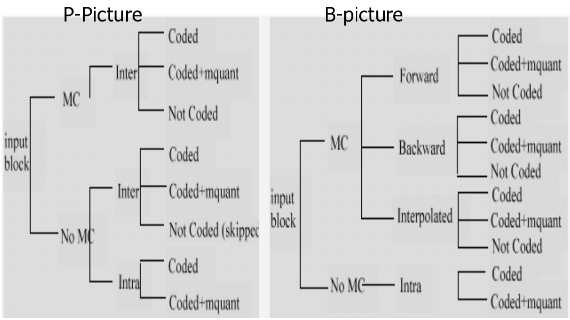


- INTRA I-frames:
  - random access
  - error robustness
- Predicted P-frames:
  - backward predicted from previous anchor picture (I or P)
- Bidirectionnally predicted B-frames:
  - forward/backward predicted from previous anchor picture (I or P)

New Features with respect to H.261

- Bi-directional motion compensation
- Motion compensation with up to half pixel accuracy
- Visually weighted quantization
- No picture size or bitrate restriction (except for constrained parameters)
- A flexible slice structure instead of group-of-blocks (GOBS)
- Two quantization characteristics: JPEG and H.261
- VLCs support a larger range of quantized DCT coefficients
- Separate VLCs for macroblock types in I, P and B pictures

MB Coding mode decisions



P-Picture

B-picture

MPEG-2 (ISO/IEC 13818) (1/2)

- Interlaced Video at 4-15 Mbit/s; DTV, CTV, DVD, Video on ATM
- Nov. 1993: Video part, stable as the Committee Draft
- MPEG-2 standard mainly consists of
  - ISO/IEC 13818-1: MPEG-2 Systems
  - ISO/IEC 13818-2: MPEG-2 Video
  - ISO/IEC 13818-3: MPEG-2 Audio
  - ISO/IEC 13818-4: MPEG-2 Conformance
  - ISO/IEC l3818-5: MPEG-2 Software
  - ISO/IEC 13818-6: MPEG-2 DSM-CC

MPEG-2 (2/2)

⇦ Only the decoder is standardized along with the bitstream syntax

⇦ Several modes are considered in order to take into account interlaced frames (field based modes)

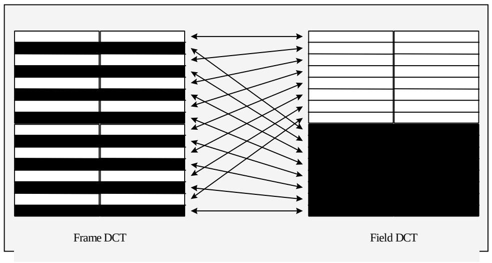
⇦ Generic structure in order to cope with several bitrates and picture formats

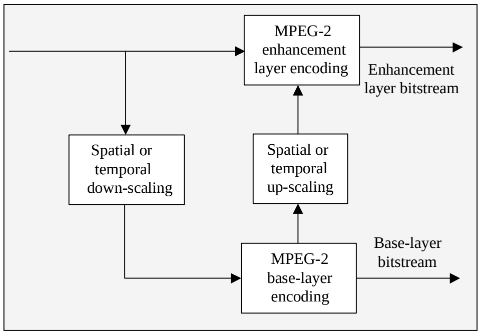
⇦ Spatial, frequency and temporal scalability

Summary of New Features in MPEG-2

- supports frame and field picture types for interlaced video
- allows 4:2:2 and 4:4:4 chroma in addition to the 4:2:0 format.
- supports new MC prediction modes for interlaced video
- supports field/frame DCT option per MB for frame pictures
- allows for finer quantization of the DCT coefficients.
- allows for finer adjustment of the quantizer scale factor.
- allows for a separate VLC table for the DCT coefficients for the intra macroblocks.
- allows alternate scan in addition to the zigzag scan.
- supports scalability/backward compatibility/error resilience
- supports six profiles and four levels

Interlaced/progressive coding in MPEG-2




MPEG-2 enhancement layer encoding
Enhancement layer bitstream
Spatial or temporal down-scaling
Spatial or temporal up-scaling
MPEG-2 base-layer encoding
Base-layer bitstream

Profiles and levels

- Profiles are a set of pre-defined tools and their configurations
- Profiles are divided into Levels each defining upper bound limits for coding parameters

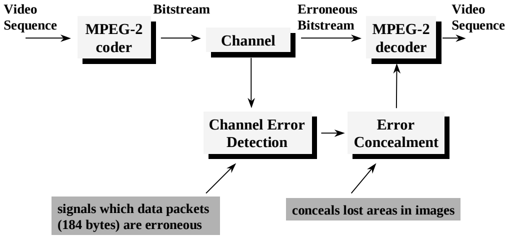
Profiles in MPEG-2

- Simple
  - Simplest profile similar to Main profile, except for the lack of B frames
- Main
  - Non scalable coding providing interlaced coding tools, random access, B mode.
- SNR Scalable
  - Similar to Main plus a 2 layer SNR scalability
- Spatial Scalable
  - Similar to SNR scalable profile plus a 2 layer spatial scalability
- High
  - Similar to Spatial Scalable profile with provisions for 3 layers in spatial and SNR scalability and 4:2:2 coding
- 4:2:2
  - Similar to Main profile with 4:2:2 coding

Levels in MPEG-2

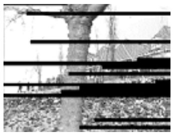
- Low
  - 352x288 pels, 30 f/s, 4Mb/s
- Main
  - 720x576 pels, 30 f/s, 15Mb/s
- High1440
  - 1440x1152 pels, 60f/s, 60Mb/s
- High
  - 1920x1152 pels, 60 f/s, 80Mb/s

Complete MPEG-2 transmission scheme



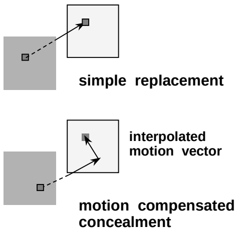
Error detection and resynchronization

- slice = basic resynchronization point (16 lines)
- channel error detection = skip detected erroneous data packets in the bitstream

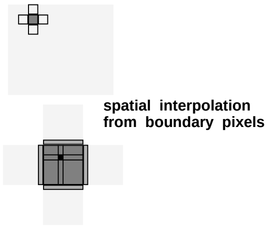


Error concealment

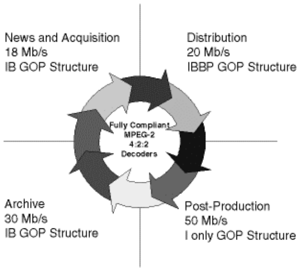
- Temporal concealment



- Spatial concealment



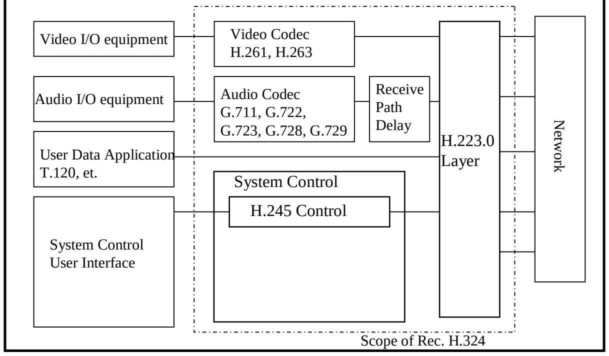
Significance of MPEG-2



Standard of Audiovisual Communication Systems

Network	GSTN	N-ISDN	Guaranteed QoS LANs	Nonguaranteed QoS LANs	ATM (B-ISDN)
Total System	H.324	H.320	H.320	H.323	H.310 H.321
Audio	G.723.1	G.711, G.722, G.728	G.711, G.722, G.728	G.711, G.722, G.723.1, G.728	G.711, G.722, G.728, ISO-IEC 11172-3
Video	H.261, H.263	H.261	H.261	H.261, H.263	H.261, H.262
Data	T.120, etc	T.120, etc	T.120, etc	T.120, etc	T.120, etc
Control	H.245	H.242	H.242	H.245	H.245, H.242
Multiplex & Sync	H.223	H.221	H.221	H.225.0, TCP/IP, etc	H.222.0, H.222.1
Call Setup Signaling	National Standard	Q.931	Q.931	Q.931, H.225.0	Q.2931

H.324 terminal equipment



Scope of H.245

- A message syntax and a set of protocols to exchange of control information between multimedia terminal
- Common control Protocol
  - Generic Recommendation employed in H.324(GSTN), H.310(ATM), H.323(Non guaranteed QoS LAN) multimedia terminals

Message Syntax

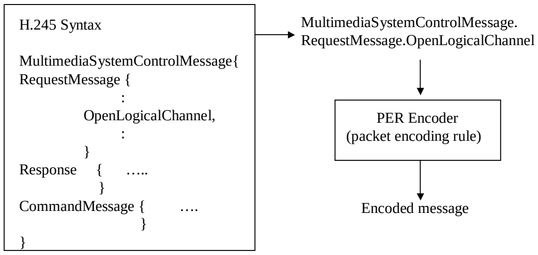
- H.245 uses ASN.1 (abstract syntax notation-one)
  - high level programming data declaration

Level 1	MultimediaSystemControlMessage			
Level 2	RequestMessage	ResponseMessage	CommandMessage	IndicationMessage
Level 3	Nonstandard Master/SlaveDetermination OpenlogicalChannel :	:	:	:

< Top Three level of the H.245 syntax >

Message Encoding

- Before transmission, Syntax must be encoded into bits



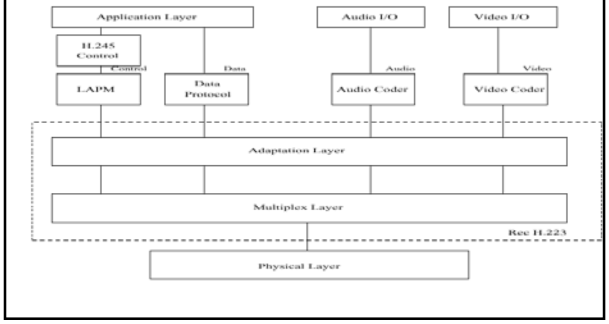
```
H.245 Syntax

MultimediaSystemControlMessage{
RequestMessage {
                  :
          OpenLogicalChannel,
                  :
          }
Response    {     …..
            }
CommandMessage {        ….
                  }
}
```

MultimediaSystemControlMessage.
RequestMessage.OpenLogicalChannel

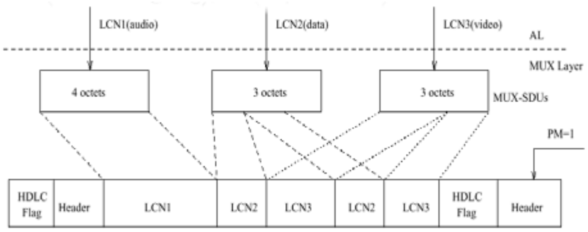
PER Encoder
(packet encoding rule)

Encoded message

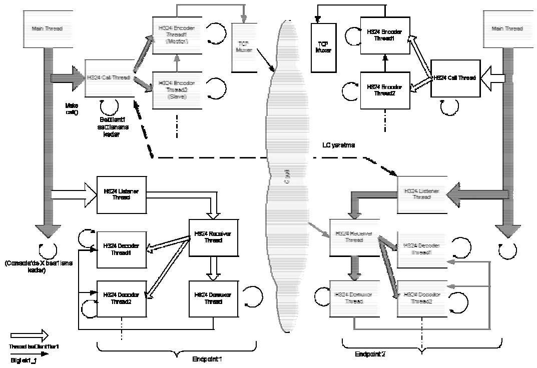
H.223 (1/2)



H.223 (2/2)



Block diagram



MPEG-2 System

- The goal of MPEG system
  - provides basic framework for integrated video, audio and data services
- Two schemes for Multiplexing process
  - Program Stream
  - Transport Stream

