MPEG-4 : ISO/IEC 14496-2

VisualObject →
- VideoObject
- StillTextureObject
- MeshObject
- FBAObject

H.261-Reference Model

H.263-Test Model Near-term

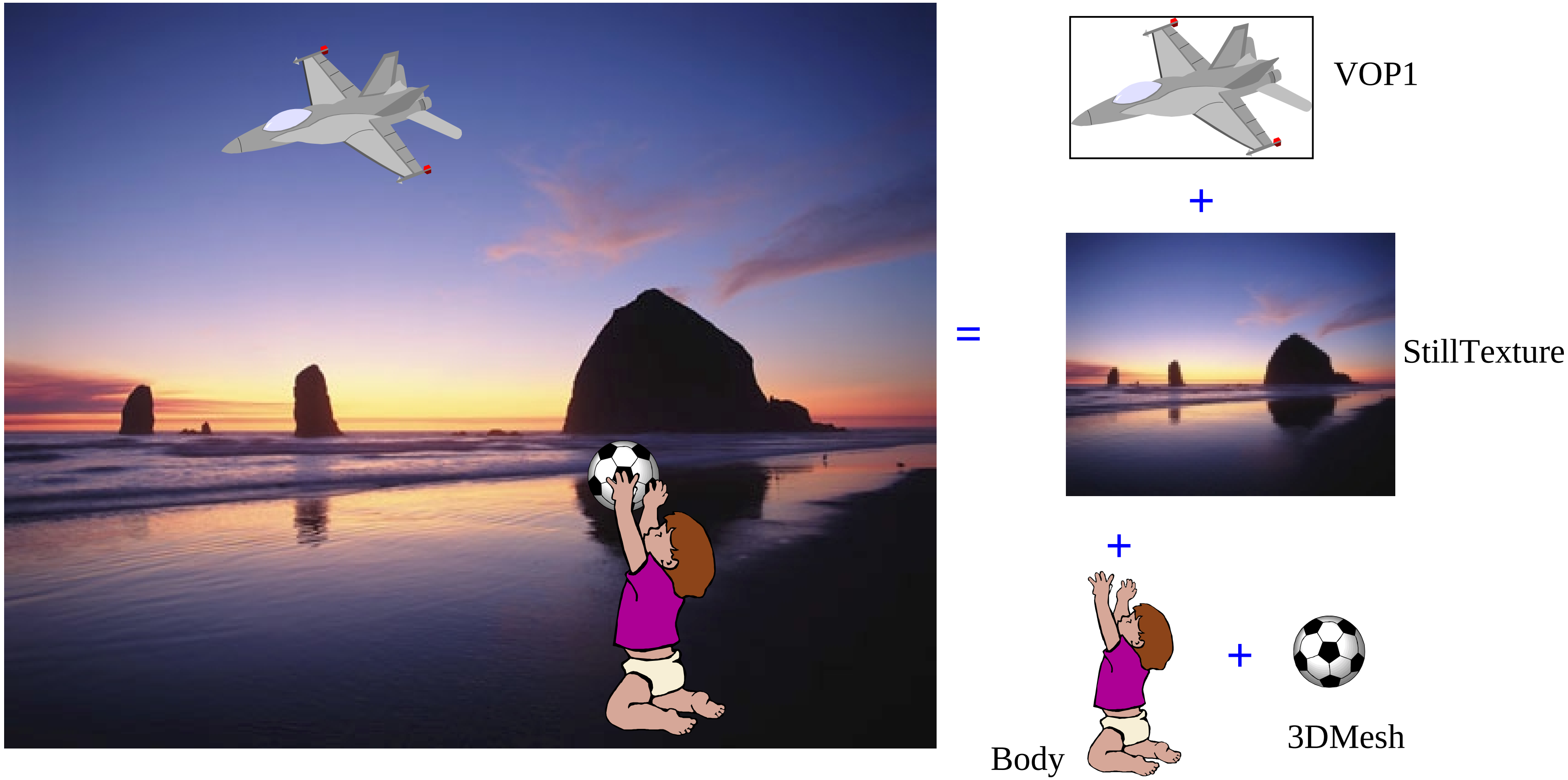
H.263+ (v.2)- Test Model Near-term

H.263L-Test Model Long-term

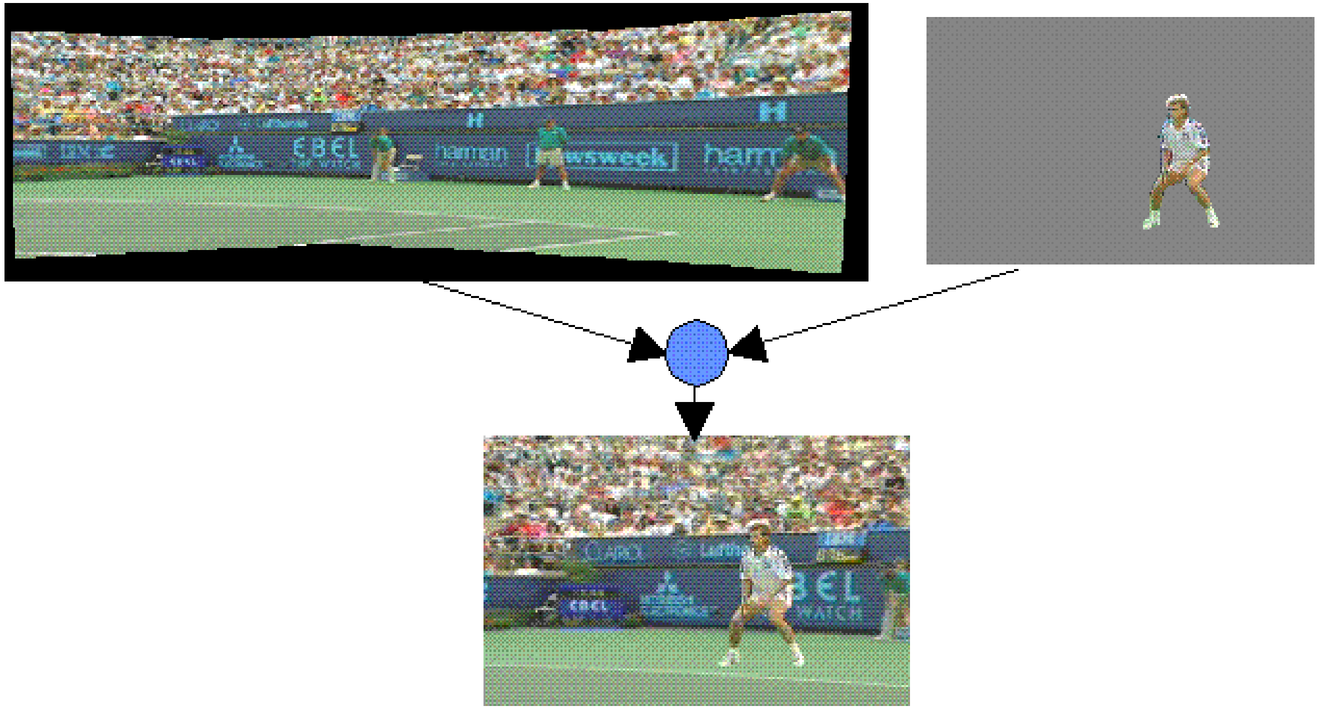
MPEG-2- Test Model

MPEG-4- Verification Model

Scene Example: 4 VOs

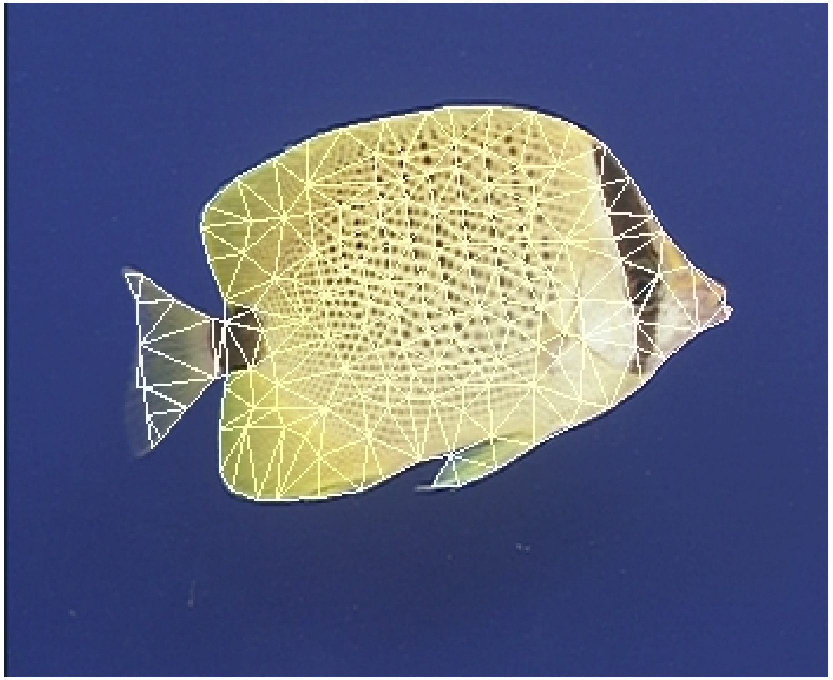


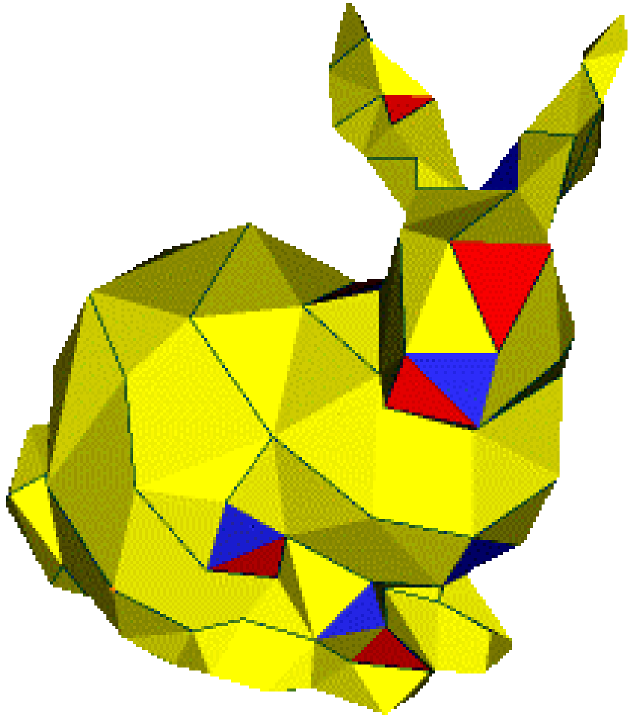
Sprite VOP (camera zoom, pan and tilt)



- Global motion compensation based on the transmission of a static “sprite”. A static sprite is a