H.264压缩编码标准

Communication University of China

H.264概述1

- ITU-T、ISO/IEC;
- ITU针对可视会议等应用分别制定了H.261、H.262、H.263、H.263+、H.263++；
- ISO/IEC主要制定了MPEG-1、MPEG-2、MPEG-4等标准;
- 由ITU-T视频编码专家组（VCEG）和ISO/IEC运动图像专家组（MPEG）联合组成的联合视频组（JVT，Joint Video Team）提出的新一代数字视频压缩标准。

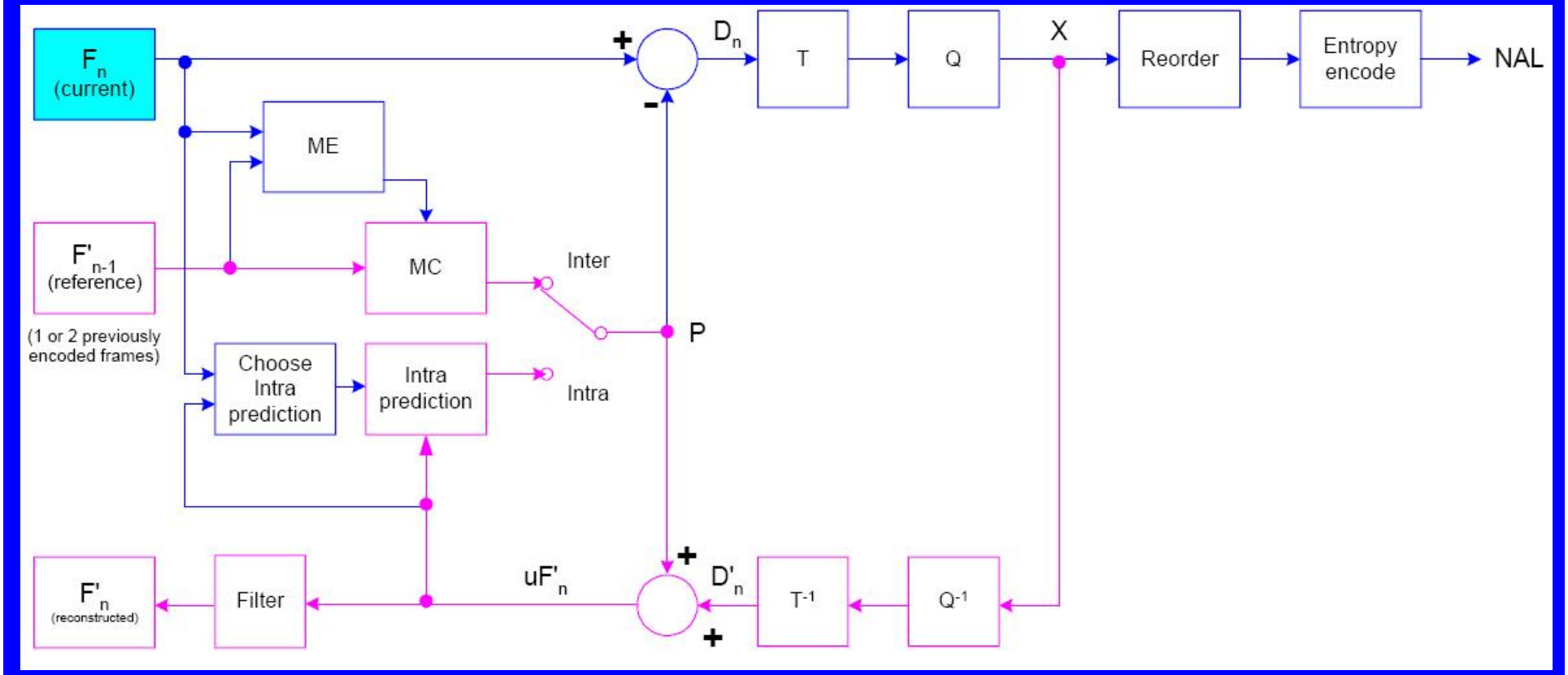
H.264概述2

- H.264/MPEG-4 AVC（Advanced Video Coding）MPEG-4第10部分；
- 注重实用，采用成熟的技术，要求更高的编码效率和简洁的表现形式；
- 注重对移动和IP网络的适应，采用分层技术，从形式上将编码和信道隔离开来，实质上是在源编码器算法中更多地考虑到信道的特点；
- 三是在混合编码器的基本框架下，对其主要关键模块都进行了重大改进；

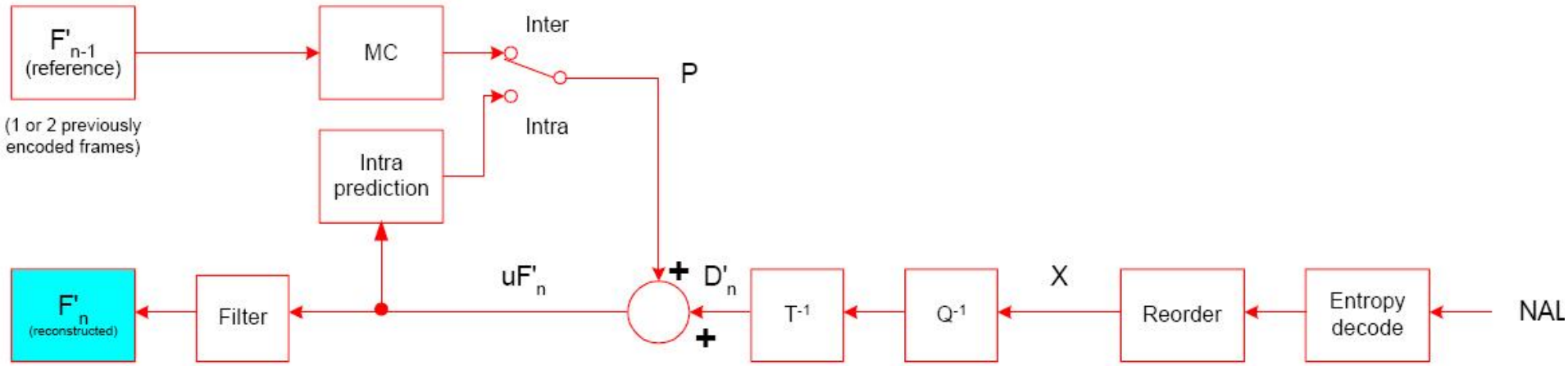
一、H.264编码原理

- H.264/AVC标准没有明确定义一个编解码器。
- 标准定义的是编码视频比特流的语法结构和对该比特流解码的方法。H.264标准中诸如预测、变换、量化、熵编码等基本功能模块与前几个标准（MPEG-1, MPEG-2, MPEG-4, H.261, H.263）并无太大区别。
- 它的变化主要体现在功能模块的具体细节上。

H.264/AVC编码器



H.264/AVC解码器



H.264/AVC与其它编码标准的差异

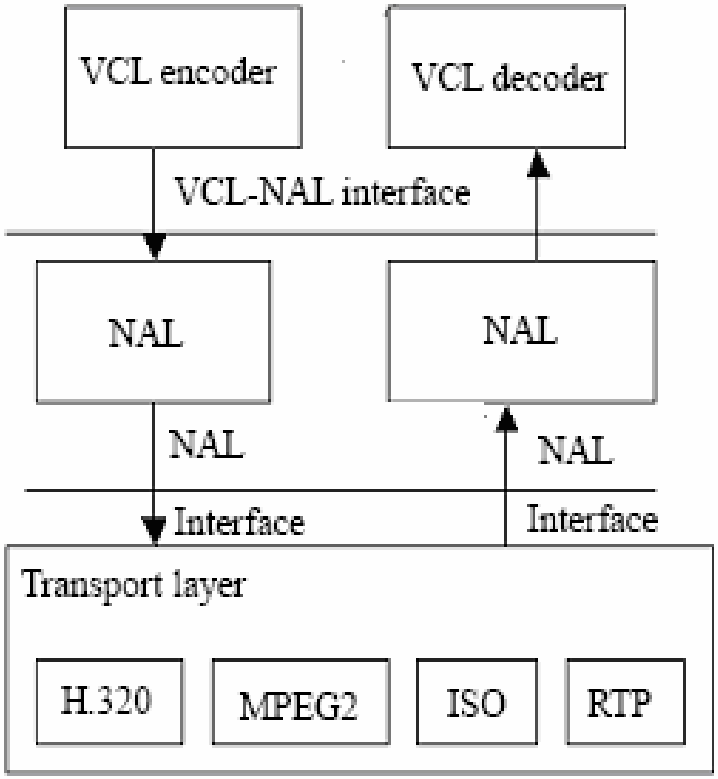
- （1）与其它标准采用的DCT变换不同的是，H.264/AVC采用基于整数的变换，其优势在于反变换时不存在误匹配的问题；
- （2）H.264/AVC使用不同尺寸的块和形状，高分辨率的子像素运动估计和选择多个参考帧来实现运动估计和运动补偿技术；
- （3）H.264/AVC中采用的熵编码技术包括通用变长编码(UVLC, universal variable length coding)、基于上下文的自适应变长码编码(CAVLC, context-based adaptive variable length coding)或基于上下文的自适应二进制算术编码(CABAC, context-based adaptive binary arithmetic coding)。

二、H.264采用的先进技术

- 分层设计；
- 帧间预测编码；
- 帧内预测编码；
- 整数变换；
- 量化处理；
- 去块效应滤波；
- 熵编码

1、分层设计

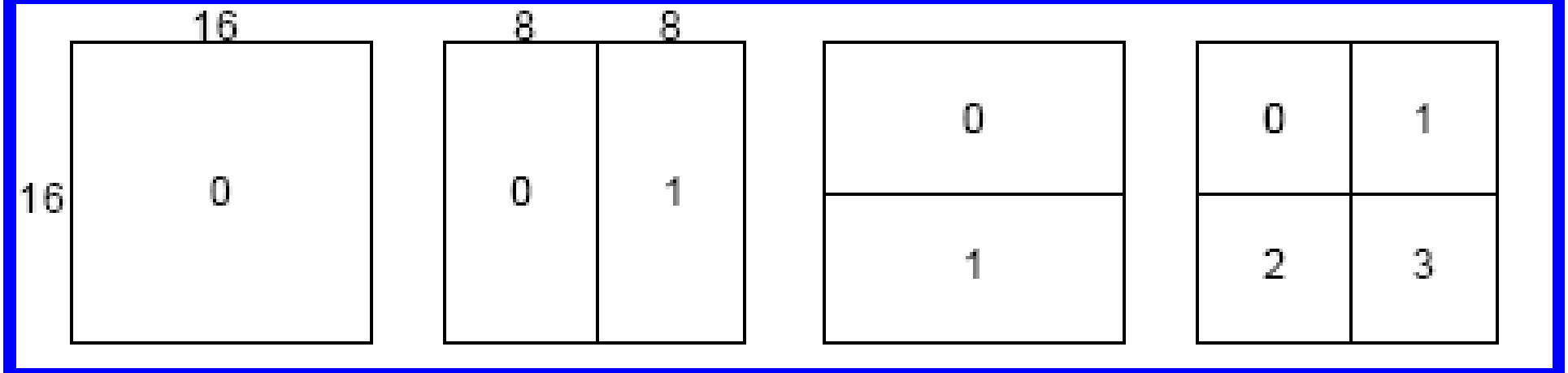
- 在网络传输环境中，其视频编码方案主要由以下两部分组成：视频编码层VCL (video coding layer)和支持视频在不同网络之间传输的网络抽象层NAL (network abstraction layer)。