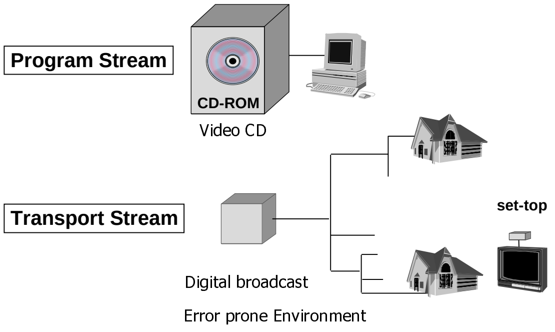
MPEG-2 Program Stream

- Primarily intended for storage and retrieval from storage media
- Grouping of video, audio, and data elementary streams that have a common time base
- Each program stream consists of only one program
- Useful in error free environments
  - Large packet size
  - Packets size may be variable (hard for decoder to predict start and end of packets)
- DVD standard uses the MPEG-2 Program Stream

MPEG-2 Transport Stream

- Combines multiple programs into a single stream
- The programs may or may not have common time base
- Fixed length packet size
  - Intended for non error free environments
  - Easier to detect start and end of frames
  - Easy to recover from packet loss/corruption
  - More difficult to produce and demultiplex than program stream

Transport and program streams


Program Stream

CD-ROM

Video CD

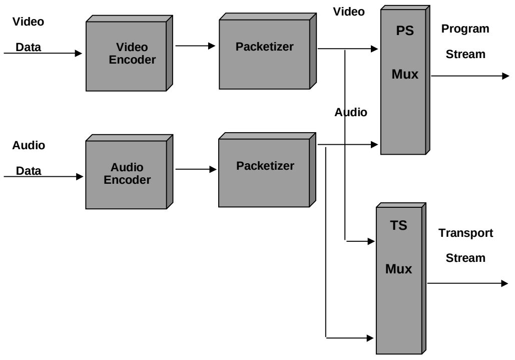
Transport Stream

set-top

Digital broadcast

Error prone Environment

Model for MPEG-2 system



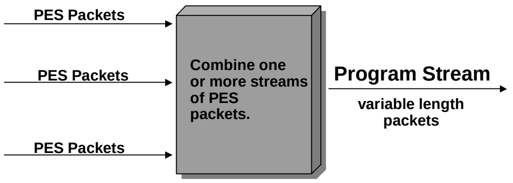
Packetized Elementary Stream (PES)

- Result of the packetization process
- The payload is the data bytes taken sequentially from the original elementary stream
- No specific format for forming the PES packet
  - Entire video frame in one PES packet (but need variable size frames)
  - Fixed size packets

Program stream

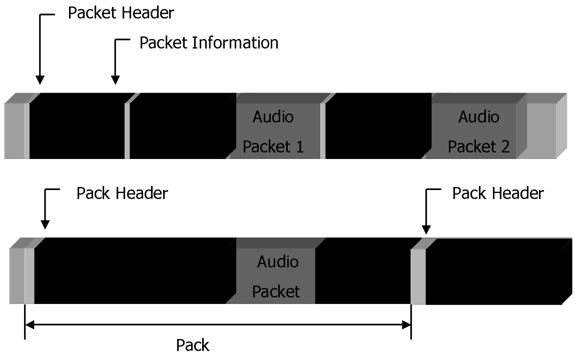
Multiplexing

For use in error free environment.

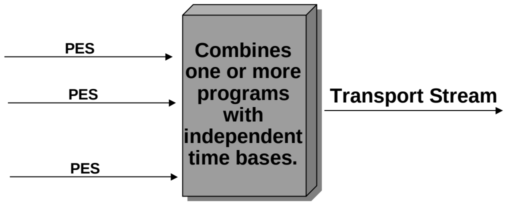


These streams of PES packets have a common base time.

Packet & Pack

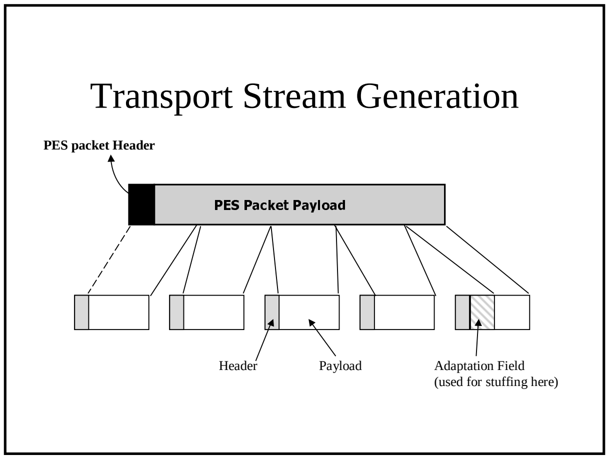


Transport stream (1/2)

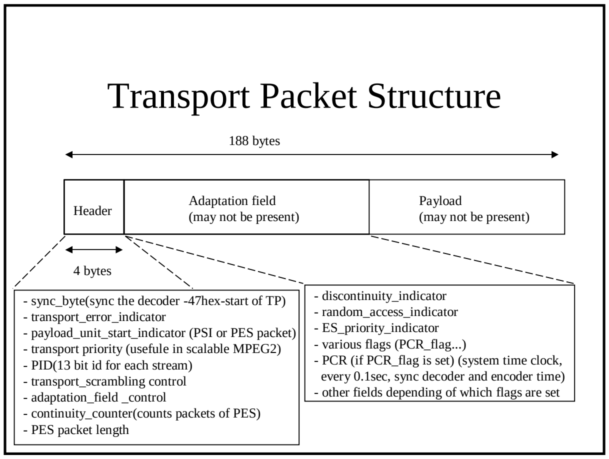


Transport Stream (2/2)

- Multiplexes various PES into one stream along with information for synchronizing between them
- Short, fixed length packets 188 bytes (4 byte header + adaptation field or payload or both)
- Constraints for forming transport packets:
  - First byte of PES packet must be first byte of transport packet payload
  - Each transport packet must contain data from only one PES packet



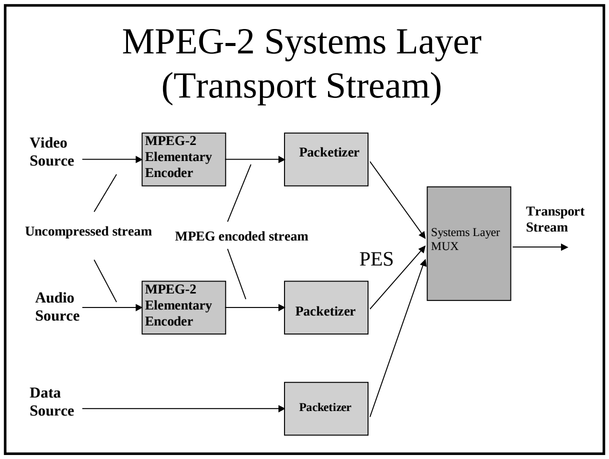
Transport Stream Generation



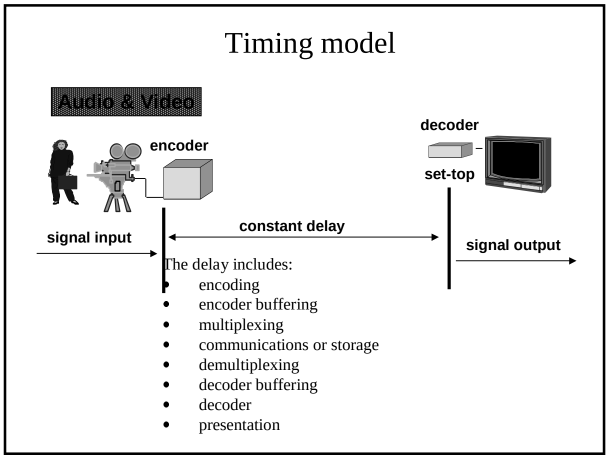
Transport Packet Structure

- sync_byte(sync the decoder -47hex-start of TP)
- transport_error_indicator
- payload_unit_start_indicator (PSI or PES packet)
- transport priority (usefule in scalable MPEG2)
- PID(13 bit id for each stream)
- transport_scrambling control
- adaptation_field _control
- continuity_counter(counts packets of PES)
- PES packet length

- discontinuity_indicator
- random_access_indicator
- ES_priority_indicator
- various flags (PCR_flag...)
- PCR (if PCR_flag is set) (system time clock, every 0.1sec, sync decoder and encoder time)
- other fields depending of which flags are set



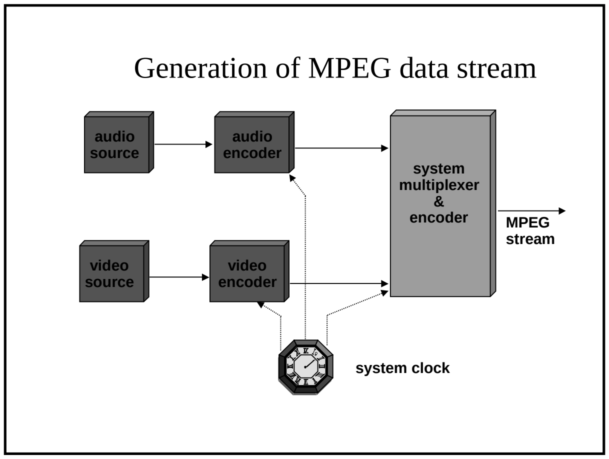
MPEG-2 Systems Layer (Transport Stream)



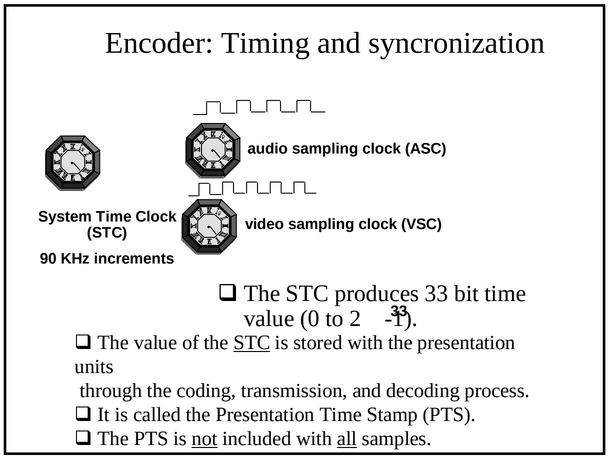
Timing model

The delay includes:

- encoding
- encoder buffering
- multiplexing
- communications or storage
- demultiplexing
- decoder buffering
- decoder
- presentation



Generation of MPEG data stream



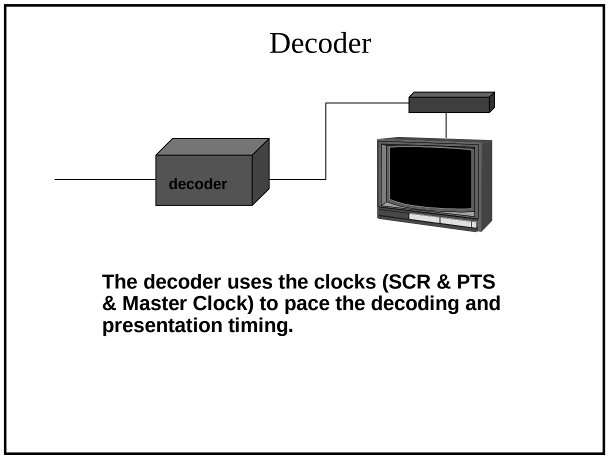
Encoder: Timing and syncronization

□ The STC produces 33 bit time value (0 to 2^{33} -1).

□ The value of the STC is stored with the presentation units through the coding, transmission, and decoding process.