possibly large still image, describing panoramic background. For each consecutive image in a sequence, only 8 global motion parameters describing camera motion are coded to reconstruct the object. These parameters represent the appropriate affine transform of the sprite transmitted in the first frame

2D and 3D meshes





Visual Bitstreams

- Visual bitstream = Configuration Information (CI) + Elementary Stream(ES)

Visual Object Sequence Header

Visual Object (VO) Header

Visual Object Layer (VOL) Header

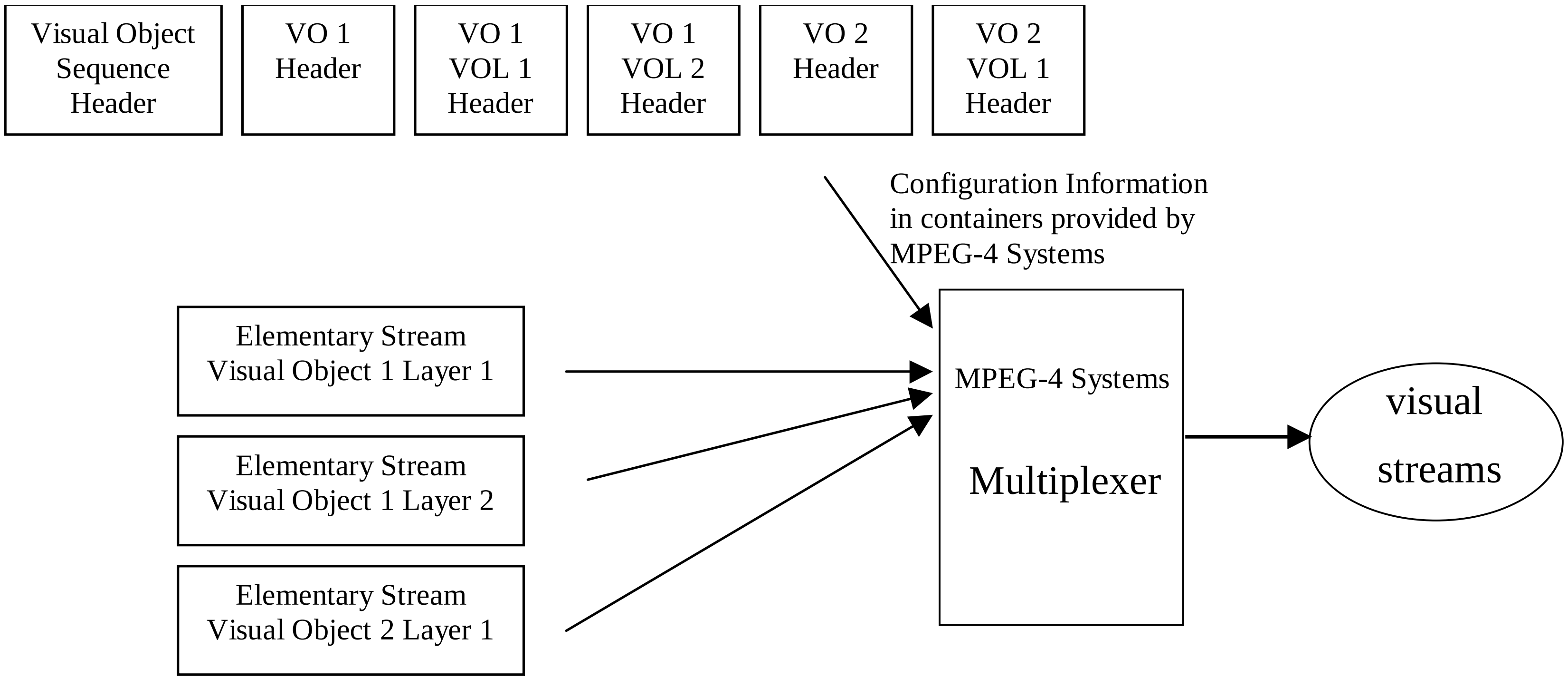
Entry point functions for elementary streams:

