



让我们越走越近...

《多媒体通信》

第三章 多媒体同步机制

2019年10月27日

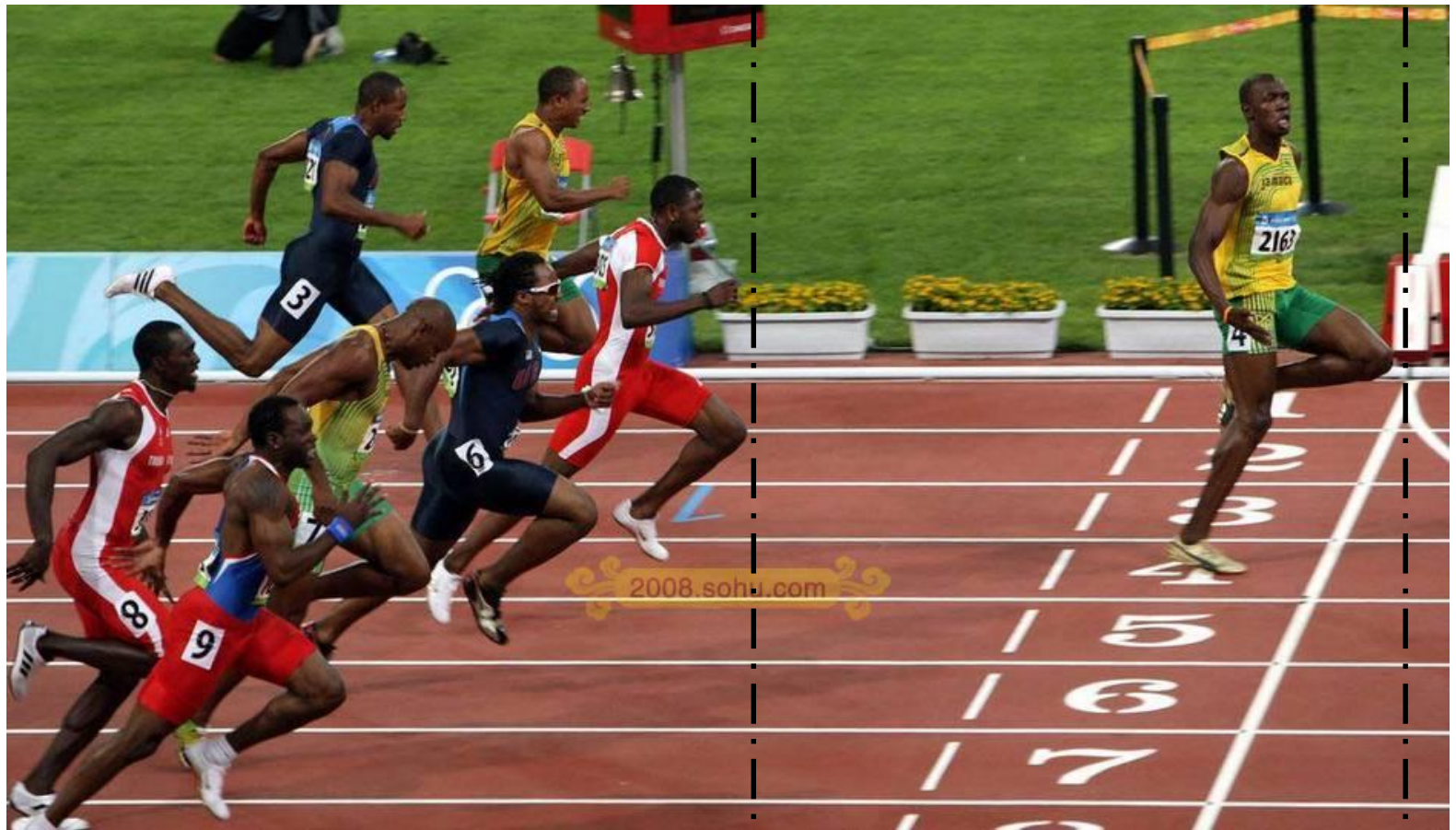


主要内容

- ◆ 什么是多媒体同步
 - 多媒体同步的定义与分类
 - 音视频的同步容限
- ◆ 如何描述多媒体同步
 - 基于Web的媒体时间关系描述
 - 基于Petri Net的多媒体时间域信息的描述
- ◆ 典型多媒体设备接口信号的同步机制
- ◆ 典型多媒体系统的媒体同步方法
- ◆ 展望未来



空间的同步还是时间的同步？



2008.sohu.com



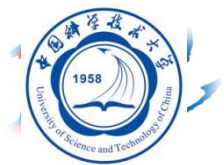
永恒的东风

(中国科学技术大学校歌)

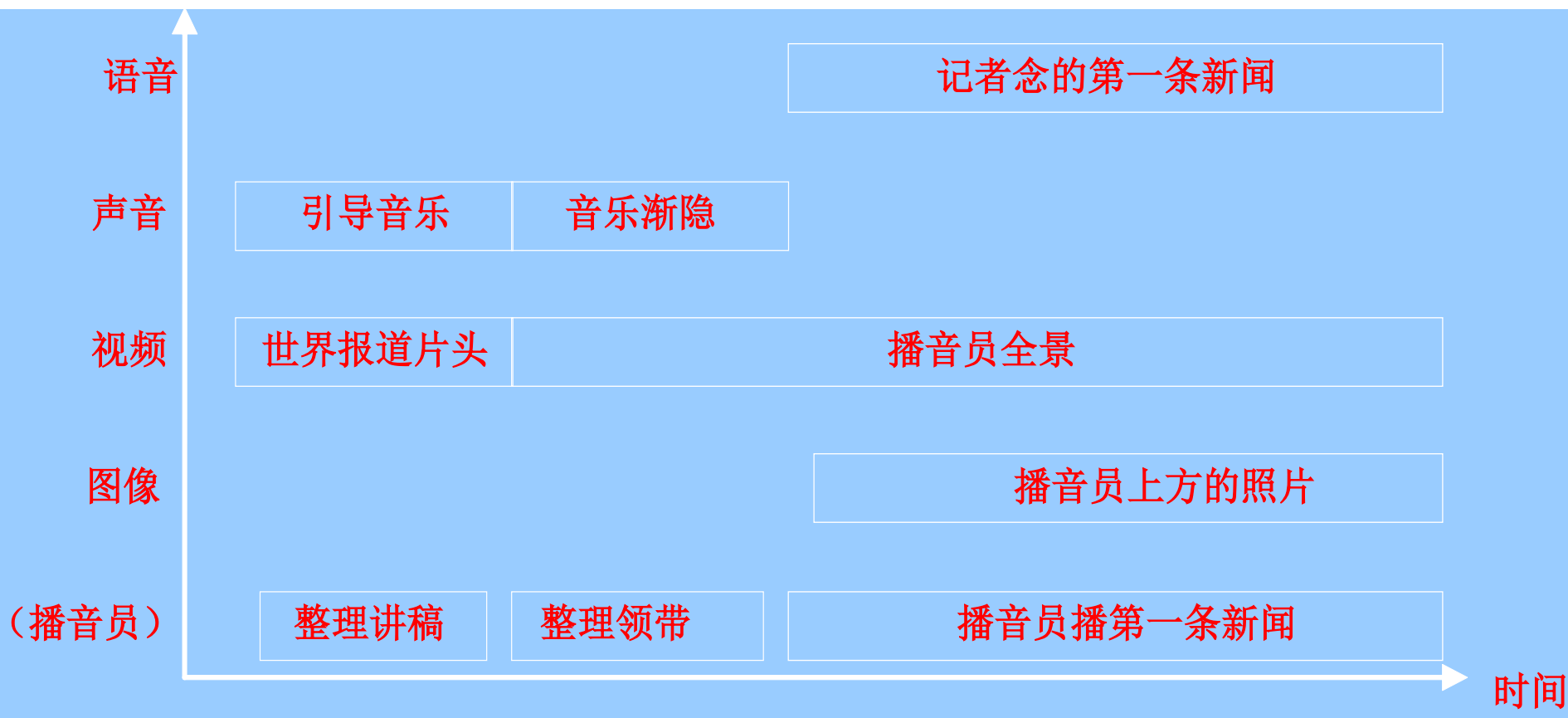
郭沫若词
吕驥曲

自由地，不太快

迎接 着 永 恒 的 东 风 把
 红 旗 高 举 起 来。
 插 上 科 学 的 高 峰。
 科 学 的 高 峰 在 不 断 创 造。
 高 峰 要 高 到 无 穷。
 红 旗 要 红 过 九 重。
 我 们 是 中 国 的 好 儿 女 要
 熬 苦 眼 炼， 辛 勤 劳 动， 在
 党 的 温 暖 抚 育， 坚 强 领 导 下， 为
 共 产 主 义 事 业 作 先 锋。
 又 红 又 专 理 实 文 融。
 团 结 互 助， 活 泼 英 勇。
 永 远 向 人 民 学 习， 学 习
 伟 大 领 袖 毛 泽 东。



多媒体同步示例：新闻





不同媒体对象的同步关系

- ◆ **空间关系**用来描述在终端显示或者输出时不同的媒体对象之间在空间域上的位置关系。
- ◆ **内容关系**是指用不同的表现形式来表示相同的数据。例如，一份数据有时候既可以用表格又可以用图表表示，那么在计算机中不管是用表格、图表或者曲线等方式表示的都是同一份数据内容。
- ◆ **时间关系**是用来描述时间相关媒体媒体流内或者媒体流之间各个数据单元的时间关系的，我们**狭义上的多媒体同步**主要就是指的这种时间同步。



多媒体同步的定义/分类

◆ 广义上的同步（Synchronization）

- 从广义上讲，同步是指多媒体系统中各个媒体对象之间存在的时间、空间、以及逻辑关系。在分布式多媒体通信中，同步是一个基于时间的概念，是指多媒体系统中各个媒体对象之间存在的的关系，同步的目的就是向用户展示多媒体信息时，保持媒体对象之间固有的时间关系。

◆ 狭义的同步分类

- 流内同步（Intra-stream Synchronization）是指在单一连接上的单一媒体流内各个媒体单元之间存在的的关系。
- 流间同步（Inter-stream Synchronization）是指在多个连接上的相关媒体流之间存在的的关系。
- 人机交互同步用于保持用户与多点传送媒体流之间的同步播映速度。



流内同步

一个 25 帧/s 的视频流，它每帧间隔的时间是 40ms

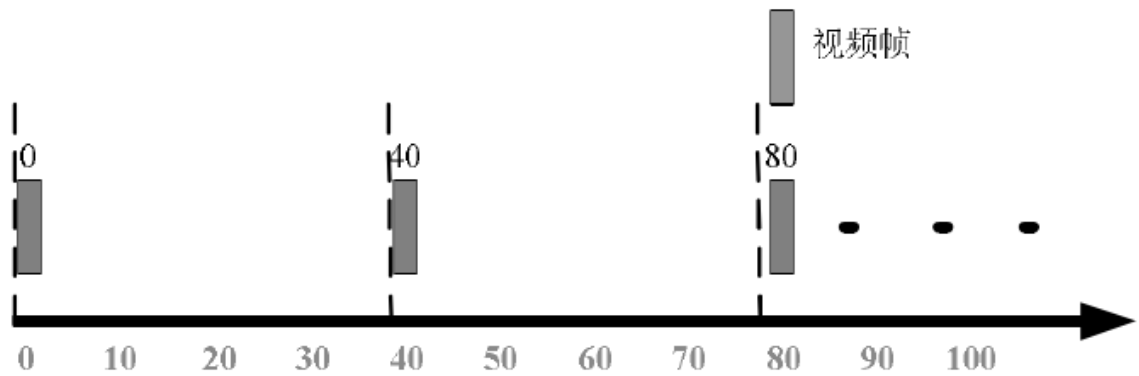


图 2-4 视频流内同步

一个用 G.729 编码的音频流，它每帧间隔时间是 10ms

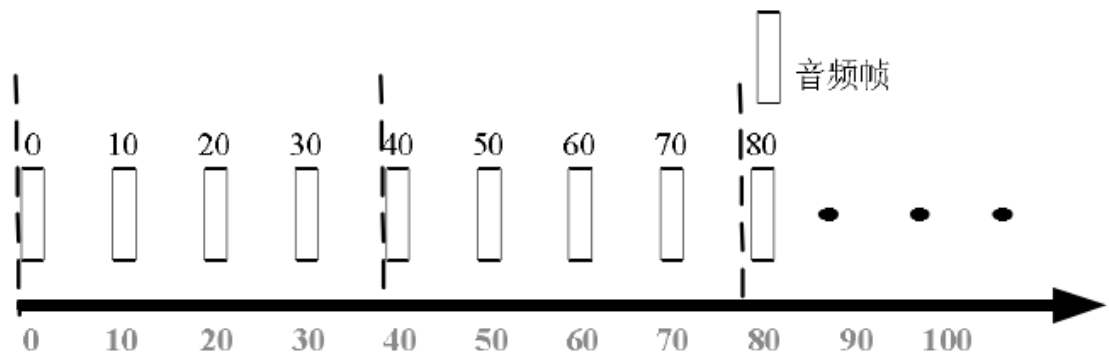


图 2-5 音频流内同步

MC 流间同步

不同步的原因：
采集端产生延时
网络传输变化
播放端产生的时间偏差

媒体间同步是指维持媒体相互之间的时间关系，其目标是消除各媒体流之间的时间偏移。

如果33.3ms产生1个视频帧，10ms产生1个音频帧，情况如何？

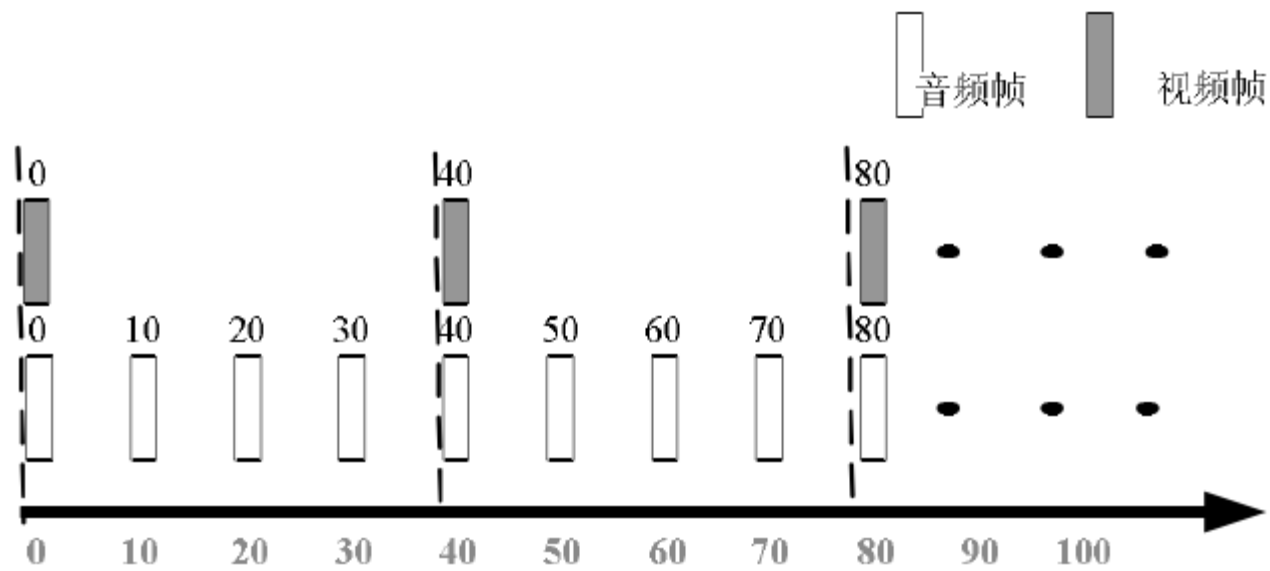


图 2-6 音频流间同步

- ◆ 人机交互同步用于保持用户与多点传送媒体流之间的同步播映速度。
- ◆ 在用户和多媒体系统之间的交互实质就是交互对象，即用户和计算机系统之间的通信。例如当用户利用输入设备控制多媒体系统的跳转、停止、变速和暂停等，系统应该及时响应用户的操作。
- ◆ 在交互系统中，用户对于鼠标、键盘等外部设备的操作必须要及时响应。例如，实时运动广播和交互式多媒体视频会议等。这类同步的特点在于同步的时间随着指令的时间所决定，因为同步发生时间具有不可预测性。

MC 常用的同步方法

◆ 复用同步法

- 将发送端的音视频数据流复用到一个数据流中，从而使音视频数据流在网络传输过程中自然保持着媒体间的相互关系，在接收端，按照音视频流的这一关系解复用，达到同步目的。

◆ 同步标记法

- 在媒体发送端发出一个同步标记，在接收端接收同步标记，用来对各个媒体进行同步处理，达到两方通信同步的目的。

◆ 时间戳同步法

- 把音视频数据流单元分别添加统一的时间信息（例如时间戳），然后尽可能的同时发送出去，保证发送端的同步。在接收过程中提取音视频单元的时间信息，根据该时间信息对音视频数据的播放先后进行调整，例如音频滞后者，就将视频播放速度减慢；如果音频超前，则加快视频的播放速率，若音频超前过多，甚至可考虑丢弃些视频帧。



可观察到的时间量

- ◆ TI DSP(C6455)中断响应: 4 μ s
- ◆ Windows线程启动: 1-10ms
- ◆ 音频压缩的样本序列长度: 5ms/10ms/20ms
- ◆ 视频中一帧图像的产生时间: 30ms-40ms
- ◆ 100M以太网传输700万像素JPEG图像: 100ms
- ◆ 网络视频播放的缓冲时间: 20s-60s

问题：多媒体同步需要的时间精度？

- ◆ 什么是多媒体同步
 - 多媒体同步的定义与分类
 - 音视频的同步容限
- ◆ 如何描述多媒体同步
 - 基于Web的媒体时间关系描述
 - 基于Petri Net的多媒体时间域信息的描述
- ◆ 典型多媒体设备接口信号的同步机制
- ◆ 典型多媒体系统的媒体同步方法
- ◆ 展望未来

MC 同步容限

- ◆ CPU抢占、缓冲区不大传输带宽不足等因素往往会导致在恢复后的时域场景中时域事件间的相对位置发生变化，我们将这种变化称为**事件间偏差(SKEW)**。
- ◆ 属于同一媒体对象的时域事件之间的偏差称为**对象内偏差**，不同媒体对象的时域事件之间的偏差为**对象间偏差**。
- ◆ **同步容限包含了用户对偏差许可范围的定义**，同步机制则需依据同步容限，保证恢复场景中事件间偏差在其许可范围内。



常见多媒体业务对传输网络的要求

要求	最大延迟 /s	最大延迟 抖动/ms	速率 /Mbps	平均吞吐 量/Mbps	可接受 误码率	可接受分 组差错率
话音	0.25	10	0.005-0.064	0.064	$<10^{-2}$	$<10^{-2}$
视频图像	0.25	10	100-1000	100	10^{-2}	10^{-3}
静止图像	1	/	2-10	2-10	10^{-4}	10^{-9}
压缩后视频 图像	0.25	1	0.02-10	2-10	10^{-6}	10^{-9}
数据文件	1	/	2-100	2-100	0	0
实时数据	0.001-1	/	<10	<10	0	0



同步容限与内容有关

◆ 人对于视频/音频之间同步程度的感知与视频内容有关



(a)

(b)

(c)

Fig. 1. (a) Head view; (b) shoulder view; (c) body view.

<http://www.kom.tu-darmstadt.de/en/people/staff/ralf-steinmetz/>

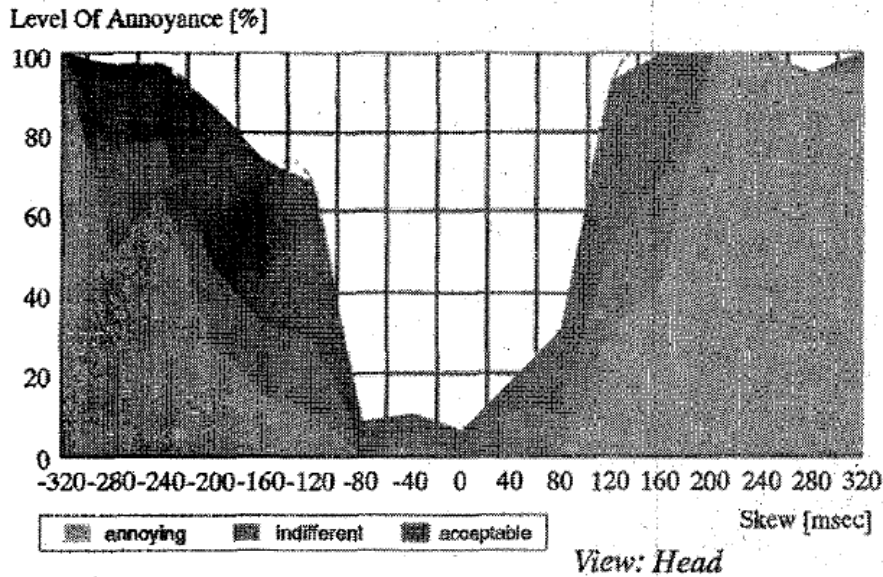
Human Perception of Jitter and Media Synchronization

IEEE JOURNAL on Selected Areas on Communications, Vol. 14, No. 1, Jan 1996

Google Scholar显示引用550次(2016.11)

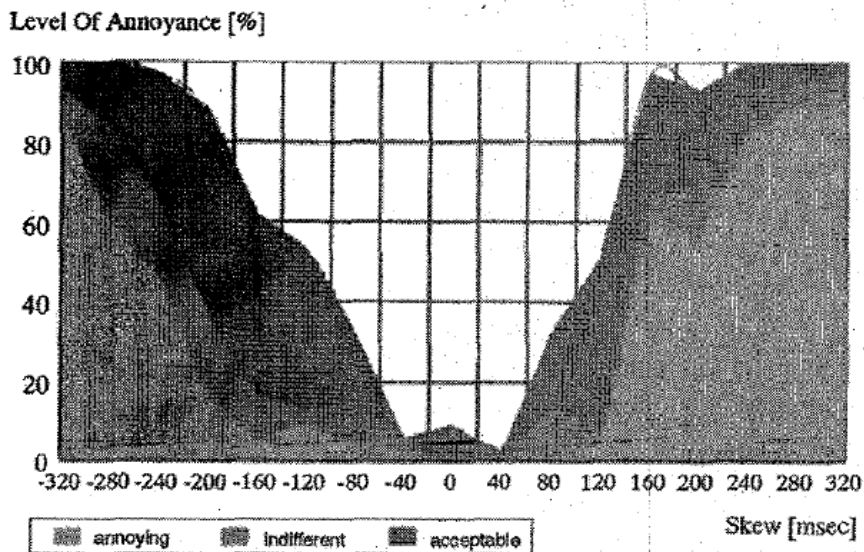


Level of Annoyance



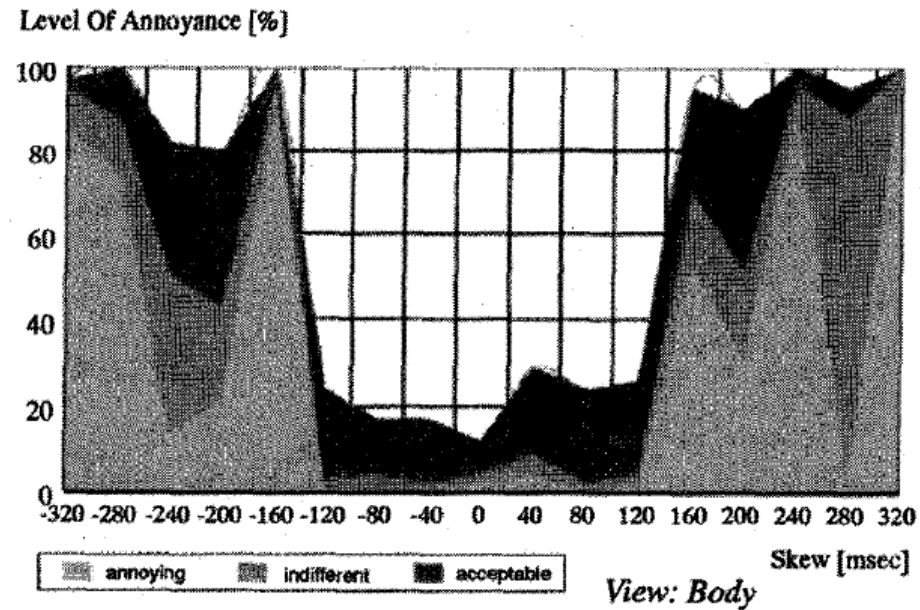
View: Head

Fig. 17. Level of Annoyance at head view—Left part: negative skew, video ahead of audio; right part: positive skew, video behind audio.