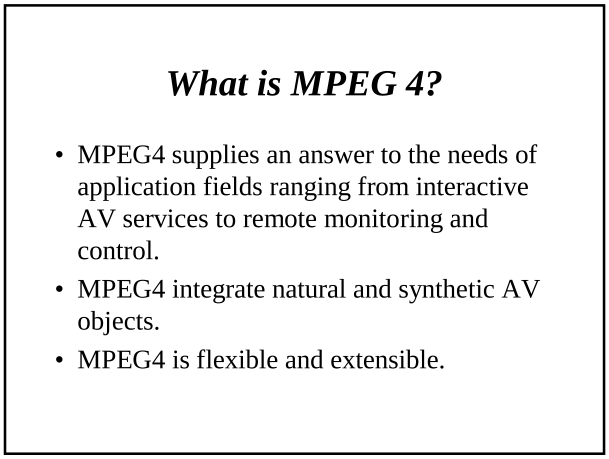
□ It is called the Presentation Time Stamp (PTS).

□ The PTS is not included with all samples.



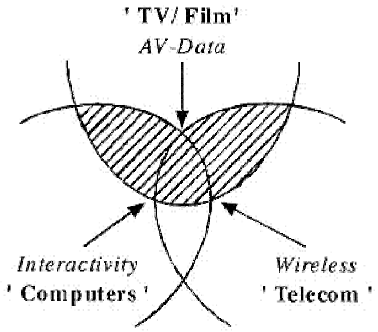
Decoder

The decoder uses the clocks (SCR & PTS & Master Clock) to pace the decoding and presentation timing.



What is MPEG 4?

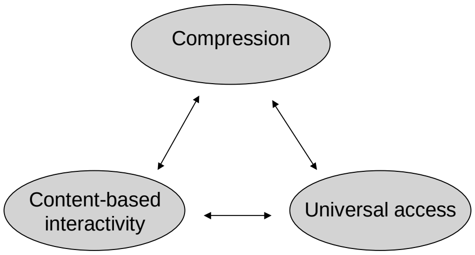
- MPEG4 supplies an answer to the needs of application fields ranging from interactive AV services to remote monitoring and control.
- MPEG4 integrate natural and synthetic AV objects.
- MPEG4 is flexible and extensible.



MPEG-4

- Started in June 1993
- Targeted for Dec. 98
- Based on digital television, interactive graphics applications and interactive multimedia

MPEG-4 : Key Functionalities



Compression

Content-based interactivity

Universal access

Content-based interactivity

- Content-based multimedia data access tools
- Content-based manipulation and bit-stream editing
- Hybrid natural and synthetic data coding
- Improved temporal random access

Compression

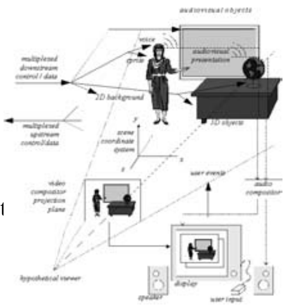
- Improved coding efficiency
- Coding of multiple concurrent data streams

Universal access

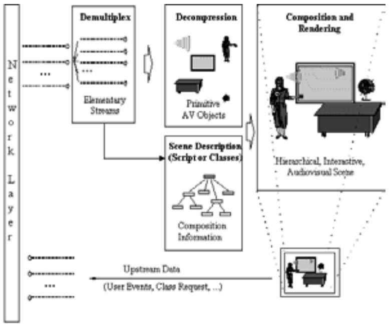
- Robustness in error-prone environments
- Content-based scalability

MPEG-4: Scope & features (1/4)

- Media objects
  - Still images
  - Video objects
  - Audio objects
- Composition of MO
  - Place anywhere
  - Change the user’s point of
    - Viewing
    - Listening



Receiver

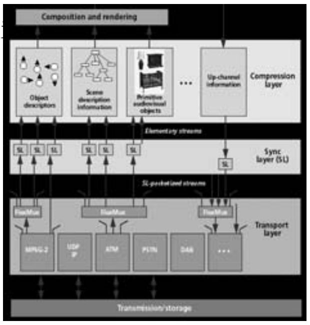

N e t w o r k L a y e r
Demultiplex
...
Elementary Streams
Decompression
Primitive AV Objects
Scene Description (Script or Classes)
Composition Information
Composition and Rendering
Hierarchical, Interactive, Audiovisual Scene
Upstream Data
(User Events, Class Request, ...)

MPEG-4: Scope & features (2/4)

- Description and Synchronization of streaming data for MO
  - Object Descriptor : hints for QOS
    - Max. bit rate
    - Bit error rate
    - Priority
  - Synchronization Layer
    - Time stamping
    - Independent of the media type

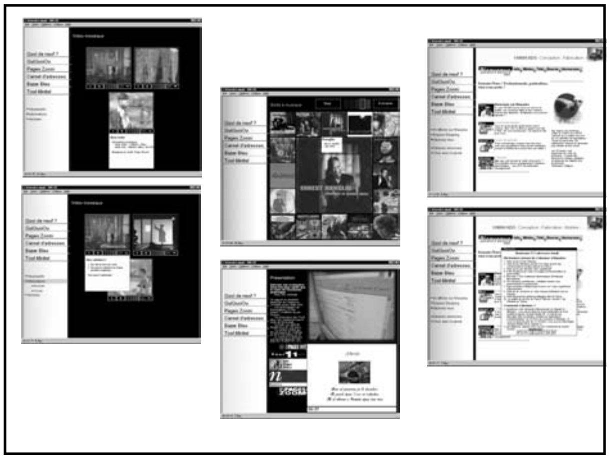
MPEG-4: Scope & features (3/4)

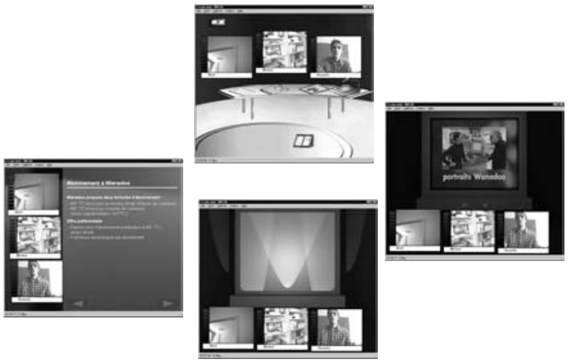
- Del

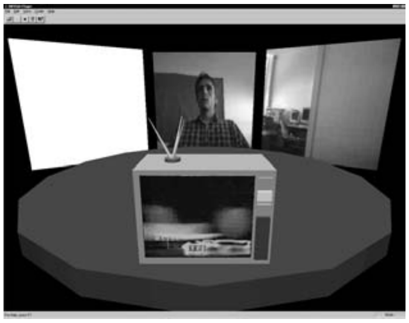


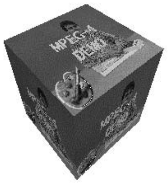
MPEG-4: Scope & features (4/4)

- Interaction with MO
  - Change the viewing/listening point
  - Drag objects
  - Select the desired language
- Management and Identification of Intellectual Property

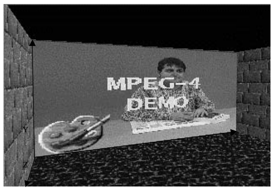




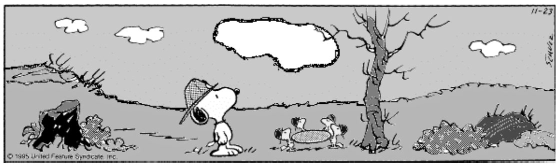




2D or 3D scene as a texture map on 3D



2D inside a 3D plane



MPEG-4: SNHC

- Synthetic and Natural Hybrid Coding
- Text & Graphics Overlay
  - May work with other standardized bitstream
- Facial & Body Animation
  - Virtual Reality
- Text to Speech
- 3D picture coding

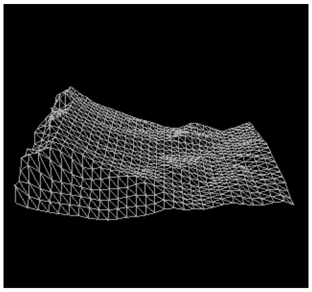
Motivation

- A new type of data is appearing in multimedia applications: Synthetic
- This new data needs to be compressed and streamed in most applications
- New technologies are needed for:
  - Compression and streaming of synthetic data

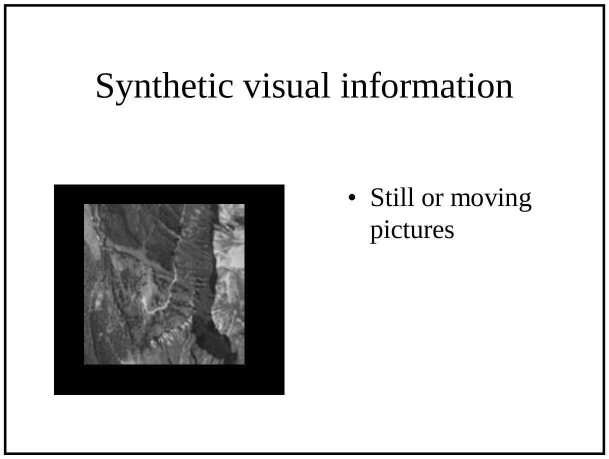
Synthetic Natual Hybrid Coding

- Version 1 (Dec. 1998):
  - Face animation
  - 2D dynamic mesh
  - Scalable coding of synthetic texture
  - View dependent scalable coding of texture
- Version 2 (Dec. 1999):
  - Body animation
  - 3D model compression

Synthetic visual information

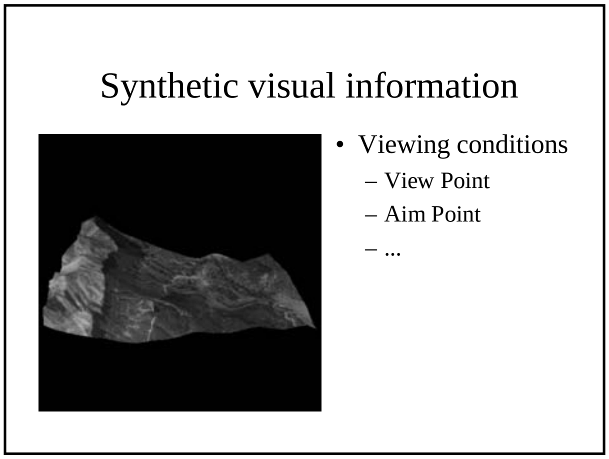


- Vertices coordinates
- Topology
- Normals
- Colors
- Texture coordinates



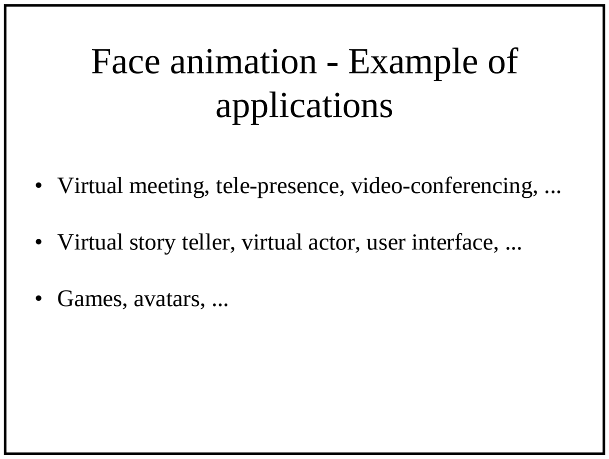
Synthetic visual information

- Still or moving pictures



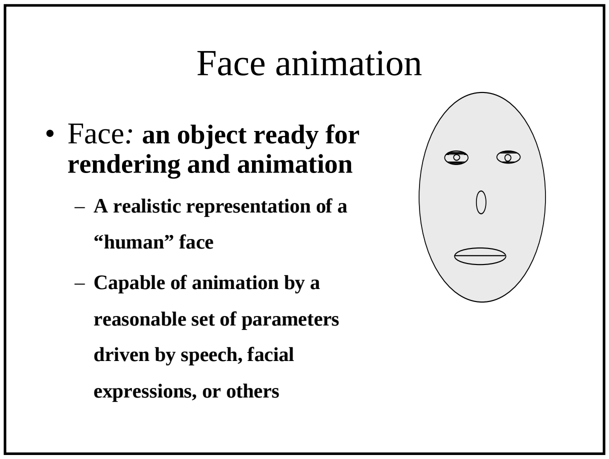
Synthetic visual information

- Viewing conditions
  - View Point
  - Aim Point
  - ...