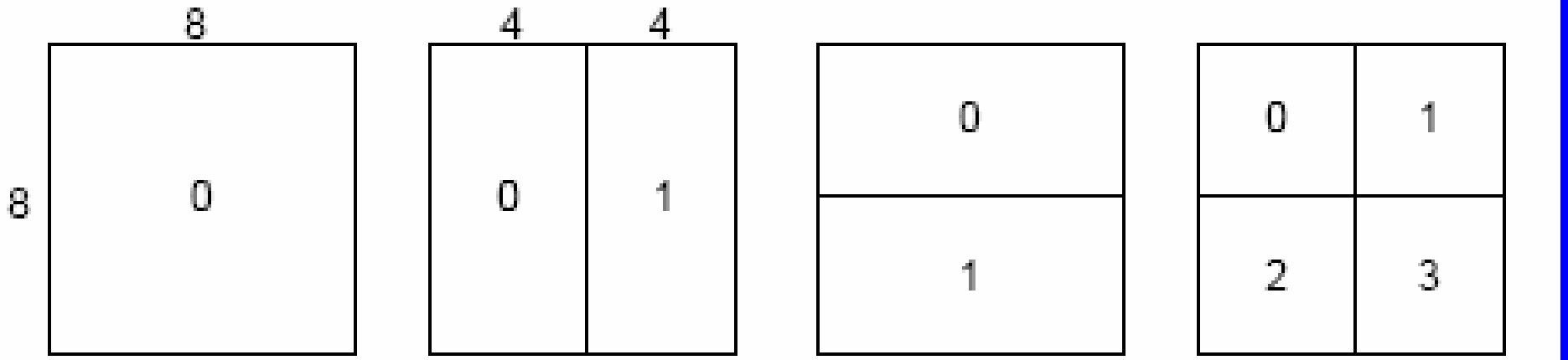


2、帧间预测编码

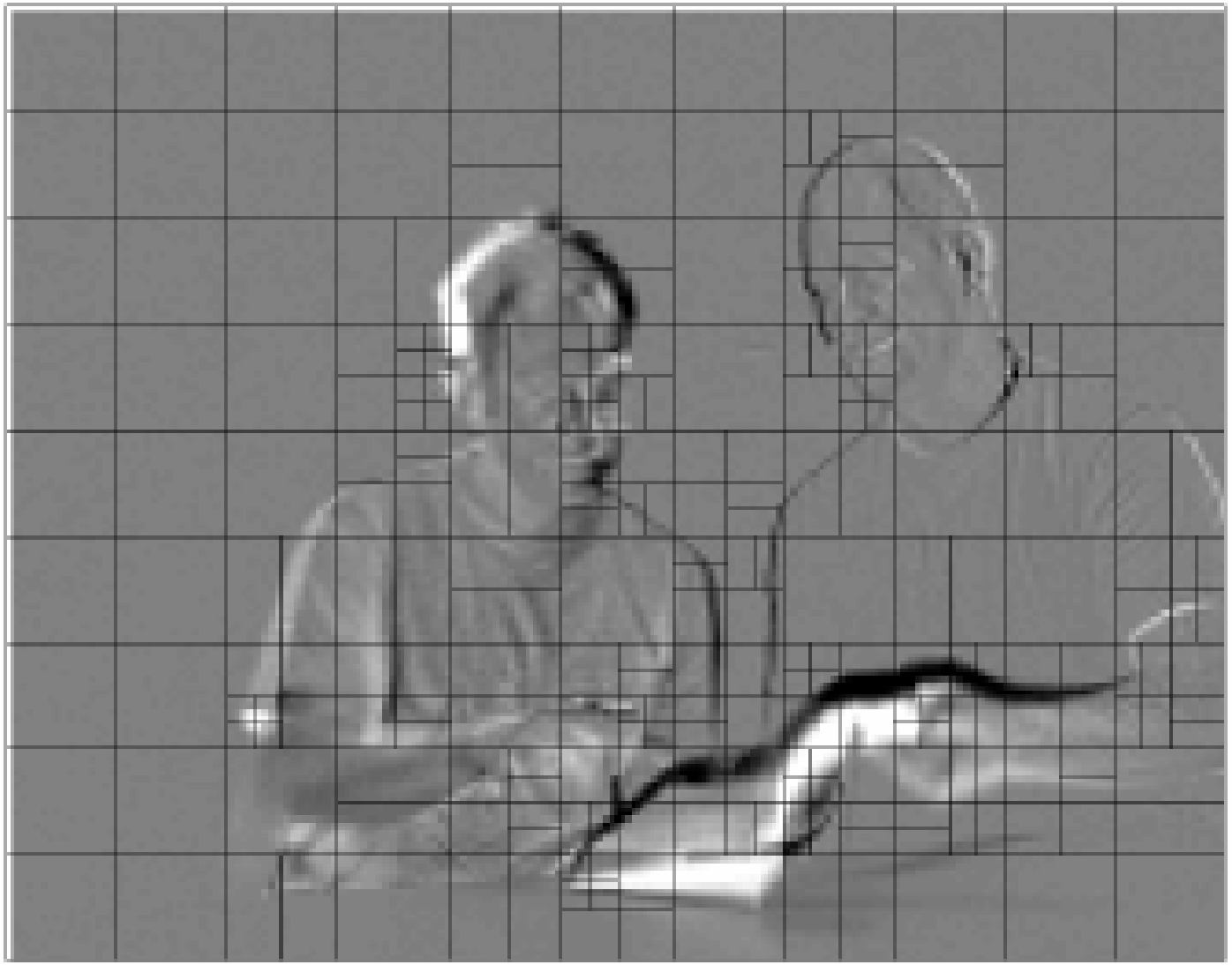
- H.264编解码器采用基于块的运动补偿。除了保留以前编码标准的主要特性外，同时还采用了一些新的特性来提高编码效率。其主要方法包括：
- (1)在运动搜索时使用不同大小和形状的块进行搜索；
- (2)使用1/4像素精度搜索，即使用高精度的运动矢量来表示图像块的运动方向和位移；
- (3)使用多个预测帧进行帧间预测；
- (4)引入SP帧和SI帧；

（1）、树型结构化运动补偿（Tree Structured Motion Compensation）



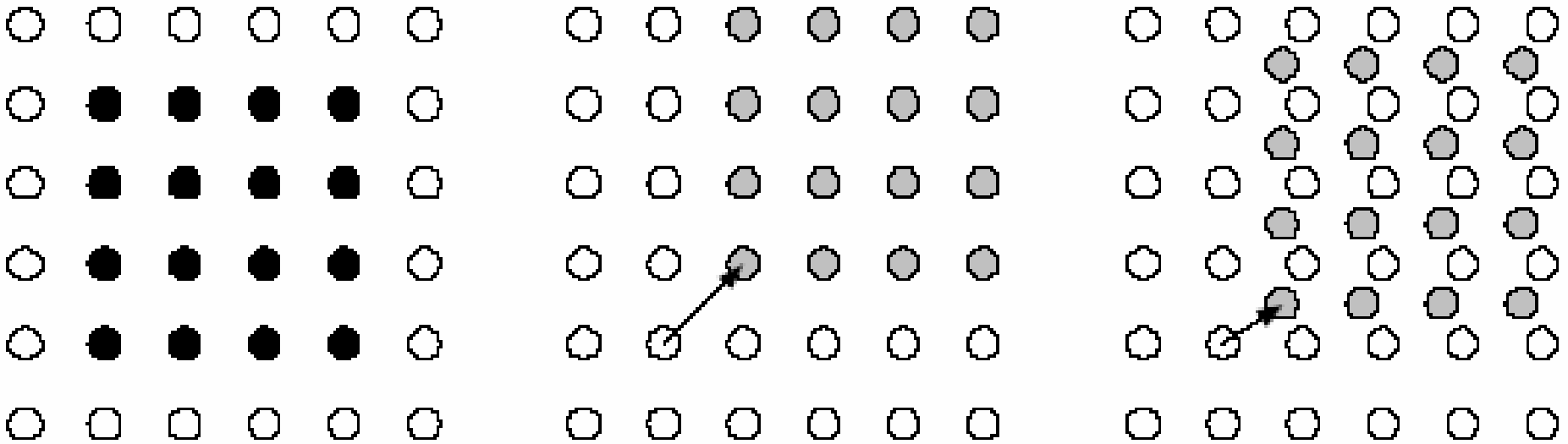


区域划分举例



（2）、1/4像素精度运动矢量

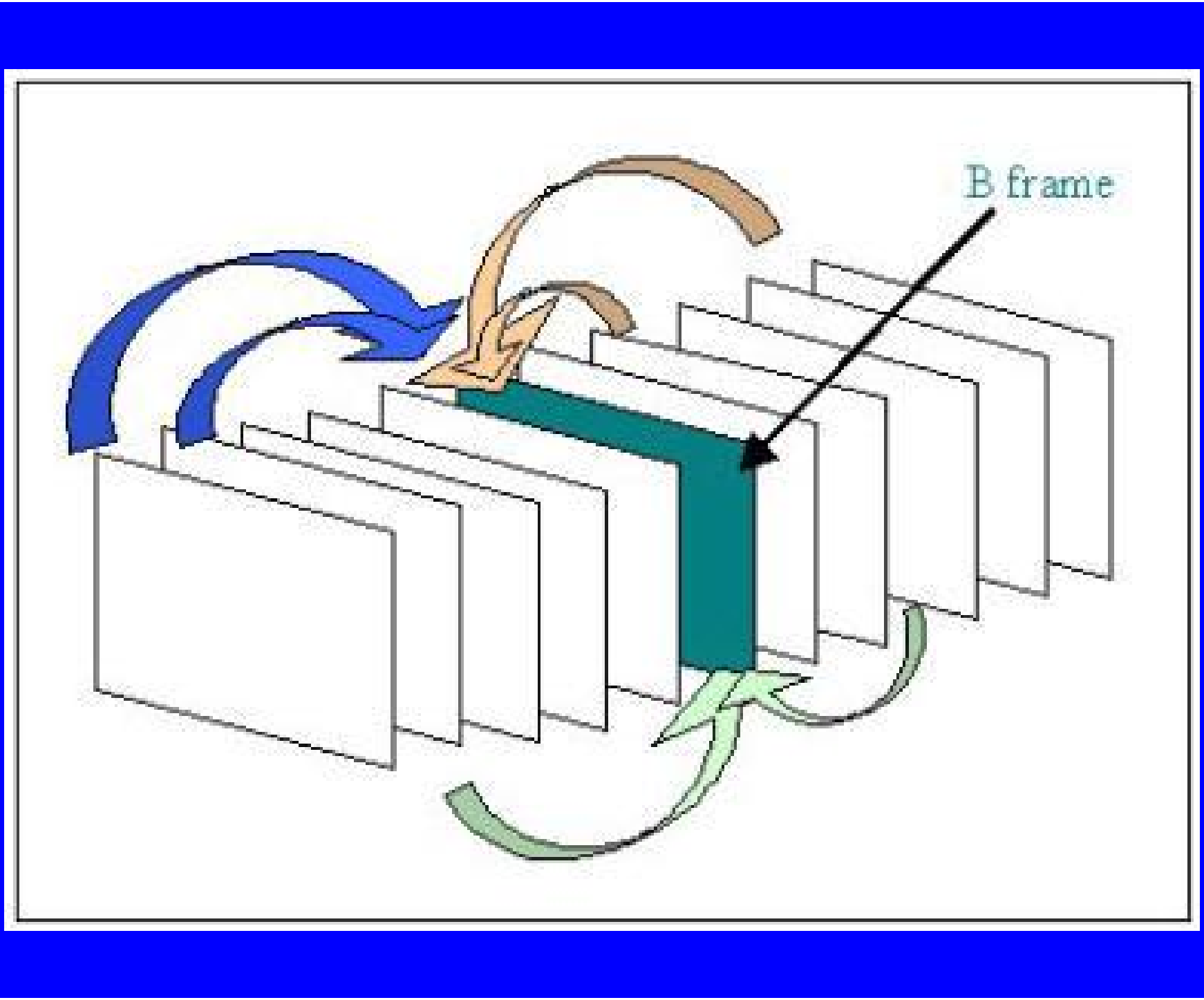
- 与以前的编码标准相比，H.264算法由于采用1/4像素精度来表示运动矢量的大小，因此能够更准确的得到预测块相对于原始块的位移，从而提高了预测精度，以达到压缩码率的效果，它能够节省20%的比特开销。



(a) 4x4 block in current frame　(b) Reference block: vector (1, -1)　(c) Reference block: vector (0.75, -0.5)