View: Shoulder

Fig. 18. Level of Annoyance at shoulder view—Left part: negative skew, video ahead of audio; right part: positive skew, video behind audio.



View: Body

Fig. 19. Level of Annoyance at body view—Left part: negative skew, video ahead of audio; right part: positive skew, video behind audio.

Correctly Detected Error

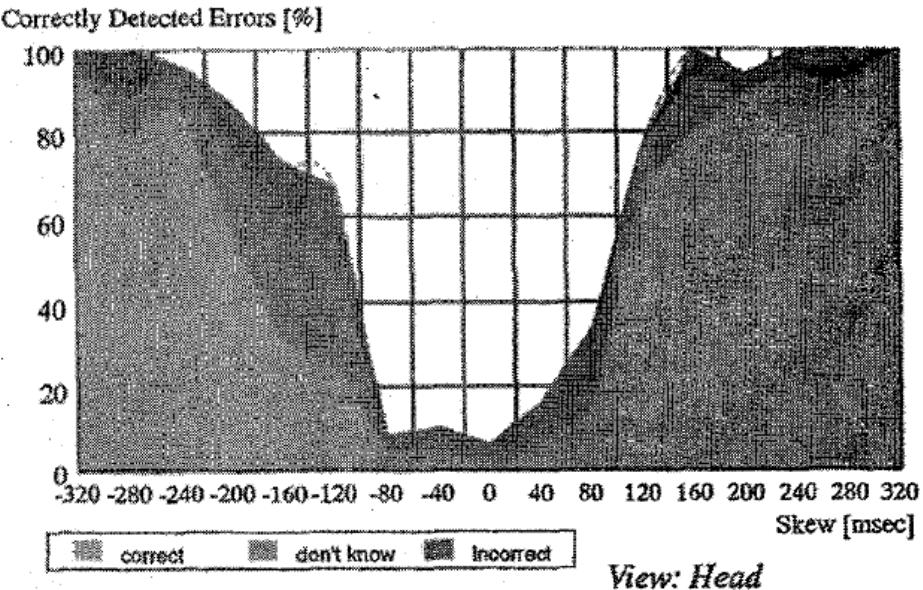


Fig. 14. Correct detection of synchronization errors at head view—Left part: negative skew, video ahead of audio; right part: positive skew, video behind audio.

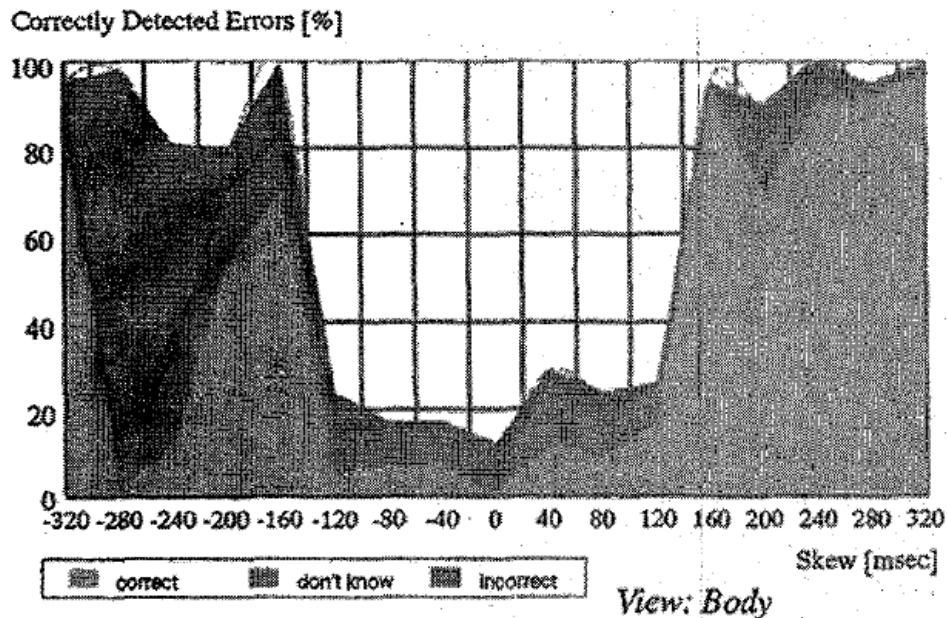


Fig. 16. Correct detection of synchronization errors at body view—Left part: negative skew, video ahead of audio; right part: positive skew, video behind audio.

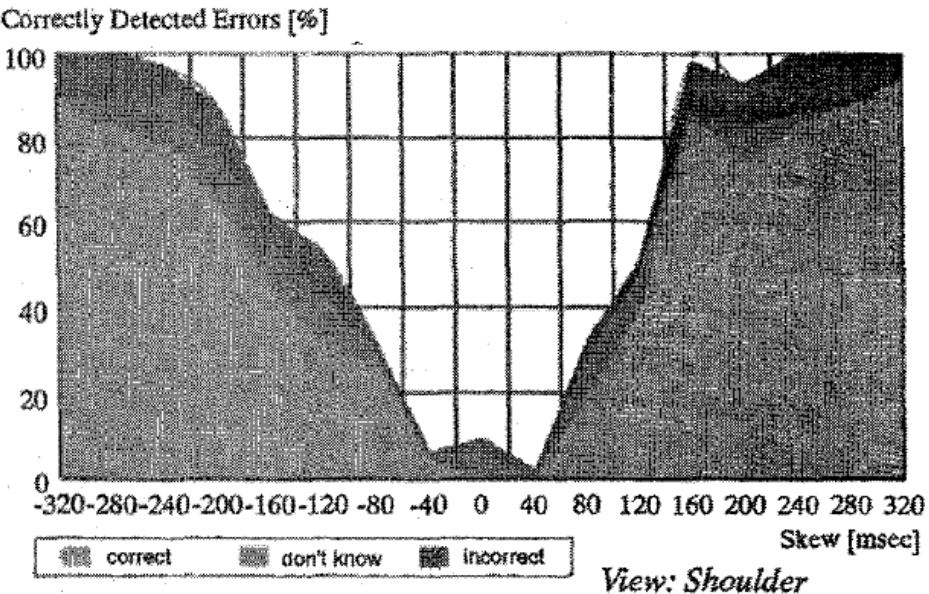


Fig. 15. Correct detection of synchronization errors at shoulder view—Left part: negative skew, video ahead of audio; right part: positive skew, video behind audio.



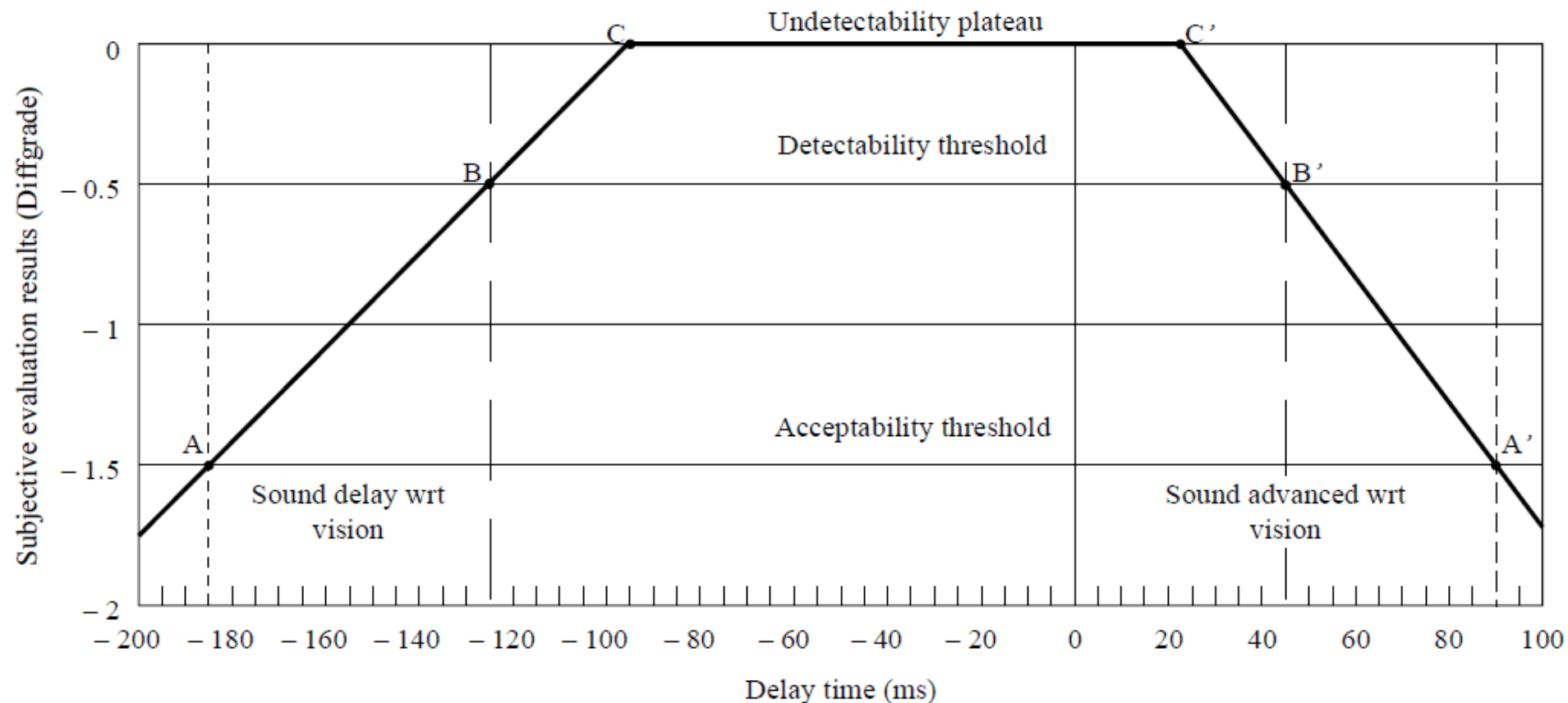
TABLE I
QUALITY OF SERVICE FOR SYNCHRONIZATION PURPOSES

Media		Mode, Application	QoS
video	animation	correlated	+/- 120 ms
	audio	lip synchronization	+/- 80 ms
	image	overlay	+/- 240 ms
		non overlay	+/-500 ms
	text	overlay	+/- 240 ms
		non overlay	+/-500 ms
audio	animation	event correlation (e.g. dancing)	+/- 80 ms
	audio	tightly coupled (stereo)	+/- 11 μ s
		loosely coupled (dialog mode with various participants)	+/- 120 ms
		loosely coupled (e.g. background music)	+/- 500 ms
	image	tightly coupled (e.g. music with notes)	+/- 5 ms
		loosely coupled (e.g. slide show)	+/- 500 ms
	text	text annotation	+/- 240 ms
	pointer	audio relates to showed item	-500 ms, + 750 ms ¹



RECOMMENDATION ITU-R BT.1359 [1998] Relative timing of sound and vision for broadcasting

◆ Detectability and acceptability thresholds



skew should be kept within +22.5 ms and -30 ms

positive skew: video²² behind audio



Skew in standards

◆ ITU BT.1359	1998	-30 ms	+22.5 ms
◆ ATSC IS/191	2003	-45 ms	+15 ms
◆ EBU R37	2007	-60 ms	+40 ms

◆ ITU: International Telecommunication Union

◆ EBU: European Broadcasting Union

◆ ATSC: Advanced Television Systems Committee

ITU-R BT.1359 1[1998] Relative timing of sound and vision for broadcasting

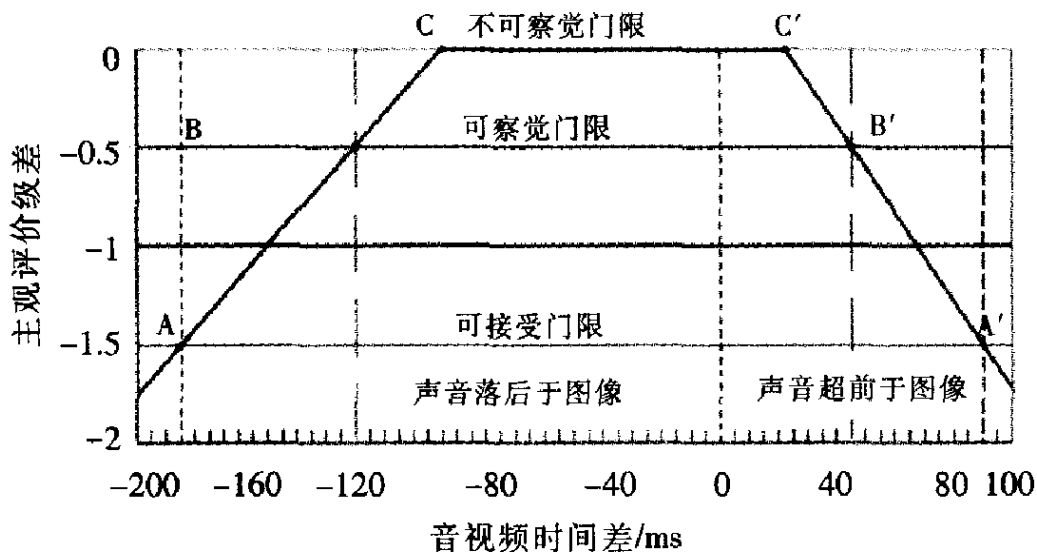
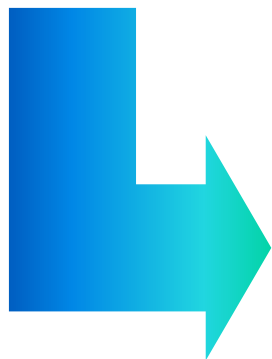
ATSC IS-191[2003] Relative Timing of Sound and Vision for Broadcast Operations

EBU R037[2007]



小结：典型音视频同步问题

- ◆ 数据流1：图像帧率是30fps，即每隔33.3ms产生一个完整的图像数据块
- ◆ 数据流2：音频采样频率是44.1k，一般每20ms的数据作为一个传输数据块



音视频同步需要的时间精度已有结论，实际系统中如何描述？

- ◆ 什么是多媒体同步
 - 多媒体同步的定义与分类
 - 音视频的同步容限
- ◆ 如何描述多媒体同步
 - 基于Web的媒体时间关系描述
 - 基于Petri Net的多媒体时间域信息的描述
- ◆ 典型多媒体设备接口信号的同步机制
- ◆ 典型多媒体系统的媒体同步方法
- ◆ 展望未来



What is SMIL

◆ SMIL 是同步多媒体集成语言（Synchronized Multimedia Integration Language）的缩写，念做smile。参与制定该标准的公司包括Compaq、Gateway、IBM、Microsoft、RealNetworks、Oratrix等等。SMIL与我们网页上用的HTML（超文本传输语言）的语法格式非常相似。后者主要针对普通的网络媒体文件进行操纵（文字、图片、声音、动画、视频的机械堆砌），而前者则操纵多媒体片断（对多媒体片断的有机的、智能的组合）。

D.C.A. Bulterman, 'SMIL 2.0-Part 1: Overview, Concepts, and Structure,' IEEE MultiMedia, vol. 8, no. 4, pp. 2001, 82-88.

D.C.A. Bulterman, 'SMIL 2.0-Part 2: Examples and Comparisons,' IEEE MultiMedia, vol. 9, no. 1, 2002, pp. 74-84.



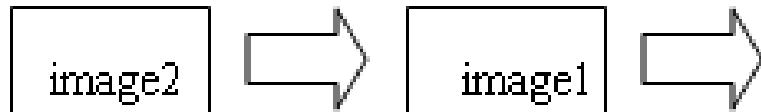
用SMIL语言连续播放rm片段

```
<smil>
  <head>
    <layout>
      <region id="a" left="0" top="0" />
      <!-- 定义区域a, 使其左上角紧挨播放窗口左上角 -->
    </layout>
  </head>
  <body>
    <seq> <!-- 定义播放序列 -->
      <video region="r" src="shushan1.rm" />
      <video region="r" src="shushan2.rm" />
      <video region="r" src="shushan3.rm" />
    </seq>
  </body>
</smil>
```

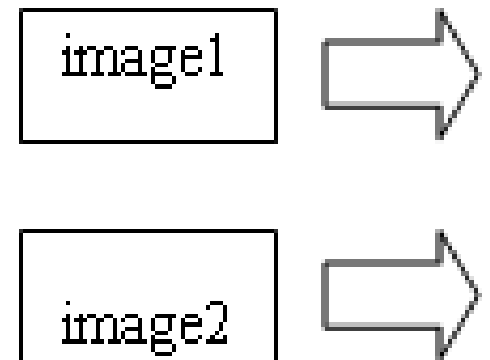


多媒体片段结构组织

```
<seq>  
  
  
</seq>
```



```
<par>  
  
  
</par>
```



MC 智能流 (surestream)

```
<switch>  
<vedio src="high.rm" system-bitrate="250000"/>  
<vedio src="mid.rm" system-bitrate="80000"/>  
<vedio src="low.rm" system-bitrate="20000"/>  
</switch>
```

◆ 当用户的联网速度大于250kbps时，播放器就从服务器下载high.rm播放；如果用户的联网速度大于80kbps小于250kbps时，播放器就从服务器下载mid.rm播放；如果用户的联网速度大于20kbps小于80kbps时，播放器就从服务器下载low.rm播放。



主要内容

- ◆ 什么是多媒体同步
 - 多媒体同步的定义与分类
 - 音视频的同步容限
- ◆ 如何描述多媒体同步
 - 基于Web的媒体时间关系描述
 - 基于Petri Net的多媒体时间域信息的描述
- ◆ 典型多媒体设备接口信号的同步机制
- ◆ 典型多媒体系统的媒体同步方法
- ◆ 展望未来

◆ 时间关系

- before, starts, meets, overlaps, during, finishes, equals, 除 equals外, 其余六种都有逆关系, 共十三种

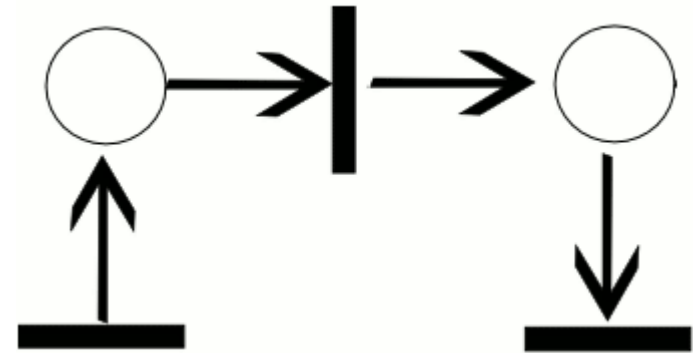
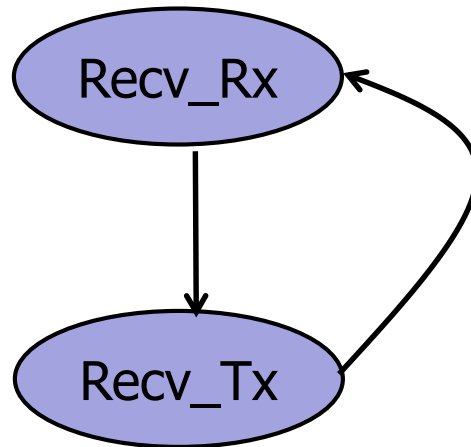
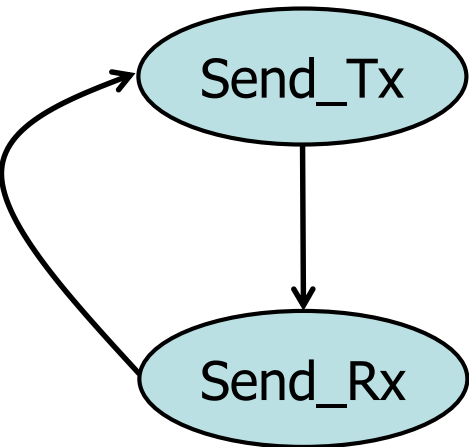
◆ 时间规范

- 绝对/相对、并行/顺序

◆ 基于图形的表示法

- 时间线、流图、Petri网、时间层次法
- 最著名的是OCPN方法

- ◆ Petri网是对离散并行系统的数学表示。1962年由C.A. Petri 发明，适合于描述**异步**的、**并发**的计算机系统模型。有严格的数学表述方式和直观的图形表达方式。
- ◆ Petri nets were invented in August 1939 by Carl Adam Petri – at the age of 13 – for the purpose of describing chemical processes.
- ◆ **对比：状态图无法表示多个状态的同时变化**