Face animation - Example of applications

- Virtual meeting, tele-presence, video-conferencing, ...
- Virtual story teller, virtual actor, user interface, ...
- Games, avatars, ...



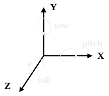
Face animation

- Face*:* **an object ready for rendering and animation**
  - **A realistic representation of a "human" face**
  - **Capable of animation by a reasonable set of parameters driven by speech, facial expressions, or others**

Face animation

- Shape, texture and expressions:
  - **specified parameters in the incoming bitstream**
  - **remote as well as local control of these parameters**

Initial face object



- Gaze along the Z axis
- All face muscles relaxed
- Eyelids tangential to the iris
- Pupil one-third of full eye
- Lips: in contact; horizontal
- Mouth: closed; upper and lower teeth touching
- Tongue: flat; tip touching front teeth

Face animation parameters

- **Three sets of parameters used to describe a face and its animation characterstics:**
  - *Facial Definition Parameters (FDPs)*
  - *Facial Animation Parameters (FAPs)*
  - *Facial Interpolation Transform (FIT)*

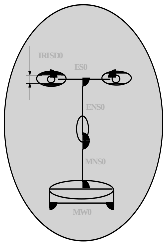
FDPs

- **Two possibilities:**
  - **To customize the default face model at the receiver to a particular face**
  - **To download a face model along with its animation information**
    - **Generally, sent once per session**
      - **for calibration**
      - **more often for “special effects”**

FAPs

- **Represent a complete set of facial actions => allow representation of most of the natural facial expressions**
- **All FAPs involving translational movement: in terms of Facial Animation Parameter Units (FAPUs)**
  - **Allows consistent interpretation of FAPs on any facial model.**

FAPUs



IRISD0	Iris diameter (by definition it is equal to the distance between upper ad lower eyelid) in neutral face	IRISD = IRISD0 / 1024
ES0	Eye separation	ES = ES0 / 1024
ENS0	Eye - nose separation	ENS = ENS0 / 1024
MNS0	Mouth - nose separation	MNS = MNS0 / 1024
MW0	Mouth width	MW0 / 1024
AU	Angle Unit	10E-5 rad

FITS

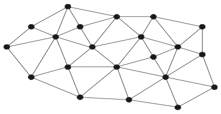
- Specification of interpolation rules for some/all FAPs
- Specified *FAP Interpolation Graph (FIG)* and set of interpolation functions
  - Allows higher degree of control over the animation and provides more realistic results.

Face animation (example)

2-D dynamic mesh - Example of applications

- *Video Object Manipulation*
  - **Augmented Reality**
  - **Object Transformation/Editing**
  - **Spatio-Temporal Interaction**
- *Video Object Compression*
  - **meshes sent regularly; send texture at key frames**

Dynamic 2D mesh



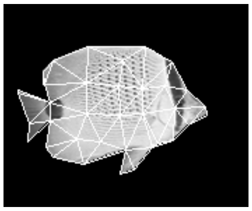
- Specifically refers to *triangular Delaunay* meshes
- Tessellation of a 2D visual object plane into a connection of triangular patches
- No addition and deletion of nodes, i.e. no change in topology

Dynamic 2D Mesh

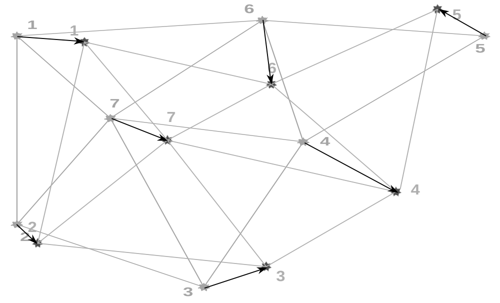
- Generation of the initial mesh
- Coding of initial node points
- Coding of the node motion vectors

Generation of initial the initial mesh

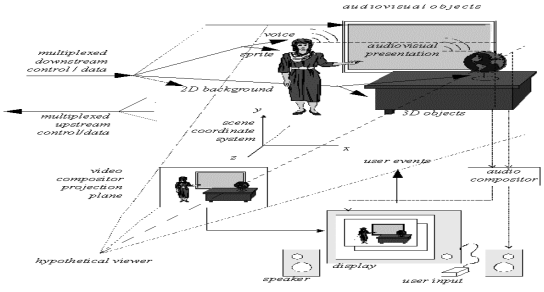
- **Any technique can be used**
- **Not imposed by the standard**



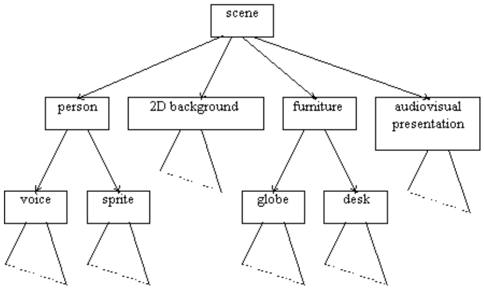
Coding of node motion vectors



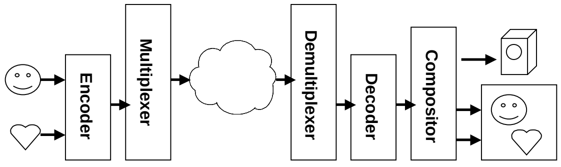
Content-based model example



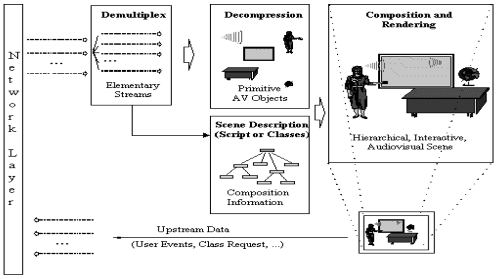
Content-based model example



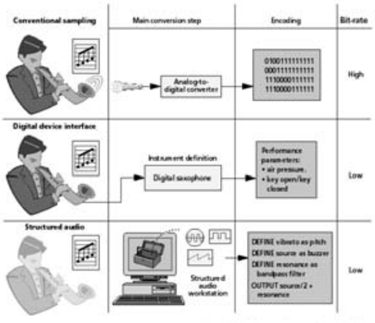
MPEG-4 Architecture



Decoding Process


Network Layer
Demultiplex
...
Elementary Streams
Decompression
Primitive AV Objects
Scene Description (Script or Classes)
Composition Information
Composition and Rendering
Hierarchical, Interactive, Audiovisual Scene
Upstream Data
(User Events, Class Request, ...)

MPEG-4: Audio Objects



Source: Eric Scheirer, Massachusetts Institute of Technology

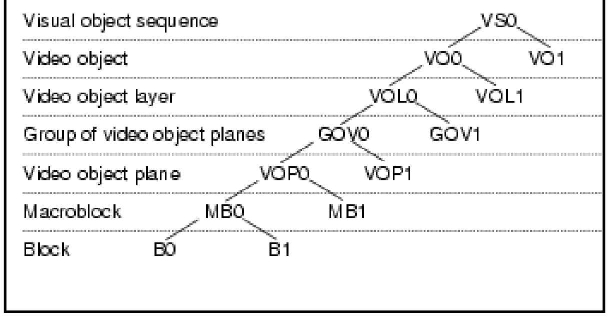
MPEG-4: Audio Objects

- Synthesized Sound
  - Text To Speech
  - Score Driven Synthesis
    - Not standardize "a method" of synthesis
    - "Scores"
      - Described in Structured Audio Orchestra Language

MPEG-4: Audio Objects

- Natural Sound
  - Coding of Audio Objects
    - HVXC (Harmonic Vector eXcitation Coding) : 2 - 4 kbps
    - CELP
      - Sampling rates 8kHz : 6 - 12 kbps
      - Sampling rates 16kHz : 18 kbps
    - MPEG-2 AAC
    - Example
      - A base layer : CELP
      - An enhancement layer : AAC

Visual Objects (1/3)



Visual object sequence VS0

Video object VO0 VO1

Video object layer VOL0 VOL1

Group of video object planes GOV0 GOV1

Video object plane VOP0 VOP1

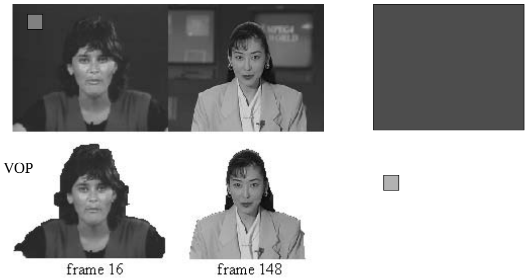
Macroblock MB0 MB1

Block B0 B1

Visual Objects(2/3)

- Video Session(VS)
  - Each VS is made up of one or more VO, corresponding to the various objects in the scene
- Video Object(VO)
  - Each one of these VOs can have several scalability layers(spatial, temporal, or SNR), corresponding to different VOL
- Video Object Layer(VOL)
  - Each VOL consists of an ordered sequence of snapshots in time, called VOP
- Video Object Plane(VOP)
  - Each VOP is a snapshot in time of a VO for a certain VOL

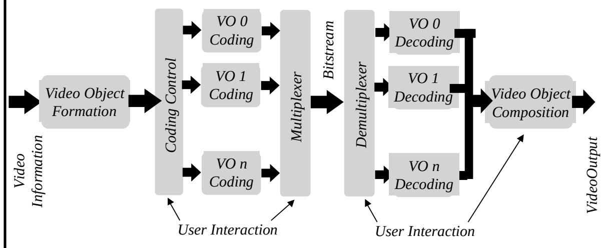
Visual Objects(3/3)



frame 16

frame 148

General video block diagram



Video Information

Video Object Formation

Coding Control

VO 0 Coding

VO 1 Coding

VO n Coding

Multiplexer

Bitstream

Demultiplexer

VO 0 Decoding

VO 1 Decoding

VO n Decoding

Video Object Composition

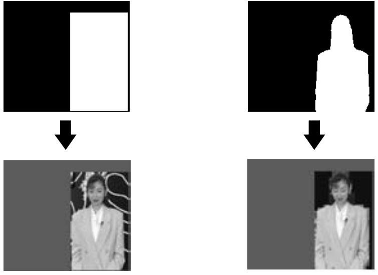
VideoOutput

User Interaction

User Interaction

Video Object Plane Formation

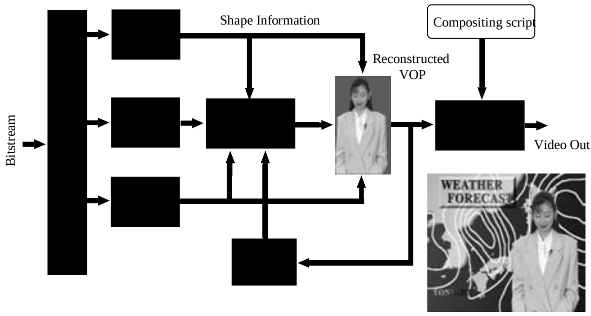
Rectangular or Arbitrary



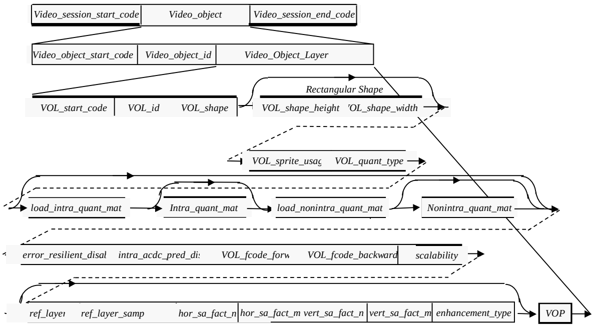
Receiver

- Receiver side must interact with the object in the scene.
  - – change of the spatial position of VOP in the scene
  - – application of a spatial scaling factor
  - – change of the `speed' with which an object moves in the scene
  - – inclusion/delition of objects
  - – change of the scene area being displayed