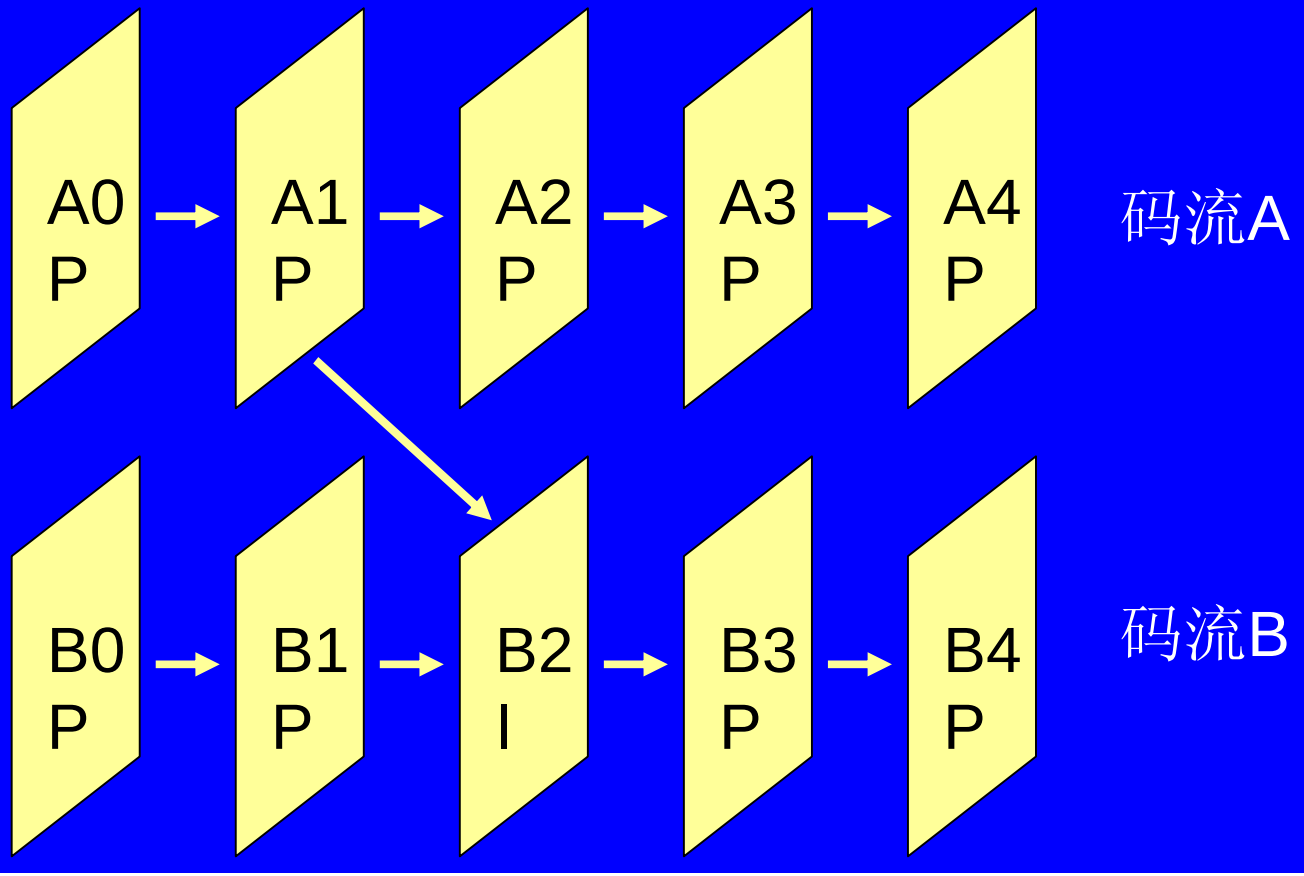
（3）、多个参考帧进行帧间预测

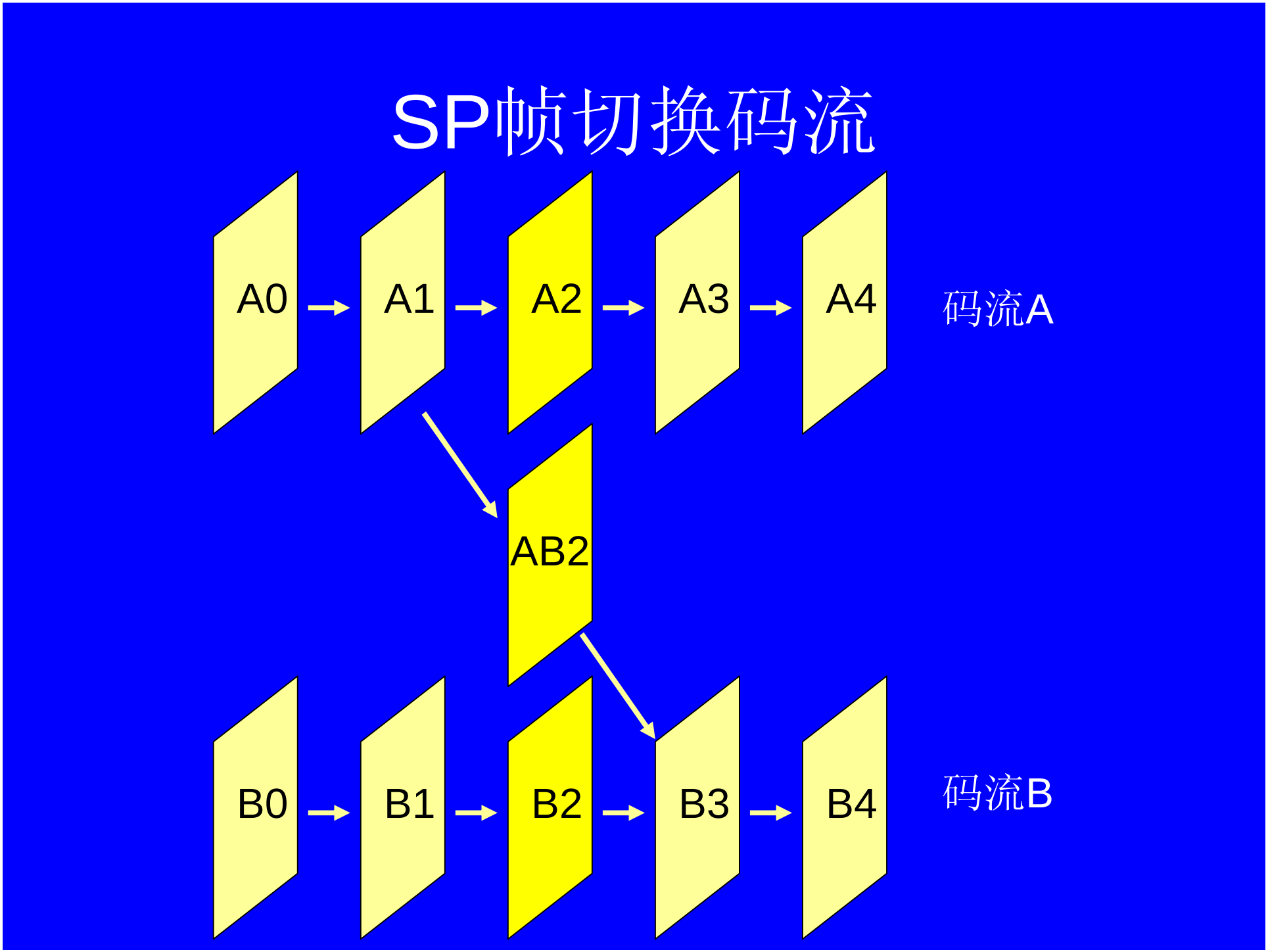
- H.264标准中，在帧间编码过程中不只使用一个参考帧进行预测。编码器可以在多个参考帧中进行运动搜索，选择一个效果最好，与编码帧最相似的一帧作为参考帧。
- 可以得到更好的主观图像质量及编码效率。
- 增加处理延时而且编解码器需要更多的内存来储存参考帧。



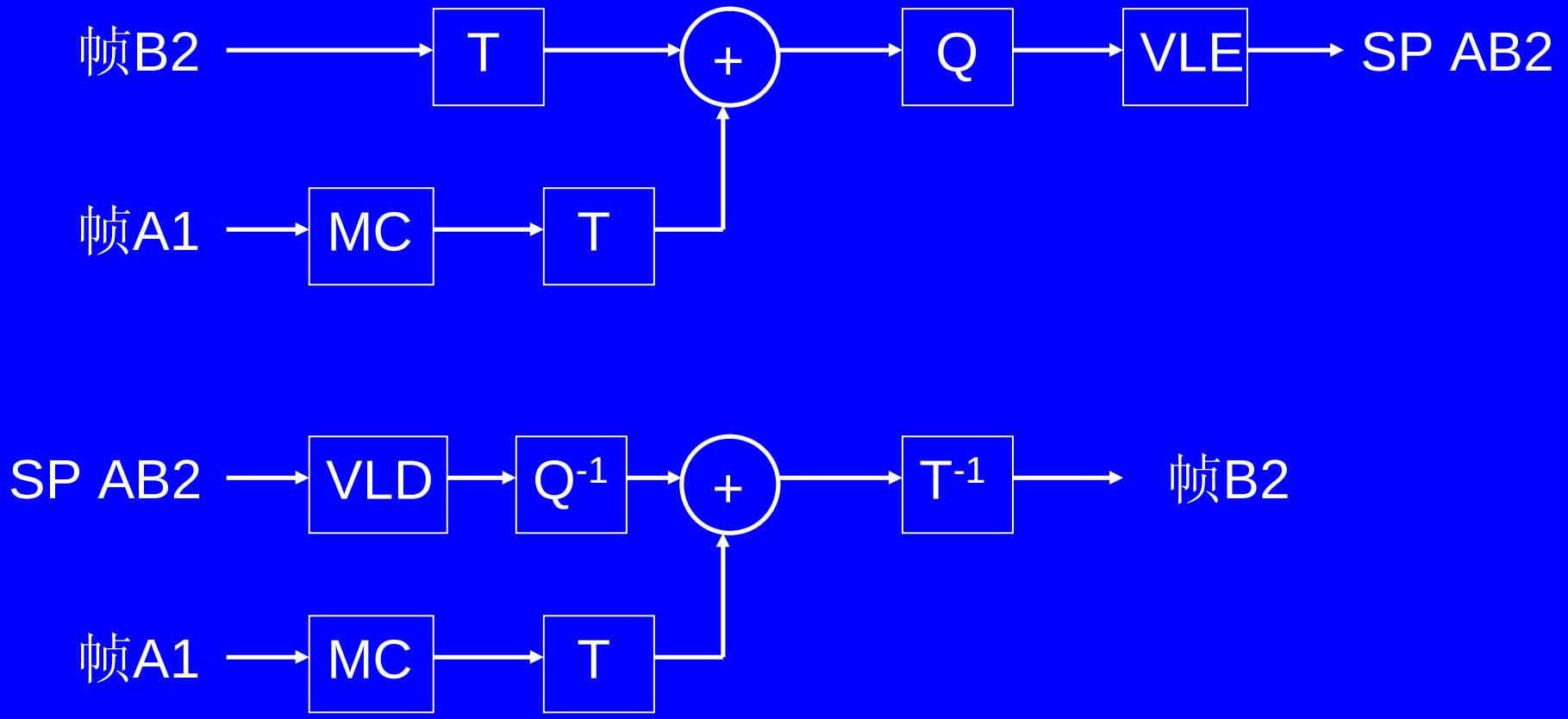
（4）、引入SP帧和SI帧

- H.264除了支持I帧、P帧和B帧外，还支持新的码流间可转换帧—SP帧和SI帧。





SP AB2的编码解码过程



3、帧内预测编码

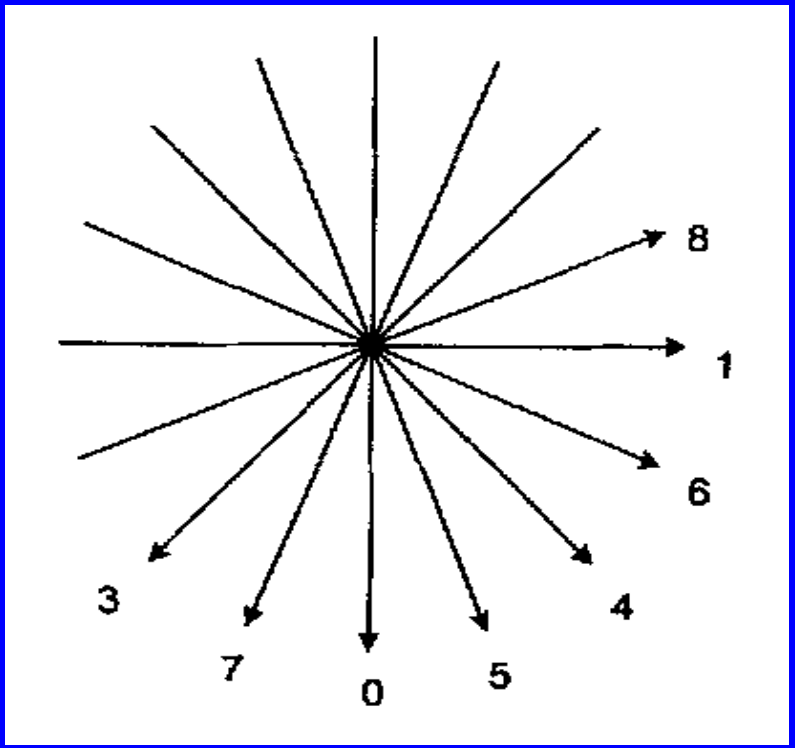
- H.264并不直接对图像块进行处理，而是根据邻近块(通常是位于当前块上面和左边相邻的已编码块)的值来预测当前宏块的值，然后再对预测值和原始值的差值进行变换、量化和编码。这种方法由于是对预测误差进行编码，因此编码效率更高。
- H.264支持两种帧内编码模式，即4×4与16×16编码模式。

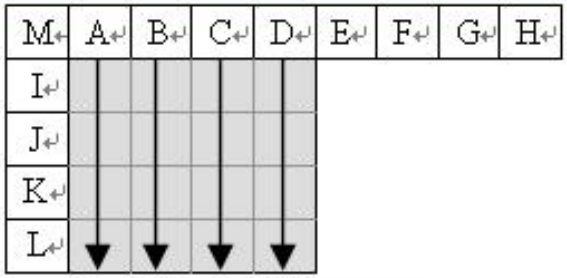
（1）、4×4帧内预测模式

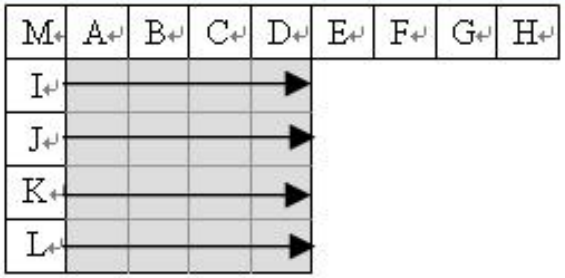
Q A B C D E F G H

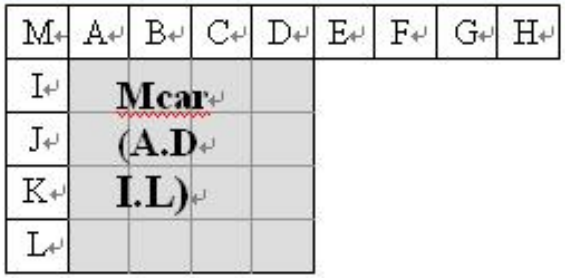
I
J
K
L
M
N
O
P

a	b	c	d	
e	f	g	h	
i	j	k	l	
m	n	o	p	

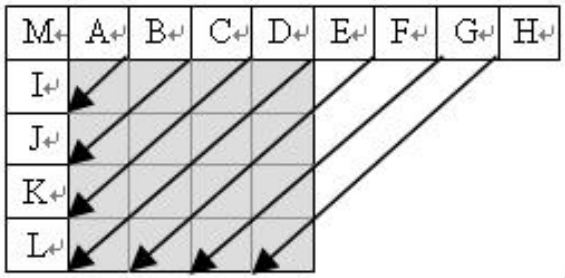


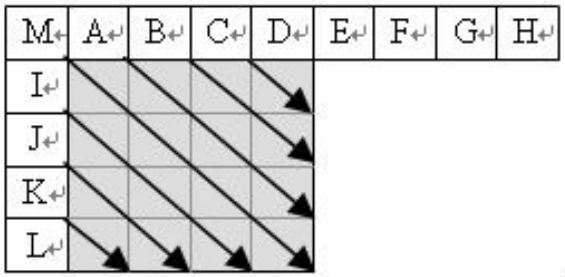


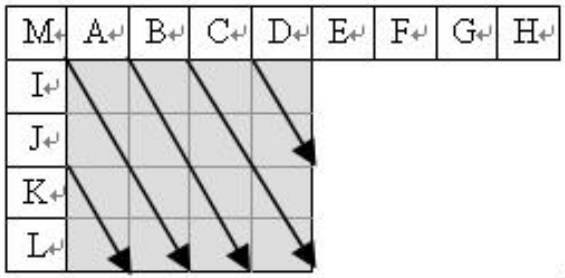




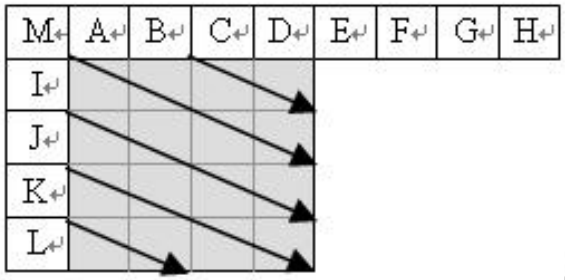
0(垂直)□□□□·····1(水平)·····2(DC)

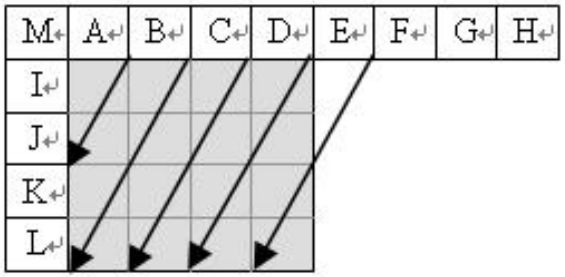


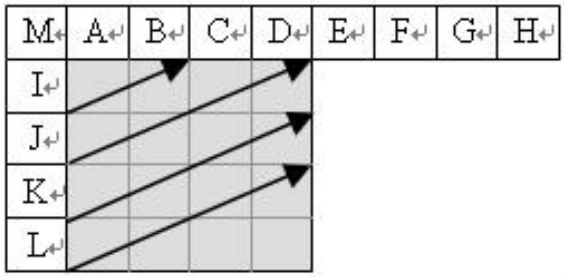




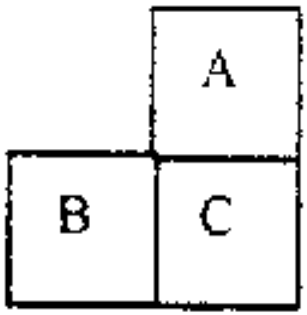
3(下左对角线)□□·····4(下右对角线)□□□·····5(右垂直)







6(下水平)□□□□□··7(左垂直)□□□□□·····8(上水平)



a

0	0	2	2
1	1	3	3
4	4	6	6
7	7	8	8

b

(2)16×16帧内预测模式

- 对于图像中的平坦区域，以16×16为单位进行预测更有助于加快处理速度和降低码率。H.264提供了垂直预测、水平预测、直流预测和平面预测四种16×16的预测模式。