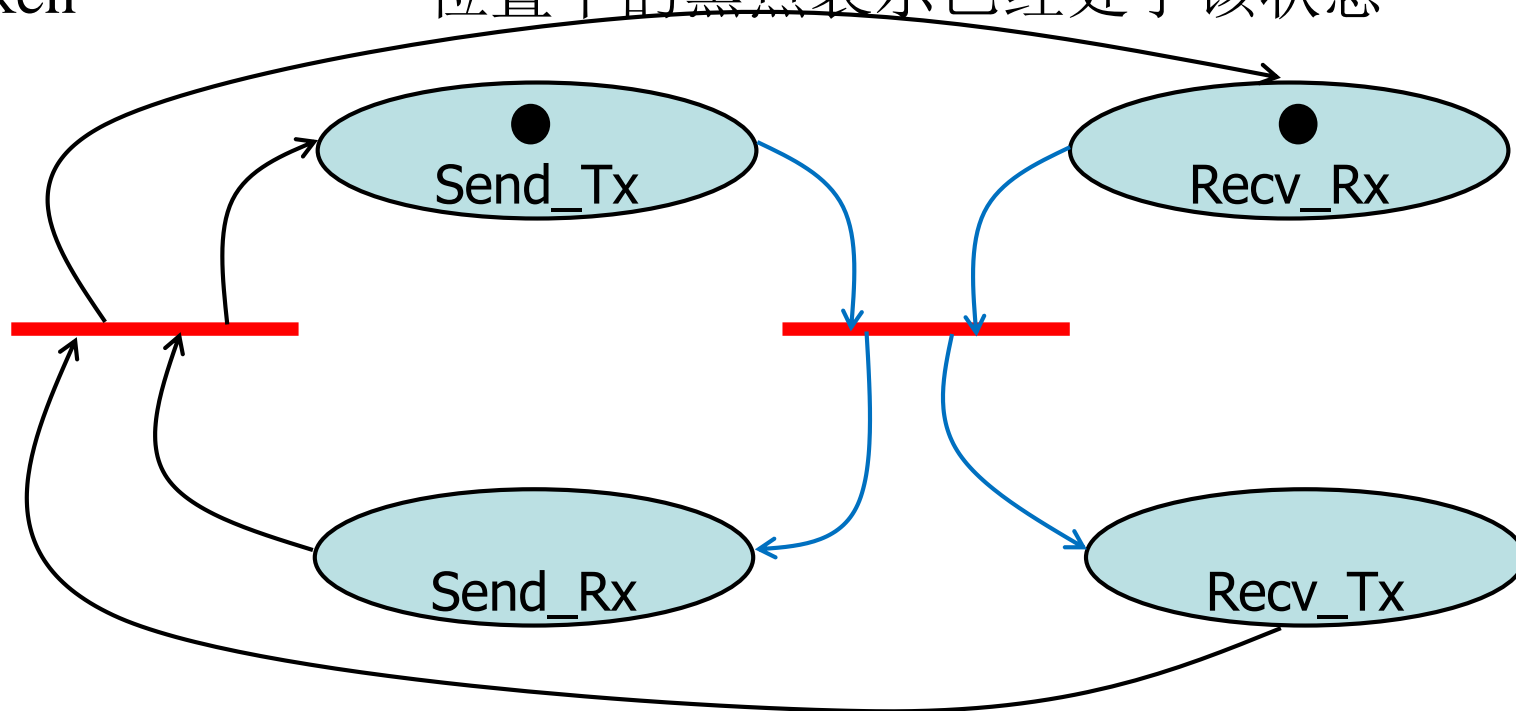




Petri网的基本要素

Petri net (known as place/transition net or P/T net)

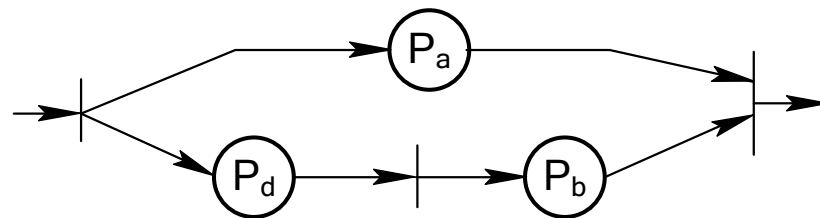
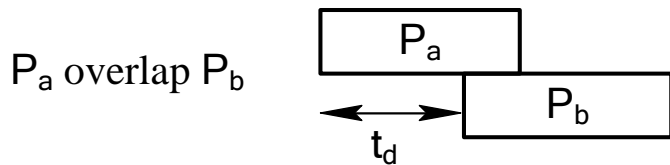
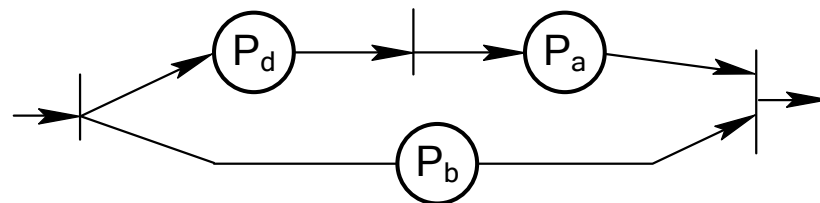
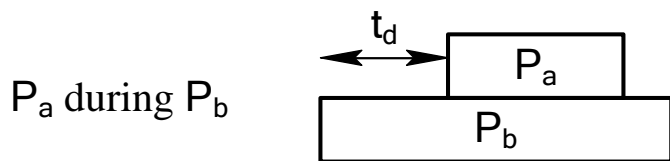
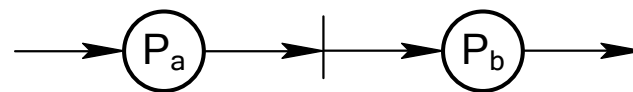
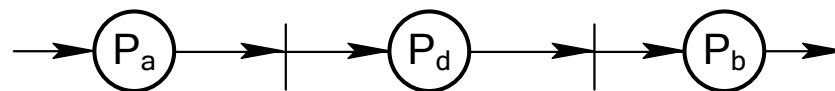
- ◆ Places 位置相当于状态图中的状态，圆圈表示
- ◆ Transitions 变迁，用直线表示
- ◆ Arcs 弧线表示变迁发生的条件和产生的结果
- ◆ Token 位置中的黑点表示已经处于该状态





用Petri网表示同步关系

◆ Petri网可用来表示同步关系





About Petri Net

Petri Net \rightarrow **MPN** (Marked PetriNet)

Transitions

Places

Arcs

Markings

Petri Net: $C_{PN} = \{T, P, A\}$

$T = \{t_1, t_2, \dots, t_n\},$

$P = \{P_1, P_2, \dots, P_m\},$

$A: \{T \times P\} \cup \{P \times T\} \rightarrow I, \quad I = \{1, 2, \dots\}$

MPN (Marked PetriNet): $C_{MPN} = \{T, P, A, M\}$

$M: P \rightarrow I' \quad , \quad I' = \{0, 1, 2, \dots\}$

M 定义了由位置集向整数集的映射，它表示集中的标记(Token)分布



About Petri Net

Petri Net \rightarrow **OCPN**(Object Composite Petri Net)

Petri Net: $C_{PN} = \{T, P, A\}$

$T = \{t_1, t_2, \dots, t_n\},$

$P = \{P_1, P_2, \dots, P_m\},$

$A: \{T \times P\} \cup \{P \times T\} \rightarrow I, \quad I = \{1, 2, \dots\}$

Transitions
Places
Arcs

MPN (Marked PetriNet): $C_{MPN} = \{T, P, A, M\}$

$M: P \rightarrow I', \quad I' = \{0, 1, 2, \dots\}$

Markings
Duration
Resource

OCPN(Object Composite Petri-Net)

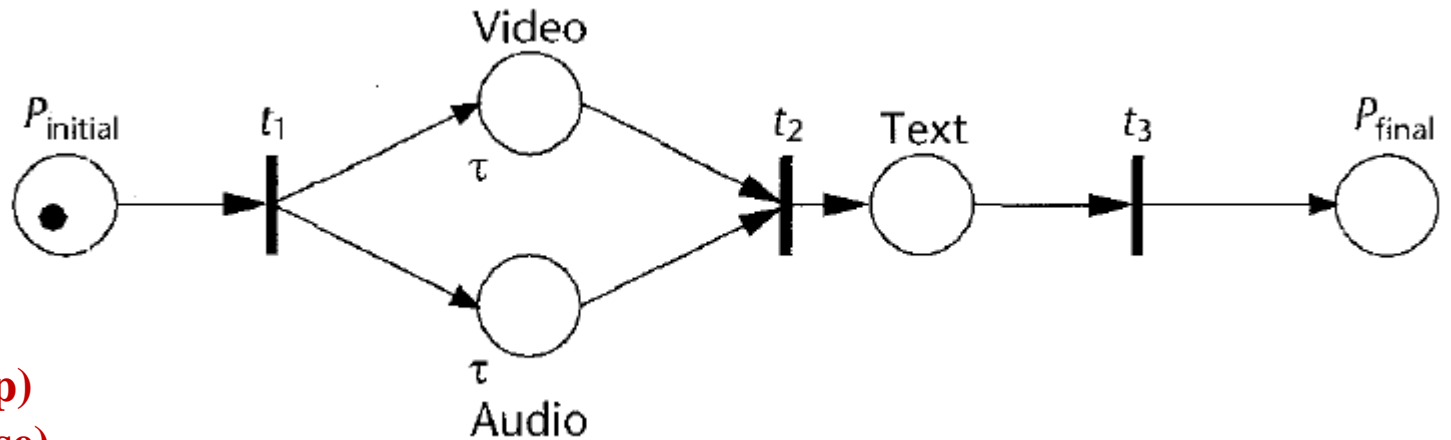
$C_{OCPN} = \{T, P, A, D, Re, M\}$

$D: P \rightarrow R$ **表示OCPN中的时间约束**

$R: P \rightarrow \{r_1, r_2, r_3, \dots, r_k\}$ **由位置集向资源集的映射**



OCPN(Object Composite Petri Net)

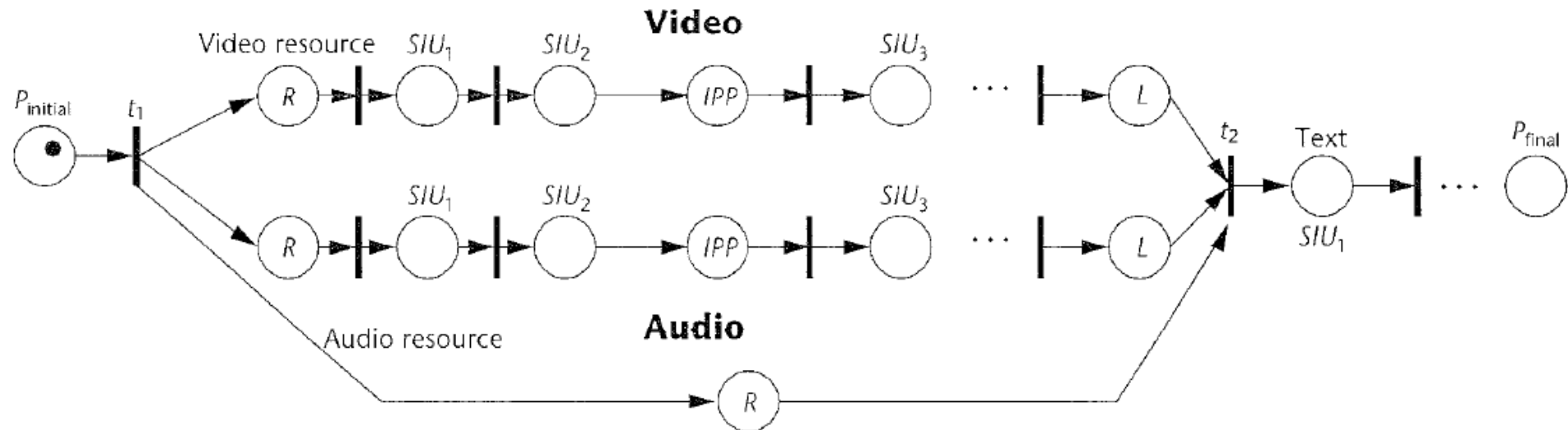


R (resource setup)

L(resource release)

IPP (inter-stream pacing point) represent control objects

SIU: places represent synchronization internal units



MC 小结：如何描述同步

◆ 时间关系

- before, starts, meets, overlaps, during, finishes, equals, 除 equals外, 其余六种都有逆关系, 共十三种

◆ 多媒体时间域信息的描述

- SMIL: 基于Web的媒体时间关系描述
- 描述并发的状态变化: Petri Net
 - OCPN(Object Composite Petri Net)

MC 主要内容

- ◆ 什么是多媒体同步
- ◆ 如何描述多媒体同步
- ◆ 典型多媒体设备接口信号的同步机制
 - CD-DA 音频的流内同步
 - 模拟电视信号的行同步与场同步
 - ITU-R BT. 601 数字视频图像帧的同步
 - HDMI 数字多媒体接口的同步
- ◆ 典型多媒体系统的媒体同步方法
 - MP3流同步
 - DVD/HDTV的流同步
 - 基于RTP的流媒体同步
- ◆ 展望未来



CD-DA每扇区持续时间1/75秒

◆螺旋型的物理光道分成多个逻辑光道。一条逻辑光道由许多扇区(sector)组成，一扇区由98帧(frame)组成。

◆CD的采样频率为44.1 kHz，每次对左右声音通道各取一个16位的样本，1秒钟的声音数据率就为

$$\square 44.1 \times 1\,000 \times 2 \times (16 \div 2) = 176\,400 \text{ 字节/秒}$$

◆由于1帧存放24字节的聲音数据，所以1秒钟所需要的帧数为

$$\square 176\,400 \div 24 = 7350 \text{ 帧/秒}$$

◆98帧构成1个扇区，所以1秒钟所需要的扇区数为

$$\square 7350 \div 98 = 75 \text{ 扇区/秒}$$

4字节的扇区地址称为HEADER			
分(MIN)	秒(SEC)	分秒(FRAC)	方式(Mode)
1字节	1字节	1字节	1字节
0~74	0~59	0~74	01

MC 主要内容

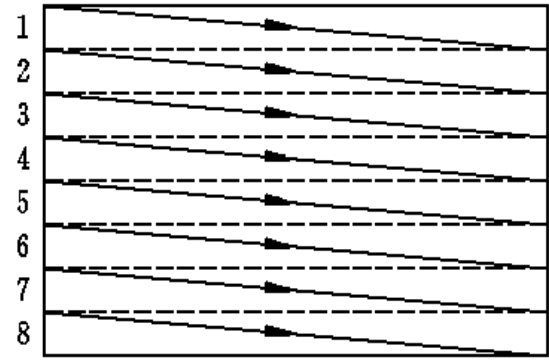
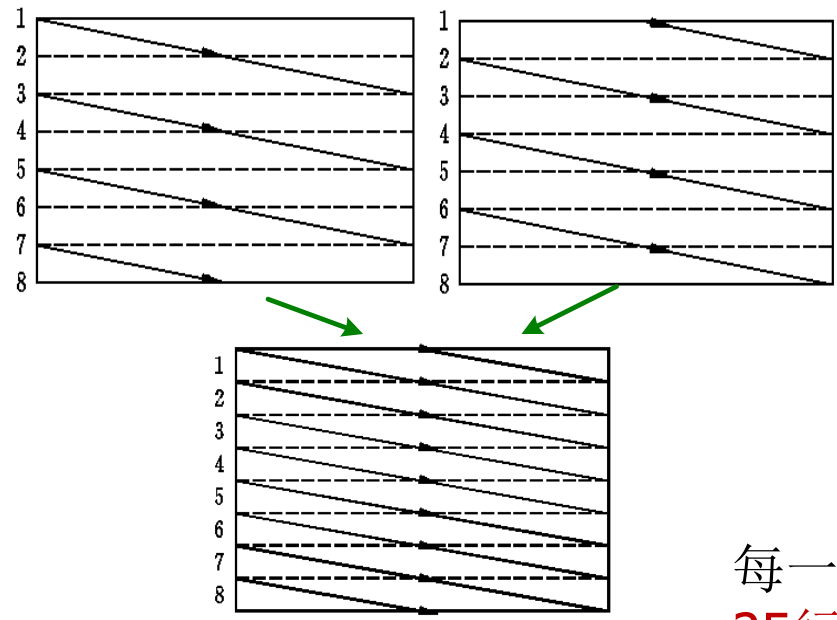
- ◆ 什么是多媒体同步
- ◆ 如何描述多媒体同步
- ◆ 典型多媒体设备接口信号的同步机制
 - CD-DA 音频的流内同步
 - 模拟电视信号的行同步与场同步
 - ITU-R BT. 601 数字视频图像帧的同步
 - HDMI 数字多媒体接口的同步
- ◆ 典型多媒体系统的媒体同步方法
 - MP3流同步
 - DVD/HDTV的流同步
 - 基于RTP的流媒体同步
- ◆ 展望未来



模拟电视扫描原理

场频/场速率(field rate) f_f
帧频/帧速率(frame rate) f_F

逐行扫描(noninterlaced scanning): 电子束从显示屏的左上角一行接一行地扫到右下角, 在显示屏上扫一遍就显示一幅完整的图像。

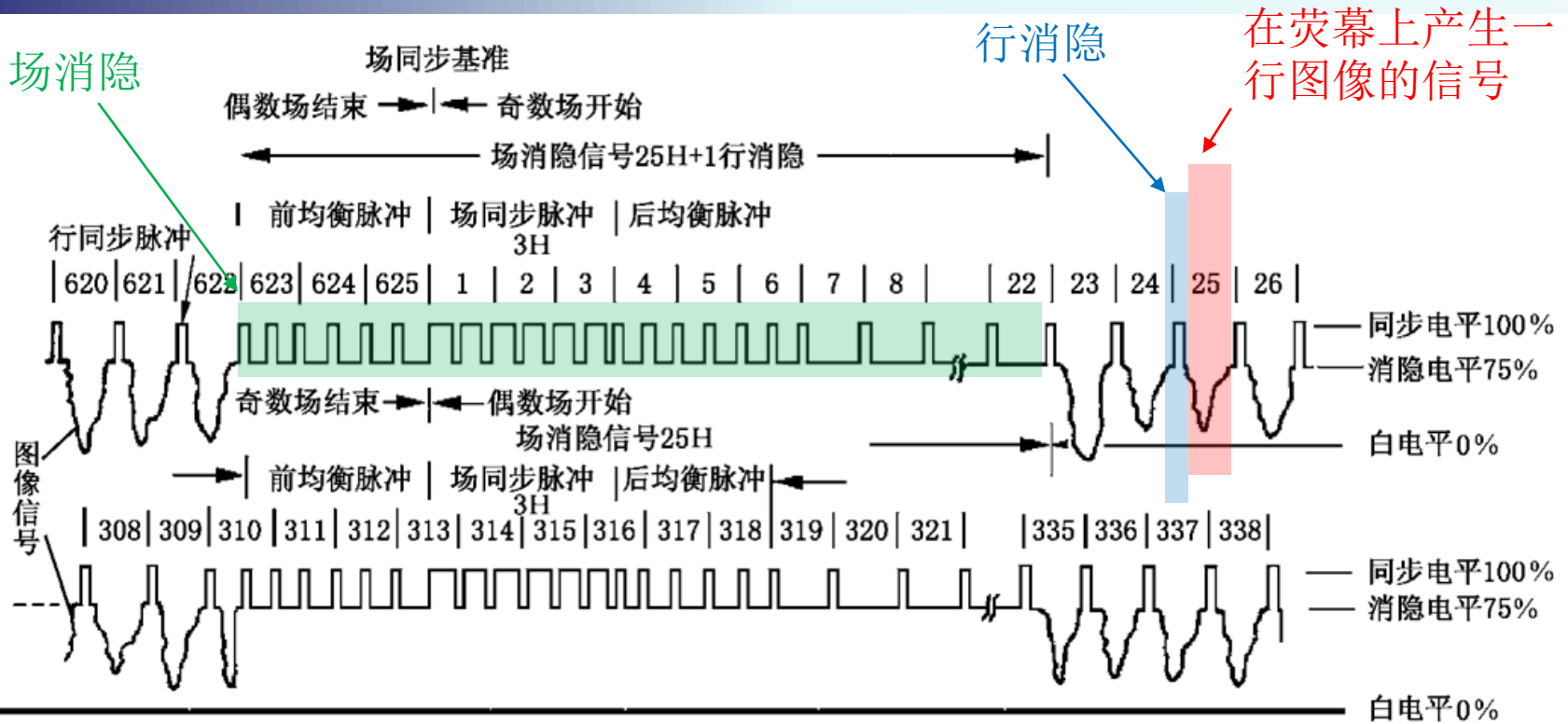


隔行扫描(interlaced scanning): 一帧图像由两部分组成: 由奇数行组成的奇数场; 由偶数行组成的偶数场。

每一场的扫描行数为 $625/2=312.5$ 行, 其中**25行作场回扫**, 传送图像的**行数每场只有287.5行**, 因此**每帧只有575行**有图像显示。