Group_ of_VideoObjectPlane()
VideoObjectPlane()
video_plane_with_short_header()
wavelet_dc_decode()
MeshObject()
FBAObject()

VideoObject -> VideoObjectLayer()
StillObject -> StillTextureObject()
MeshObject -> MeshObject()
FBAObject -> FBAObject()

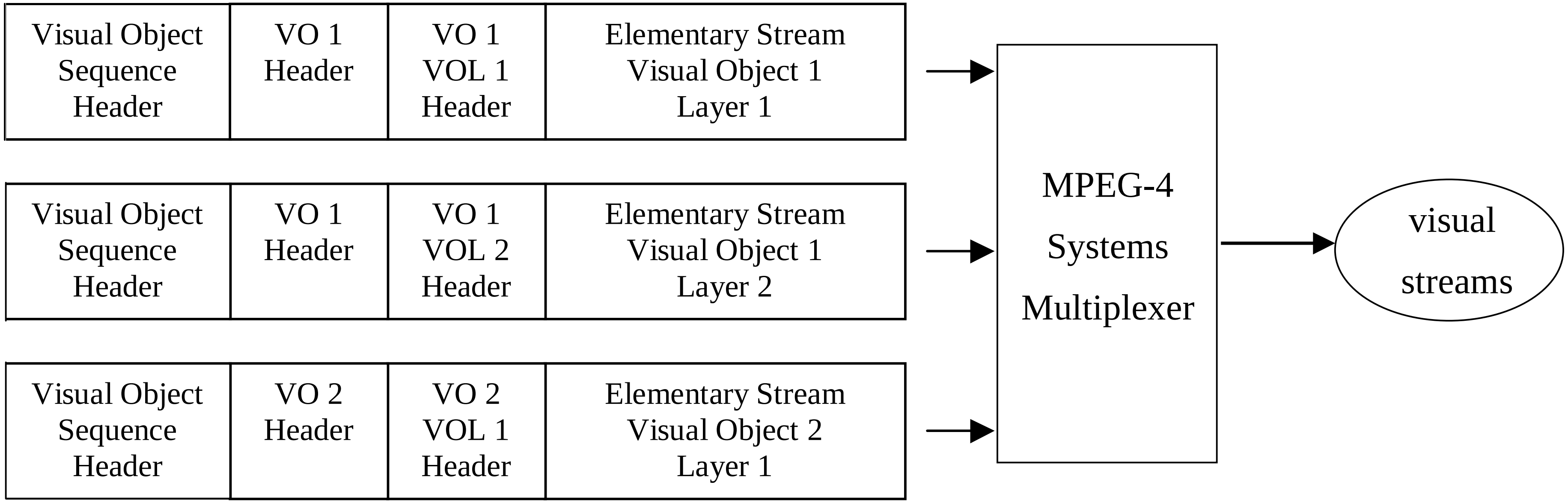
- Mesh and FBA Objects do not have VOL Header

MPEG-4 Bit stream Multiplexing: Separate CI/ES



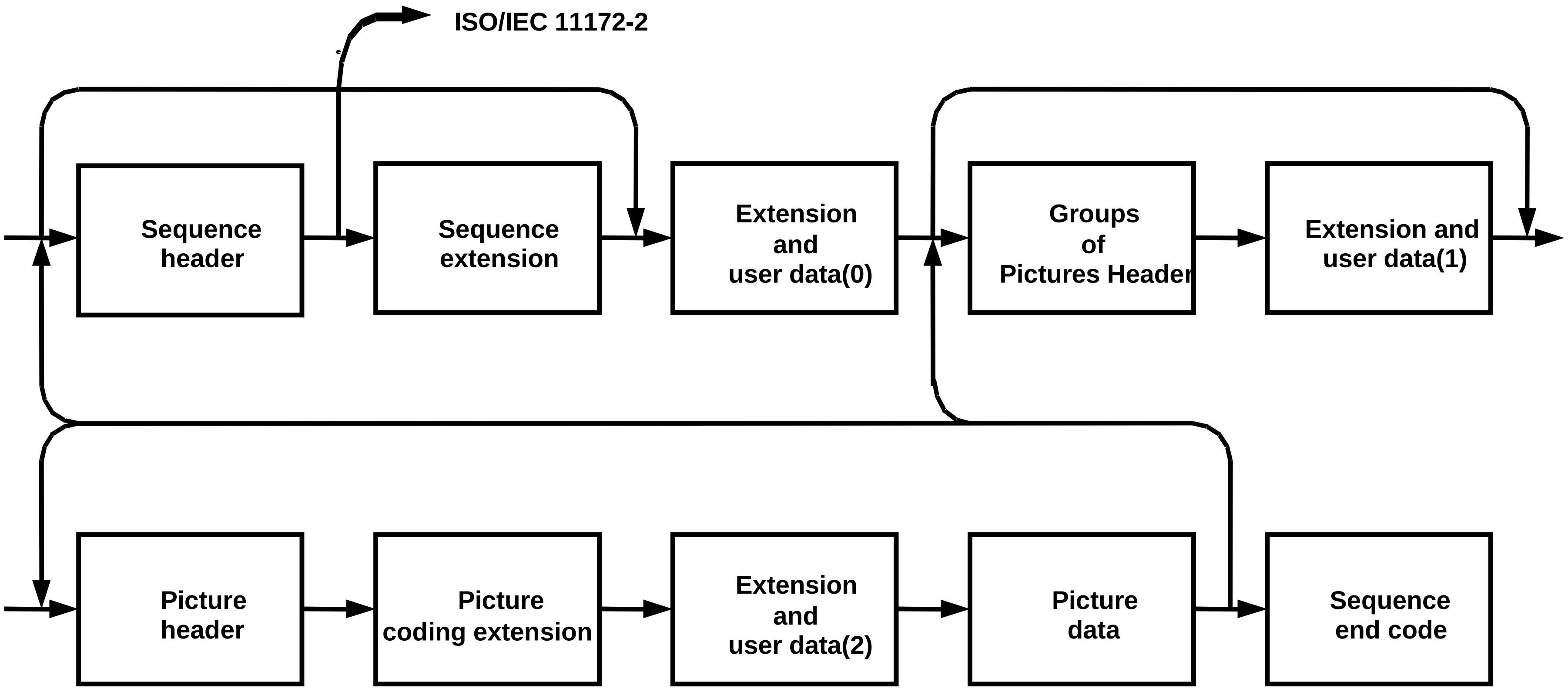
- **VOL1 and VOL2 mean the lower (Layer1) and the enhancement (Layer2) layers in scalability (temporal or spatial) coding respectively. Hybrid scalability supports up to 4 layers**