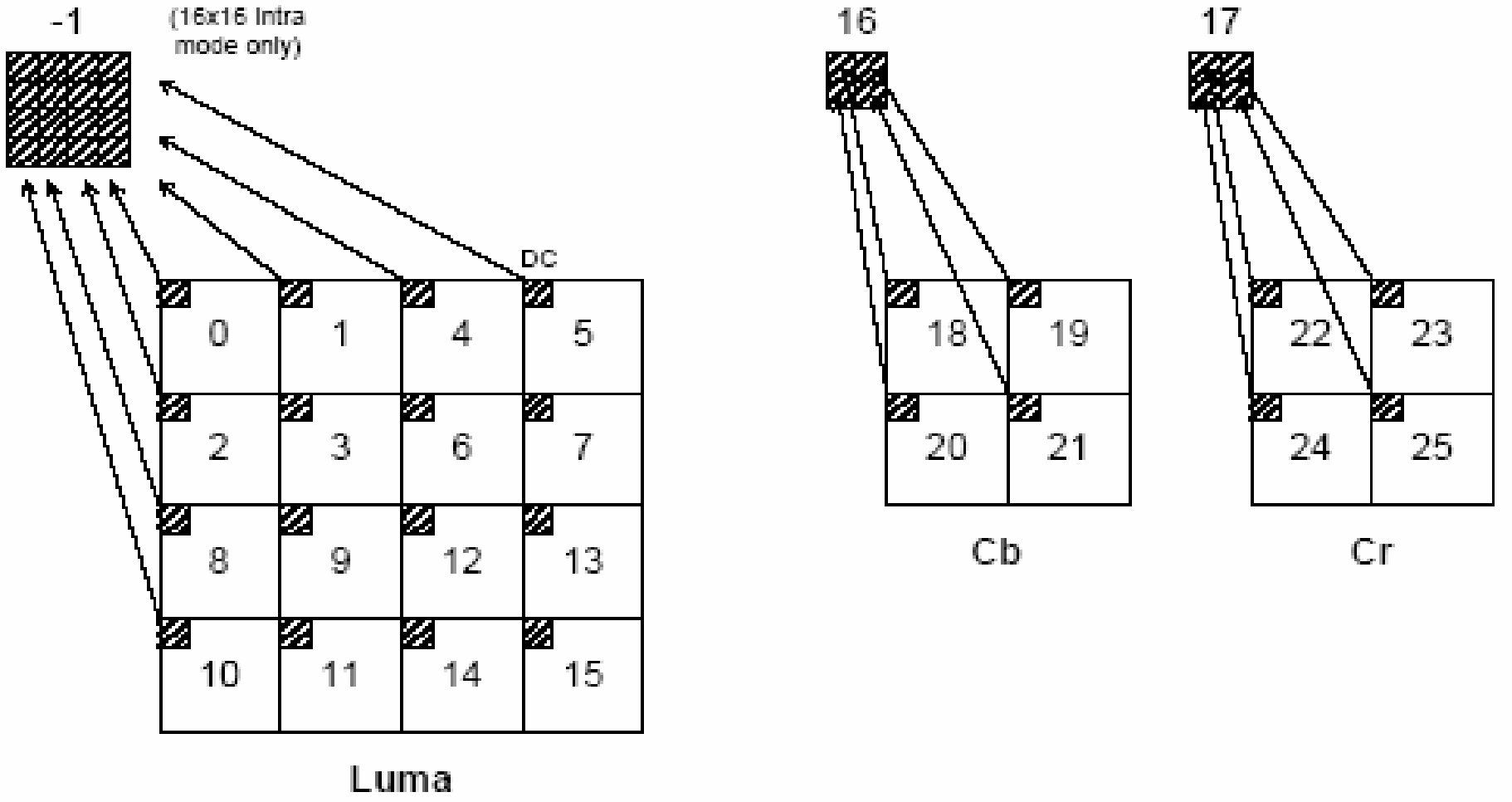
(3)色度块的帧内预测及编码

- 每个帧内编码宏块的8×8色度分量由已编码左上方色度像素的预测而得，两种色度分量常用一种预测模式。4种预测模式类似于帧内16×16的4种预测模式，只是模式编号不同，其中DC为模式0,水平为模式1,垂直为模式2,平面为模式3。

4、整数变换

- 对每个宏块的预测误差都将进行变换，量化和编码。MPEG-1、MPEG-2、MPEG-4、H.263等标准都以8×8的离散余弦（DCT）作为基本变换。H.264的“baseline”型（profile）根据所压缩的数据类型不同而采用三种不同的变换方式。
- 适用于帧内预测宏块4×4亮度DC系数块的变换；
- 适用于任何宏块色度2×2 DC系数块的变换；
- 适用于其他4×4残差数据块的变换；
- 如果需要，还可以选择与运动补偿块大小（4×8、8×4、8×8、16×8）相对应的变换。

宏块扫描和传输顺序



(1)4×4残差变换

- 变换适用于运动补偿或帧内预测后的4×4残差数据块（0—15和18—25）。虽然该变换的基础是DCT变换但二者之间有根本的差别：
- 1)该变换是一种整数变换，所有的运算都是整数运算并且没有精度损失。
- 2)H.264标准中对该变换的反变换有详细的说明，如果完全按照说明正确执行，编解码器之间不会出现误匹配。
- 3)该变换的核心部分不需要乘法，仅仅需要加法和移位运算。

5、量化处理

- H.264标准支持52个量化步长，对应于不同的量化参数（QP）如表1-1所示，QP值每增加6，Qstep值增加一倍。QP值每增加1，Qstep值增加12.5%。量化步长取值范围很广，这就为编码中兼顾比特率和编码质量提供了足够多的灵活度和准确度。

QP	0	1	2	3	4	5	6	7	8	9	10	11	12	….
QStep	0.625	0.6875	0.8125	0.875	1	1.125	1.25	1.375	1.625	1.75	2	2.25	2.5	….
QP	…	18	…	24	…	30	…	36	…	42	…	48	…	51
QStep		5		10		20		40		80		160		224

6、去块效应滤波

- H.264/AVC定义了一个自适应循环滤波器，滤波的强度通过几个语法元素控制。
- 滤波的基本思想是：如果块边沿的绝对差值相对比较大，出现块人工瑕疵的可能性就很大，因此需要进行相应处理。

7、熵编码

- H.264标准提供的熵编码方案有：
- 通用变长码编码(UVLC)；
- 基于上下文的自适应变长码编码(CAVLC)；（Exponential Golomb Codes）
- 基于上下文的自适应二进制算术编码(CABAC)。

(1)、Exp-Golomb编码

- 对所有的句法元素，除了量化系数外，使用单一无限扩展的码字表。
- 是有规则结构的可变长编码。

码字序号	码字
0	1
1	010
2	011
3	00100
4	00101
5	00110
6	00111
7	0001000
…	…

(2)、CAVLC编码

- 基于上下文的自适应变字长编码是对经过之字形扫描的4×4块变换系数进行编码的方法。利用了4×4块的一些特性：
- 预测变换量化后的块一般是稀疏的。
- 之字形扫描后的最高零系数是+1/-1的序列。
- 相邻块的非零系数是相关的。
- 非零系数的幅度在重排数组的开始处比较高，在高频系数比较低。

Example1

- 重排后数据是：-2，4，3，-3，0，0，-1，0，…
- 对系数的数目和绝对值为“1”的数值进行编码；
- 对各个T1的符号进行编码。T1的符号用1个比特编码0表示+，1表示-。从最高频率的T1开始反向编码。

-2	4	0	-1
3	0	0	0
-3	0	0	0
0	0	0	0

Example1

- 按逆向扫描的次序对余下的数值进行编码。
- 对最后一个系数之前的零的总数进行编码。
- 对零的游程进行编码，说明零是如何分布的。

Example1

TotalCoeffs = 5 (indexed from highest frequency [4] to lowest frequency [0])
TotalZeros = 2
T1s = 1

Encoding:

Element	Value	Code
coeff_token	TotalCoeffs=5, T1s=1	0000000110
T1 sign (4)	-	1
Level (3)	Sent as –2 (**see note 1**) (use Level_VLC0)	0001
Level (2)	3 (use Level_VLC1)	0010
Level (1)	4 (use Level_VLC1)	00010
Level (0)	-2 (use Level_VLC2)	111
TotalZeros	2	0011
run_before(4)	ZerosLeft=2; run_before=2	00
run_before(3..0)	0	No code required

The transmitted bitstream for this block is 000000011010001001000010111001100.

Example1

Note 2: After encoding level (3), the level_VLC table is incremented because the magnitude of this level is greater than the first threshold (which is 0). After encoding level (1), with a magnitude of 4, the table number is incremented again because level (1) is greater than the second threshold (which is 3). Note that the final level (-2) uses a different code from the first encoded level (also –2).

Decoding:

Code	Element	Value	Output array
0000000110	coeff_token	TotalCoeffs=5, T1s=1	Empty
1	T1 sign	-	-1
0001	Level	-2 decoded as -3	-3, -1
0010	Level	+3	+3, -3, -1
00010	Level	+4	+4, 3, -3, -1
111	Level	-2	-2, 4, 3, -3, -1
0011	TotalZeros	2	-2, 4, 3, -3, -1
00	run_before	2	-2, 4, 3, -3, 0, 0, -1

All zeros have now been decoded and so the output array is:
-2, 4, 3, -3, 0, 0, -1

Example2

4x4 block:

0	3	-1	0
0	-1	1	0
1	0	0	0
0	0	0	0

Reordered block:
0,3,0,1,-1,-1,0,1,0…

TotalCoeffs = 5 (indexed from highest frequency [4] to lowest frequency [0])
TotalZeros = 3
T1s = 3 (in fact there are 4 trailing ones but only 3 can be encoded as a “special case”)

Example2

Encoding:

Element	Value	Code
coeff_token	TotalCoeffs=5, T1s=3	0000100
T1 sign (4)	+	0
T1 sign (3)	-	1
T1 sign (2)	-	1
Level (1)	+1 (use Level_VLC0)	1
Level (0)	+3 (use Level_VLC1)	0010
TotalZeros	3	111
run_before(4)	ZerosLeft=3; run_before=1	10
run_before(3)	ZerosLeft=2; run_before=0	1
run_before(2)	ZerosLeft=2; run_before=0	1
run_before(1)	ZerosLeft=2; run_before=1	01
run_before(0)	ZerosLeft=1; run_before=1	No code required; last coefficient.

The transmitted bitstream for this block is 000010001110010111101101 .

The output array is "built up" from the decoded values as shown below. Va
array at each stage are underlined.

Code	Element	Value	Output array
0000100	coeff_token	TotalCoeffs=5, T1s=3	Empty
0	T1 sign	+	1
1	T1 sign	-	-1, 1
1	T1 sign	-	-1, -1, 1
1	Level	+1	1, -1, -1, 1
0010	Level	+3	3, 1, -1, -1, 1
111	TotalZeros	3	3, 1, -1, -1, 1
10	run_before	1	3, 1, -1, -1, 0, 1
1	run_before	0	3, 1, -1, -1, 0, 1
1	run_before	0	3, 1, -1, -1, 0, 1
01	run_before	1	3, 0, 1, -1, -1, 0, 1

The decoder has inserted two zeros; however, TotalZeros is equal to 3 and s
before the lowest coefficient, making the final output array:
0, 3, 0, 1, -1, -1, 0, 1

（3）、CABAC编码

- 算术编码是把整个信源表示为实数中0~1之间的一个区间，起长度等于该序列的概率；
- 在区间内选择一个代表性的小数，转化为二进制作为实际的编码输出；
- 算术编码的平均编码长度为小数；

8、H.264的类

- Baseline Profile，BP版本简单、应用广；
- Main Profile，MP，采用多项提高图像质量和增加压缩比的技术措施，可用于SDTV，HDTV，DVD等。编码效率高，复杂度高。熵编码采用CABAC。
- Extended Profile，EP，可用于网络的视频流传输，编码效率高，实时性不高，流的播放实时性要求较高。去掉了CABAC，增加了流的工具群。

4、作业

- ASI接口有哪些特点？它与SDI接口有什么异同？
- MPEG-4采用了哪些技术措施？
- 分析MPEG-4与MPEG-2在视频压缩方面的主要联系和区别。
- H264与目前广泛应用的视频压缩标准相比有哪些突出的优点？有无缺点？
- H264采用了哪些先进的技术措施？

MPEG-4

Communication University of China

MPEG-4

- MPEG-4是由运动图像专家组(Moving Picture Experts Group, MPEG) 制定的压缩标准；
- MPEG-4于1999年1月正式发布，2000年正式成为一项国际标准；
- 旨在为机顶盒、互联网、移动设备等应用实现更高质量的压缩和更灵活的格式，提供更加丰富的选择。
- 网上（流媒体）及光盘分发，语音传送（视像电话），与及电视广播。

标准组成

- 第一部（ISO/IEC 14496-1）：系统：描述视频和音频的同步以及混合方式(multiplexing)。
- 第二部（ISO/IEC 14496-2）：视频：定义了一个对各种视觉信息(包括视频，静止纹理，计算机合成图形等等)的编解码器。对视频部分来说，众多"Profiles"中很常用的一种是Advanced Simple Profile (ASP)。
- 第三部（ISO/IEC 14496-3）：音频：定义了一个对各种音频信号进行编码的编解码器的集合。包括高级音频编码(AAC for Advanced Audio Coding)的若干变形和其他一些音频/语音编码工具。
- 第四部（ISO/IEC 14496-4）：一致性：定义了对本标准其他的部分进行一致性测试的程序。
- 第五部（ISO/IEC 14496-5）：参考软件:提供了用于演示功能和说明本标准其他部分功能的软件。

标准组成

- 第六部（ISO/IEC 14496-6）：多媒体传输集成框架(DMIF for Delivery Multimedia Integration Framework)。
- 第七部（ISO/IEC 14496-7）：优化的参考软件：提供了对实现进行优化的例子。(这里的实现指的是第五部分)。
- 第八部（ISO/IEC 14496-8）：在IP网络上传输：定义了在IP网络上传输MPEG-4内容的方式。
- 第九部（ISO/IEC 14496-9）：参考硬件：提供了用于演示怎样在硬件上实现本标准其他部分功能的硬件设计方案。
- 第十部（ISO/IEC 14496-10）：进阶视讯编码(AVC for Advanced Video Coding)：定义了一个被称为AVC的视频编解码器。从技术上讲，它和ITU-T H.264标准是一致的。

标准组成

- 第十二部（ISO/IEC 14496-12）：基于ISO的媒体文件格式：定义了一个存储媒体内容的文件格式。
- 第十三部（ISO/IEC 14496-13）：知识产权管理和保护(IPMP for Intellectual Property Management and Protection)拓展。
- 第十四部（ISO/IEC 14496-14）：MPEG-4文件格式：定义了基于第十二部分的用于存储MPEG-4内容的容器文件格式。
- 第十五部（ISO/IEC 14496-15）：AVC文件格式：定义了基于第十二部分的用于存储第十部分的视频内容的文件格式。
- 第十六部（ISO/IEC 14496-16）：动画框架扩展(AFX for Animation Framework eXtension)。