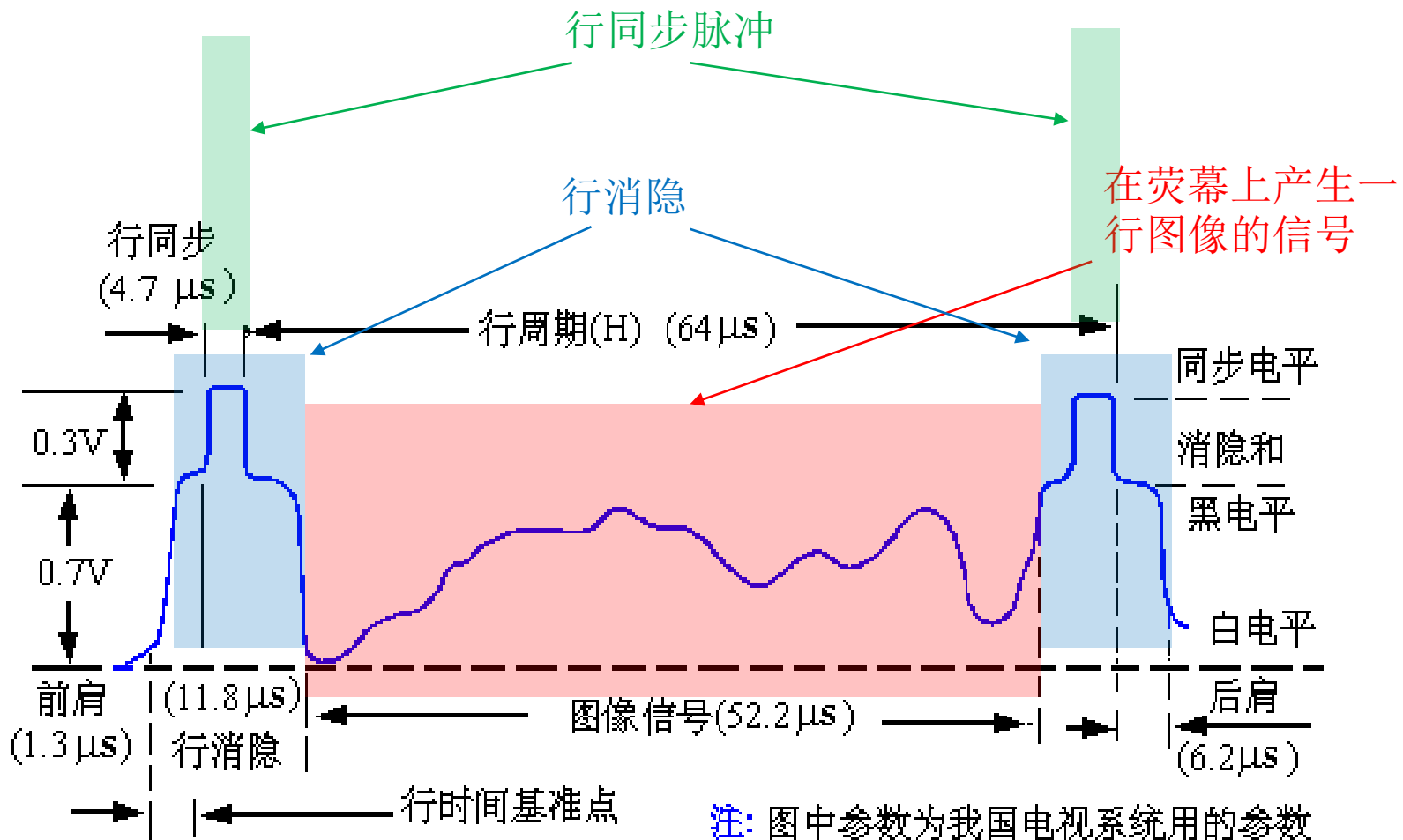


扫描定时与同步信号

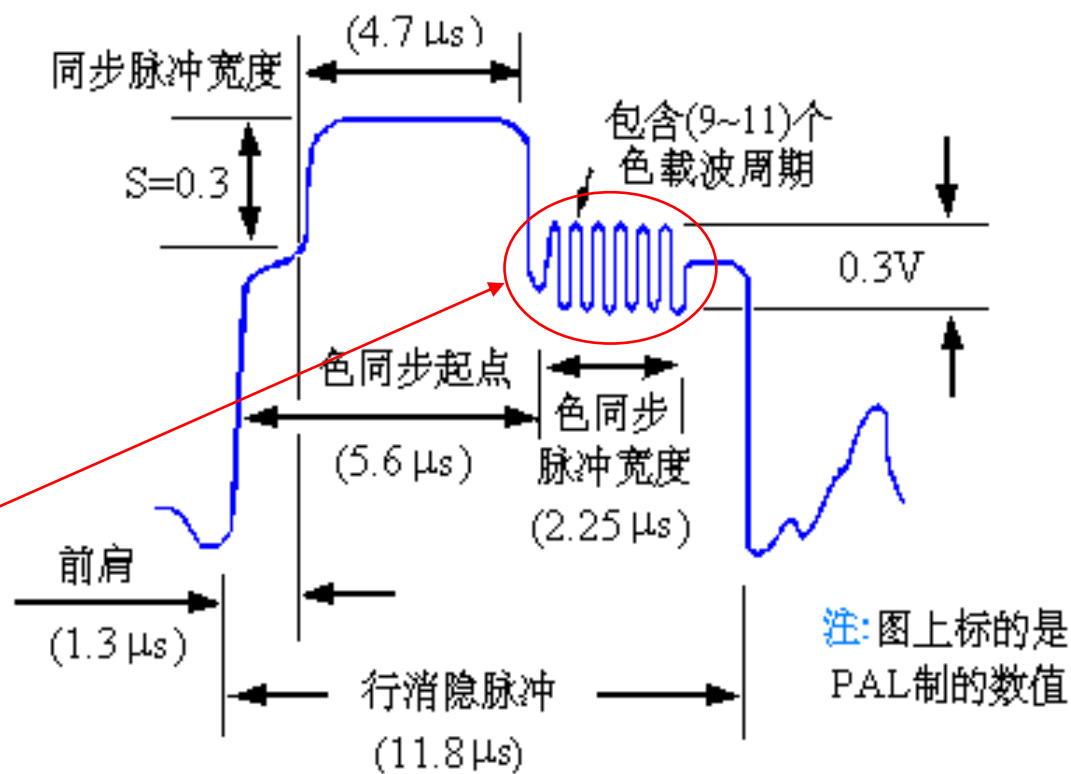


每次扫描都是从左到右，扫描完第一行时，电子束要从第一行的末尾回到第二行的行首，这段时间会产生一条回扫线，影响图像质量，所以要用水平消隐（**行消隐**，Hblank，持续 $12\mu\text{s}$ ）使电子枪截止不发送电子。同理，电子枪从右下角移动到左上角的那段时间就是垂直消隐（**场消隐**，Vblank，持续 $1612\mu\text{s}$ ）。

一个行周期的黑白全电视信号



在彩色电视系统中，色差信号通过色载波信号调制之后加到每一扫描行的水平消隐信号上，再与图像亮度信号复合后就得到彩色全电视信号。





小结：模拟电视信号的行同步与场同步

- ◆ 行扫描信号周期 $64\mu\text{s}$
 - 图像信号持续 $52\mu\text{s}$
 - 行消隐，Hblank，持续 $12\mu\text{s}$
- ◆ 一帧完整图像扫描行数为625行
 - 575行有图像显示
 - 25行作场回扫：场消隐，Vblank，持续 $1612\mu\text{s}$
 - 隔行扫描：每一场的扫描行数为 $625/2=312.5$ 行
- ◆ 黑白电视→彩色电视
 - 行消隐信号上调制了色度信号
- ◆ 图像和声音的同步：FDM

MC 主要内容

- ◆ 什么是多媒体同步
- ◆ 如何描述多媒体同步
- ◆ 典型多媒体设备接口信号的同步机制
 - CD-DA 音频的流内同步
 - 模拟电视信号的行同步与场同步
 - ITU-R BT. 601 数字视频图像帧的同步
 - HDMI 数字多媒体接口的同步
- ◆ 典型多媒体系统的媒体同步方法
 - MP3流同步
 - DVD/HDTV的流同步
 - 基于RTP的流媒体同步
- ◆ 展望未来



CCIR 601 / ITU-R BT. 601

- ◆ 数字视频信号是在模拟视频信号的基础上经过取样、量化和编码形成的，不同的空间垂直分辨率，会有不同的空间水平分辨率，这给不同制式的数字视频信号的互通带来不便。故有必要在世界范围内建立统一的数字视频标准。
- ◆ 1982年10月，国际无线电咨询委员会（CCIR, Consultative Committee for International Radio）通过了第一个关于演播室彩色电视信号数字编码的建议，即1993年变更成的国际电信联盟无线电通信部门**ITU-R BT. 601**分量数字系统建议，该标准规定了图像屏幕宽高比（IAR, Image Aspect Ratio）为4：3和16：9两种数字视频格式。该建议考虑到现行的多种彩色电视制式，提出了一种世界范围内兼容的数字编码方式，它是标准数字电视（SDTV, Standard Digital TV）的图像格式。



ITU-R BT.601 数字化规范

◆ 采样频率

- 对于数字分量编码，CCIR601建议亮度抽样频率为525/60和625/50三大制式行频公倍数2.25MHz的6倍，即13.5MHz。对现行电视制式而言，亮度信号的最大带宽是6MHz，13.5MHz大于2倍信号带宽，符合奈奎斯特定理。而色差信号的带宽比亮度信号窄得多，同时考虑到抽样的样点结构应满足正交结构要求，两个色差信号的抽样频率均选为6.75MHz，这样亮度信号与两个色差信号的抽样频率之比为4:2:2。

◆ 数据量

- 4:2:2的格式采样，10Bit量化
 - $13.5 \times 10 + 6.75 \times 10 + 6.75 \times 10 = 270\text{Mbps}$
- 4:2:2的格式采样，8Bit量化
 - $13.5 \times 8 + 6.75 \times 8 + 6.75 \times 8 = 216\text{Mbps}$

◆ 数据时分复用

- 经量化、编码后的视频数据字有3个样本值，CCIR601规定进行多路复用后按27Mwords/s来传递，其中字按以下顺序传递:Cb、Y、Cr、Y、Cb、Y、Cr、Y、Cb、Y、Cr、Y.....

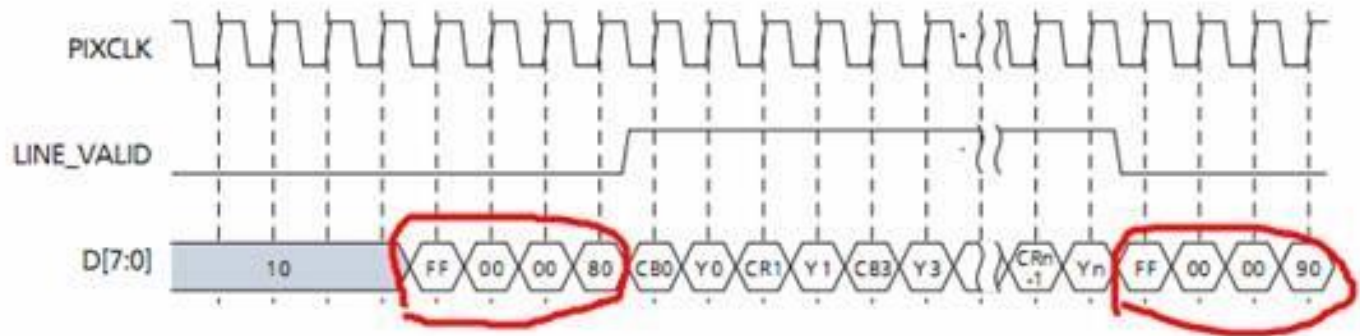
- ◆ BT.601规划了对视频进行数字编码的方法
- ◆ BT.656则实际定义了实施BT.601所必需的物理接口和数据流，以及位并行和位串行模式
 - 位并行模式
 - 只需要27MHz的时钟以及8或10条连线（取决于像素的深度）。
 - 同步化信号都嵌入到数据流中，因此无需额外添加硬件连线。
 - 位串行模式
 - 只需要在单个通道上传输像素串行数据流，不过它需要运用复杂的同步化、频谱整形和时钟恢复调理等技术手段。
 - 此外，其位时钟速率高达接近300MHz，因此要在很多系统中实施基于采用串行位形式的BT.656是极富挑战性的任务。



ITU-R BT.656 并行传输模式

ITU-R BT.656包含三部分信号：视频信号、定时基准信号、辅助信号。辅助数据信号可在行消隐期间传送，或在场消隐中的行的有效期间传送。

27 MHz →



有两个定时基准信号，一个在每个视频数据块的开始(Start of ActiveVideo, SAV), 另一个在每个视频数据块的结束(End of Active Video, EAV); 每个定时基准信号由4个字的序列组成, 格式如下: FF 00 00 XY (16进制) 头三个是固定前缀, 第4个字包含定义第二场标识、场消隐状态和行消隐状态的信息。

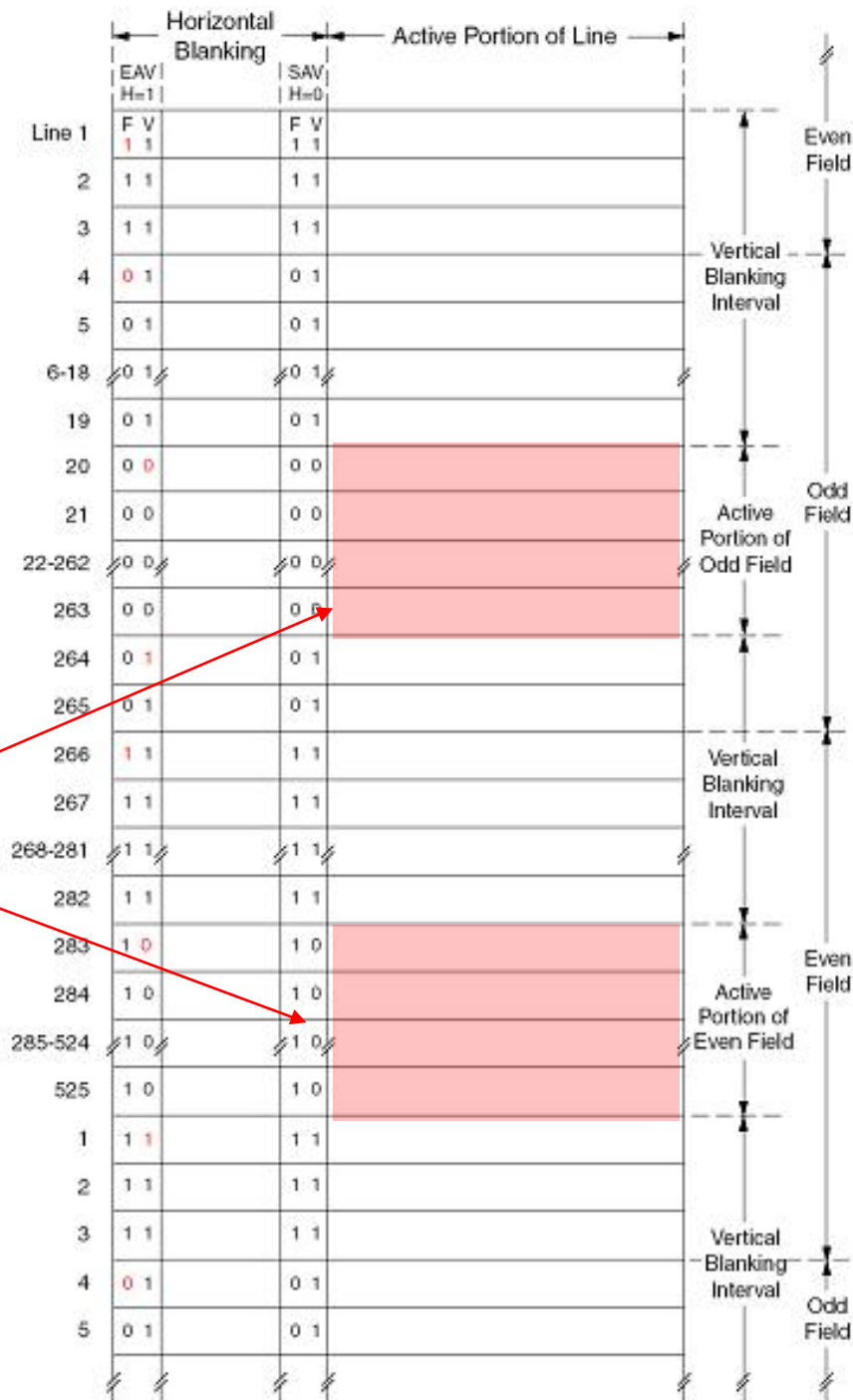


ITU-R BT.656 数据格式

◆ 数字化后的数据包括

- 行消隐信号
- 有效的图像像素信号
- 场消隐信号

接收端需要从上述信号中提
取出有效的图像像素信号





ITU-R BT.656 串行传输模式

◆ 串行比特流每个10比特字的LSB（最不重要比特）应最先发射。发射之前，通过补充编码进行频谱整形、字同步，并促进时钟的恢复。

◆ 未编码的串行比特流中低频成分很丰富，不利于接收端的时钟提取，故使用生成器多项式 $G1(x) \times G2(x)$ 进行扰码

□ $G1(x) = x^9 + x^4 + 1$ 用于生成扰码NRZ信号

□ $G2(x) = x + 1$ 用于生成无极性NRZI序列

码流中当出现常“0”和常“1”
DC-balanced / DC-Free ?



小结：ITU-R BT. 601 数字视频

- ◆ BT.601规划了对视频进行数字编码的方法
- ◆ BT.656则定义了实施BT.601所必需接口
 - 并行传输模式：27MHz时钟信号线、8或10根数据信号线
 - 串行传输模式：扰码/解扰；接收端做时钟恢复
- ◆ 这是一个典型的基带传输系统
 - 接收端需要做比特同步：并行接口有时钟线，串行接口要恢复时钟
 - 接收端需要做帧同步：定时基准信号SAV/EAV
- ◆ 此接口上如何传输音频？
 - BT.601仅是一个视频信号的数字编码规范
 - BT.656未提供音频信号线

消隐信号对应的时钟周期可用来传输视频以外的数据

MC 主要内容

- ◆ 什么是多媒体同步
- ◆ 如何描述多媒体同步
- ◆ 典型多媒体设备接口信号的同步机制
 - CD-DA 音频的流内同步
 - 模拟电视信号的行同步与场同步
 - ITU-R BT. 601 数字视频图像帧的同步
 - **HDMI 数字多媒体接口的同步**
- ◆ 典型多媒体系统的媒体同步方法
 - MP3流同步
 - DVD/HDTV的流同步
 - 基于RTP的流媒体同步
- ◆ 展望未来



HDMI

High Definition Multimedia Interface

◆ 2002年4月，日立、松下、飞利浦、Silicon Image、索尼、汤姆逊、东芝七家公司共同组建了HDMI高清多媒体接口组织，开始着手制定一种符合高清时代标准的全新数字化视频/音频接口技术。经过半年多时间的准备工作，HDMI组织在2002年12月9日正式发布了HDMI 1.0版标准。

HDMI version	1.0	1.1	1.2	1.3	1.4	2.0
Date initially released	December 9, 2002	May 20, 2004	August 8, 2005	June 22, 2006	May 28, 2009	September 4, 2013
Maximum clock rate (MHz)	165	165	165	340	340 ^[109]	600
Maximum TMDS throughput per channel (Gbit/s) including 8b/10b overhead	1.65	1.65	1.65	3.40	3.40	6
Maximum total TMDS throughput (Gbit/s) including 8b/10b overhead	4.95	4.95	4.95	10.2	10.2	18
Maximum throughput (Gbit/s) with 8b/10b overhead removed	3.96	3.96	3.96	8.16	8.16	14.4
Maximum audio throughput (Mbit/s)	36.86	36.86	36.86	36.86	36.86	49.152 (IEC61937 and DST audio)
Maximum color depth (bit/px)	24	24	24	48 ^[a]	48	48
Maximum consumer resolution over single link at 24-bit/px ^[b]	1920×1200p60	1920×1200p60	1920×1200p60	2560×1600p75	3840×2160p30	3840×2160p60

HDMI Version

	1.0–1.2a	1.3–1.3a	1.4–1.4b	2.0–2.0b	2.1
Release Date	Dec 2002 (1.0) ^[124] May 2004 (1.1) Aug 2005 (1.2) ^[125] Dec 2005 (1.2a) ^[126]	Jun 2006 (1.3) ^[127] Nov 2006 (1.3a) ^[6]	Jun 2009 (1.4) ^[128] Mar 2010 (1.4a) ^[106] Oct 2011 (1.4b)	Sep 2013 (2.0) ^[110] Apr 2015 (2.0a) ^[129] Mar 2016 (2.0b)	Nov 2017 ^[130]
Signal Specifications					
Transmission Bandwidth ^[a]	4.95 Gbit/s	10.2 Gbit/s	10.2 Gbit/s	18.0 Gbit/s	48.0 Gbit/s
Maximum Data Rate ^[b]	3.96 Gbit/s	8.16 Gbit/s	8.16 Gbit/s	14.4 Gbit/s	42.6 Gbit/s
Maximum Character Rate ^[c]	165 MHz ^{[93](\$3)}	340 MHz ^[127]	340 MHz	600 MHz ^{[111](\$6.1.1)}	1.2 GHz
Data Channels	3	3	3	3	4
Encoding Scheme ^[d]	8b/10b ^{[93](\$5.1)}	8b/10b	8b/10b	8b/10b	16b/18b ^[123]
Compression (Optional)	-	-	-	-	DSC 1.2 ^[131]
Color Depth Support					
8 bpc (24 bit/px)	Yes ^{[93](\$3)}	Yes	Yes	Yes	Yes
10 bpc (30 bit/px)	Yes ^[f]	Yes	Yes	Yes	Yes
12 bpc (36 bit/px)	Yes ^[f]	Yes	Yes	Yes	Yes
16 bpc (48 bit/px)	No	Yes ^{[6](\$6.5)}	Yes	Yes	Yes
Audio Specifications					
Max. Sample Rate Per Channel	192 kHz ^{[93](\$7.3)}	192 kHz	192 kHz	192 kHz	192 kHz
Max. Aggregate Sample Rate	?	?	768 kHz ^{[99](\$7.3)}	1536 kHz ^{[111](\$9.2)}	1536 kHz
Sample Size	16–24 bits ^{[93](\$7.3)}	16–24 bits	16–24 bits	16–24 bits	16–24 bits
Maximum Audio Channels	8 ^{[93](\$7.3.1)}	8	8	32 ^{[111](\$8.3.1)}	32



High Definition Multimedia Interface, HDMI 机械特性

HDMI™
HIGH-DEFINITION MULTIMEDIA INTERFACE



HDMI type A receptacle connector



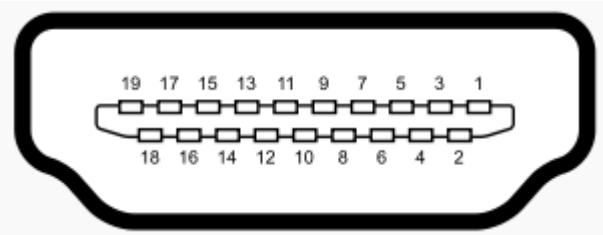
A HDMI to DVI adapter



HDMI type A plug connector



A DVI to HDMI adapter





High Definition Multimedia Interface, HDMI

功能特性

最小化传输差分信号(TMDS:Transition Minimized Differential signal)。通过异或及异或非等逻辑算法将原始信号**8bit数据转换成10bit**，用来**实现直流平衡（DC-balanced）**，转换后的数据以差分传动方式传送。这种算法使得被传输信号过渡过程的上冲和下冲减小，传输的数据趋于直流平衡，使信号对传输线的电磁干扰减少，提高信号传输的速度和可靠性。