MPEG-4 Bit stream Multiplexing: Combine CI/ES



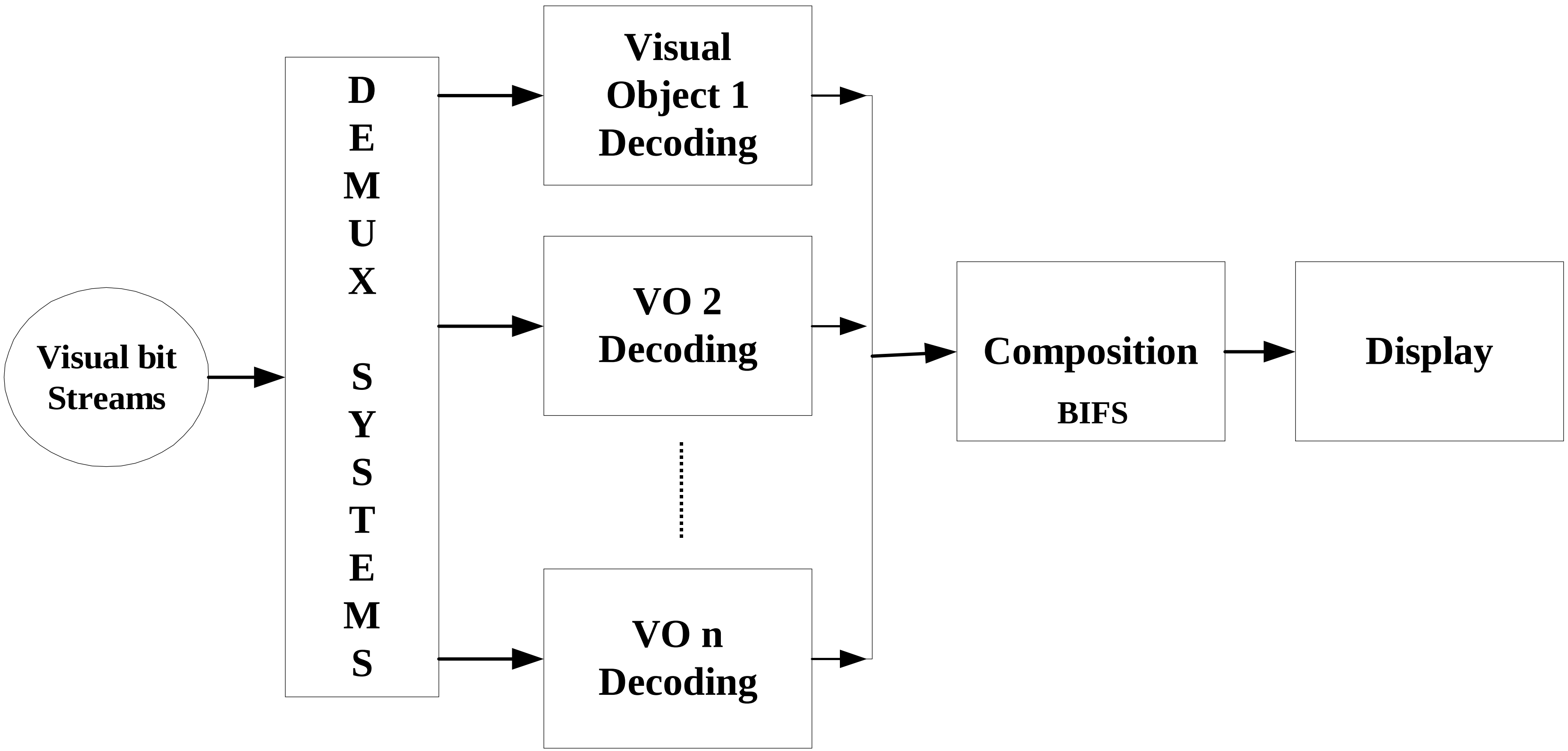
- **The Visual Object Sequence Header must be identical for all streams input simultaneously to a decoder. The Visual Object Headers for each layer of a multilayer (scalability) object must be identical**

MPEG2 Comparison

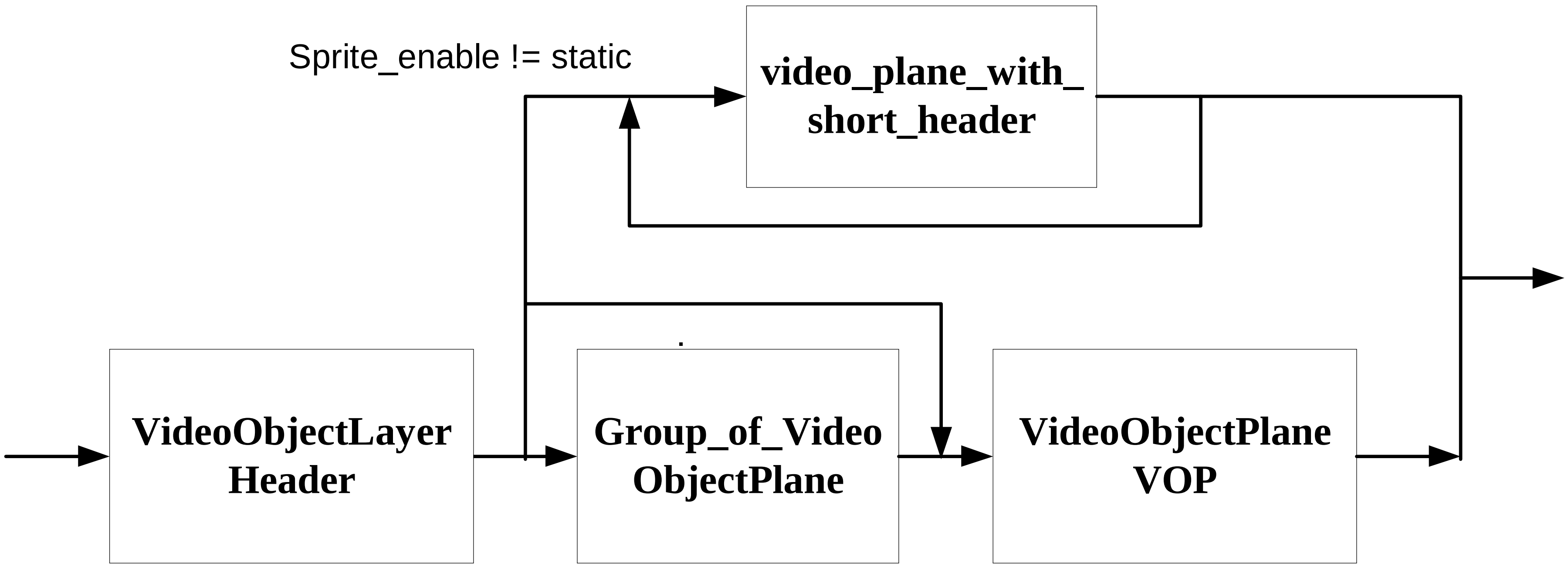


- - Initially the “bit stream” should include sequence header and sequence extension
  - In MPEG-4, VisualObjectSequence defines the profile and level