标准组成

- 第十七部（ISO/IEC 14496-17）：同步文本字幕格式（尚未完成 - 2005年1月达成"最终委员会草案"(FCD for Final Committee Draft)。
- 第十八部（ISO/IEC 14496-18）：字体压缩和流式传输(针对公开字体格式)。
- 第十九部（ISO/IEC 14496-19）：综合用材质流(Synthesized Texture Stream)。
- 第二十部（ISO/IEC 14496-20）：简单场景表示(LASeR for Lightweight Scene Representation)(尚未完成 - 2005年1月达成"最终委员会草案"(FCD for Final Committee Draft)。
- 第二十一部（ISO/IEC 14496-21）：用于描绘(Rendering)的MPEG-J拓展(尚未完成 - 2005年1月达成"委员会草案"(CD for Committee Draft)。

一、MPEG-4的特点

（1）基于内容的交互性

- MPEG-4提供了基于内容的多媒体数据访问工具，如索引、超级链接、上下载、删除等。利用这些工具，用户可以方便地从多媒体数据库中有选择地获取自己所需的与对象有关的内容；
- MPEG-4提供了内容的操作和位流编辑功能；
- MPEG-4提供了高效的自然或合成的多媒体数据编码方法。它可以把自然场景或对象组合起来成为合成的多媒体数据。

一、MPEG-4的特点

- （2）高效的压缩性
- MPEG-4基于更高的编码效率。
- MPEG-4还能对同时发生的数据流进行编码。

一、MPEG-4的特点

（3）通用的访问性

- MPEG-4提供了易出错环境的鲁棒性，来保证其在许多无线和有线网络以及存储介质中的应用；
- MPEG-4还支持基于内容的的可分级性，即把内容、质量、复杂性分成许多小块来满足不同用户的不同需求，支持具有不同带宽，不同存储容量的传输信道和接收端。

二、MPEG-4的设计思想

- 开放性
- 引入基于对象表达的概念（Object-Based Representation）

基于对象表达的概念（Object-Based Representation）

The Object-Based Representation Vision

- Audiovisual Scene is seen as a composition of 'objects'
- Integration of objects from different nature: A&V, natural and synthetic, text & graphics, animated faces, arbitrary and rectangular shapes, ...
- Interaction with objects is possible
- Object-based content may be re-used in different contexts
- Object-based processing and coding
- Composition principle is independent of bitrate: from low bitrates to (virtually) lossless quality

MPEG-4 VIDEO OBJECTS

MPEG-4 VLBV Core Coder

Video Object Plane → Motion (MV) | Texture (DCT) → bitstream

(Similar to H.263/MPEG-1)