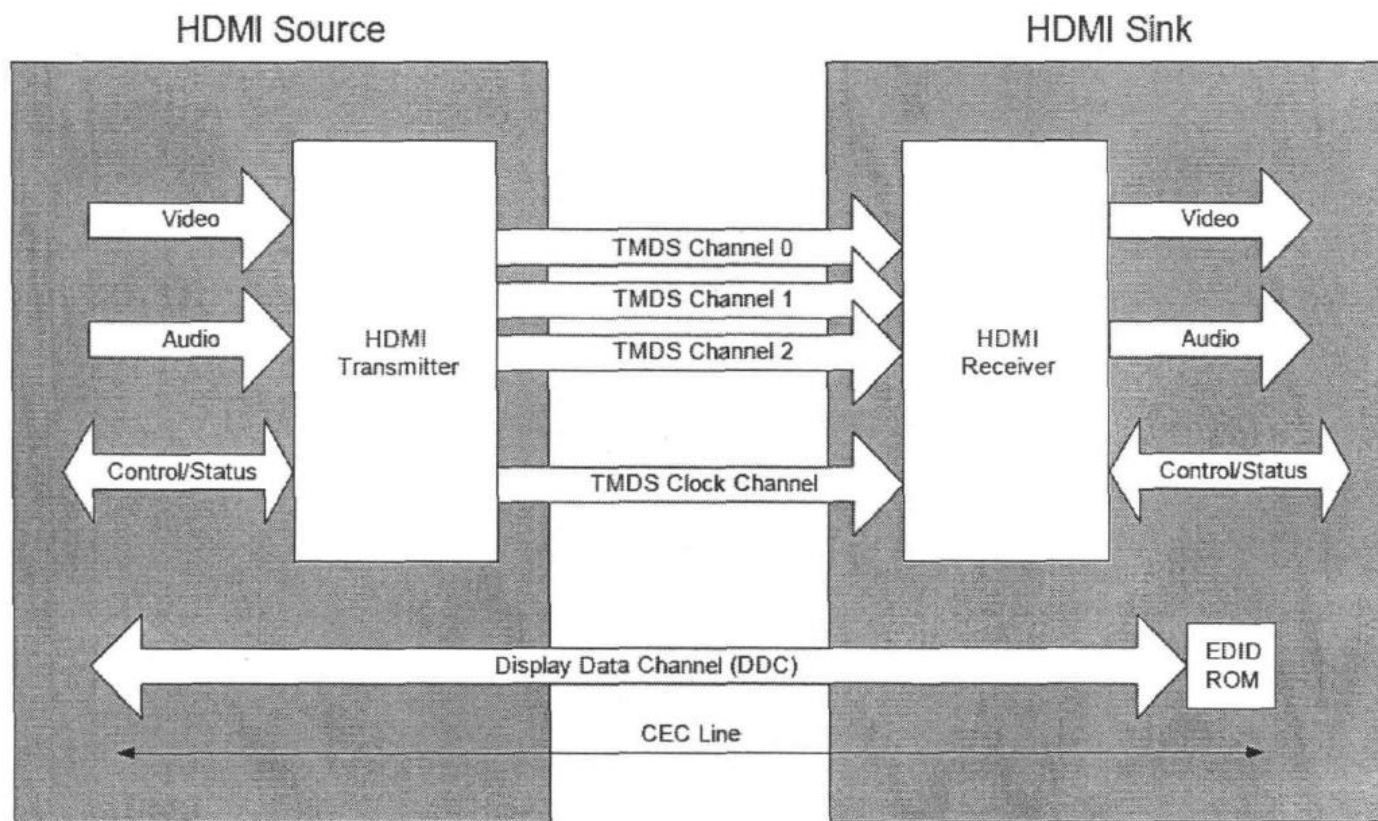
Table 4-1 Type A Connector Pin Assignment

PIN	Signal Assignment
1	TMDS Data2+
3	TMDS Data2-
5	TMDS Data1 Shield
7	TMDS Data0+
9	TMDS Data0-
11	TMDS Clock Shield
13	CEC
15	SCL
17	DDC/CEC Ground
19	Hot Plug Detect

PIN	Signal Assignment
2	TMDS Data2 Shield
4	TMDS Data1+
6	TMDS Data1-
8	TMDS Data0 Shield
10	TMDS Clock+
12	TMDS Clock-
14	Reserved (N.C. on device)
16	SDA
18	+5V Power

MC HDMI系统架构

音频、视频和其他辅助的数据通过三对TMDS数据通道来传输，一对TMDS时钟，通常是等于视频像素的比特率，通过一对TMDS时钟通道来发送。



DDC (Display Data Channel)
I²C总线

HDMI源端设备与接收端设备交互EDID数据及HDCP密钥。

EDID用于了解接收端设备音视频的接收能力；HDCP Key用于数据内容保护。

CEC (Consumer Electronics Control)

one-wire总线，用户通过一个遥控器即可对连接的设备进行控制



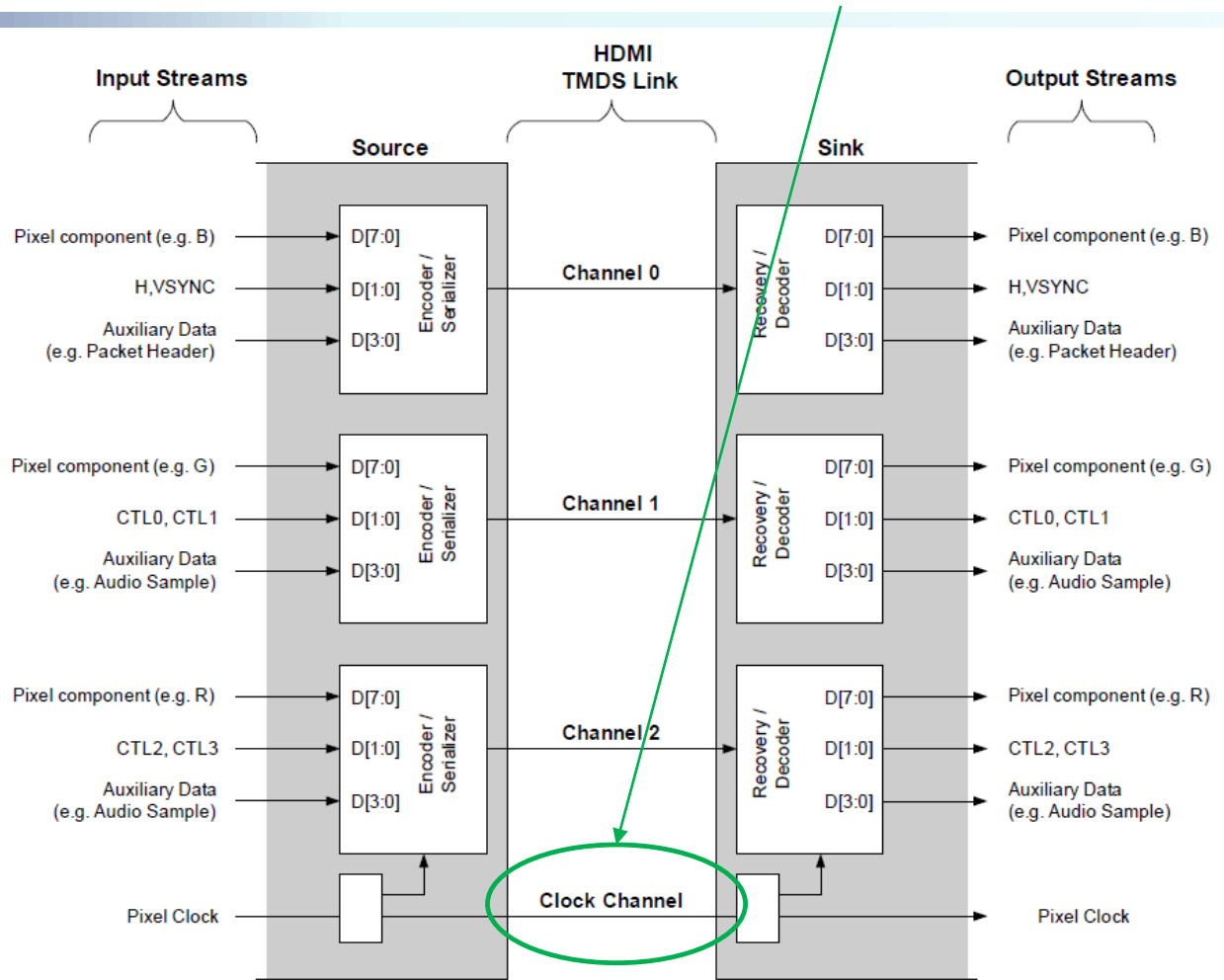
High Definition Multimedia Interface, HDMI

TMDS通道

Pixel Clock的10倍频

BT.656: 每个Clock
传输Y或Cr或Cb一个
样本值

HDMI: 默认像素
深度24bits情形下
一个Pixel Clock之
内可完成一个像素
值的传输



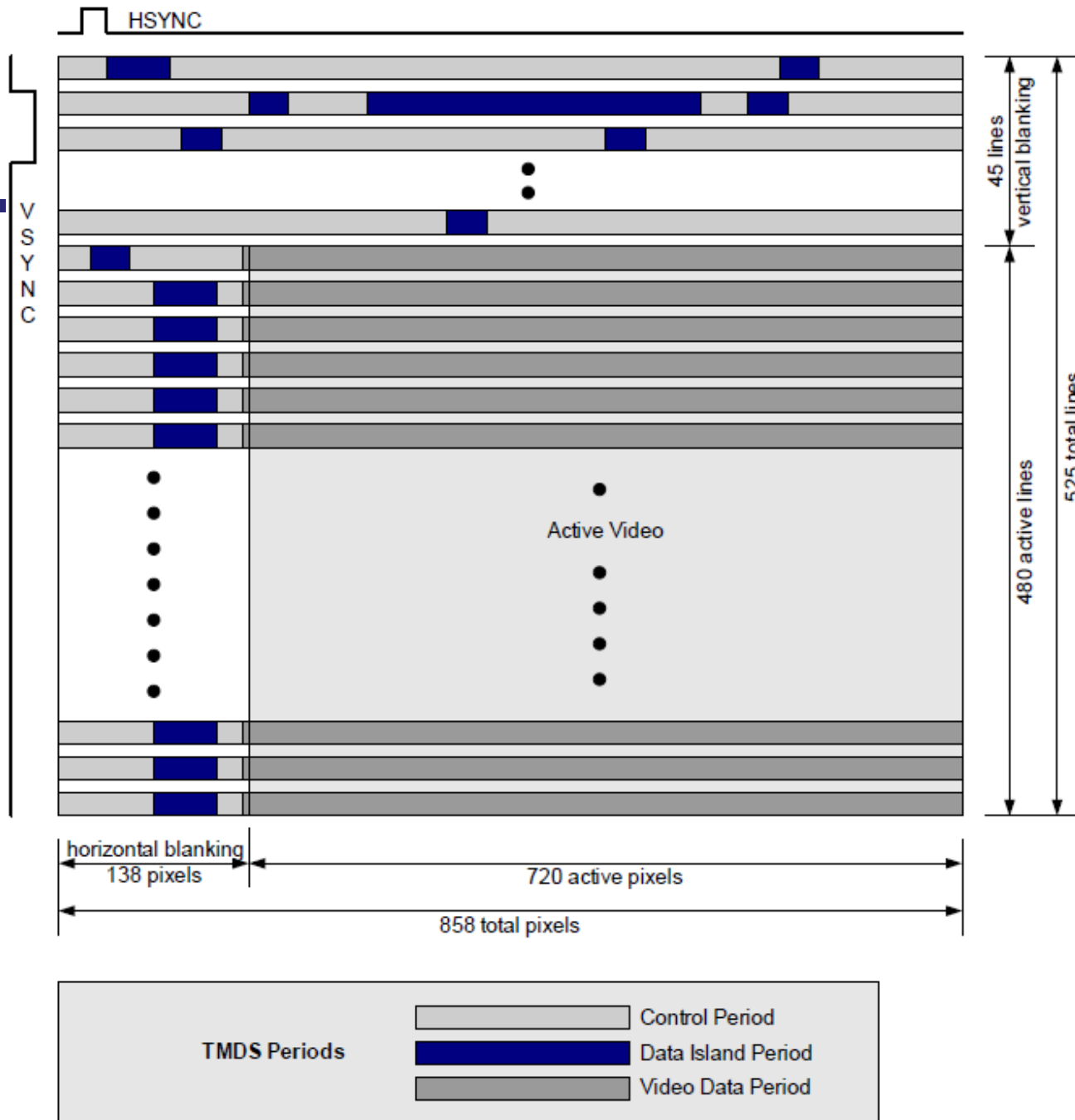
HDMI不同版本的maximum pixel clock rate不同

165MHz(HDMI 1.0) → 340MHz(HDMI 1.3) → 600MHz(HDMI 2.0)

HDMI 规程特性

HDMI 连接可处于三种操作模式之一：
■ 视频数据周期、
■ 数据岛周期、
■ 控制周期。

- 视频数据周期发送视频像素。
- 数据岛周期中，音频和辅助数据使用一系列数据包来发送。
- 两个非控制周期之间是控制周期，传输引导数据





High Definition Multimedia Interface, HDMI 不同传输周期的编码方式

周期	数据发送	编码类型
视频周期	防护带	固定的 10 位模式
	视频像素	视频数据编码 (8 位转 10 位)
数据岛	数据包 (包括音频采样数据, 信息帧, 行同步, 场同步)	TERC4 编码 (4 位转 10 位)
	防护带	固定的 10 位模式
控制	控制 (引导, HSYNC, VSYNC)	控制周期编码 (2 位转 10 位)

表 2-2 编码类型和数据发送内容

控制周期传输引导数据, 同时也被接收端用于字符同步

TMD5 Channel	D0	D1
0	HSYNC	VSYNC
1	CTL0	CTL1
2	CTL2	CTL3

表 2-3 控制周期传输内容

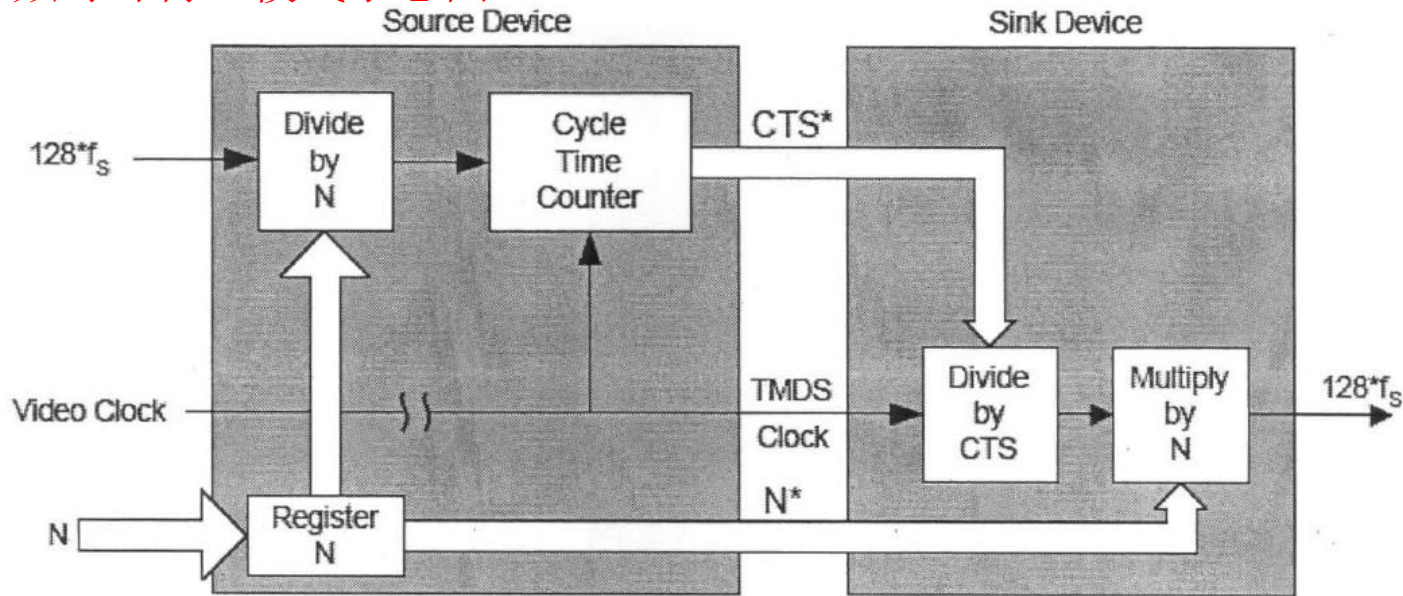
CTL0	CTL1	CTL2	CTL3	数据周期类型
1	0	0	0	视频数据周期
1	0	1	0	数据岛周期

表 2-4 控制位意义



High Definition Multimedia Interface, HDMI 音频同步的实现方法

音频时钟再生模式示意图



音频样本会在数据岛周期利用3个TMDSClock通道传输，采用4bits → 10bits的编码。音频的采样时钟 f_s 需要在接收端恢复。通过 N 和 CTS (Cycle Time Stamp) 的传输可以恢复 f_s 。满足下式：

$$128 \times f_s = t_{TMDSclock} \times N / CTS$$



High Definition Multimedia Interface, HDMI

音频时钟信息通过DDC传递

- ◆ DDC(Display Data Channel)就是一个I²C总线(SCL,SDA)
- ◆ HDMI源端可通过DDC访问HDMI接收端的EDID(Extended Display Identification Data)信息，从而获知接收端的接受能力和一些关于接收端的信息。
- ◆ HDMI接收端还通过DDC通道探测信息帧：辅助视频信息帧、音频信息帧。

Audio Infoframe
包含音频格式的基本信息如编码方式、采样频率等

Table 18. Audio InfoFrame format

InfoFrame Type Code	InfoFrame Type = 04 ₁₆ (new)							
InfoFrame Version Number	Version = 01 ₁₆							
Length of Audio InfoFrame	Length of Audio InfoFrame (10)							
Data Byte 1	CT3	CT2	CT1	CT0	Rsvd	CC2	CC1	CC0
Data Byte 2	Reserved (shall be 0)			SF2	SF1	SF0	SS1	SS0
Data Byte 3	Format depends on coding type (i.e., Data Byte 1)							
Data Byte 4	CA7	CA6	CA5	CA4	CA3	CA2	CA1	CA0
Data Byte 5	DM_I NH	LSV3	LSV2	LSV1	LSV0	Reserved (shall be 0)		
Data Byte 6	Reserved (shall be 0)							
Data Byte 7	Reserved (shall be 0)							
Data Byte 8	Reserved (shall be 0)							
Data Byte 9	Reserved (shall be 0)							
Data Byte 10	Reserved (shall be 0)							



High Definition Multimedia Interface, HDMI 视频和音频如何同步?

◆ HDMI的流内同步

- 视频同步: TMDS在控制周期传HSync和VSync
- 音频同步: 通过DDC传输音频采样频率等基本信息

◆ HDMI音视频流间同步

- HDMI 1.3(2006年)添加了automatic audio syncing 支持
- HDMI 2.0 (2013年)添加了Dynamic auto lip-sync支持
- 其基本原理应该是给音频/视频增加一个延时控制电路, 因HDMI为收费标准, 未查到具体规范内容



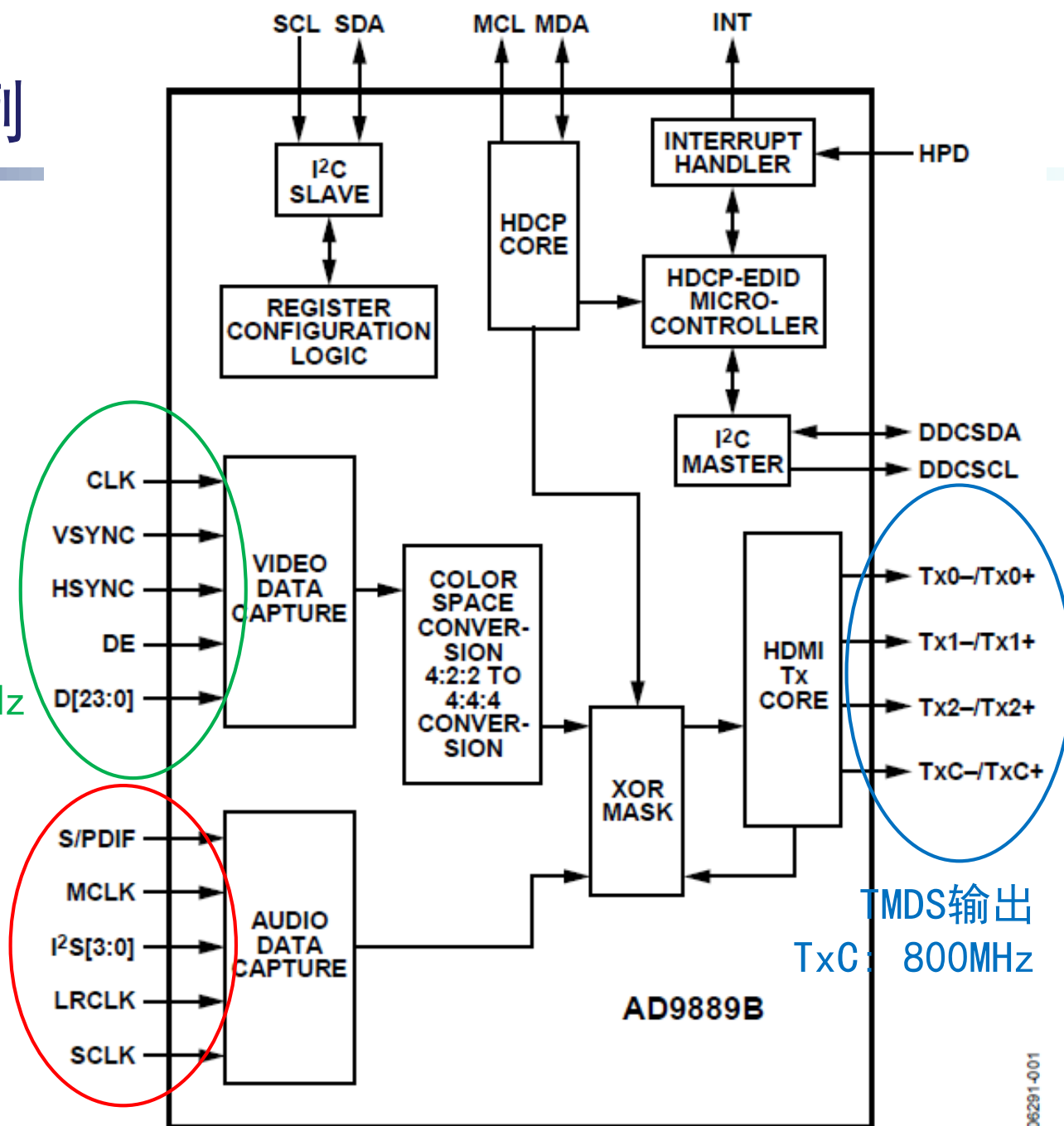
HDMI 芯片示例

AD公司的AD9889B
是一款HDMI/DVI驱
动芯片

视频输入

CLK: 13.5-165MHz

音频输入



TMDs输出
TxC: 800MHz



小结： HDMI 数字视频接口

◆ TMDS (Transition Minimized Differential Signal)通道

- 视频数据周期
- 数据岛周期： 音频流
- 控制周期

◆ DDC(Display Data Channel)通道

- I²C总线(SCL,SDA): 音频流采样频率

◆ CEC(Consumer Electronics Control)通道

- 1-Wire总线

◆ TMDS时钟是Pixel时钟的10倍！

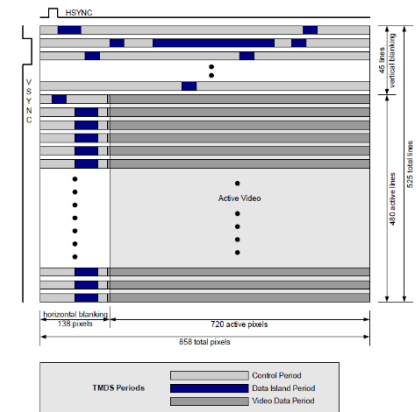


Figure 5-2 Informative: Example: TMDS periods in 720x480p video frame



小结：多媒体设备接口信号的同步机制

◆ CD-DA 音频的流内同步

- 比特同步 → 帧同步（每帧持续 $1/7350\text{s}$ ）

◆ 模拟电视信号的行同步与场同步

- 每个扫描行持续 $64\mu\text{s}$ ，625行持续 40ms
- 图像和伴音信号调制2个不同的载波（FDM）

◆ ITU-R BT. 601 数字视频图像帧的同步

- 每行720像素 → $720/64\mu\text{s} = 11.25\text{MHz} \rightarrow 13.5\text{MHz}$
- 比特同步（仅串行需要） → 帧同步（SAV/EAV）