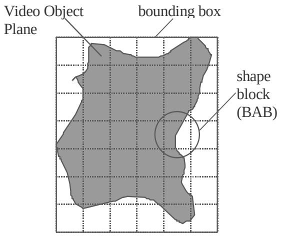
Decoder



Video Bitstream Syntax

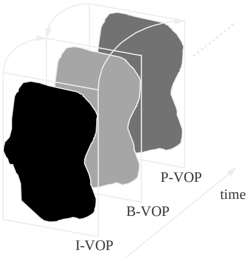


Shape coding tool



Every VOP is coded by dividing it into smaller macroblocks

Motion compensation tools



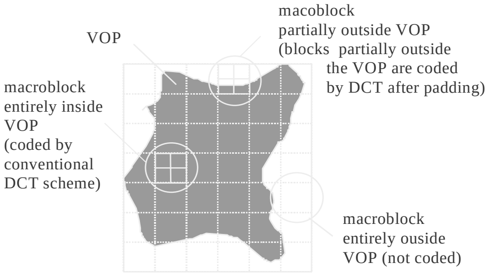
Motion compensated coding modes (I, B, P)

Motion Compensation

What's new: Specialized tools for motion compensation

- The common technique: Block-based motion estimation; uni-directional (P-VOPs) or bidirectional (B-VOPs; relative to I or P VOPs)
- Global motion compensation techniques
  - Sprite-based
  - Mesh-based
- Predictive coding for motion vectors

Texture coding tools



Coefficients scanning

0	1	2	3	10	11	12	13
4	5	8	9	17	16	15	14
6	7	19	18	26	27	28	29
20	21	24	25	30	31	32	33
22	23	34	35	42	43	44	45
36	37	40	41	46	47	48	49
38	39	50	51	56	57	58	59
52	53	54	55	60	61	62	63

Alternate-Horizontal scan

0	4	6	20	22	36	38	52
1	5	7	21	23	37	39	53
2	8	19	24	34	40	50	54
3	9	18	25	35	41	51	55
10	17	26	30	42	46	56	60
11	16	27	31	43	47	57	61
12	15	28	32	44	48	58	62
13	14	29	33	45	49	59	63

Alternate-Vertical scan

0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

zig-zag scan

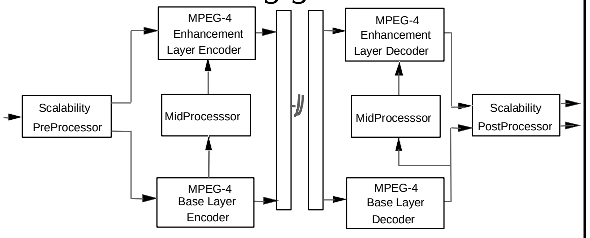
Quantization

- Method 1: Similar to that of H.263
- Method 2: Similar to that of MPEG-2
- Optimized non-linear quantization of DC coefficients
- Quantization matrices and loading mechanism

Scalability

- Object scalability
  - Achieved by the data structure used and the shape coding
- Temporal scalability
  - Achieved by generalized scalability mechanism
- Spatial scalability
  - Achieved by generalized scalable mechanism

Scalable coding general scheme

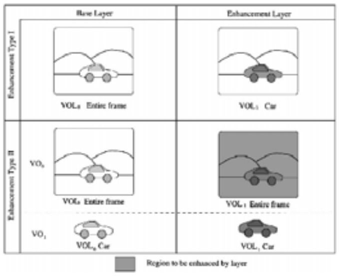


Scalability PreProcessor
MPEG-4 Enhancement Layer Encoder
MidProcesssor
MPEG-4 Base Layer Encoder
MPEG-4 Enhancement Layer Decoder
MidProcesssor
MPEG-4 Base Layer Decoder
Scalability PostProcessor

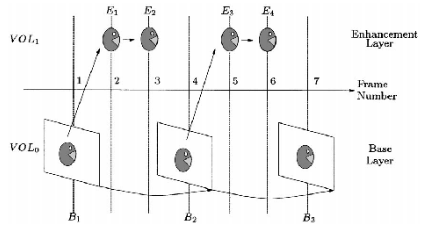
Temporal scalability

- The temporal scalability is achievable for both rectangular frames and arbitrarily shaped VOPs
- The base layer is encoded conventional MPEG-4 video
- The enhancement layer is encoded using one of the following two mechanisms:
  - Type 1
  - Type 2

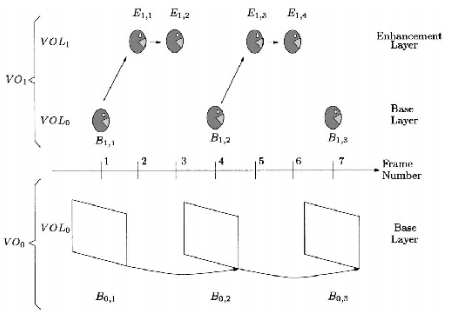
Temporal enhancement types



Temporal saclability type 1



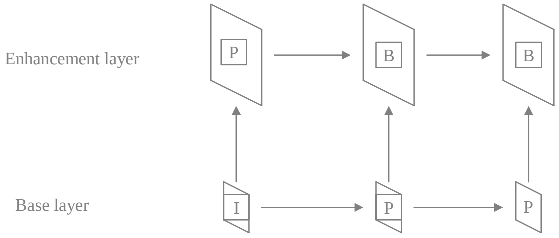
Temporal scalability type 2



Spatial scalability

- The base layer is coded as conventional MPEG-4 video
- The enhancement layer is encoded using prediction mechanism from the base layer

Spatial scalability



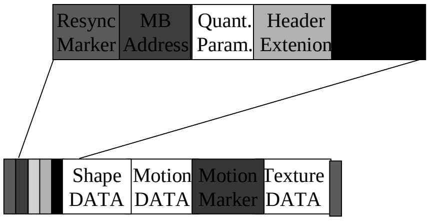
Scalability

- Object scalability
  - Achieved by the data structure used and the shape coding
- Temporal scalability
  - Achieved by generalized scalability mechanism
- Spatial scalability
  - Achieved by generalized scalable mechanism

Error resilience tools

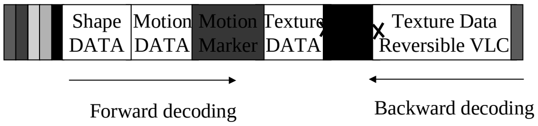
- Resynchronization markers
- Extended header code
- Data partitioning
- Reversible VLCs

EHC, DP

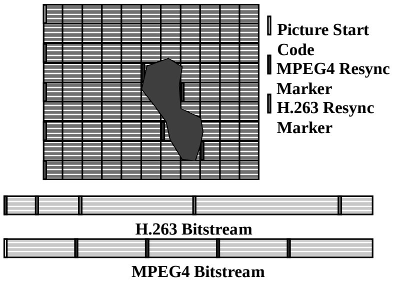


RVLC

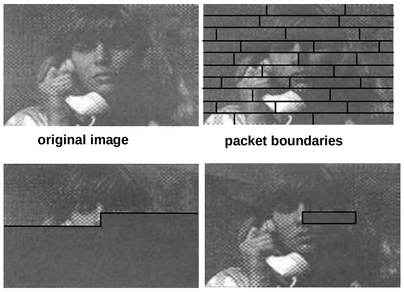
- •Data Recovery
  - –to recover data that in general would be lost
  - –not simply error correcting codes but instead techniques which encode the data in an error resilient manner
  - –RVLC(Reversible Variable Length Code)



Resynchronization markers (1/2)



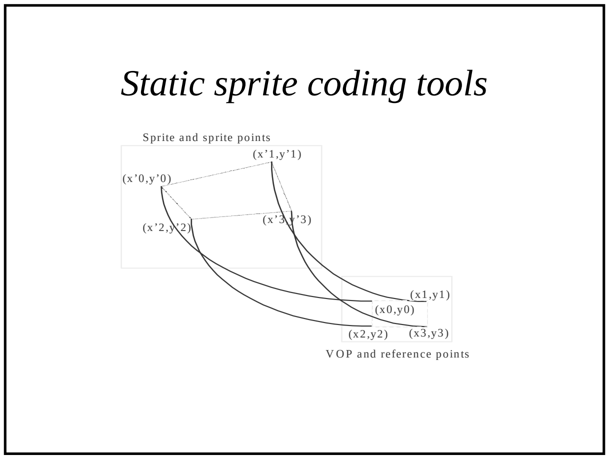
Resynchronization markers (2/2)


original image

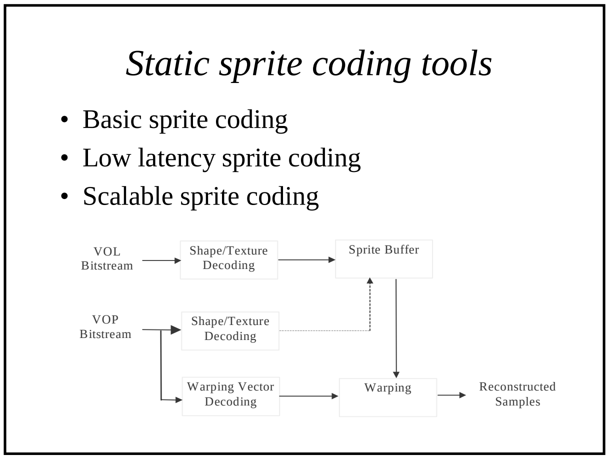
packet boundaries

propagation of an error w/o video packet

same error with video packet



Static sprite coding tools



Static sprite coding tools

- Basic sprite coding
- Low latency sprite coding
- Scalable sprite coding

VOL Bitstream

Shape/Texture Decoding

Sprite Buffer

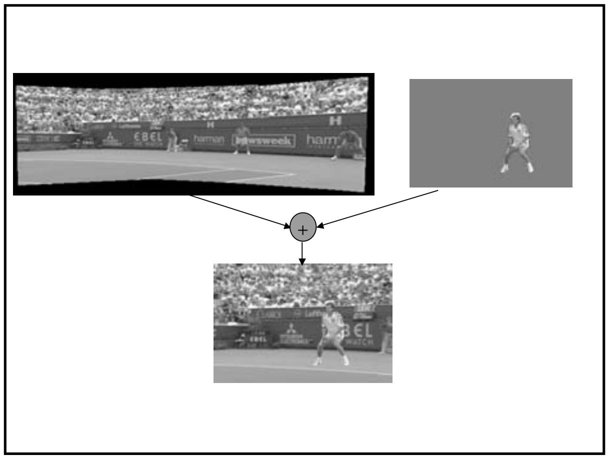
VOP Bitstream

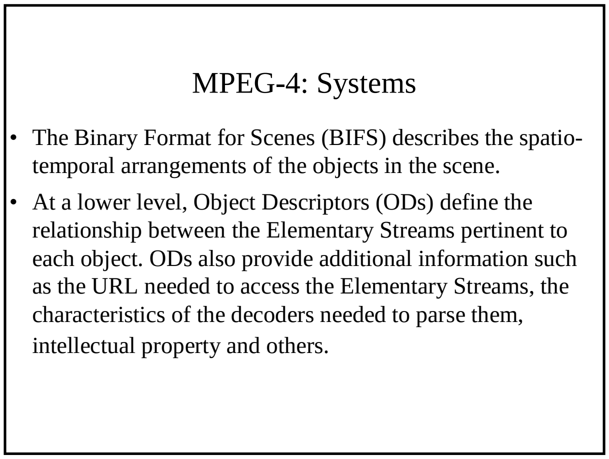
Shape/Texture Decoding

Warping Vector Decoding

Warping

Reconstructed Samples

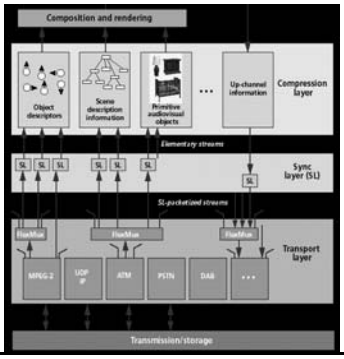




MPEG-4: Systems

- The Binary Format for Scenes (BIFS) describes the spatio-temporal arrangements of the objects in the scene.
- At a lower level, Object Descriptors (ODs) define the relationship between the Elementary Streams pertinent to each object. ODs also provide additional information such as the URL needed to access the Elementary Streams, the characteristics of the decoders needed to parse them, intellectual property and others.