Generic MPEG-4 Coder

Video Object Plane → Shape | Motion (MV) | Texture (DCT) → bitstream

Example of an Object-Based Scene

Sports results: Paris S. G. - Benfica

Example of an Object-Based Scene

Sports results: Paris S. G. - Benfica

Example of an Object-Based Scene

Sports results: Paris S. G. - Benfica

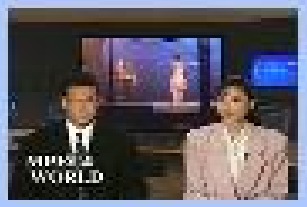
Example of an Object-Based Scene

Sports results: Paris S. G. - Benfica

Example of an Object-Based Scene

Sports results: Paris S. G. - Benfica

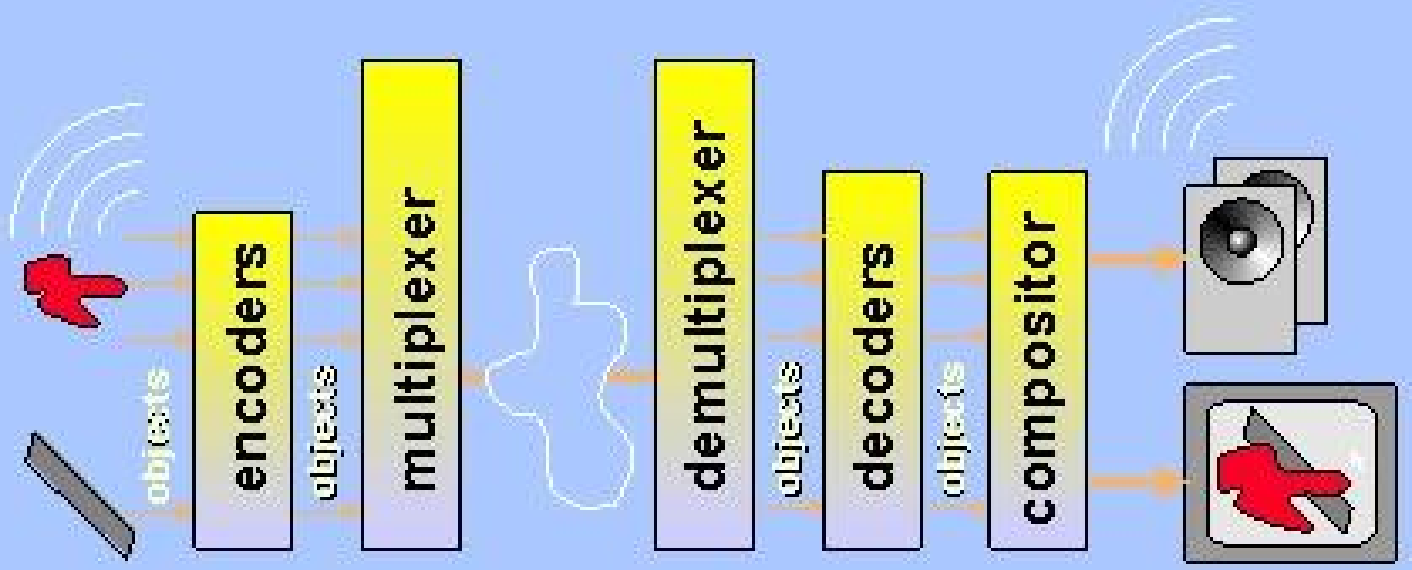
Example of an Object-Based Scene



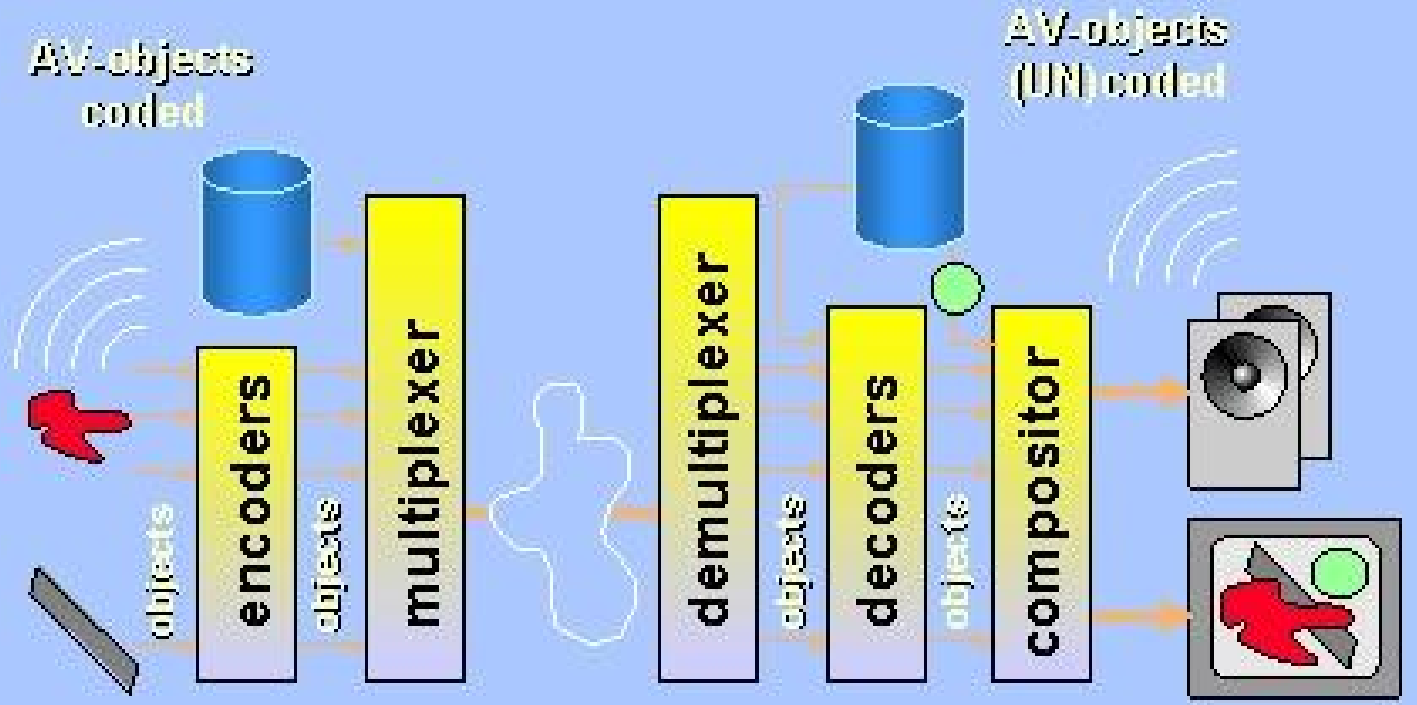
Sports results: Paris S. G. - Benfica

Sports results: Paris S. G. - Benfica

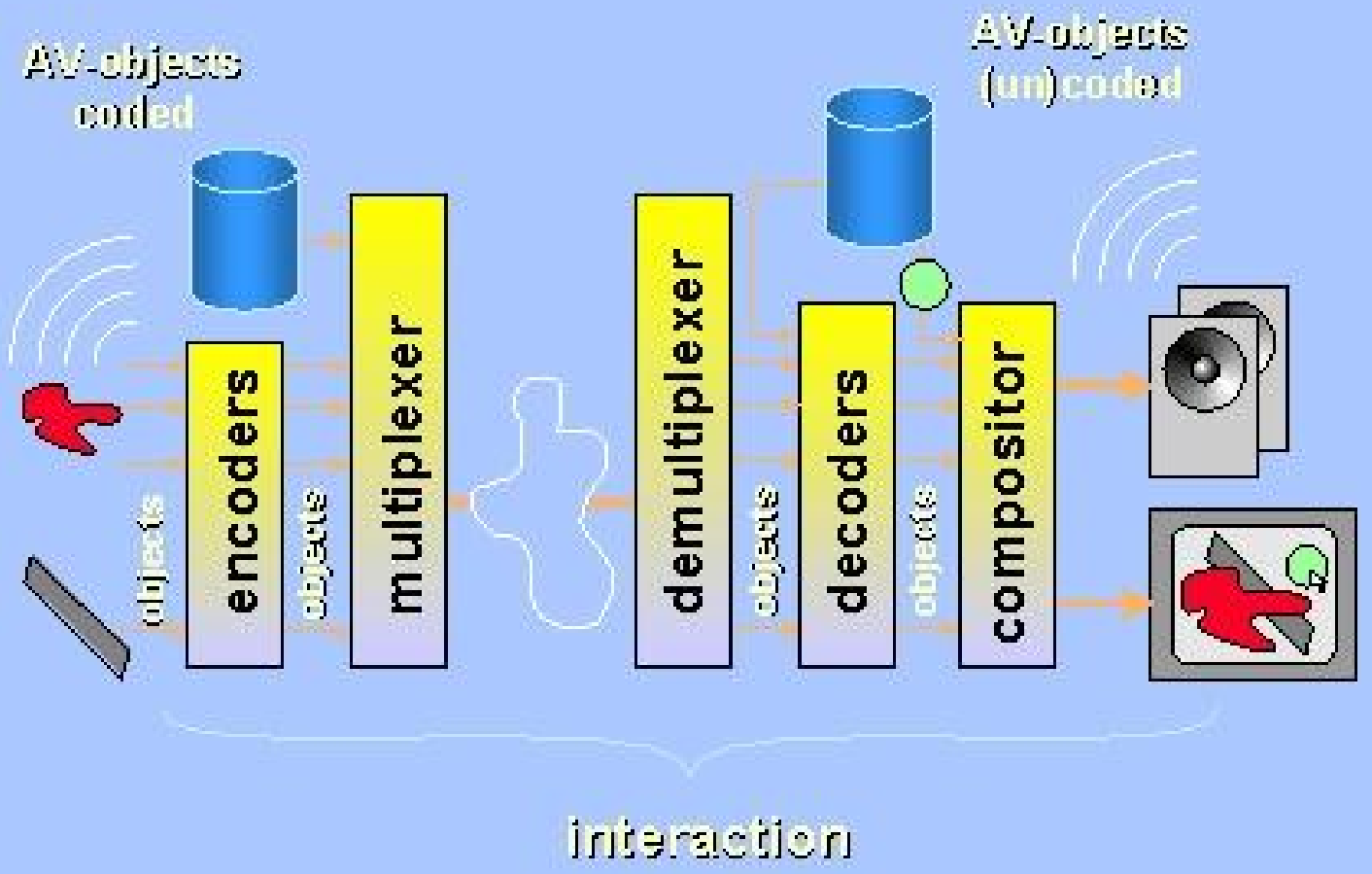
Object-based Audiovisual System



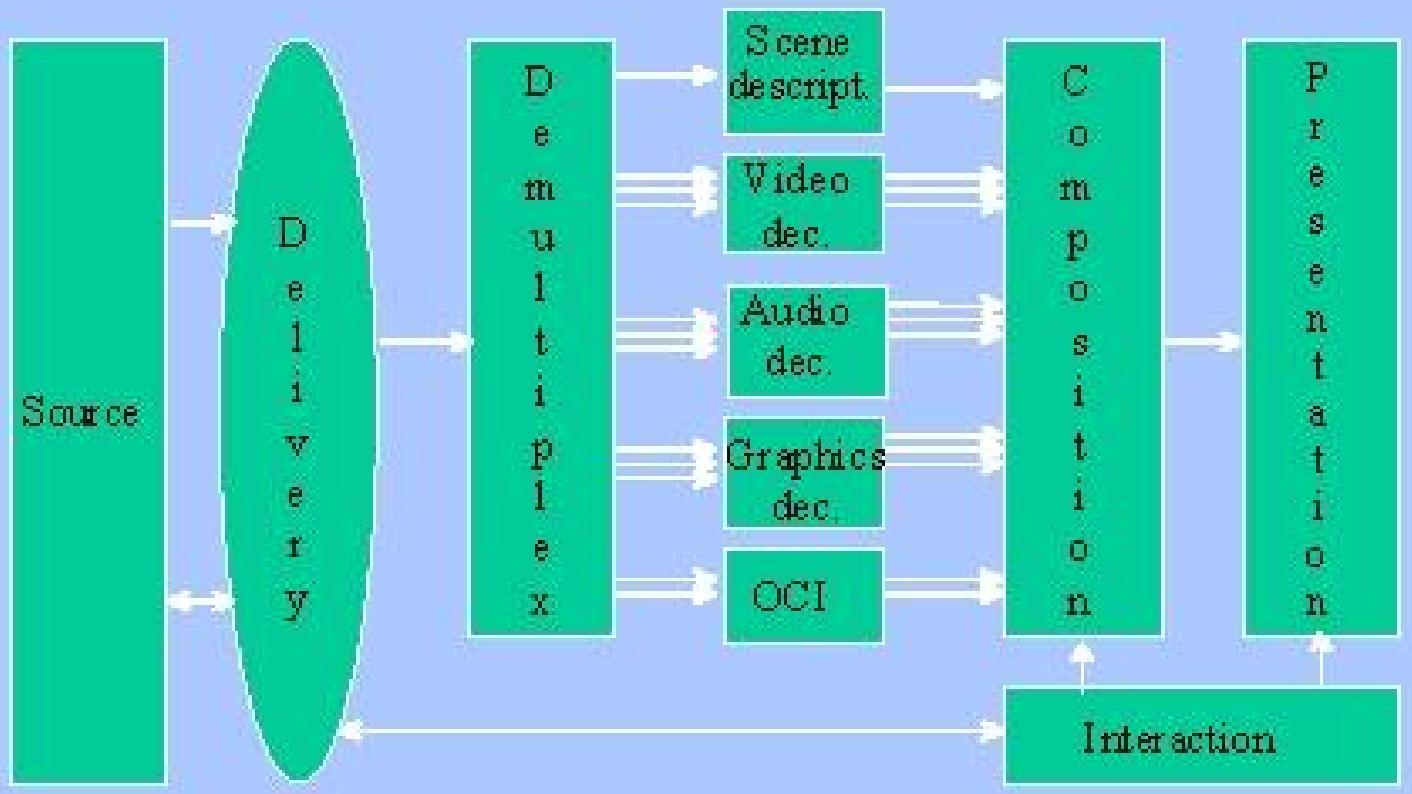
Object-based Audiovisual System



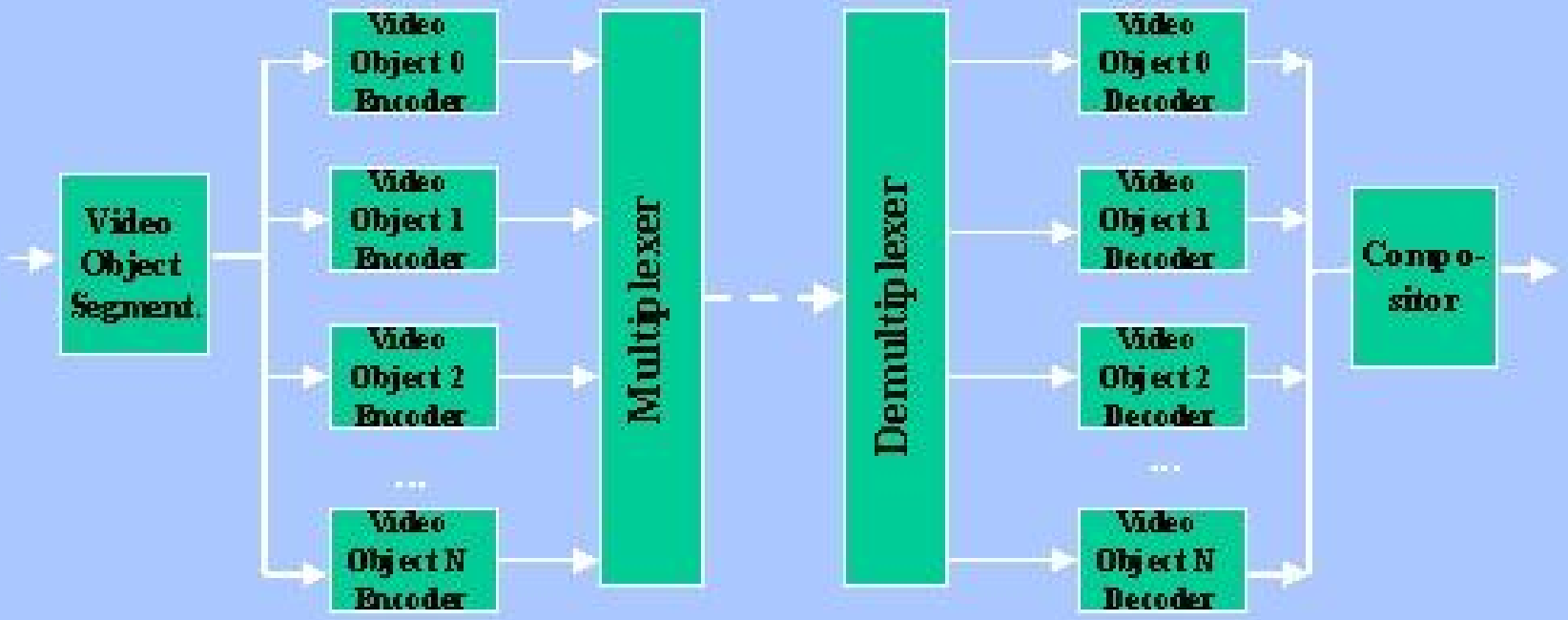
Object-based Audiovisual System

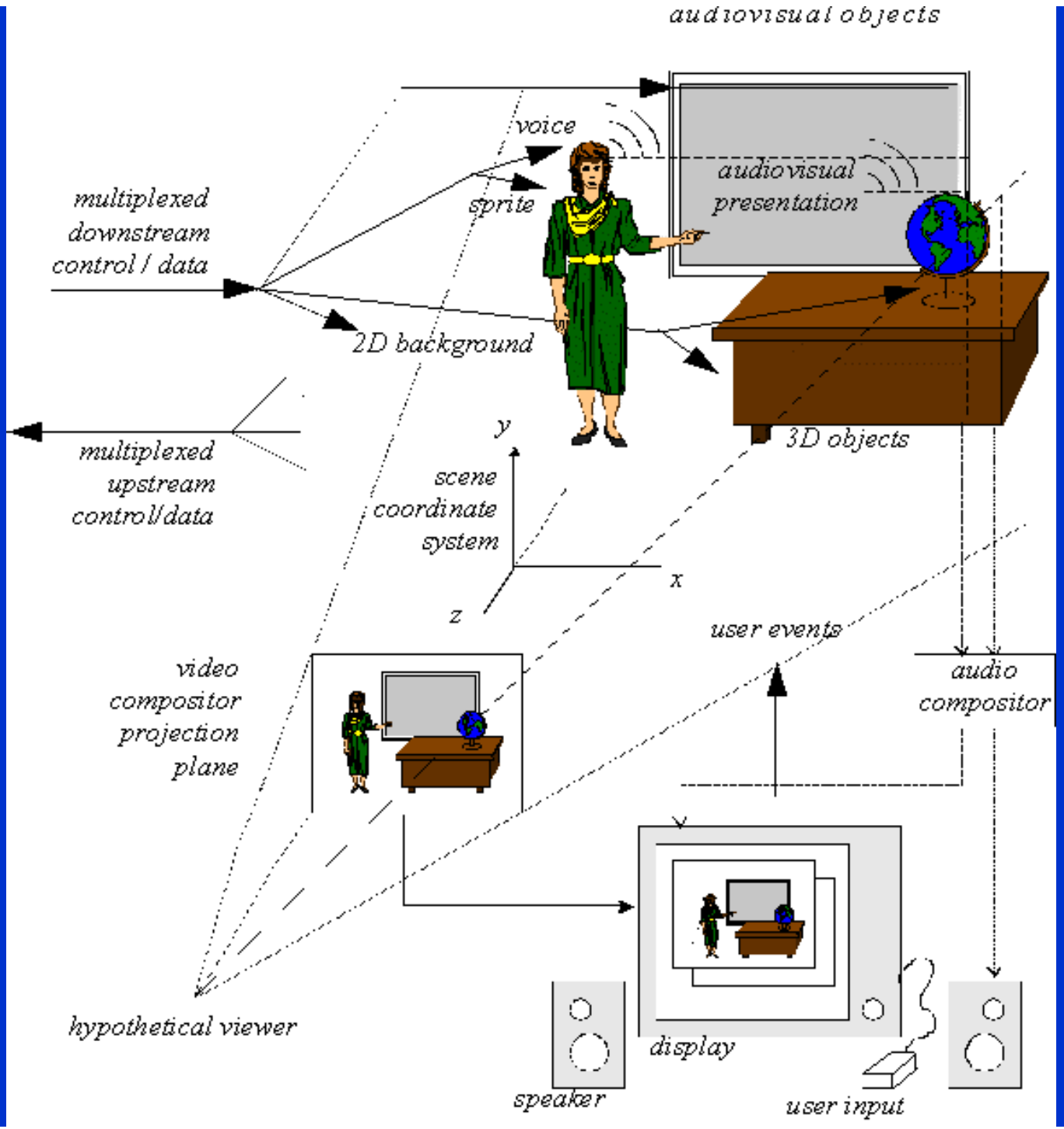


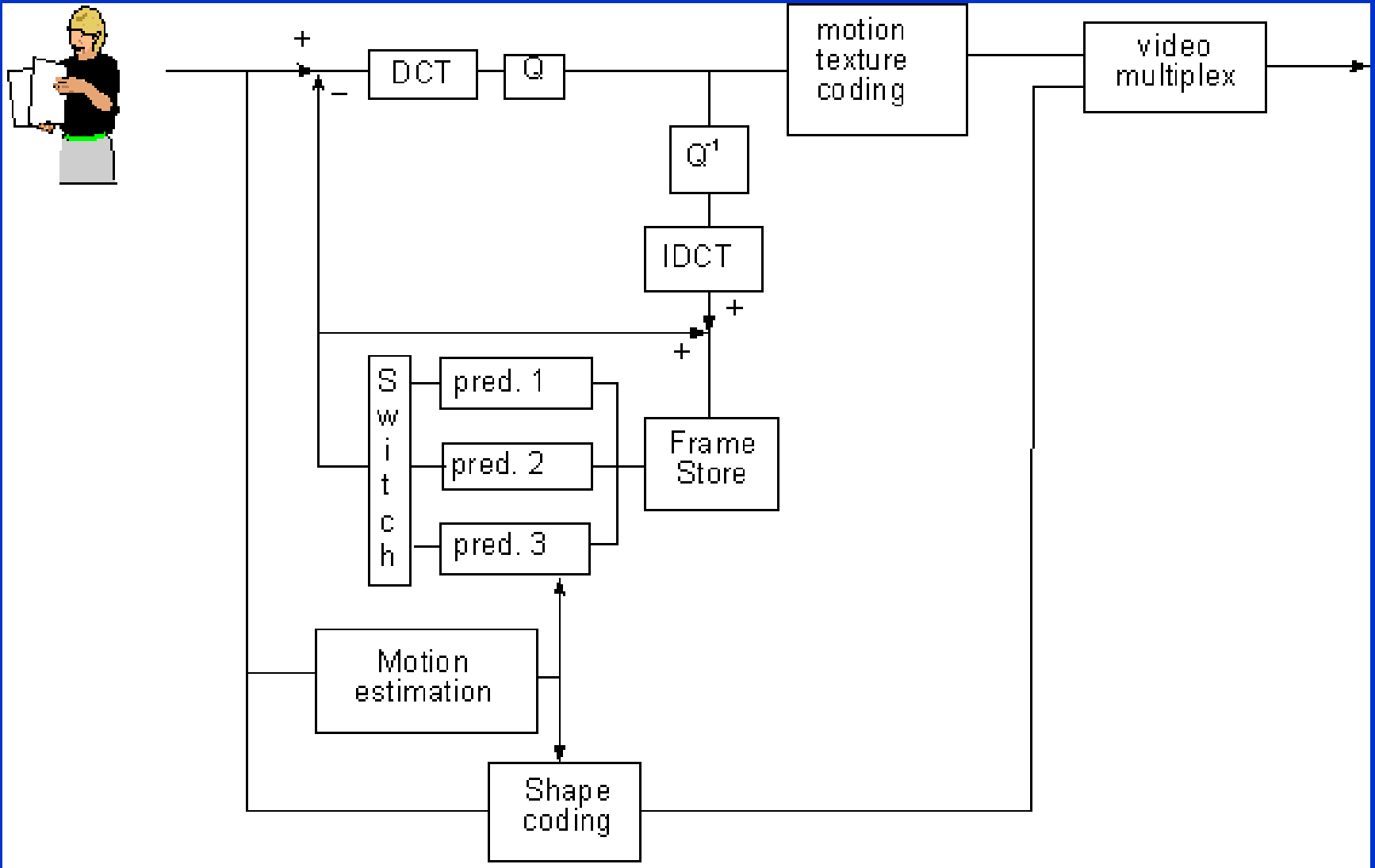