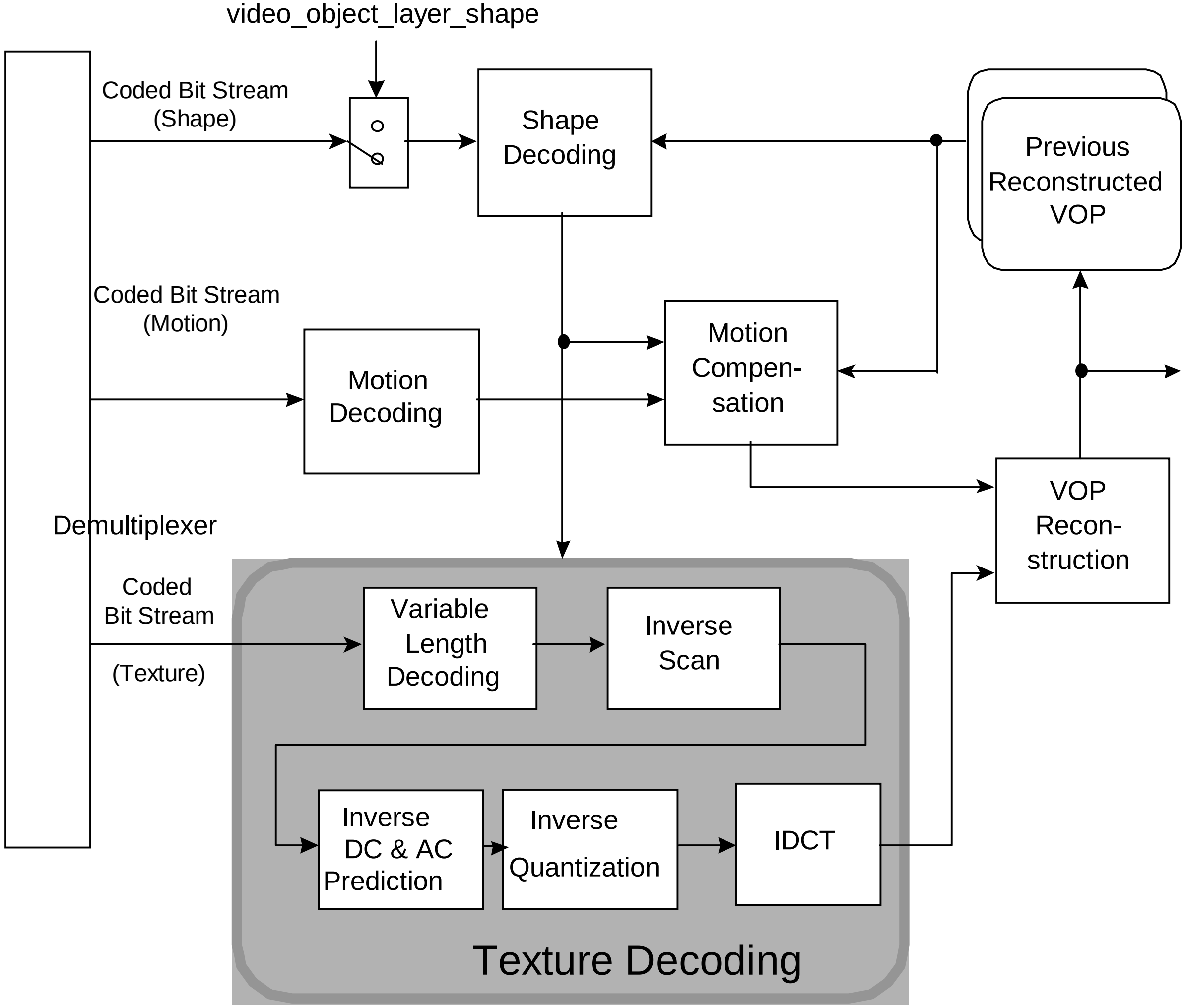
MPEG-4 Decoder



VideoObjectLayer() Sintaxe



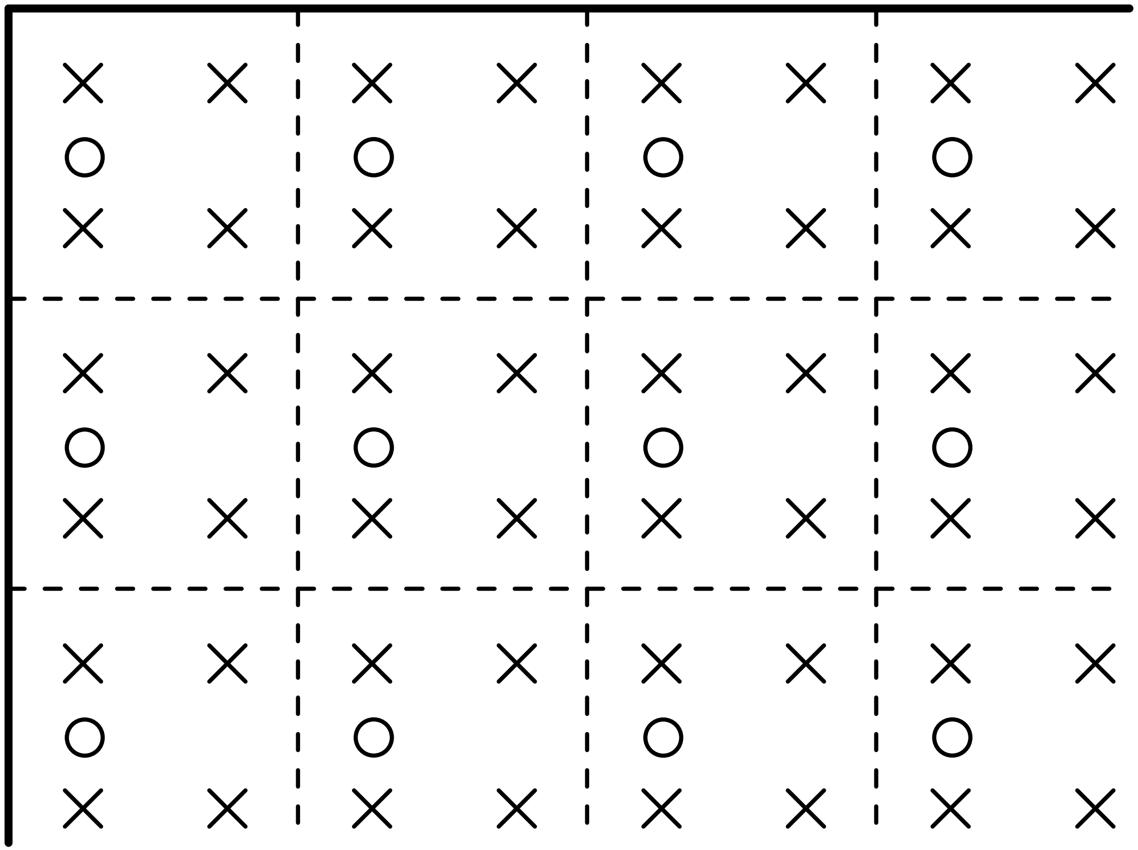
Decoding of VOPs



VOP definitions

- **A reconstructed VOP is obtained by decoding a coded VOP. A coded VOP may have been derived from either a progressive or interlaced frame.**
- **There are four types of VOPs that use different coding methods:**
  - An Intra-coded (I) VOP is coded using information only from itself.
  - A Predictive-coded (P) VOP is a VOP which is coded using motion compensated prediction from a past reference VOP.
  - A Bidirectionally predictive-coded (B) VOP is a VOP which is coded using motion compensated prediction from a past and/or future reference VOP(s).
  - A sprite (S) VOP is a VOP for a sprite object or a VOP which is coded using prediction based on global motion compensation from a past reference VOP. -> S(GMC)