◆ HDMI 数字多媒体接口的同步

- BT. 601类似的数据块结构
- 消隐信号持续时间 → 数据岛周期
- $\text{TMDs Clock} = 10 \times \text{Pixel Clock}$

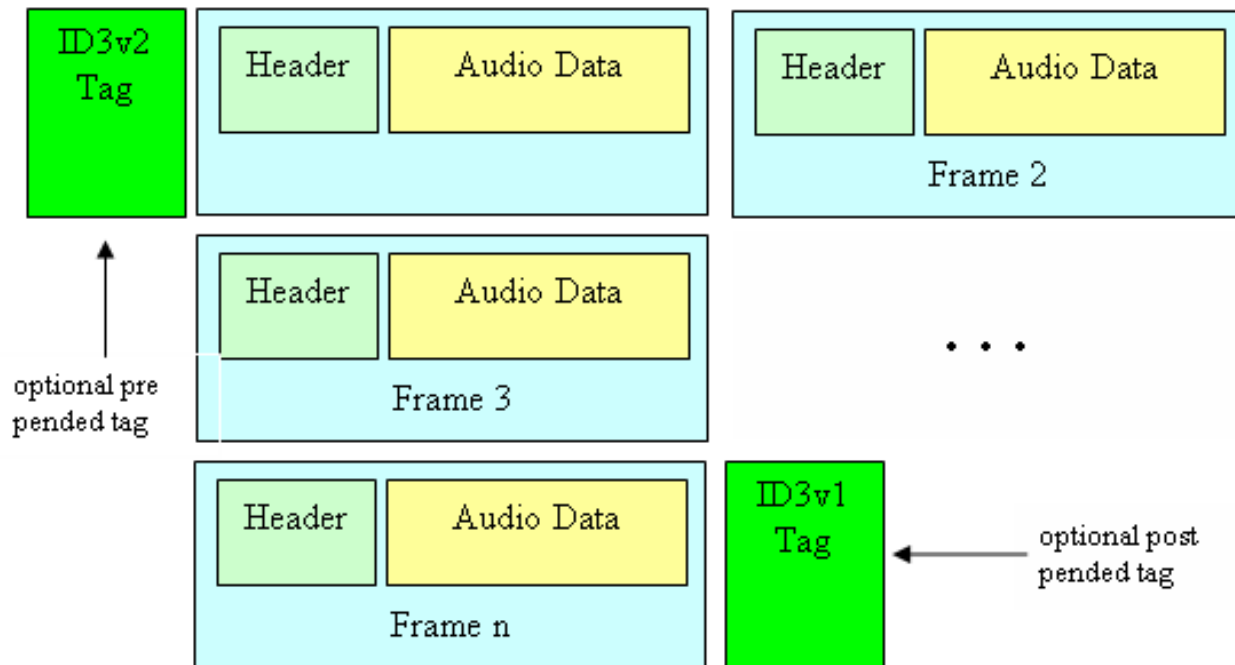
MC 主要内容

- ◆ 什么是多媒体同步
- ◆ 如何描述多媒体同步
- ◆ 典型多媒体设备接口信号的同步机制
 - CD-DA 音频的流内同步
 - 模拟电视信号的行同步与场同步
 - ITU-R BT. 601 数字视频图像帧的同步
 - HDMI 数字多媒体接口的同步
- ◆ 典型多媒体系统的媒体同步方法
 - MP3流同步
 - DVD/HDTV的流同步
 - 基于RTP的流媒体同步
 - MMT: MPEG Media Transport
- ◆ 展望未来



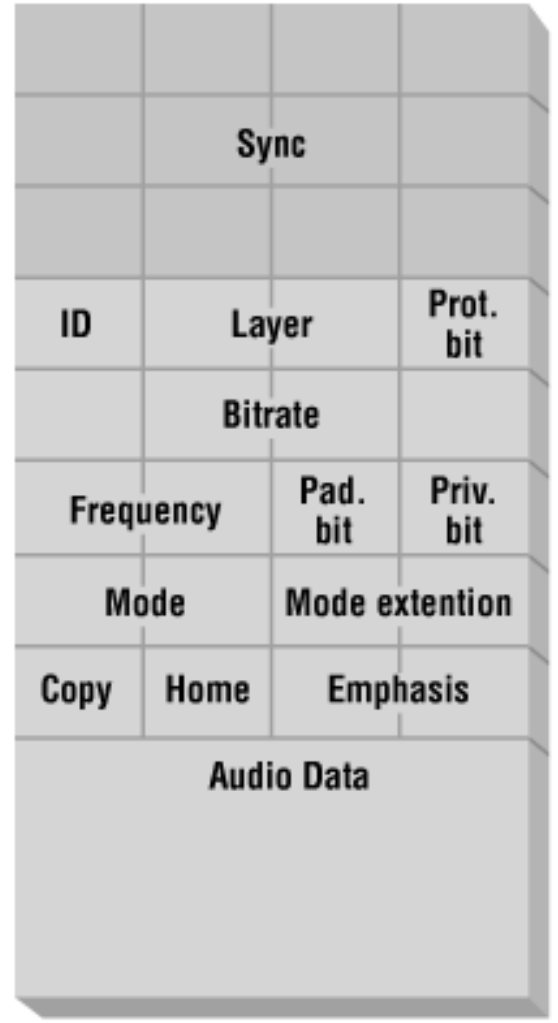
MP3文件结构：Tag+Frame

- ◆ **Tag:** MP3除了保存声音信息外，还可以同时保存一些和曲目相关的文本信息，比如歌名、演唱者、专辑名等等，这些信息就称为tag（标签）信息。
- ◆ **Frame:** 用于存放声音数据。



MC MP3帧格式

- ◆ Frame sync (11bits ‘1’)
 - ◆ MPEG version
 - ◆ MPEG layer
 - ◆ Protection
 - ◆ Bitrate index
 - ◆ Sampling rate frequency
 - ◆ Padding bit
 - ◆ Private bit
 - ◆ Channel mode
 - ◆ Mode extension
 - ◆ Copyright
 - ◆ Original
 - ◆ Emphasis
- 同步信息
 - 版本
 - 层
 - CRC校验
 - 比特率索引
 - 采样频率索引
 - 帧长调节
 - 保留字
 - 声道模式
 - 扩充模式
 - 版权
 - 原版标志
 - 强调方式





比特率索引表

128Kbps是常见MP3的比特率

下面是比特率索引表，（单位：Kbit/sec）。

比特率索引	MPEG 1			MPEG 2, 2.5 (LSF)	
	Layer I	Layer II	Layer III	Layer I	Layer II & III
0000	空闲				
0001	32	32	32	32	8
0010	64	48	40	48	16
0011	96	56	48	56	24
0100	128	64	56	64	32
0101	160	80	64	80	40
0110	192	96	80	96	48
0111	224	112	96	112	56
1000	256	128	112	128	64
1001	288	160	128	144	80
1010	320	192	160	160	96
1011	352	224	192	176	112
1100	384	256	224	192	128
1101	416	320	256	224	144
1110	448	384	320	256	160
1111	保留				



采样率索引表

每帧的持续时间 = 样本数 / 采样频率

◆ Frame Size = ((Samples Per Frame / 8 * Bitrate) / Sampling Rate) + Padding Size

	MPEG 1	MPEG 2 (LSF)	MPEG 2.5 (LSF)
Layer I	384	384	384
Layer II	1152	1152	1152
Layer III	1152	576	576

表 3 每帧的采样数

根据 MPEG 帧头里面的索引，对应的，不同版本的 MPEG 的采样率也是不同的：

采样率索引	MPEG 1	MPEG 2 (LSF)	MPEG 2.5 (LSF)
00	44100 Hz	22050 Hz	11025 Hz
01	48000 Hz	24000 Hz	12000 Hz
10	32000 Hz	16000 Hz	8000 Hz
11	保留		

表 4 MPEG 不同版本对应的采样率



MP3的流间同步 (Lyrics)

[ti:我们不一样]

[ar:大壮]

[00:00.00]大壮 - 我们不一样

....

[03:33.68]我们不一样

[03:36.41]每个人都有不同的境遇

[03:40.43]我们在这里

[03:42.58]在这里等你

[03:47.30]我们不一样

[03:49.95]虽然会经历不同的事情

[03:54.11]我们都希望

[03:56.20]来生还能相遇

[04:01.01]我们不一样

[04:03.68]虽然会经历不同的事情

[04:07.87]我们都希望

[04:09.90]来生还能相遇

[04:14.71]我们都希望

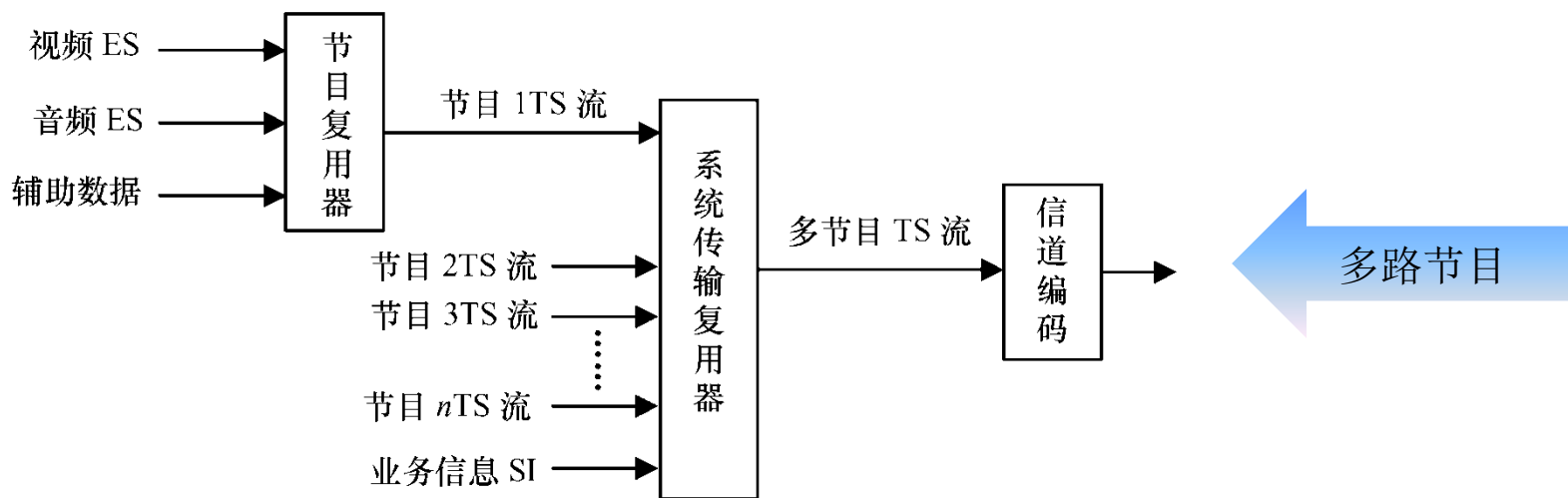
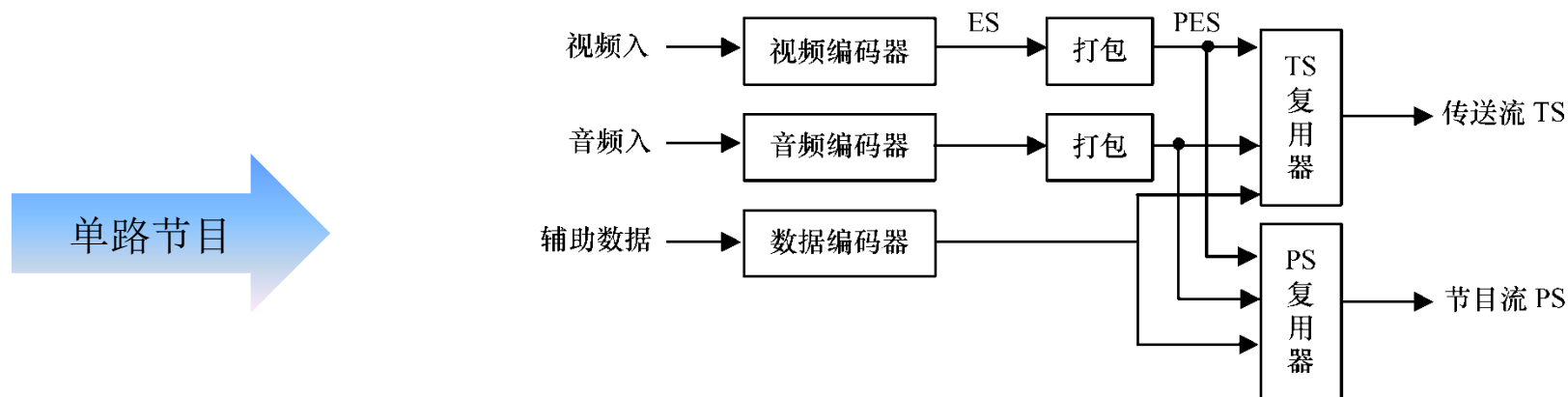
[04:16.82]来生还能相遇

MC 主要内容

- ◆ 什么是多媒体同步
- ◆ 如何描述多媒体同步
- ◆ 典型多媒体设备接口信号的同步机制
 - CD-DA 音频的流内同步
 - 模拟电视信号的行同步与场同步
 - ITU-R BT. 601 数字视频图像帧的同步
 - HDMI 数字多媒体接口的同步
- ◆ 典型多媒体系统的媒体同步方法
 - MP3流同步
 - DVD/HDTV的流同步
 - 基于RTP的流媒体同步
 - MMT: MPEG Media Transport
- ◆ 展望未来



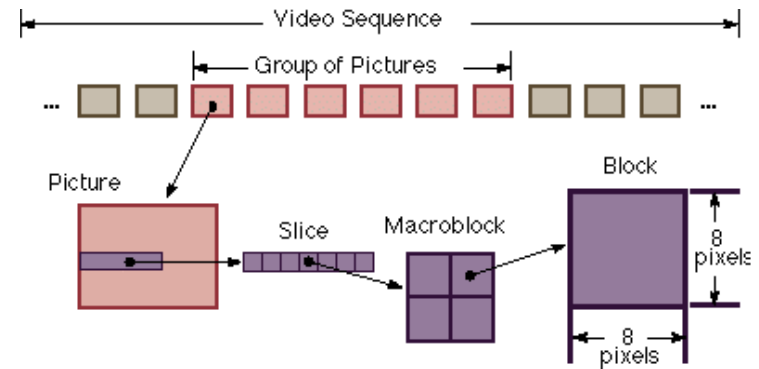
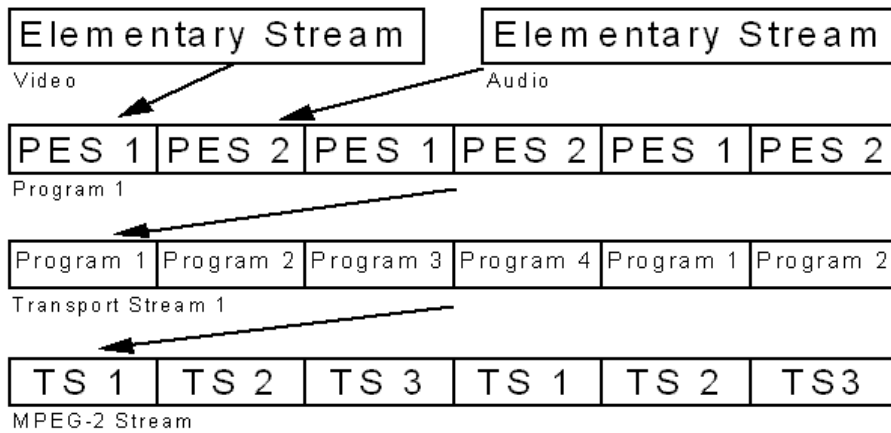
MPEG-2音视频复用





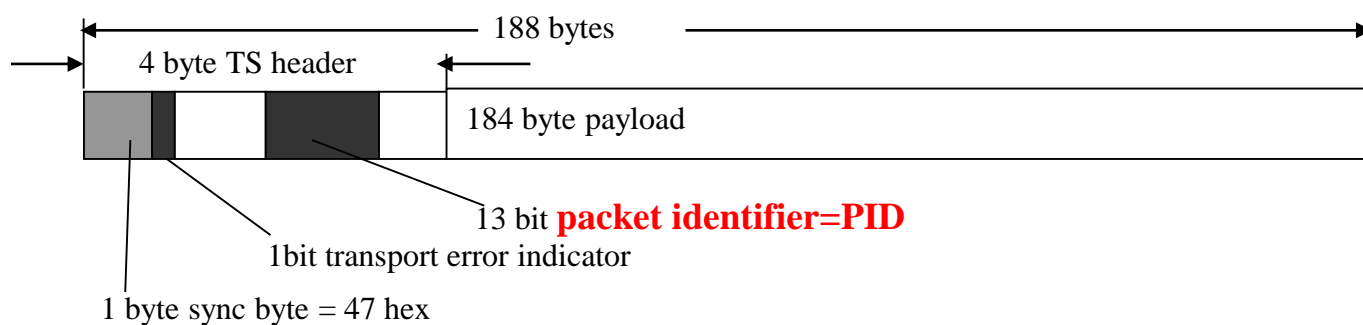
MPEG-2数据流格式

- ◆ Elementary Stream (ES)
- ◆ Packetised Elementary Stream (PES)
- ◆ Transport Streams (TS)



MC MPEG-2 TS包

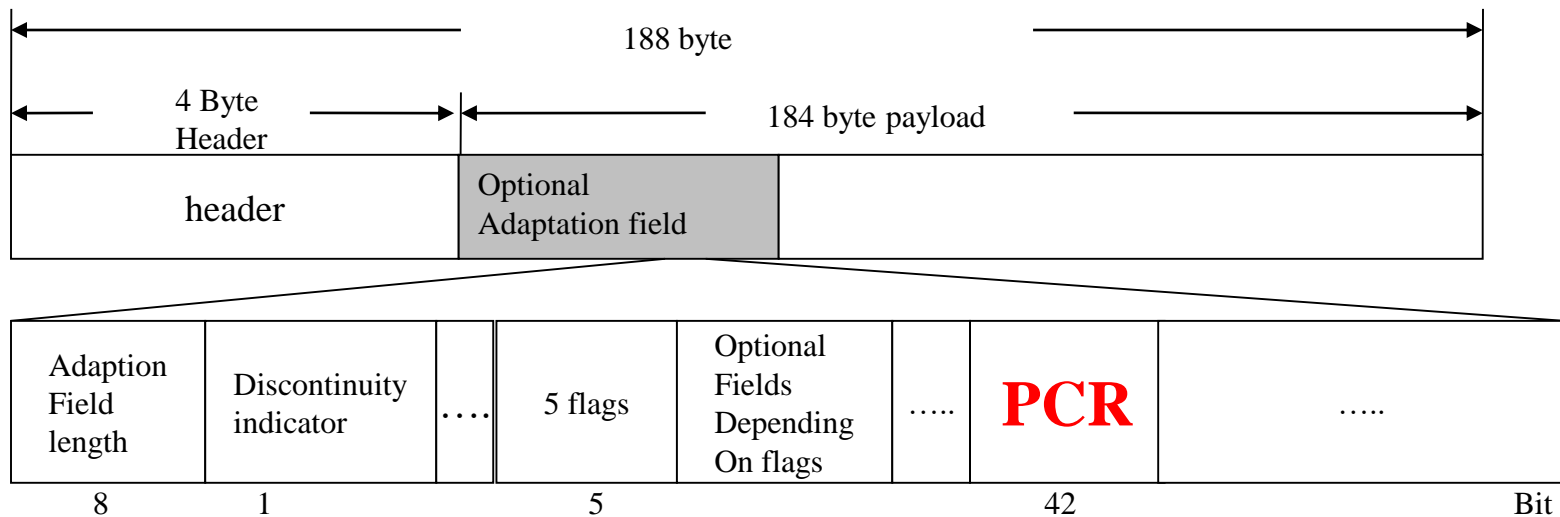
- ◆ 固定长度188字节，4字节header和184字节payload
- ◆ Header包括：**同步**字节、传输**差错标志**、13bit的**PID**（描述该包中payload的内容以及该包属于哪个ES流）



MC TS扩展包头

◆ 有时TS包的头长度必须大于4个byte:

- 包头扩展到payload中，称作adaptation field
- 在4bytes的包头中由Adaptation control bits标志是否有adaptation field。





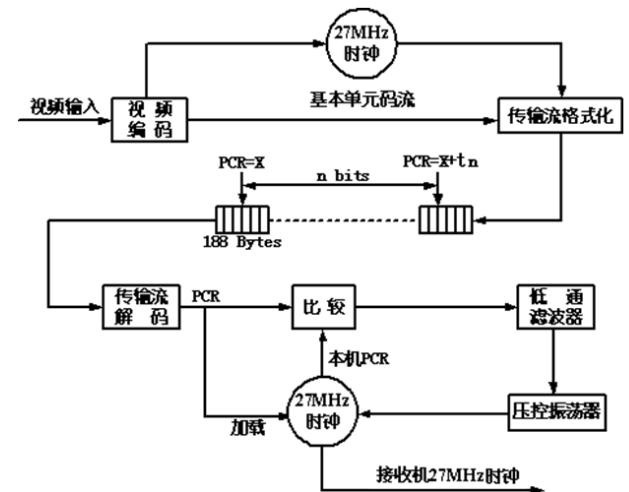
PCR(Program Clock Reference)