三、MPEG-4编解码系统



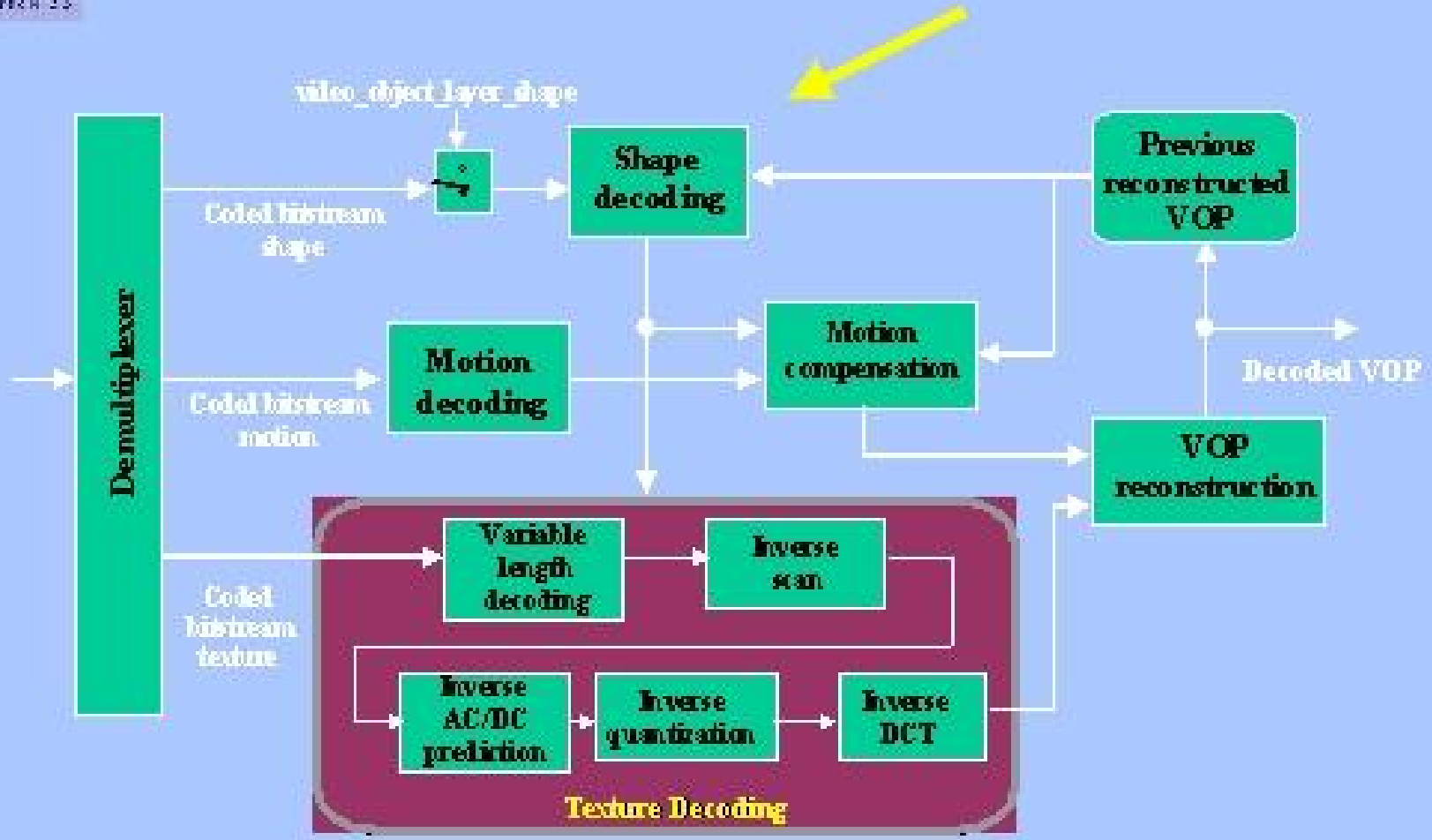
MPEG-4 Object-Based Video Coding







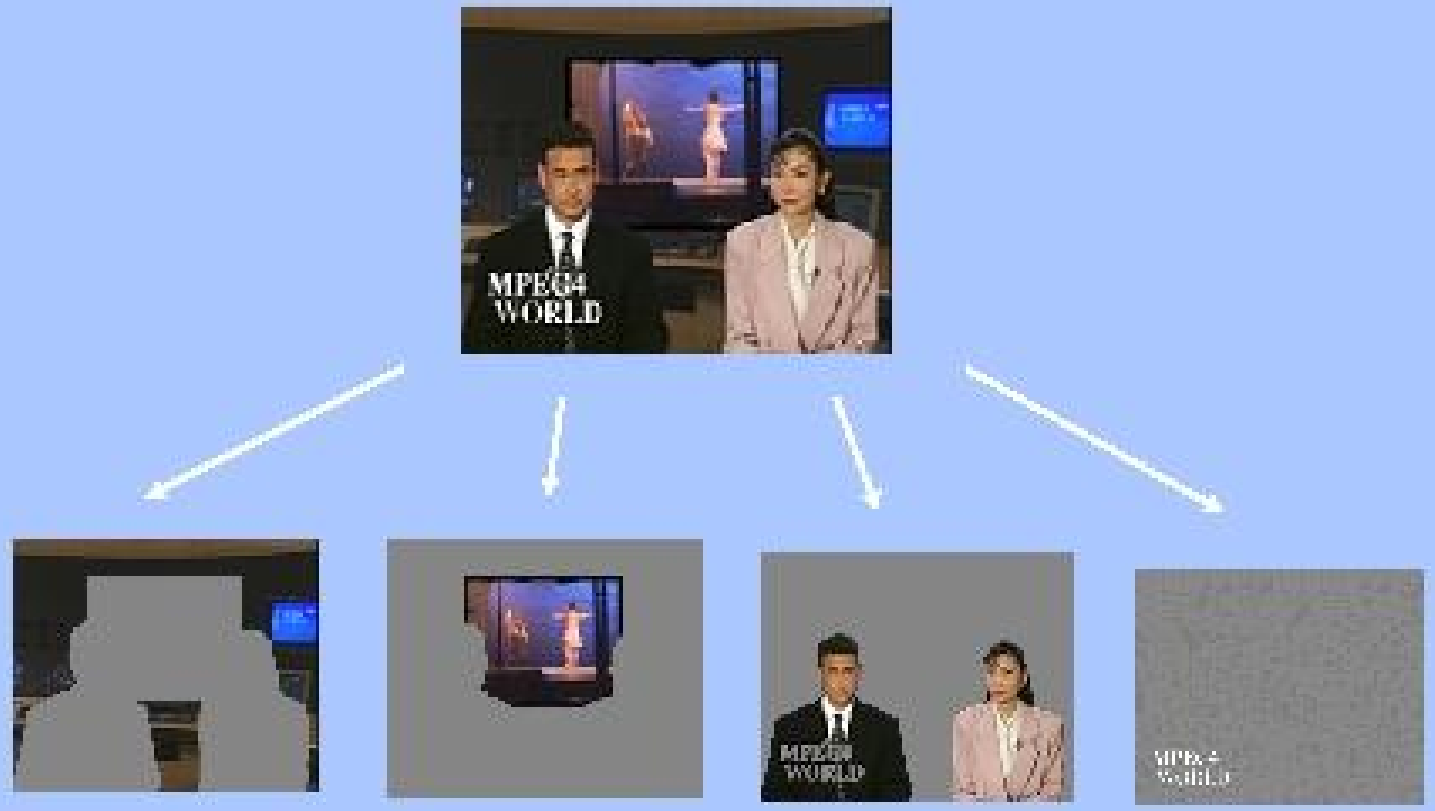
MPEG-4 Video (VOP) Decoding



四、MPEG-4的编码技术

- Video analysis for segmentation and object-based selective coding
- Object-based rate control, real-time and off-line
- Shape coding
- Object-based scalability, e.g. shape, spatial, temporal, SNR
- Object-based error resilience and concealment
- Object-based composition and multiplexing
- 2D and 3D synthetic objects coding
- Integration of natural and synthetic content
- Tools for object-based content creation, management and playback

1、分割Segmentation



Segmentation: The Way Around

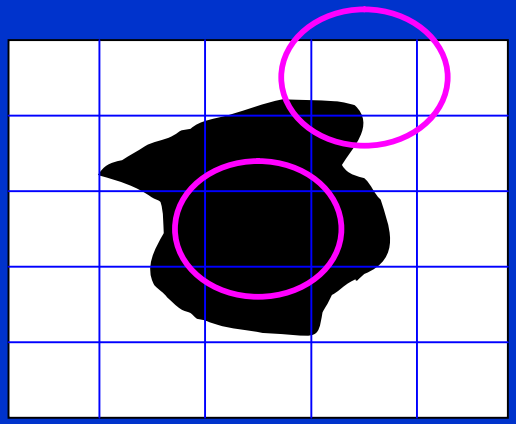
Segmentation: Real-time, Automatic Case !

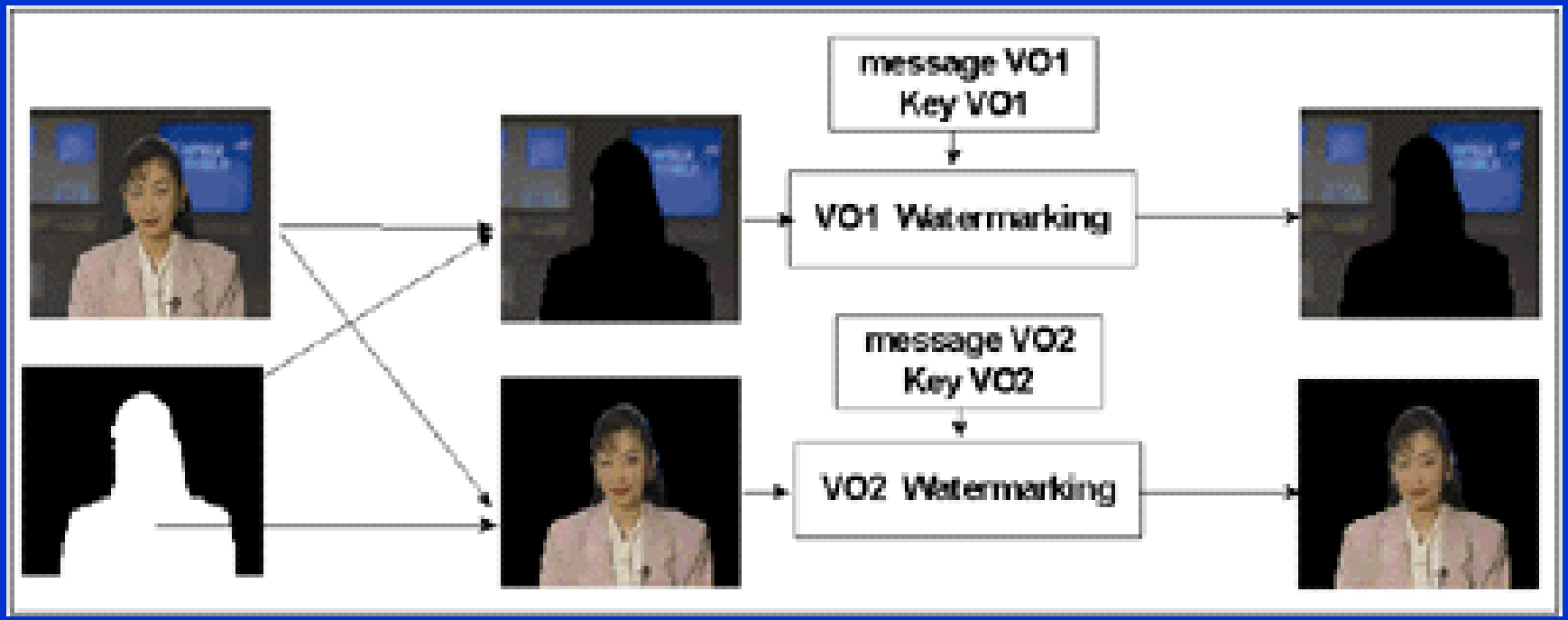
2、Synthetic Content

- **Model independent**
- **Model Animation**: performed by means of 68 Facial Animation Parameters (FAP)
- **Model Configuration**: performed by means of Facial Definition Parameters (FDP)