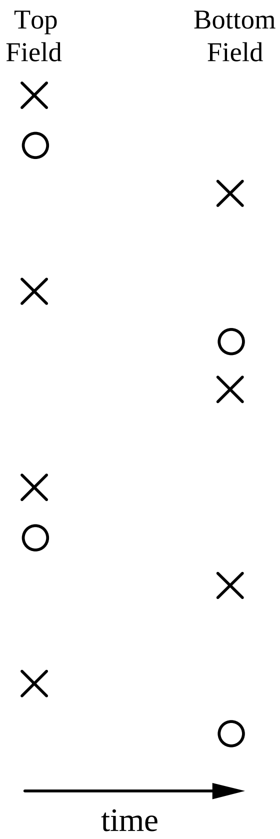
VOP Format 4:2:0



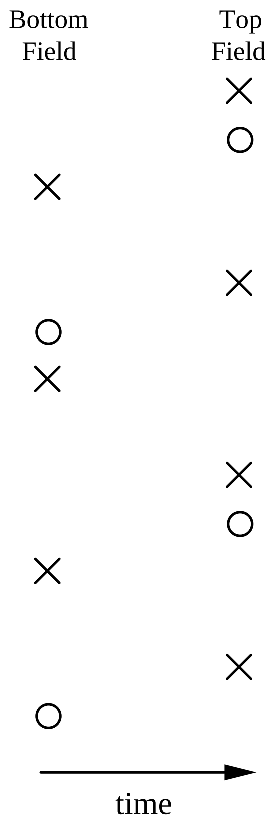
x Represent luminance samples

O Represent chrominance samples

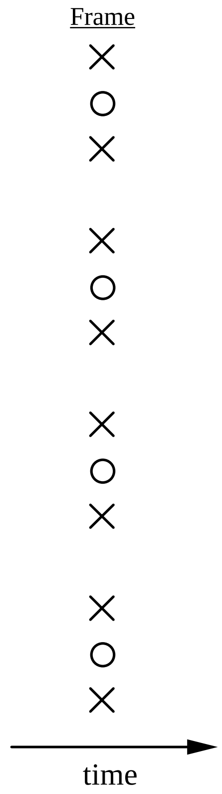
VOP Format 4:2:0



Vertical and temporal positions of samples in an interlaced frame with top_field_first=1



Vertical and temporal positions of samples in an interlaced frame with top_field_first=0



Vertical and temporal positions of samples in a progressive frame

VOP Ordering

- **At the encoder input:**

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
I	**B**	**B**	**P**	**B**	**B**	**P**	**B**	**B**	**I**	**B**	**B**	**P**

- **At the encoder output, in the coded bitstream, and at the decoder input:**

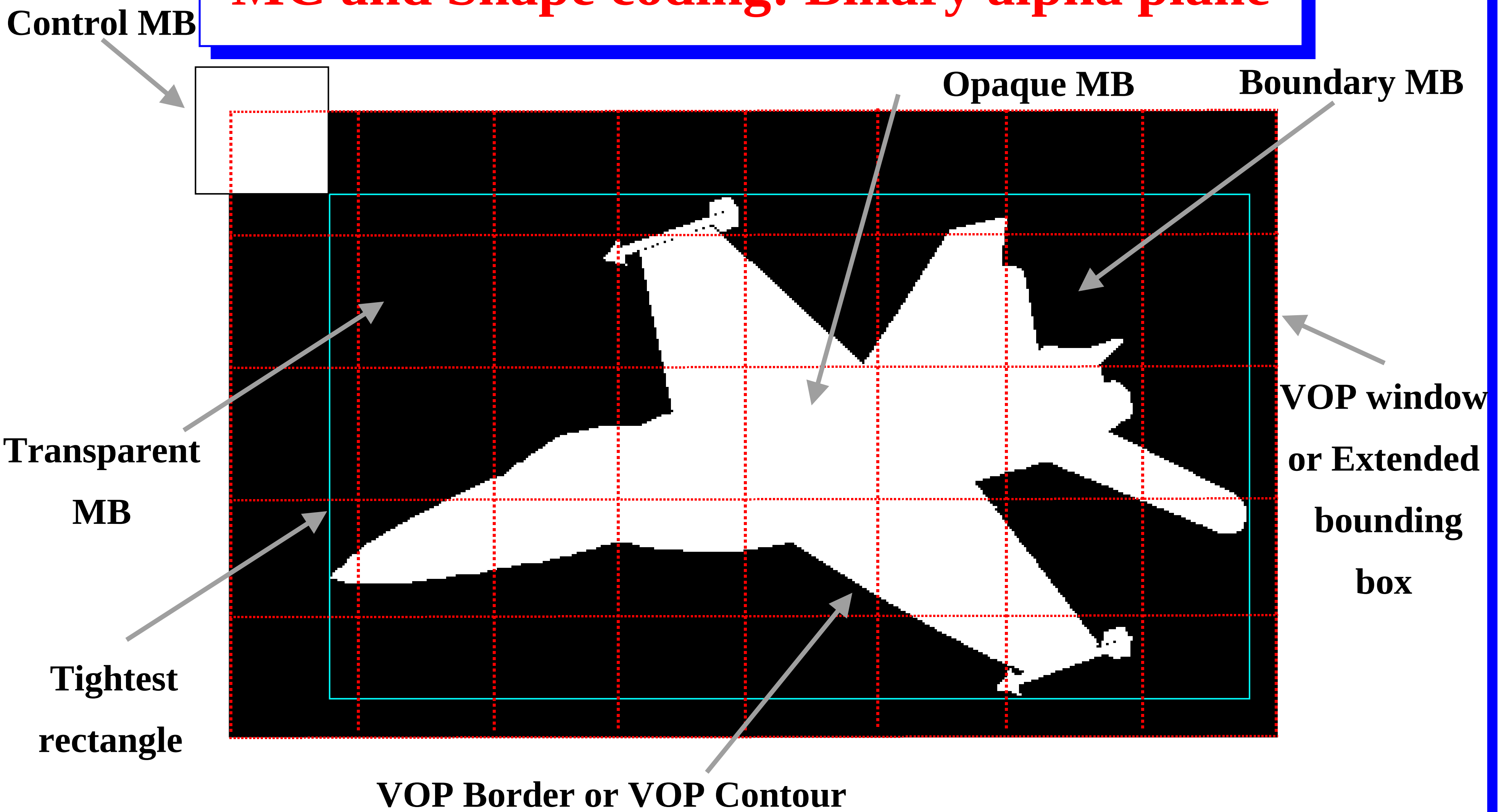
1	**4**	**2**	**3**	**7**	**5**	**6**	**10**	**8**	**9**	**13**	**11**	**12**
I	**P**	**B**	**B**	**P**	**B**	**B**	**I**	**B**	**B**	**P**	**B**	**B**