- ◆ PCR是由对系统时钟脉冲触发的计数器状态抽样而来
- ◆ PCR共占6Bytes，其中6bits预留，42bits有效位。有效位分为两部分：33bits的PCR-Base和9bits的PCR-Ext。
- ◆ PCR-Base是由27MHz脉冲经300分频后的90kHz脉冲触发计数器，再对计数器状态进行取样得到的。
- ◆ PCR-Ext是由27MHz脉冲直接触发计数器，再对计数器状态进行取样得到的。

$$\text{PCR-Base}(i) = 90\text{kHz} \times t(i), \quad \text{mod}(2^{33})$$

$$\text{PCR-Ext}(i) = 27\text{MHz} \times t(i), \quad \text{mod}(300)$$

$$\text{PCR}(i) = \text{PCR-Base}(i) \times 300 + \text{PCR-Ext}(i)$$



用PCR重建远地27MHz时钟



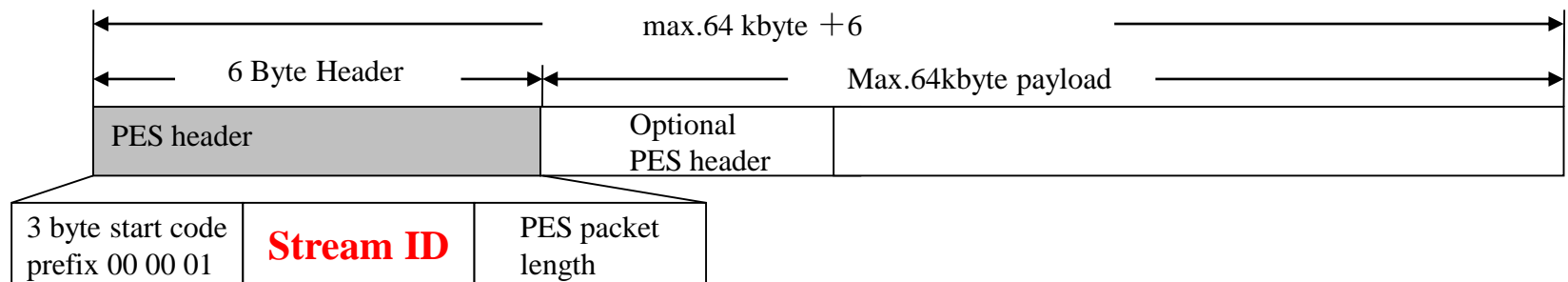
MPEG-2 PES包

◆ 6byte的PES Header

- 前3个byte是起始码前缀00 00 01
- 第4个byte标识payload是视频、音频还是数据
- 后两个byte是包长度，说明后面还有多少字节。如果长度为0，表示PES包大于64kbytes。

◆ 可选PES Header

◆ 传送ES流的净负荷数据(payload)





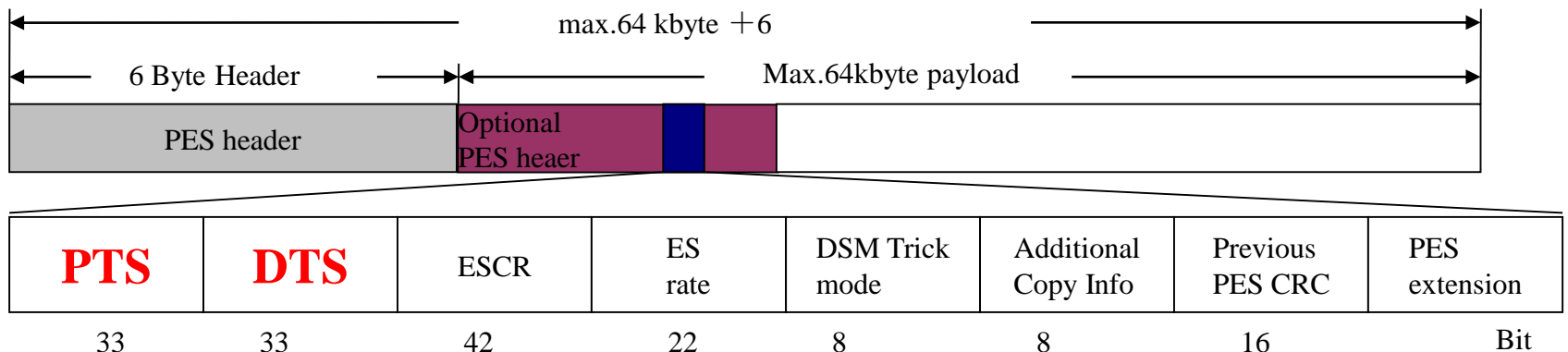
Optional PES header

◆ PTS(presentation time stamps)

- 预定义的播放时间:由27MHz脉冲经300分频后的90kHz脉冲触发计数器

◆ DTS(decoding time stamps)

- 预定义的解码时间:由27MHz脉冲经300分频后的90kHz脉冲触发计数器





MPEG2同步小结

◆ 时钟同步

- TS

- PCR(Program Clock Reference) ←27MHz

◆ 解码同步

- PES

- DTS(decoding time stamps) ←90kHz

◆ 显示同步

- PES

- PTS(presentation time stamps) ←90kHz

MC 主要内容

- ◆ 什么是多媒体同步
- ◆ 如何描述多媒体同步
- ◆ 典型多媒体设备接口信号的同步机制
 - CD-DA 音频的流内同步
 - 模拟电视信号的行同步与场同步
 - ITU-R BT. 601 数字视频图像帧的同步
 - HDMI 数字多媒体接口的同步
- ◆ 典型多媒体系统的媒体同步方法
 - MP3流同步
 - DVD/HDTV的流同步
 - 基于RTP的流媒体同步
 - MMT: MPEG Media Transport
- ◆ 展望未来



RTP(Real-time Transport Protocol)

◆ RTP提供端对端网络传输功能，适合通过组播和点播传送实时数据。RTP没有涉及资源预订和质量保证等实时服务。RTP报文格式中包括固定的RTP报文头，可选用的作用标识（CSRC）和负载数据。

V	P	X	CC	M	PT	序列号(Sequence Number)
时戳(Timestamp)						
同步源标识符(SSRC Identifier)						
其它同步源标识符(CSRC Identifier)						
静荷(PayLoad)						

SSRC帮助接收方利用发送方生成的唯一的数值来区分多个同时的数据流。SSRC必须是一个严格的随机数。



RTP Header中的Timestamp

- ◆ 时间戳单位：时间戳计算的单位往往不用秒之类的单位，而是由采样周期所代替，这样做可简化时间戳计算。
 - 如一个音频的采样频率为8000Hz，可设时间戳单位为1/8000s
- ◆ 时间戳计算的单位也可以采用设备的脉冲周期
 - MPEG2系统有90kHz的计数器，可设时间戳单位为1/90000
- ◆ 时间戳增量：相邻两个RTP包之间的时间差。
 - 假设时间戳单位为1/90kHz，若每秒发送25帧，那么每一个帧的发送占多少个时间块呢？当然是 $90000/25 = 3600$ 。根据定义“时间戳增量是发送第二个RTP包相距发送第一个RTP包时的时间间隔”，时间戳增量应该为3600。
- ◆ 注：时间戳也可以用其他方式来定义。



RTCP(Real Time Control Potocol)

◆ RTP本身并不能为按顺序传送数据包提供可靠的传送机制，也不提供流量控制或拥塞控制，它依靠RTCP传输控制协议提供这些服务。

□ RTP本身仅提供了同步需要的信息，但不具备同步的功能

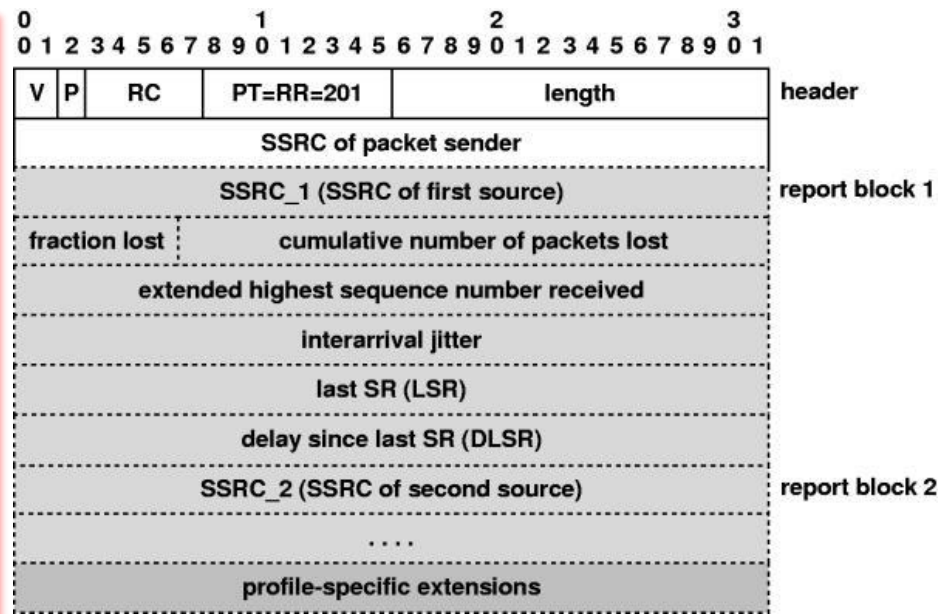
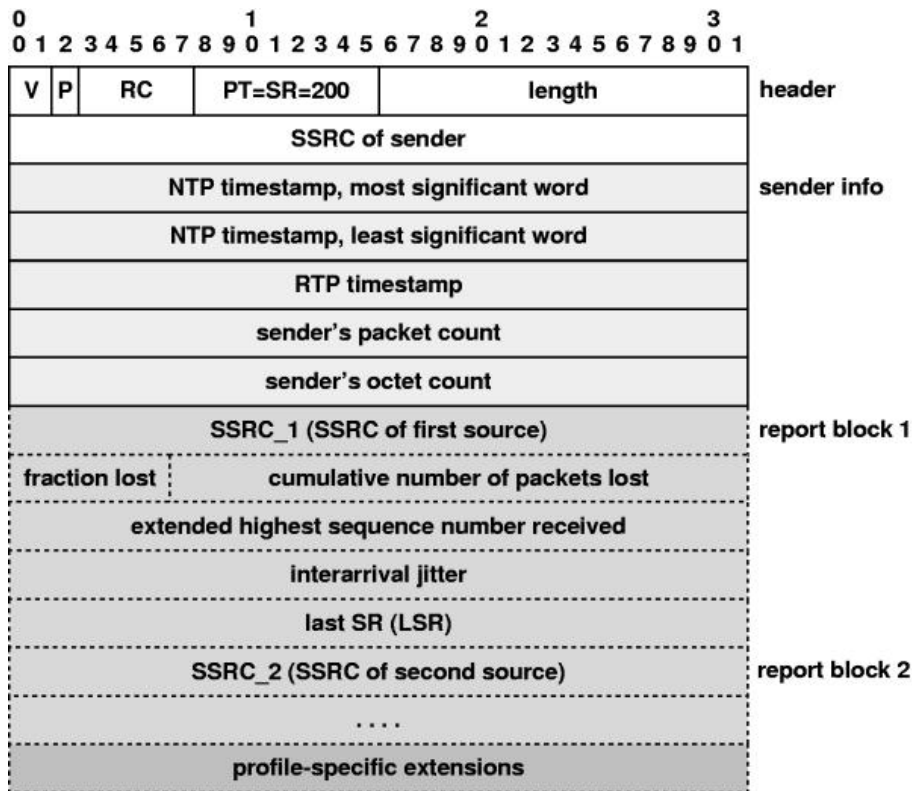
◆ RTP的控制协议RTCP通过在会话用户之间周期性地递交控制报文来完成监听服务质量和交换会话用户信息等功能。

□ RTCP数据报有几种类型如：**发送端报告（SR）**、接收端报告（RR）、源描述项（SDES）、结束标识（BYE）、应用程序说明（APP）等等。



RTCP SR(Sender Report)

RTCP RR(Receiver Report)



RTCP RR Header

RTCP SR Header

```

{
.....
NTP timestamp: 64 bits
RTP timestamp: 32 bits ←这里的timestamp与RTP报文中一致
.....
}

```

NTP (Network Time Protocol, 网络时间协议)



RTCP的NTP时间戳

◆ NTP的时间戳有MSW和LSW组成（共64bits）

□ MSW表示秒，LSW表示 $1/2^{32}$ 秒（约200ps）

□ NTP时间是绝对时间

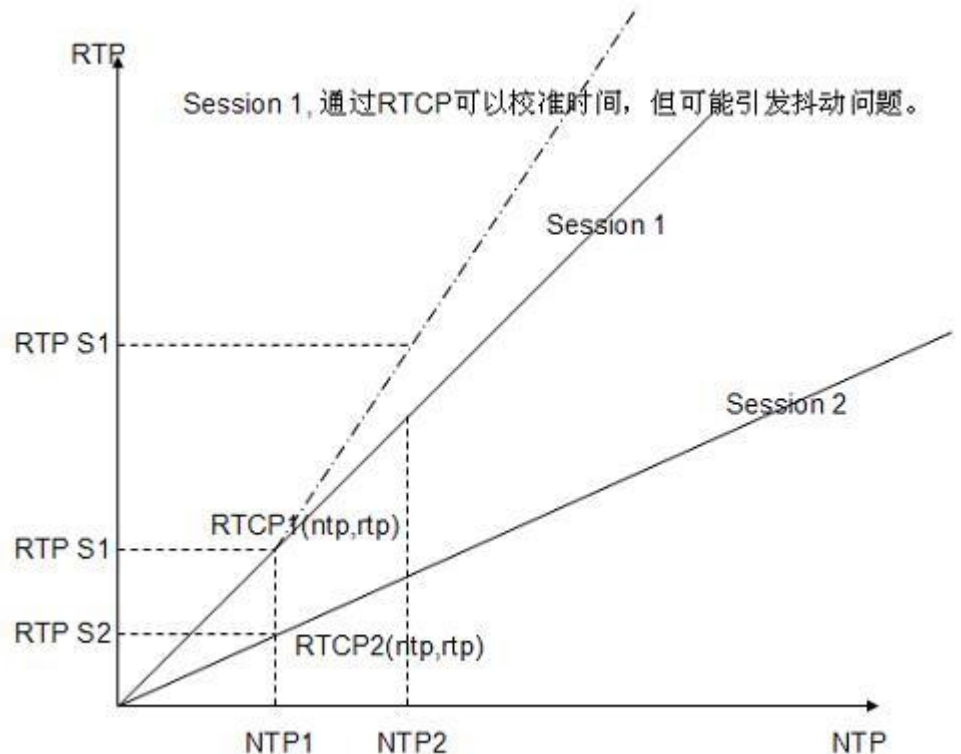
```
⊕ User Datagram Protocol, Src Port: 44721 (44721), Dst Port: 28902 (28902)
⊖ Real-time Transport Control Protocol (Sender Report)
  ⊕ [Stream setup by RTSP (frame 217)]
    10.. .... = Version: RFC 1889 Version (2)
    ..0. .... = Padding: False
    ...0 0000 = Reception report count: 0
    Packet type: Sender Report (200)
    Length: 6 (28 bytes)
    Sender SSRC: 0x6a3708ca (1781991626)
    Timestamp, MSW: 2212707919 (0x83e33e4f)
    Timestamp, LSW: 2662914985 (0x9eb8dba9)
    [MSW and LSW as NTP timestamp: Feb 13, 1970 01:05:19.6200 UTC]
    RTP timestamp: 1884142806
    Sender's packet count: 227
```



根据RTCP SR可实现流的时间同步

在每个RTCP SR包中对应有一个RTP时间和一个NTP时间。不同stream的RTP时间尽管不同，但可以通过NTP时间映射到同一个时间轴上，从而实现同步。

如图所示，RTP session 1 发送H.264 使用90000Hz，而RTP session 2 发送G.711 使用8000HZ。





H.264可基于RTP/RTCP实现流内同步

◆ H.264 Payload 格式定义了三种不同的基本的负载 (Payload) 结构。可能的结构类型分别有：

- 单一NAL单元模式：一个 RTP 包仅由一个 NALU 组成。
- 组合封包模式：即可能是由多个 NAL 单元组成一个 RTP 包。分别有4种组合方式：STAP-A、STAP-B、MTAP16、MTAP24。对应类型值分别是 24, 25, 26 以及 27。
- 分片封包模式：把一个 NALU 单元封装成多个 RTP 包。存在两种类型 FU-A 和 FU-B。类型值分别是 28 和 29。

RTP Payload的第一个字节包含Payload的Type

- 24 STAP-A** 单一时间的组合包
- 25 STAP-B** 单一时间的组合包
- 26 MTAP16** 多个时间的组合包
- 27 MTAP24** 多个时间的组合包
- 28 FU-A** 分片的单元
- 29 FU-B** 分片的单元

H.264的Network Friendly特性：
H.264视频流与其他流的复用可以方便地基于RTP实现



小结：基于RTP/RTCP实现流同步

- ◆ RTP提供必要的流识别与同步信息
 - 时戳(Timestamp)
 - 同步源标识符(SSRC Identifier)
- ◆ RTCP SR(Sender Report)
 - 给接收方提供了同步的完整信息
- ◆ NTP timestamp: 64 bits
 - MSW表示秒，LSW表示 $1/2^{32}$ 秒（约200ps）
- ◆ RTP timestamp: 32 bits
 - 时间戳计算的单位一般是对应流的采样周期

MC 主要内容

- ◆ 什么是多媒体同步
- ◆ 如何描述多媒体同步
- ◆ 典型多媒体设备接口信号的同步机制
 - CD-DA 音频的流内同步
 - 模拟电视信号的行同步与场同步
 - ITU-R BT. 601 数字视频图像帧的同步
 - HDMI 数字多媒体接口的同步
- ◆ 典型多媒体系统的媒体同步方法
 - MP3流同步
 - DVD/HDTV的流同步
 - 基于RTP的流媒体同步
 - **MMT: MPEG Media Transport**
- ◆ 展望未来



MMT: MPEG Media Transport

Developed by the MPEG as a replacement of MPEG-2 TS

◆ MPEG视频流在不同网络中广泛使用，然而，188bytes的MPEG2-TS格式并不适合在IP网络中传输。为适应在IP网络中传输视频，MPEG组织开发了新的标准，名为MPEG Media Transport (MMT)。并在2014年5月发布了第一个正式版本。

- ISO/IEC 23008 High Efficiency Coding and Media Delivery in Heterogeneous Environments (MPEG-H) standard suite.
- ISO/IEC 23008-1 defines MMT Protocol and related technologies, for example, (AL-FEC) framework, buffer model, signaling messages and the data model.
- ISO/IEC 23008-10 defines AL-FEC (Application Layer Forward Error Correction) codes.
- ISO/IEC 23008-11 defines Composition Information, technologies for spatio-temporal layout of the multimedia components in a multiscreen, multi-device environment.

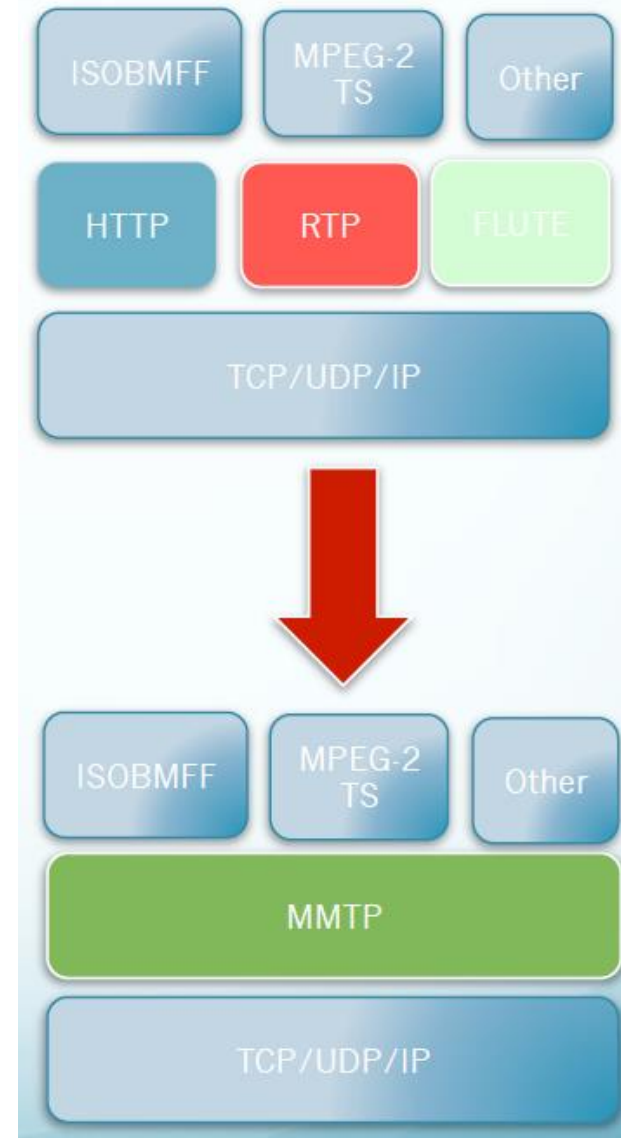


从RTP到MMTP

MMTP: MPEG Media Transport Protocol

Why not RTP?

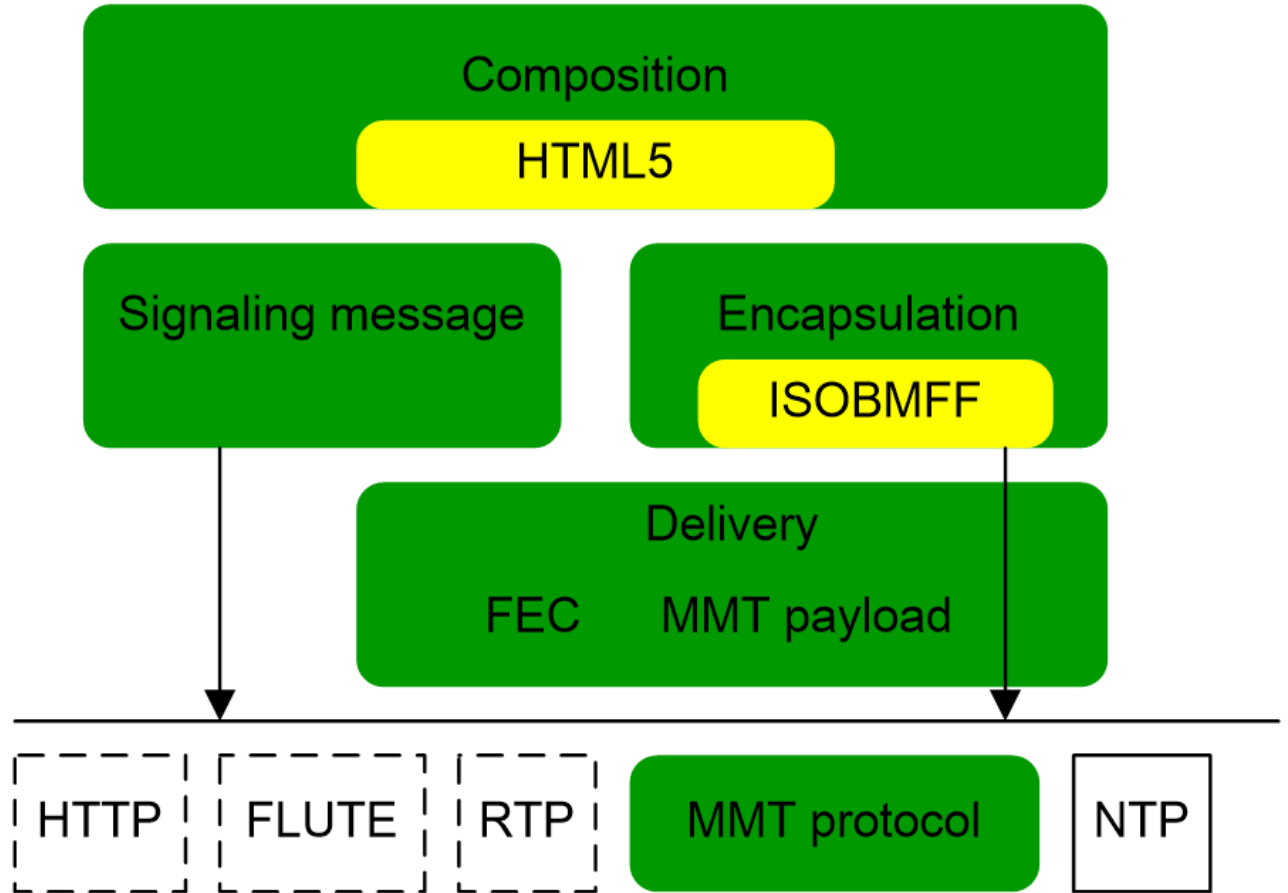
- Lack of Multiplexing: One media session per component and without RTP multiplexing, 2 ports per session
- Server Maintenance: RTP Payload Format for every new media codec; Support needs to be added to the media server.
- Coupling of Presentation and Delivery: RTP carries presentation and synchronization information at the transport level
- Limited support for Non-Real Time Media: Presentations consist of timed and non-timed media; Need other protocol or countless number of payload formats to support NRT.