3、形状编码

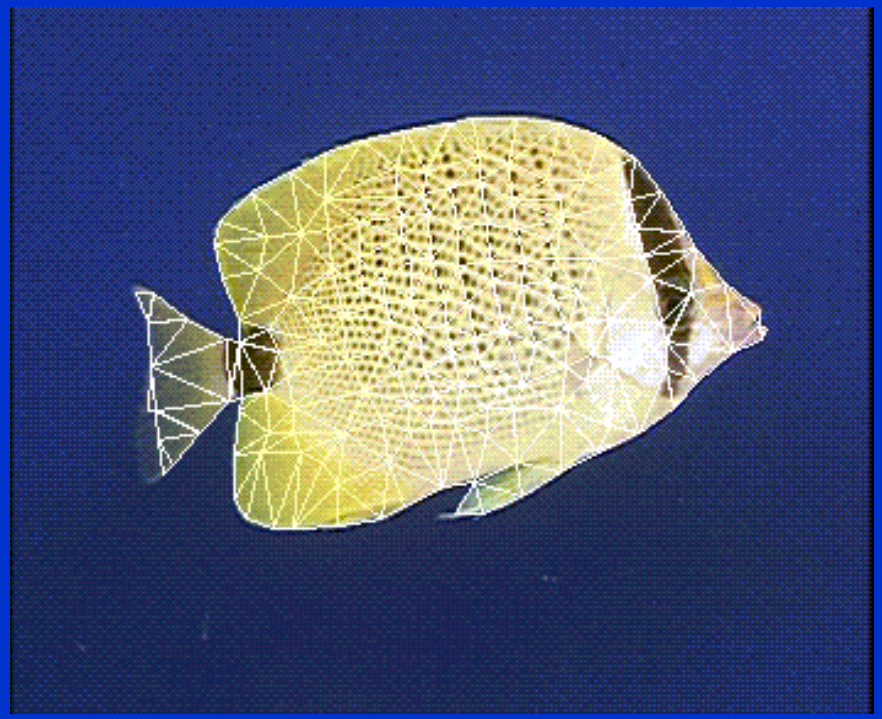
- VO形状信息：
- 二值；
- 灰度形状值；



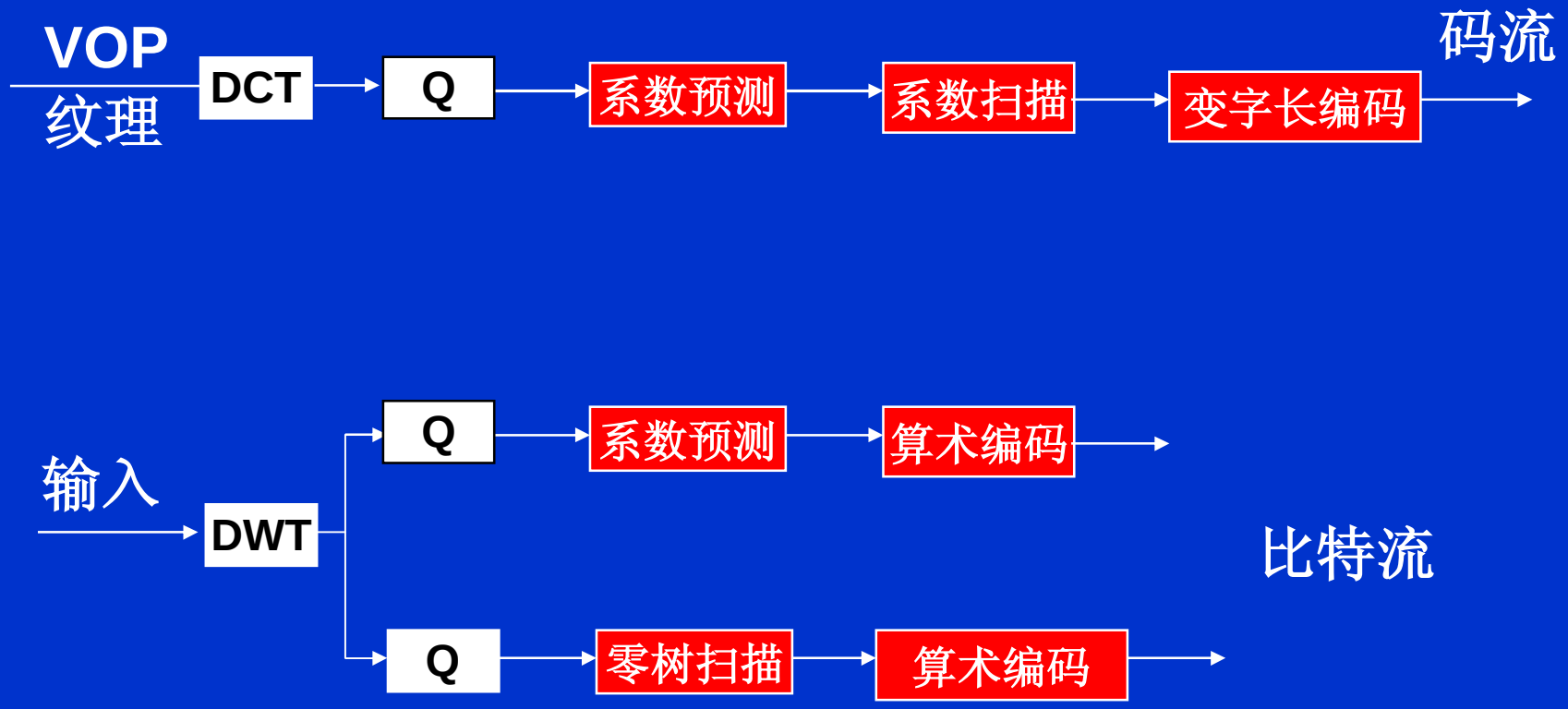


4、纹理编码

- 在帧内编码模式中,采用DCT变换
- 静态纹理编码基于小波变换和算术编码



4、纹理编码

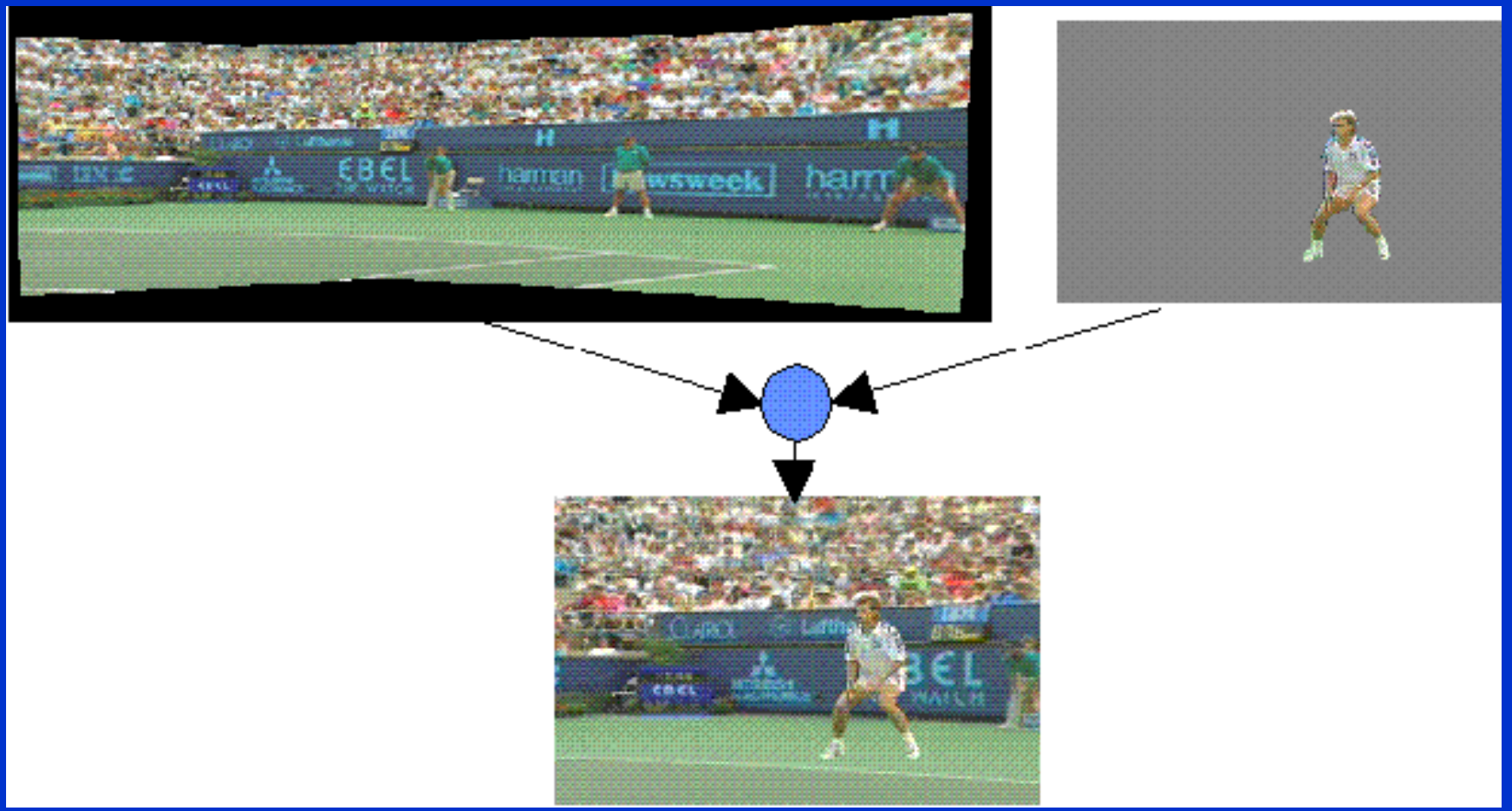


5、可分级编码

- 基于对象的可分级编码
- 空间
- 时间
- 图象的复杂性

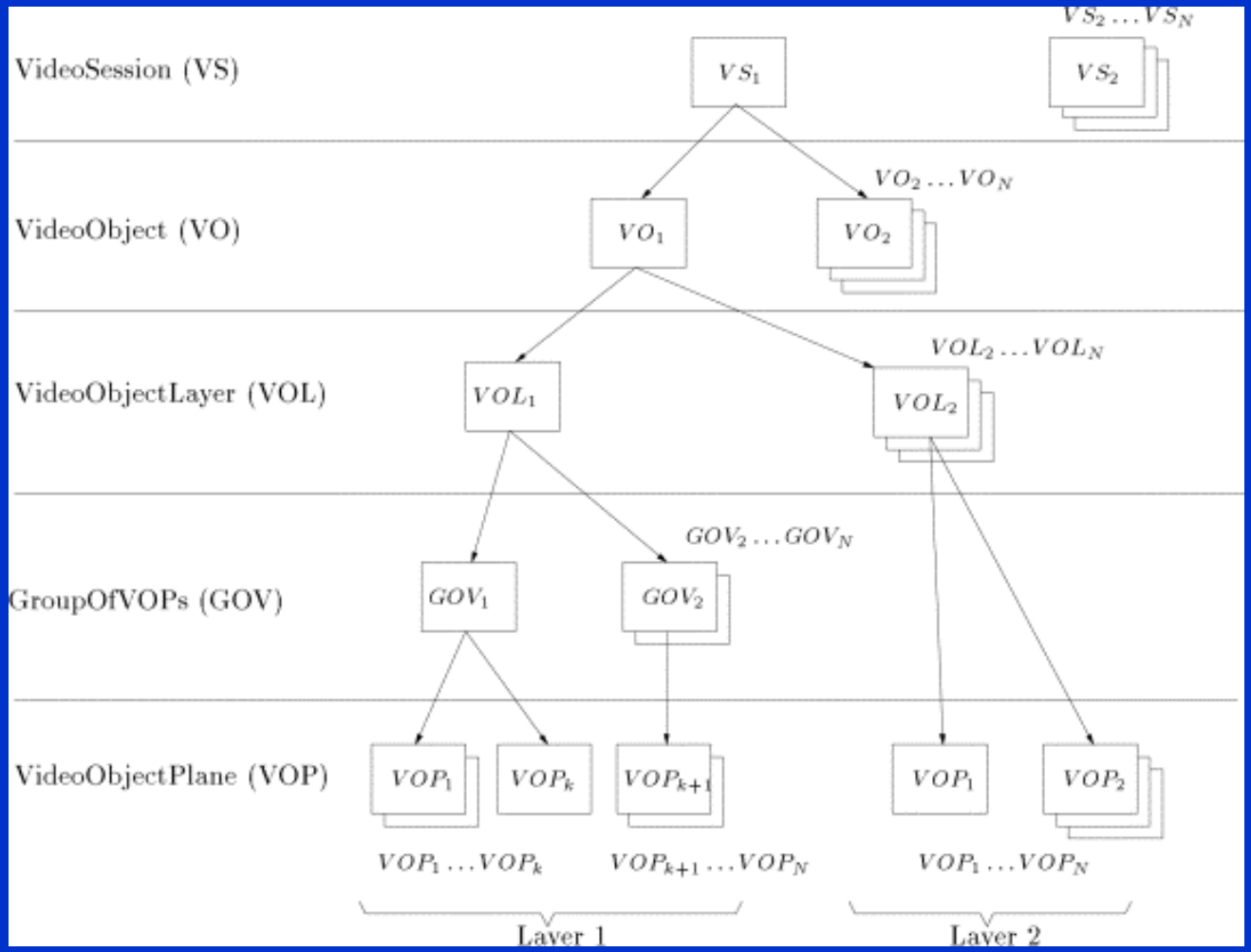
6、Sprite编码

- 指相对静止的长背景



五、MPEG-4视频

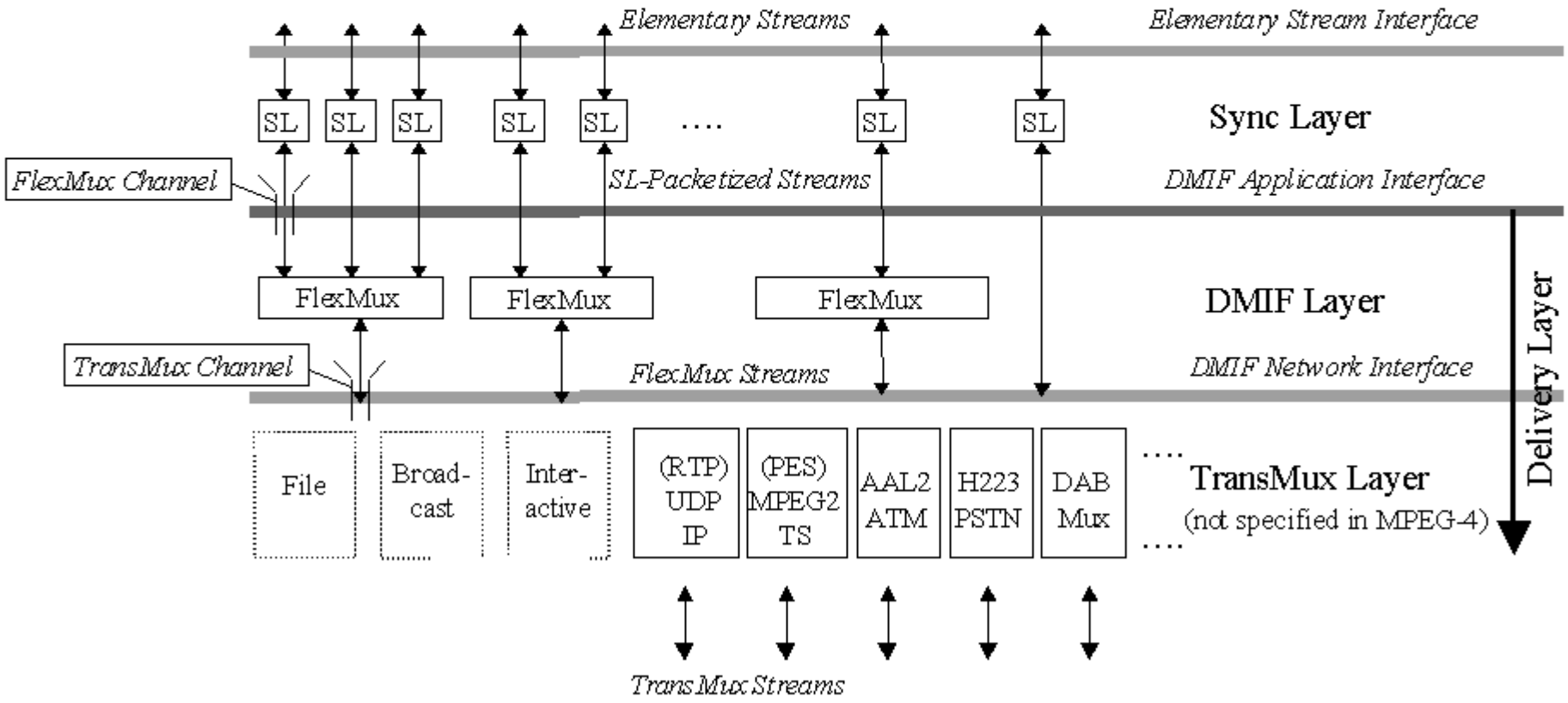
- 视频序列、视频对象、视频对象层、视频对象平面、视频包、宏块和块；



六、MPEG-4系统

- 解码器模型
- 场景描述
- 对象描述框架
- 基本码流同步层
- 基本码流的两层复用

MPEG-4 SYSTEM LAYER



七、MPEG-4容错性

- MPEG-4视频流语法结构

Resync Marker	MB Address	Quant Param	Header Extension	Temporal Reference	Shape Data	Motion Data	Motion Marker	Texture Data

1、重新同步标志

- MPEG-4在固定长度的数据流上插入同步标志。

起始码

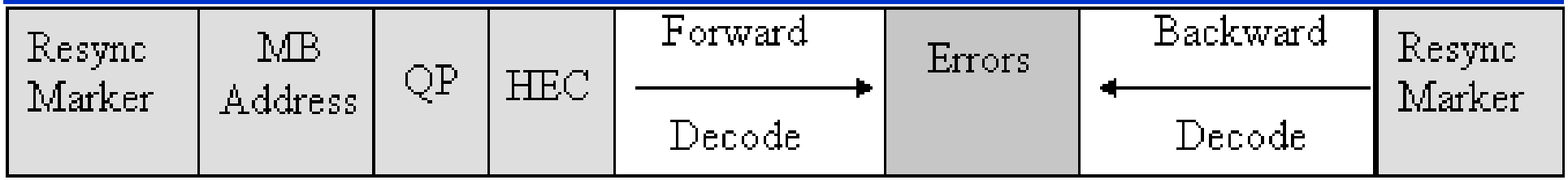
重新同步标志

2、数据分割

- 视频流中各数据在解码时的作用不一样。
- 将重要的数据与编码产生的数据分开，选择合适信道，使重要数据传输得到保证。

3、可反向编码的变长码

- 利用可反向解码的变长码确认错误位置，将错误的影响降到最低。



八、MPEG-4视频的类和级

- Profiles定义了MPEG-4标准中码流语法的子集,用于限制解码设备中采用的工具集;
- Level定义了码流中加在参数上的限制的集合,用于限制复杂性;

MPEG-4 Profiling: Basic Concepts

Profiles and levels are defined to allow easy interoperability and reduced complexity.

Two types of profiles are defined:

- **MEDIA PROFILES** - Visual, Audio and Graphics - specify the syntax/semantics for the coding of the media objects.
- **SCENE DESCRIPTION PROFILES** specify the syntax/semantics for the description the scene structure.

MPEG-4 Profiling: the Organization