MPEG-4: components



What Is BIFS

- BIFS is based on VRML :
  - it is a set of nodes to represent the primitive scene objects to be composed, the scene graph constructs, the behavior and interactivity

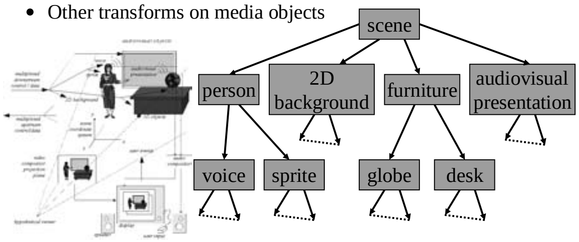
- Additionally to VRML, BIFS defines :
  - 2D capabilities,
  - Integration of 2D and 3D,
  - Advanced Audio Features,
  - a timing model,
  - BIFS - Update protocol to modify the scene in time,
  - BIFS - Anim protocol to animate the scene in time,
  - a binary encoding for the scene.

- BIFS Update Protocol
  - Used to modify of the Scene.
  - Supported commands :
    - Replace Scene
    - Add/Remove object
    - Change Scene Properties
    - Add/Remove Behaviors

- BIFS Anim Protocol
  - Used to continuously animate objects
  - Support continuous animation of :
    - Position, size, colors of objects,
    - Face and Body parameters,
    - 2D and 3D Mesh parameters.

MPEG-4: Scene Description

- How objects are grouped together
- How objects are positioned in space & time
- Attribute Value Selection
- Other transforms on media objects



Example

```
.
Group {
   DEF position Transform {
        translation 1 0 0 {
               children [
                       Shape {
                               geometry DEF MyText Text {
                                      string “My Text”
   } } ] } } }
DEF PosInterp PositionInterpolator {
   Key[0 1]
   KeyValue [1 0 0, -1 0 0]
}
DEF MyTime TimeSensor {
   loop TRUE
   stopTime -1
   }
ROUTE MyTime.fraction_changed TO PosInterp.key
ROUTE PosInterp.keyValue TO position.translation
```

Animations in BIFS

Two ways:

- Interpolator nodes, which may be used to hold the key positions of the object to be animated (a bit static)
- BIFS-anim tool, which is using BIFS-command to modify the fields of the corresponding transform node (more dynamic)

Version 1 Tools

Representation of Multimedia Content :

- Identification of Elementary Streams : Object Descriptor
- Scene Description : BIFS (BIFS, BIFS Update)
- Animation of the Scene : Animation Streams
- Object Content Information : OCI
- Identification of Intellectual Property

Version 1 Tools

Management of Elementary Streams

- Time and Buffer Management : Systems Decoder Model
- Synchronization : Sync. Layer and AccessUnit
- Multiplexing of Elementary Streams : FlexMux

Generic Representation Tools

Syntactic Representation : Syntactic Description Language

Object Descriptor

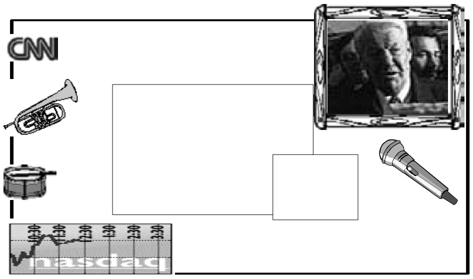
- Identification of Elementary Streams
  - Stream types and decoder configuration,
  - Streams identification and location,
  - IPR and object content description.

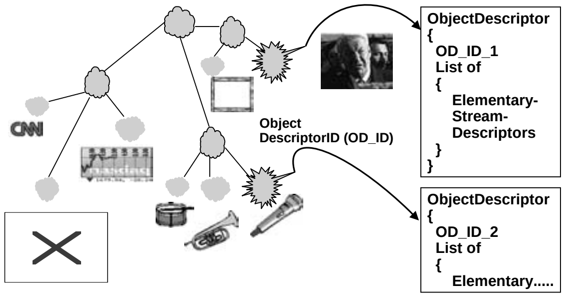
Association between Streams

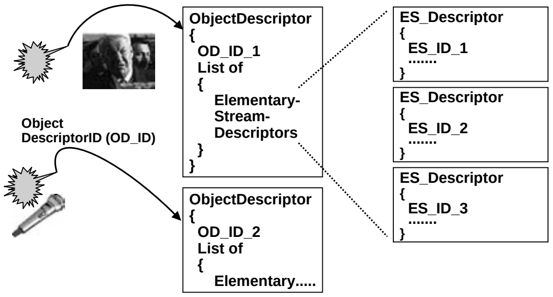
Coding dependency between streams,

Clock dependencies between streams,

Association between Scene streams and Media streams.







Elementary Stream Descriptors

- They contain pointers to :
  - – scalably coded content streams
  - – alternate quality content streams
  - – Object Content Information
  - – IPR Information
- And sub descriptors for :
  - – Decoder Configuration
  - – Sync. Layer Header Configuratio
  - – Quality of Service Information
  - – Extension Information

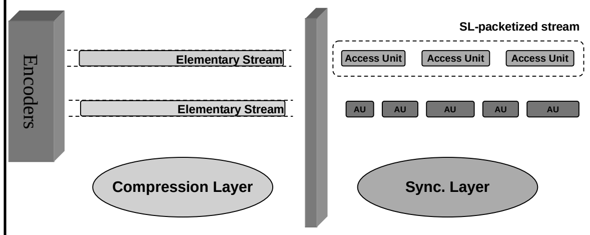
Allows the terminal to select suitable streams

"Stream Type"
"Stream Header"

Configures parser for Sync. Unit Layer

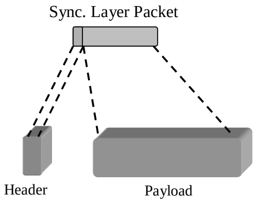
Needed in heterogeneous networks only

Synchronization Layer

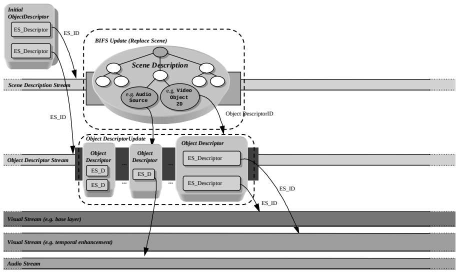


Synchronization Layer

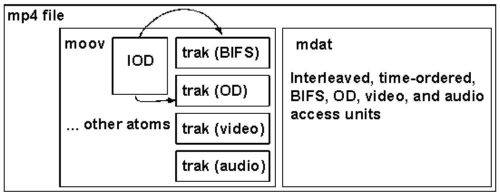
- Boundaries of AccessUnit. Access Units may use more than one SL-Packet.
- Provides consistency checking for lost packets.
- Carries Object Clock Reference.
- Carries Decoding and Composition time stamps.
- Carries Wall Clock time stamps.



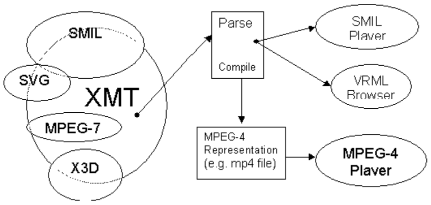
Walk Through of a Session

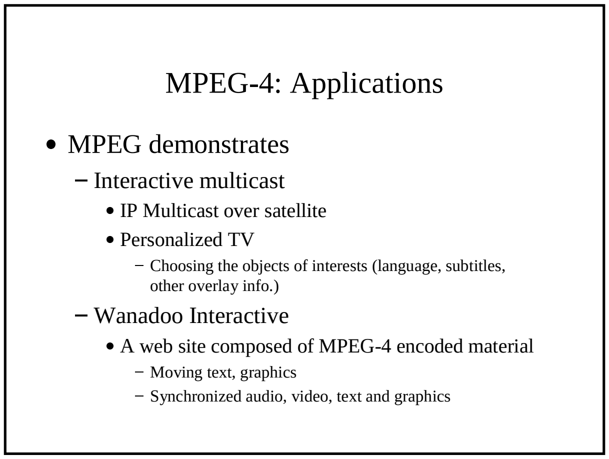


MPEG-4 File Format



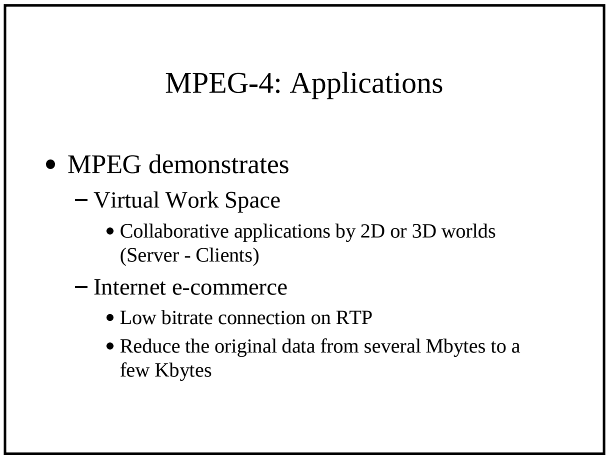
Textual format (XMT)





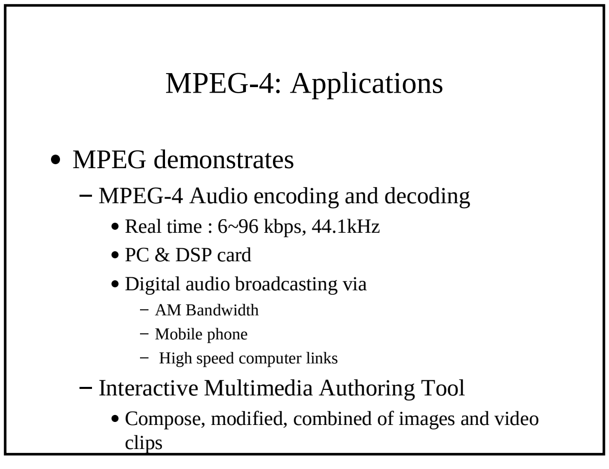
MPEG-4: Applications

- MPEG demonstrates
  - Interactive multicast
    - IP Multicast over satellite
    - Personalized TV
      - Choosing the objects of interests (language, subtitles, other overlay info.)
  - Wanadoo Interactive
    - A web site composed of MPEG-4 encoded material
      - Moving text, graphics
      - Synchronized audio, video, text and graphics



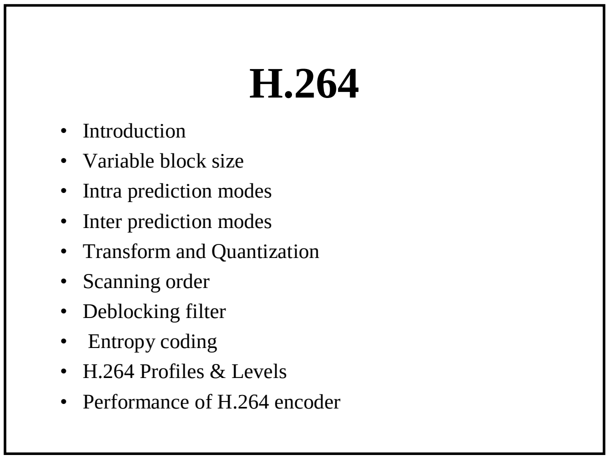
MPEG-4: Applications

- MPEG demonstrates
  - Virtual Work Space
    - Collaborative applications by 2D or 3D worlds (Server - Clients)
  - Internet e-commerce
    - Low bitrate connection on RTP
    - Reduce the original data from several Mbytes to a few Kbytes