- **At the decoder output:**

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
I	**B**	**B**	**P**	**B**	**B**	**P**	**B**	**B**	**I**	**B**	**B**	**P**

➩ **In terms of VOP reording, an S(GMC)-VOP can be regarded as a P-VOP.**

MC and Shape coding: Binary alpha plane



Binary Alpha Block (BAB) Types

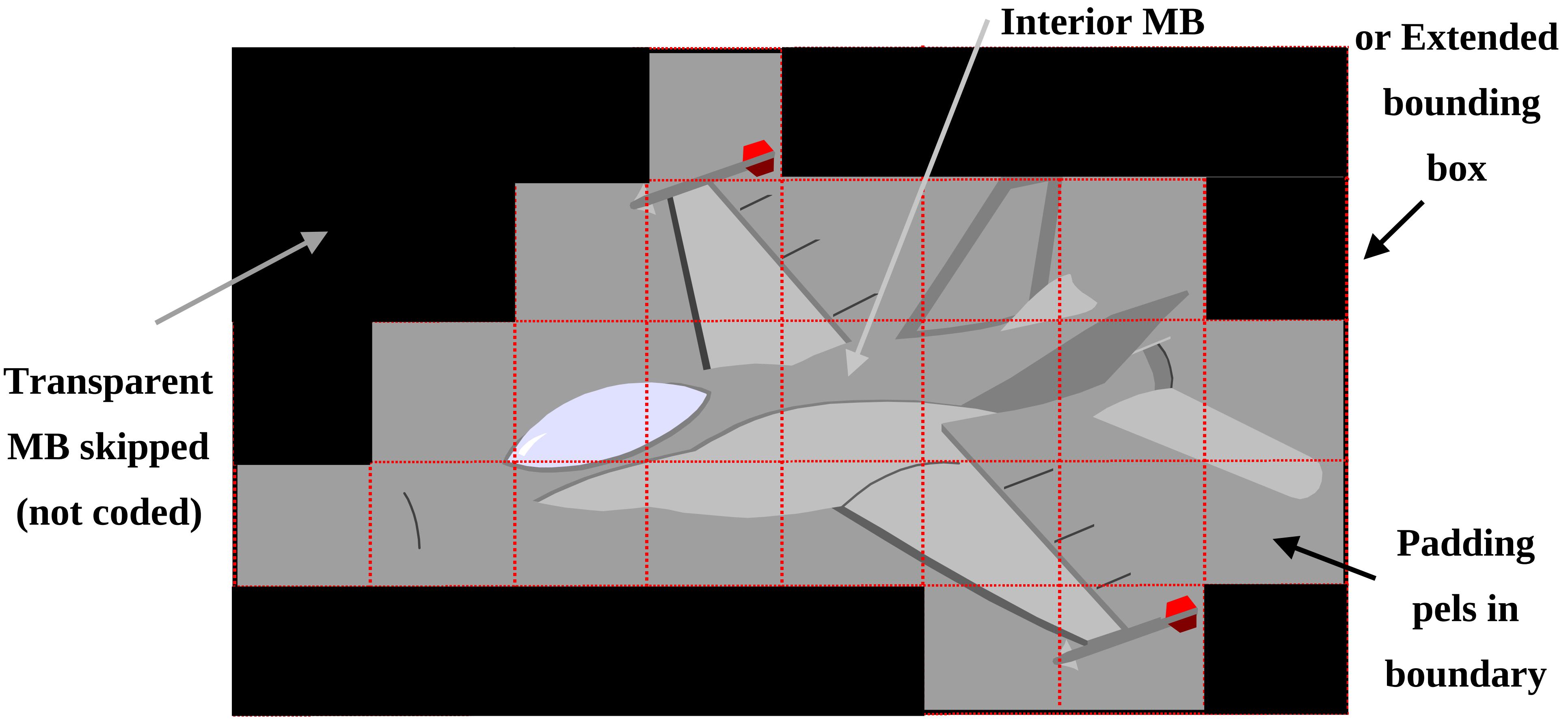
bab_type	Semantic	Used in
0	MVDs==0 && No Update	P,B, and S(GMC)- VOPs
1	MVDs!=0 && No Update	P,B, and S(GMC)-VOPs
2	transparent (All_0)	All VOP types
3	opaque (All_255)	All VOP types
4	intraCAE	All VOP types
5	MVDs==0 && interCAE	P,B, and S(GMC)-VOPs
6	MVDs!=0 && interCAE	P,B, and S(GMC)-VOPs

- **CAE - Context based Arithmetic Coding algorithm. All types, except 2 and 3, are blocks placed on the VOP borders.**

BAB size: 16*16 or 8*8 pels

- Opaque means that all pixels of the bab are part of the object.
- Transparent means that none of the bab pixels belong to the object. Opaque and Transparent blocks are not coded
- IntraCAE means the intra-mode CAE decoding will be required to reconstruct the pixels of the bab.
- No_update means that motion compensation is used to copy the bab from the previous VOP's binary alpha map.
- InterCAE means the motion compensation and inter_mode CAE decoding are used to reconstruct the bab. MVDs refers to the motion vector difference for shape.

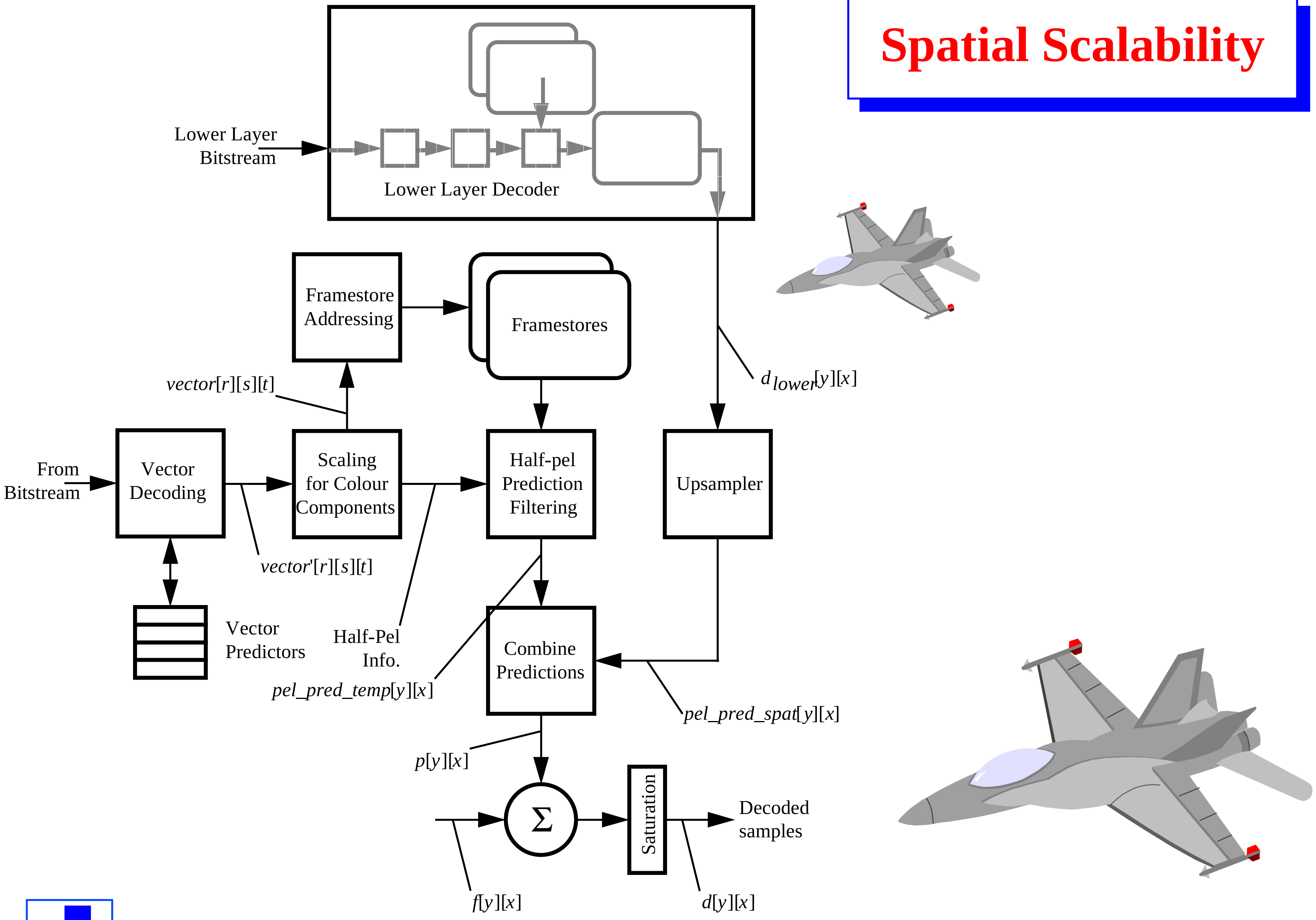
MC and Texture coding



In intra MB, it is padded with horizontal and vertical repetition

In inter MB, it is padded with zeros

Spatial Scalability



Definition of Natural Visual Profiles@Levels

Visual Profile	Level	Typical Visual Session Size	Max MBs per VOP	Max ob-jects[1]	Maximum number per type	Max unique Quant Tables	Max. VMV buffer size (MB units)[2]	Max VCV buffer size (MB)[8]	VCV decoder rate (MB/s)[4]	VCV Boundary MB decoder rate (MB/s)[9]	Max total VBV	Max vbv buffer size (units of 16384 bits)	Max. video packet length	Max sprite size (MB units)	Wavelet restric-tions	Max bitrate (kbit/s)	Max. enhance-ment layers per object
N-Bit	L2	CIF	396	16	16 x Core or Simple or N-Bit	4	2376	792	23760	11880	80	40	4096	N. A.[7]	N. A.	2000	1
Main	L4	1920 x 1088	8160	32	32 x Main or Core or Simple	4	48960	16320	489600	**244800**	760	380	16384	65280	1 taps default integer filter	38400	1 temporal, 2 spatial
Main	L3	CCIR 601	1620	32	32 x Main or Core or Simple	4	9720	3240	97200	48600	320	160	16384	6480	1 taps default integer filter	15000	1
Main	L2	CIF	396	16	16 x Main or Core or Simple	4	2376	1188	23760	11880	80	40	8192	1584	1 taps default integer filter	2000	1
Core	L2	CIF	396	16	16 x Core or Simple	4	2376	792	23760	11880	80	40	8192	N. A.	N. A.	2000	1
Core	L1	QCIF	99	4	4 x Core or Simple	4	594	198	5940	2970	16	8	4096	N. A.	N. A.	384	1

Definition of Natural Visual Profiles@Levels

Visual Profile	Level	Typical Visual Session Size	Max MBs per VOP	Max ob-jects [1]	Maximum number per type	Max unique Quant Tables	Max. VMV buffer size (MB units) [2]	Max VCV buffer size (MB) [8]	VCV decoder rate (MB/s) [4]	VCV Bound-ary MB decoder rate (MB/s) [9]	Max total VVB	Max vbv buffer size (units of 16384 bits)	Max. video packet length	Max sprite size (MB units)	Wavelet restric-tions	Max bitrate (kbit/s)	Max. enhance-ment layers per object
Simple Scalable [3]	L2	CIF	396	4	4 x Simple or Simple Scalable	1	3168	792	23760	N.A.	40	20	4096	N. A.	N. A.	256	1 spatial or temporal enhancement
Simple Scalable	L1	CIF	396	4	4 x Simple or Simple Scalable	1	1782	495	7425	N. A.	40	20	2048	N. A.	N. A.	128	1 spatial or temporal enhancement
Simple	L3	CIF	396	4	4 x Simple	1	792	396	11880	N. A.	40	20	8192	N. A.	N. A.	384	N. A.
Simple	L2	CIF	396	4	4 x Simple	1	792	396	5940	N. A.	40	20	4096	N. A.	N. A.	128	N. A.
Simple	L1	QCIF	99	4	4 x Simple	1	198	99	1485	N.A.	10	5	2048	N. A.	N. A.	64	N. A.

Highlighting new Transform Algorithms

- **Shape Adaptive DCT (SA-DCT)**
- **Shape Adaptive Wavelet Transform (SA-Wavelet)**

MPEG-4

Prof. António Navarro

Instituto de Telecomunicações

Universidade de Aveiro

3800 Aveiro- PORTUGAL

Tel: +351 34 383089 Fax: +351 34 383091

Email: navarro@av.it.pt

http://www.av.it.pt/colaboradores/Navarro/