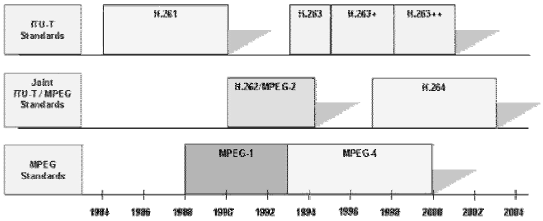
MPEG-4: Applications

- MPEG demonstrates
  - MPEG-4 Audio encoding and decoding
    - Real time : 6~96 kbps, 44.1kHz
    - PC & DSP card
    - Digital audio broadcasting via
      - AM Bandwidth
      - Mobile phone
      - High speed computer links
  - Interactive Multimedia Authoring Tool
    - Compose, modified, combined of images and video clips



H.264

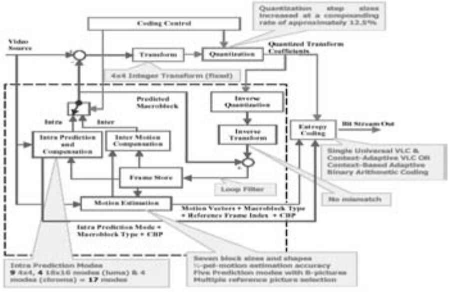
- Introduction
- Variable block size
- Intra prediction modes
- Inter prediction modes
- Transform and Quantization
- Scanning order
- Deblocking filter
- Entropy coding
- H.264 Profiles & Levels
- Performance of H.264 encoder



Goals

- Improved coding efficiency
  - Average bit rate reduction of 50% compared to any others standard.
- Improved network friendliness
  - Major target are mobile network and internet.
- For low bit rate and real time application (<1 Mbps, low latency)
- For broadcasting serial storage on optical and magnetic devices (CD & DVD) (1-8 Mbps and higher latency)

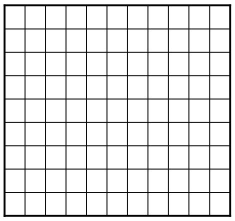
Block diagram


Quantization step sizes increased at a compounding rate of approximately 12.5%
Coding Control
Video Source
Transform
Quantization
Quantized Transform Coefficients
4x4 Integer Transform (fixed)
Predicted Macroblock
Inverse Quantization
Intra
Inter
Inverse Transform
Entropy Coding
Bit Stream Out
Intra Prediction and Compensation
Inter Motion Compensation
Single Universal VLC & Context-Adaptive VLC OR Context-Based Adaptive Binary Arithmetic Coding
Frame Store
Loop Filter
No mismatch
Motion Estimation
Motion Vectors + Macroblock Type + Reference Frame Index + CBP
Intra Prediction Mode + Macroblock Type + CBP
Intra Prediction Modes 9 4x4, 4 16x16 modes (luma) & 4 modes (chroma) = 17 modes
Seven block sizes and shapes ¼-pel-motion estimation accuracy Five Prediction modes with B-pictures Multiple reference picture selection

Variable block size

macroblock

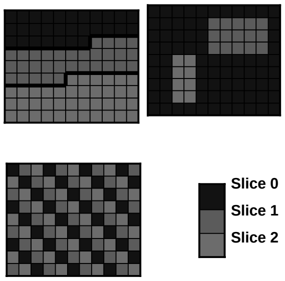
- ❑ A picture is divided into a number of 16x16 macroblock
- ❑ Example:
  - o A QCIF (176x144) picture is divided into 99=11x9 macroblock



Variable block size

slice

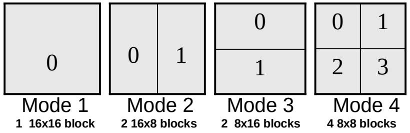
- **Slices**
  - ▪A picture split into 1 or several slices
  - ▪Slices are a sequence of macroblocks
- **Macroblock**
  - ▪Contains 16x16 luminance samples and two 8x8 chrominance samples
  - ▪Macroblocks within a slices depend on each others
  - ▪Macroblocks can be further partitioned



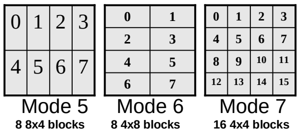
Variable block size

block and subBlock

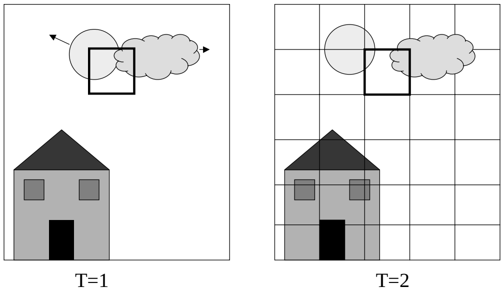
❑Block sizes of 16x8, 8x16, 8x8, 8x4 , 4X8 and 4X4 are available.



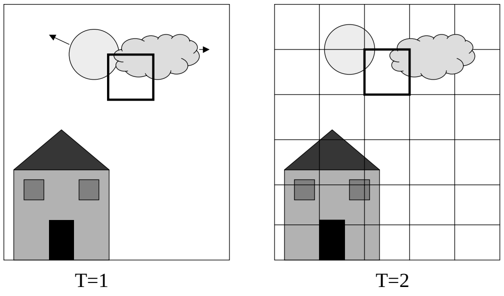
❑Using seven different block sizes can translate into bit rate savings of more than 15% as compared to using only a 16x16 block size.



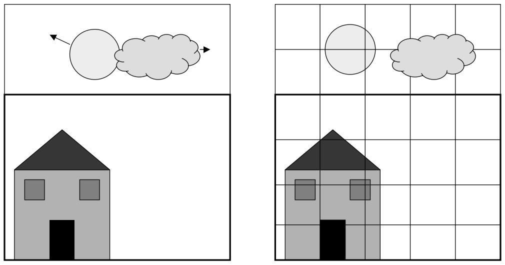
Motion Scale Example



Motion Scale Example

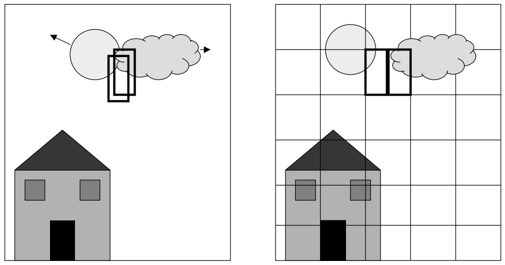


Motion Scale Example



T=1 T=2

H.264 VBS Example



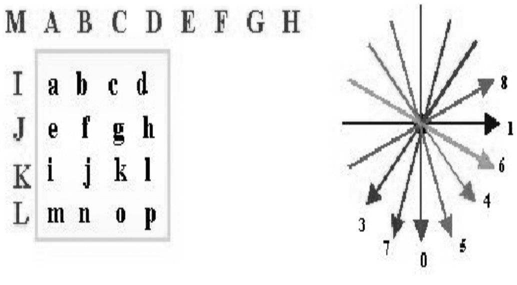
T=1 T=2

Intra prediction modes

- For I macroblocks
- For luminance samples
  - – 4x4 prediction process
  - – 16x16 prediction process
- For chrominance samples
  - – 8x8 prediction process

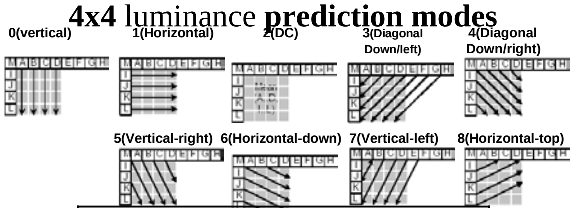
Intra prediction modes

4x4 luminance **prediction modes**



Intra prediction modes

4x4 luminance **prediction modes**



Mode 2 (DC)

Predict all pixels from

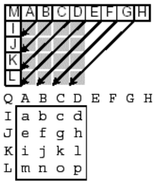
(A+B+C+D+I+J+K+L+4)/8 or

(A+B+C+D+2)/4 or (I+J+K+L+2)/4

Intra prediction modes

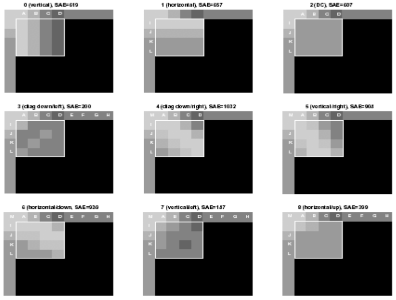
4x4 luminance **prediction modes**

- For example in Mode 3 (Diagonal-Down-Left prediction) is chosen, the values of **a** to **p** are given as follows:
  - a is equal to (A+2B+C+2)/4
  - b, e are equal to (B+2C+D+2)/4
  - c, f, i are equal to (C+2D+E+2)/4
  - d, g, j, m are equal to (D+2E+F+2)/4
  - h, k, n are equal to (E+2F+G+2)/4
  - l, o are equal to (F+2G+H+2)/4
  - p is equal to (G+3H+2)/4



Intra prediction modes

4x4 luminance **prediction modes**



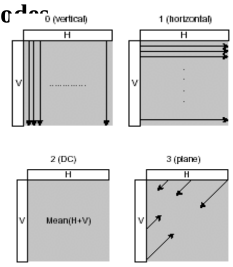
0 (vertical), SAE=619
1 (horizontal), SAE=657
2 (DC), SAE=607
3 (diag down/left), SAE=200
4 (diag down/right), SAE=1032
5 (vertical/right), SAE=908
6 (horizontal/down, SAE=939
7 (vertical/left), SAE=187
8 (horizontal/up), SAE=399

Intra prediction modes

Intra 16x16 luminance and 8x8 chrominance **prediction modes**

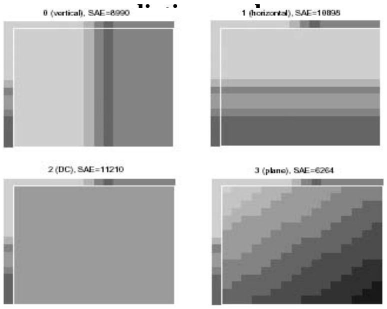
- Mode 0 (**Vertical**)
- Mode 1 (**Horizontal**)
- Mode 2 (**DC**)
- Mode 4 (**Plane**): a linear "plane" function is fitted to the upper and left-hand samples H and V.

If any of the 8x8 blocks in the luminance component are coded in Intra mode, both chrominance blocks (Cr,Cb) are also intra coded



Intra prediction modes

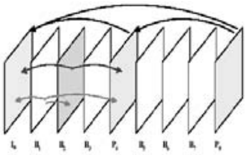
Intra 16x16 luminance and 8x8 chrominance



Inter prediction modes

P and **B** pictures

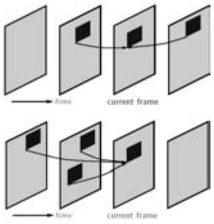
- **Temporal prediction can refer to several pictures (classical 5)**
- **B pictures can be used as reference pictures**
- **P and B pictures can refer to pictures from the future**

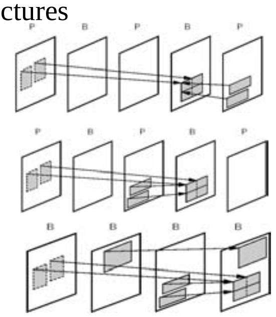


Output order :
I0 B1 B2 B3 P4
Decoding order :
I0 P4 B2 B1 B3

Inter prediction modes

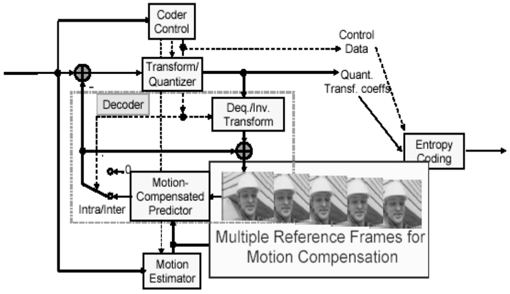
B pictures





Inter prediction modes

Multiple frame reference

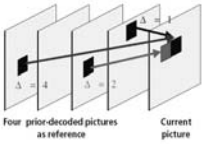


Inter prediction modes

Multiple frame reference

- 5 reference frames ==> 5-10% bit rate savings
- Multiple picture buffer (MPB)
  - – **Short-term picture buffer**
    - **FIFO** or **Sliding Window**
    - **Adaptive Memory Control**
  - – **long-term picture buffer**
- Picture number (PN):

  between **0** to **MAX_PN** (detected loss pictures)



Inter prediction modes

motion vectors

- MVs for neighboring partitions are often highly correlated.
- So we encode MVDs instead of MVs
- MVD = predicted MV – MVp
- ¼ pixel accurate motion compensation

Inter prediction modes

luminance Pixel interpolation

- Using ¼-pixel spatial accuracy ==> 20% bit rate savings as compared to using integer-pixel spatial accuracy.
- **Half-pixel samples** are calculated by applying a 6-tap filter F(1 , -5 , 20 , 20 , -5 , 1) and scaling

G	a	**b**	c	**H**
d	e	f	g	
h	i	**j**	k	**m**
n	p	q	r	
M		**s**		**N**

Inter prediction modes

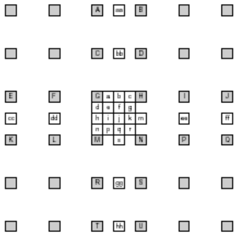
luminance Pixel interpolation (cont.)

b1 = (E–5 F+20G+20H–5I+J)
h1 = (A–5C+20G+20M–5R+T)

b = (b1+16) >> 5
h = (h1+16) >> 5

j1 = cc–5dd+20h1+20m1–5ee+ff

j = (j1+512) >> 10



Inter prediction modes

luminance Pixel interpolation (cont.)

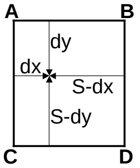
- **a, c, d, n, f, i, k,** and **q** are derived by averaging with upward rounding of the two nearest samples at integer and half sample:
  a = (G+b+1) >> 1
- **e, g, p** and **r** are derived by averaging with upward rounding of the two nearest samples at half sample positions in the diagonal direction
  e = (b+h+1) >> 1

G	a	b	c	H
d	e	f	g	
h	i	j	k	m
n	p	q	r	
M		s		N

Inter prediction modes

chrominance Pixel interpolation

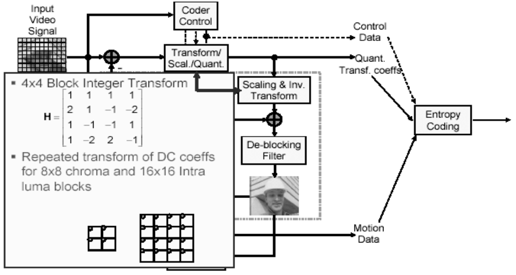
Quarter chrominance Pixels are interpolated by tacking weighted averages of distance from the new pixel to four surrounding original pixels.



$$V=\frac{(s-dx)(s-dy)A+dx(s-dy)B+(s-dx)dyC+dxdyD+s^2/2}{S^2}$$

Transform and Quantization

Transform (Hadamad transform)



- 4x4 Block Integer Transform

$$\mathbf{H}=\begin{bmatrix}1 & 1 & 1 & 1\\ 2 & 1 & -1 & -2\\ 1 & -1 & -1 & 1\\ 1 & -2 & 2 & -1\end{bmatrix}$$

- Repeated transform of DC coeffs for 8x8 chroma and 16x16 Intra luma blocks

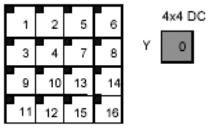
Transform and Quantization

Transform (Hadamad transform)

- A 4x4 integer transform
- multiplier free (16-bit arithmetic computation)

$$\begin{bmatrix}1 & 1 & 1 & 1\\ 2 & 1 & -1 & -2\\ 1 & -1 & -1 & 1\\ 1 & -2 & 2 & -1\end{bmatrix} \begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$$

- For 16x16 **intra coded MB** and for 8x8 **chrominance blocks** a second transform is applied on DC coefficients



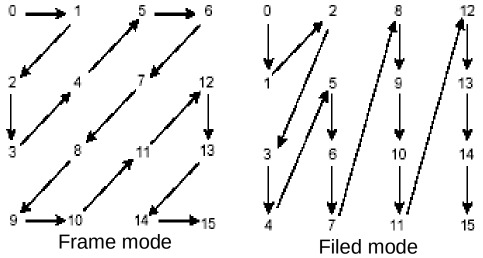
Transform and Quantization

Quantization

- **52 values of QP** ranging from 0 to 51
- Scaling magnitude **increases of 12%** when QP is increased by 1
- **QP can be adapted** with a delta: at the slice and/or at the macroblock level

Scanning order

zig-zag scan

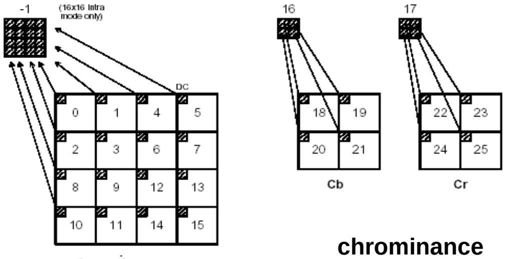


Frame mode

Filed mode

Scanning order

residual blocks within a macroblock

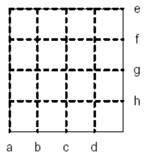


chrominance

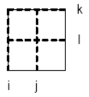
luminance

Deblocking filter

- First: the **3 vertical** edge between 4x4 blocks
- Second: **3 horizontal** edges
- Finally: the **top** and the **left** side of macroblock



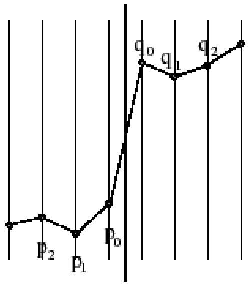
Boundary filtering: 16x16 luma



8x8 chroma

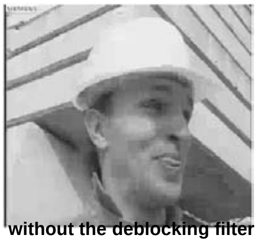
Deblocking filter

- | p0 - q0 | < α(QP)
- | p1 - p0 | < β(QP)
  | q1 - q0 | < β(QP)
- | p2 – p1 | < β(QP)
  | q2 – q1 | < β(QP)

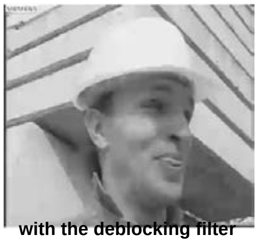


Deblocking filter

example

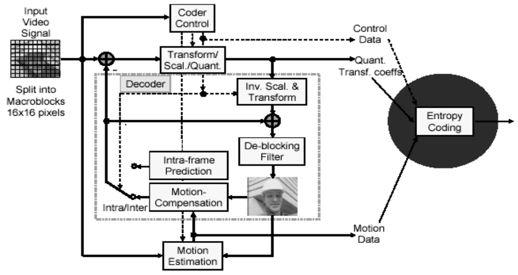


without the deblocking filter



with the deblocking filter

Entropy coding



Entropy coding

CAVLC(**Context-based variable Length Coding**)

- Probability distribution is **static**
- Code words must have integer number of bits (Low coding efficiency for highly peaked pdfs)
- H.264 offers a single Universal VLC (UVLC) table for all symbol

H.264 Profiles & Levels

- A Profileis a set of algorithmic features
- A Levelis a degree of capability
- H.264/AVC currently has three Profiles
  - Baseline (broad range of applications, low latency)
  - Main (adds interlace, B-Slices and CABAC efficiency gains)
  - Extended (the so-called streaming profile)
- H.264/AVC has many (14) Levels

H.264 Profiles & Levels

Baseline Profile

- Progressive
- I and P slices types
- 1/4-sample Inter prediction
- Deblocking filter
- VLC-based entropy coding
- 4:2:0 chroma format
- Arbitrary Slice Order (ASO)
- Flexible Macroblock Ordering (FMO)
- Redundant slices

H.264 Profiles & Levels

Baseline: Arbitrary Slice Order (ASO)

- Decoder can be process slices in an arbitrary order as they arrive to the decoder.
- The decoder **dose not have a wait** for all slices to be properly arranged before it starts processing them.
- Reduces the processing **delay** at the decoder.

H.264 Profiles & Levels

Baseline: FMO & Redundant slices

- FMO: Flexible Macroblock Ordering
  - With FMO, macroblocks are coded according to a macroblock allocation map that groups, within a given slice.
  - Macroblocks from **spatially different locations** in the frame.
  - Enhances **error resilience**
- Redundant slices:
  - allow the transmission of duplicate slices.

H.264 Profiles & Levels

Main Profile (broadcast)

- All Baseline features Plus
  - Interlace
  - B slice types (bi directional reference)
  - CABAC (**Context-based Adaptive Binary Arithmetic Coding** for the transform coefficients)
  - Weighted prediction
- All features included in the Baseline profile **except**:
  - Arbitrary Slice Order (ASO)
  - Flexible Macroblock Order (FMO)
  - Redundant Slices

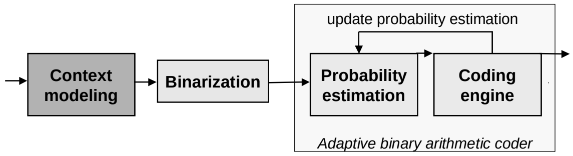
H.264 Profiles & Levels

Main: CABAC (Context-based Adaptive Binary Arithmetic Coding)

- Good performance by
  - Selecting models by context
  - Adapting estimates by local statistics
  - Arithmetic coding reduces computational complexity
- Improve computational complexity more than 10%~20% of the total decoder execution time at medium bitrate
- Average bit-rate saving over CAVLC 10-15%

H.264 Profiles & Levels

Main: CABAC: Technical Overview



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

H.264 Profiles & Levels

CABAC Performance Gain

- Test model of ITU H.26L (TML4)
- QCIF
  - Saving 4.5%~15% bit-rate for whole sequence
  - Saving 3.5%~17% bit-rate for pure intra coding
- CIF
  - Saving 5%~32% bit-rate for whole sequence
  - Saving 4.5%~28% bit-rate for pure intra coding

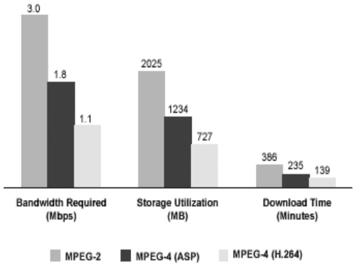
H.264 Profiles & Levels

Extended Profile

- All Baseline features plus
  - Interlace
  - B slice types
  - Weighted prediction

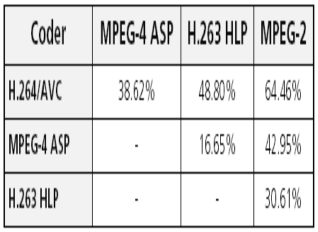
Performance of H.264 encoder

Performance comparison for 90-minute DVD-quality movie (Download time at 700 Kbps)



Performance of H.264 encoder

Average bit-rate savings compared with various prior decoding schemes



Coder	MPEG-4 ASP	H.263 HLP	MPEG-2
H.264/AVC	38.62%	48.80%	64.46%
MPEG-4 ASP	-	16.65%	42.95%
H.263 HLP	-	-	30.61%