MMT protocol Stack

- **Encapsulation**
- **Delivery**
- **Signaling**
- **Composition**



 : Defined by MMT

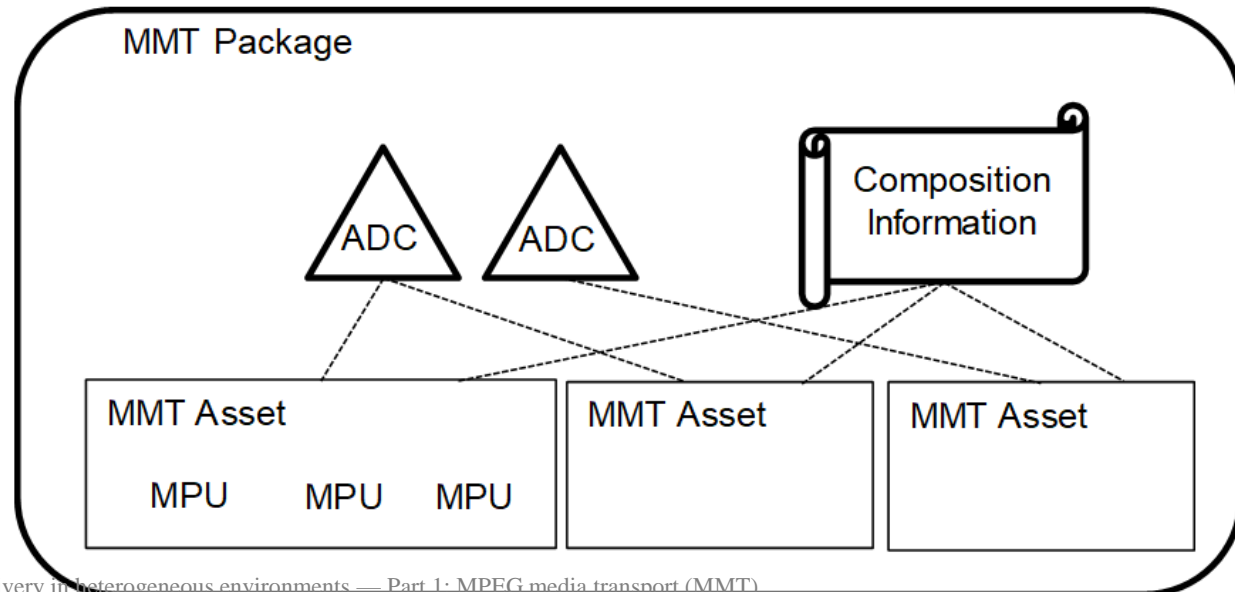
 : Referenced by MMT



Encapsulation

Logical structure of MMT Content

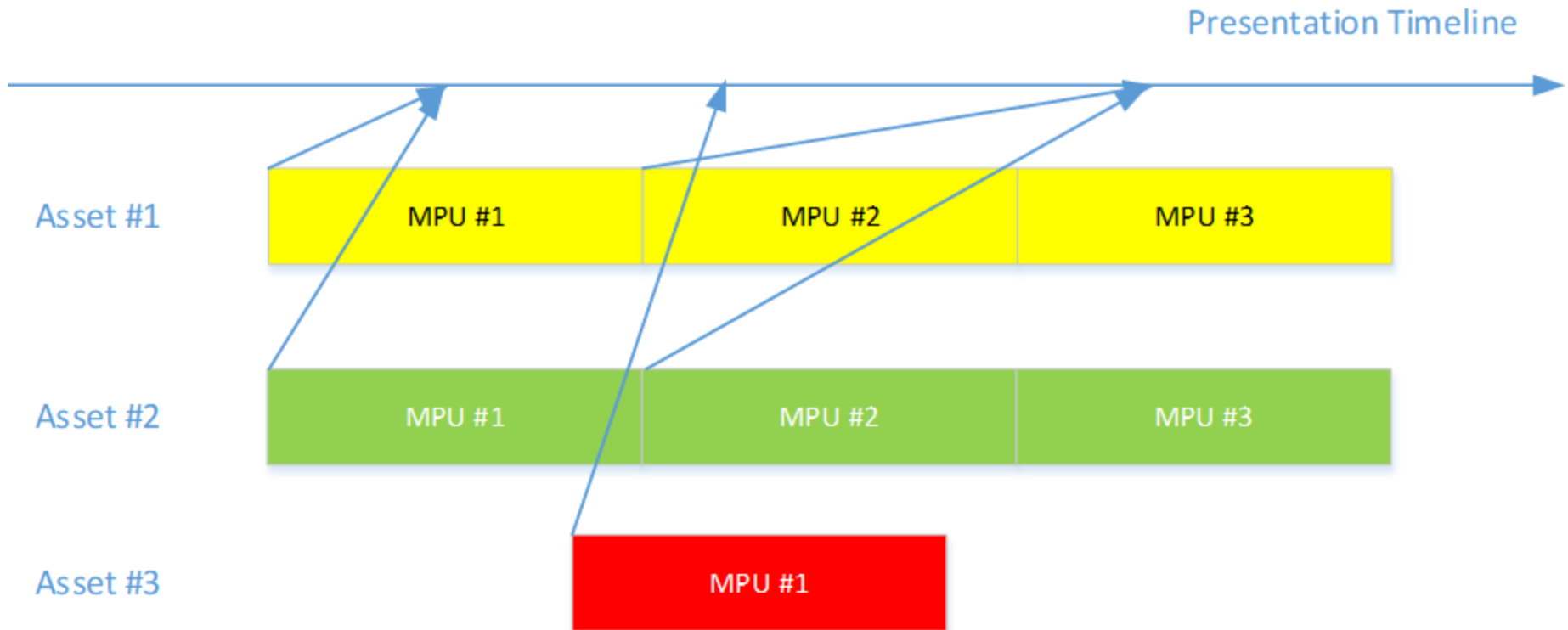
- Composition Information (**CI**) shall specify the **spatial and temporal relationship** among the Assets for consumption and presentation by using HTML5.
- Asset Delivery Characteristics (**ADC**) shall provide the required **QoS information** for transmission of Assets.
- A Media Processing Unit (**MPU**) is **media data** which may be independently and completely processed by an MMT entity and consumed by the media codec layer. Processing of an MPU by an MMT entity includes encapsulation, decapsulation and (de)packetization.





Encapsulation

Mapping MPUs to the presentation timeline

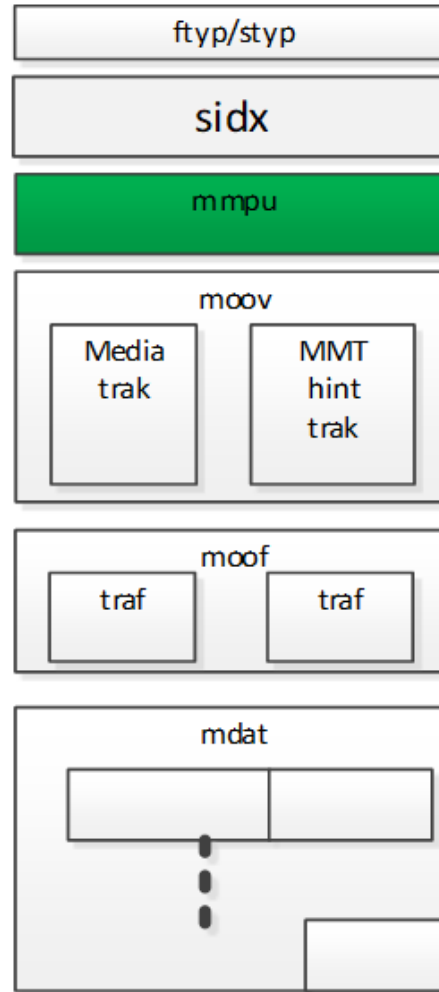




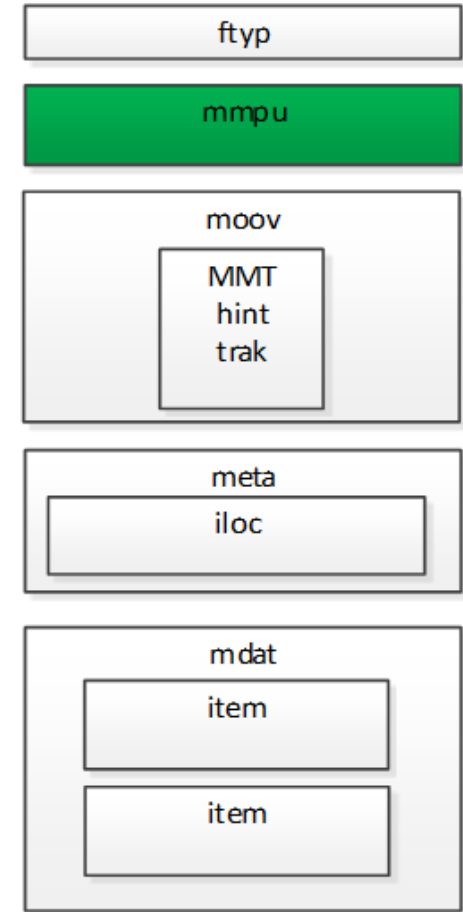
Encapsulation

Encapsulation of MPU

An MPU shall be **encapsulated as a single ISOBMFF file**. The sequence number and Asset ID of the MPU is provided in the 'mmpu' box to uniquely indentify the MPU encapsulated in the file. Additionally, in case of timed media, a 'sidx' box may be present to index movie fragments comprising MPU. As each MPU is self-contained, each file encapsulating MPU shall include the initialization data by using the 'ftyp', 'styp', and 'moov' boxes. The 'moov' box shall contain all codec configuration information for decoding and presentation of media data in MPU.



a) MPU with Timed Media



b) MPU with Non-timed Media



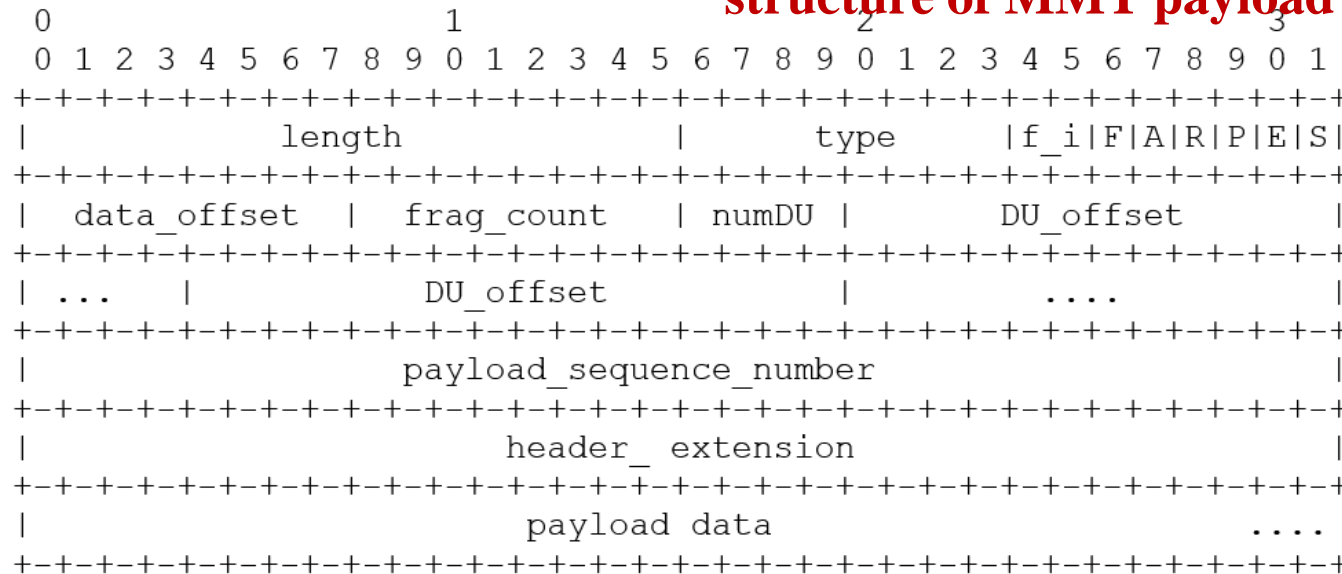
Delivery

MMTP packet: MMT payload header

An MMTP packet is mainly composed of three parts:

- i) packet header
- ii) payload data
- iii) AL-FEC (Application Layer Forward Error Correction) information

structure of MMT payload header



number_data_unit (numDU: 4 bits) – This field specifies the number of data unit within this MMT payload. This field shall be ‘0’, if fragmentation_flag is set to ‘1’.

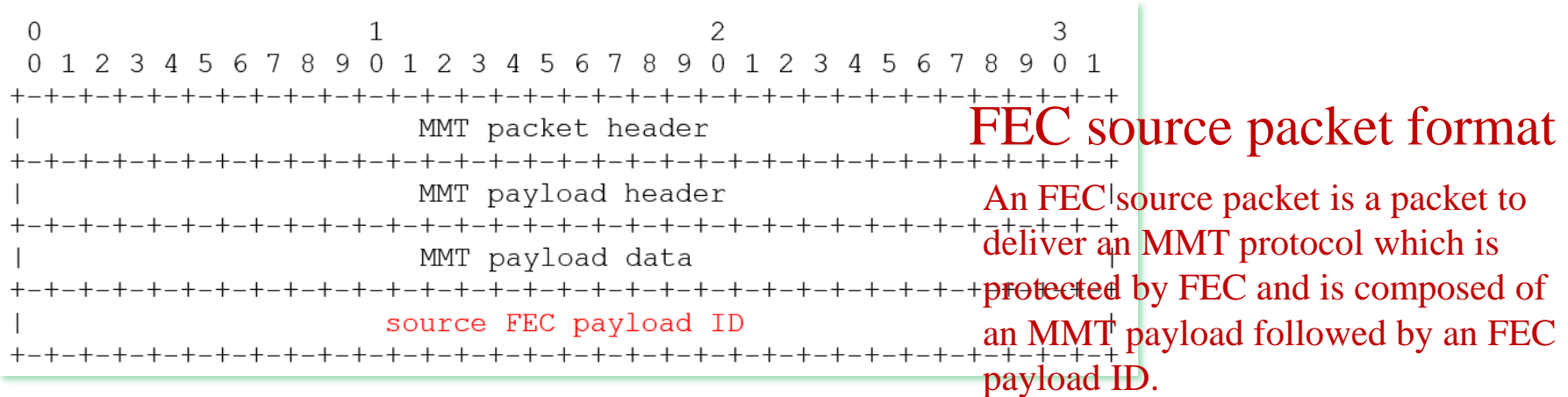
DU_offset (16bits) – This field specifies location of each data unit from the byte indicated by data_offset. This field shall be used, when aggregation_flag is set to ‘1’.



Delivery

MMTP packet: AL-FEC (Application Layer Forward Error Correction)

MMT provides the application level forward error correction (FEC) mechanism for reliable delivery in IP network environments that are prone to packet losses.

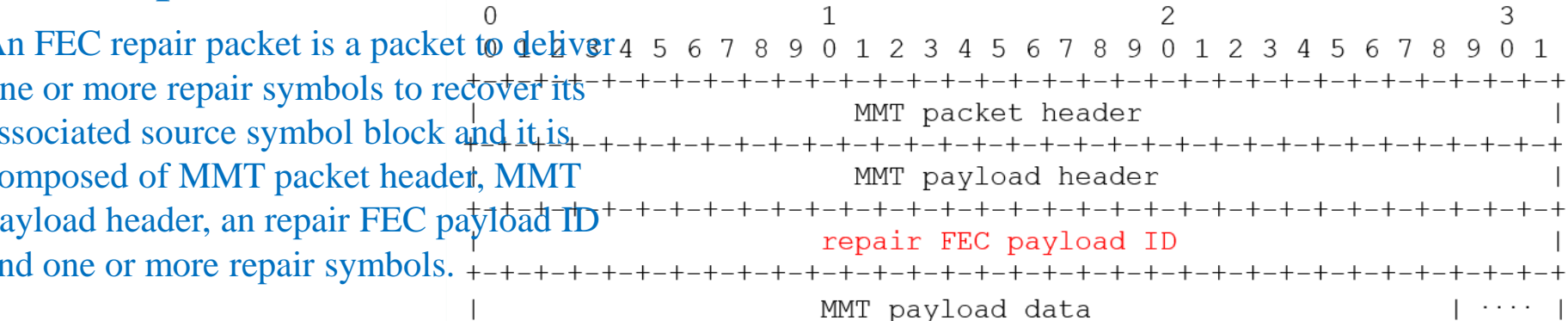


FEC source packet format

An FEC source packet is a packet to deliver an MMT protocol which is protected by FEC and is composed of an MMT payload followed by an FEC payload ID.

FEC Repair Packet Format

An FEC repair packet is a packet to deliver one or more repair symbols to recover its associated source symbol block and it is composed of MMT packet header, MMT payload header, an repair FEC payload ID and one or more repair symbols.





Signaling

Message for Consumption

- ◆ Six messages are defined that relate to the consumption of the Package;
 - **Package Access (PA)** message: It includes all tables required for Package access including MMT Package Table and Composition Information Table;
 - **MMT Composition Information (MCI)** message: It includes MCI table encapsulating a complete CI or a subset of CI. It may also include MPT corresponding MCI table for fast Package consumption;
 - **MMT Package Table (MPT)** message : It includes MMT Package Table providing a whole information or a part of information required for a single Package consumption;
 - **Clock Relation Information (CRI)** message: it includes CRI Table providing the clock relation information used for the **mapping between the NTP Clock and MPEG-2 System Time Clock**;
 - **Device Capability Information (DCI)** message: it includes DCI table providing the required device capability information for a Package consumption;
 - **Security Software Request (SSWR)** message: it is used to request security software for consuming MMT Package or Asset by a MMT receiving entity. It can also include PA table or MPT.



Signaling Messages for Delivery

◆ Six messages are defined that relate to the delivery of the Package:

- ❑ **Measurement Configuration (MC)** message: it provides information to configure a measurement;
- ❑ **Application Layer Forward Error Correction (AL-FEC)** message: it provides AL-FEC configuration information;
- ❑ **Hypothetical Receiver Buffer Model (HRBM)** message: it provides information to configure HRBM operation;
- ❑ **Automatic Repeat-Request (ARQ)** message: It provides information required for ARQ operation;
- ❑ **Reception Quality Feedback (RQF)** message: It defines a format of measurement report from a receiving entity;
- ❑ **Network Aware Media Feedback (NAMF)** message: It defines a format of NAM parameter report from a receiving entity.

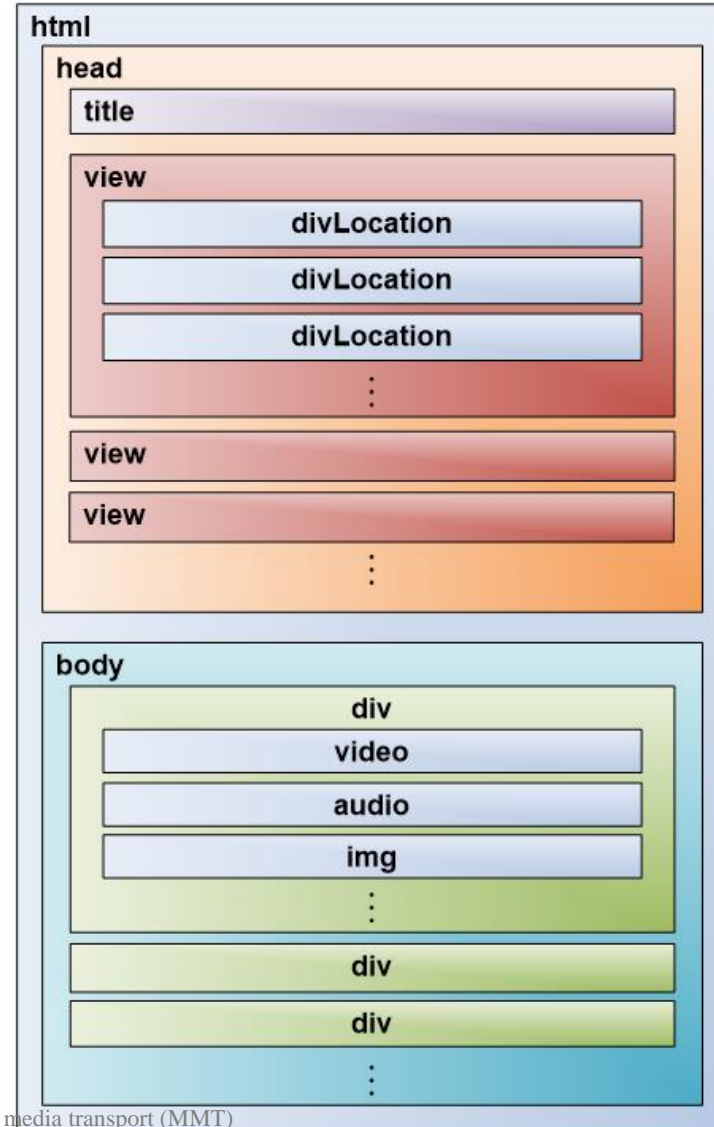
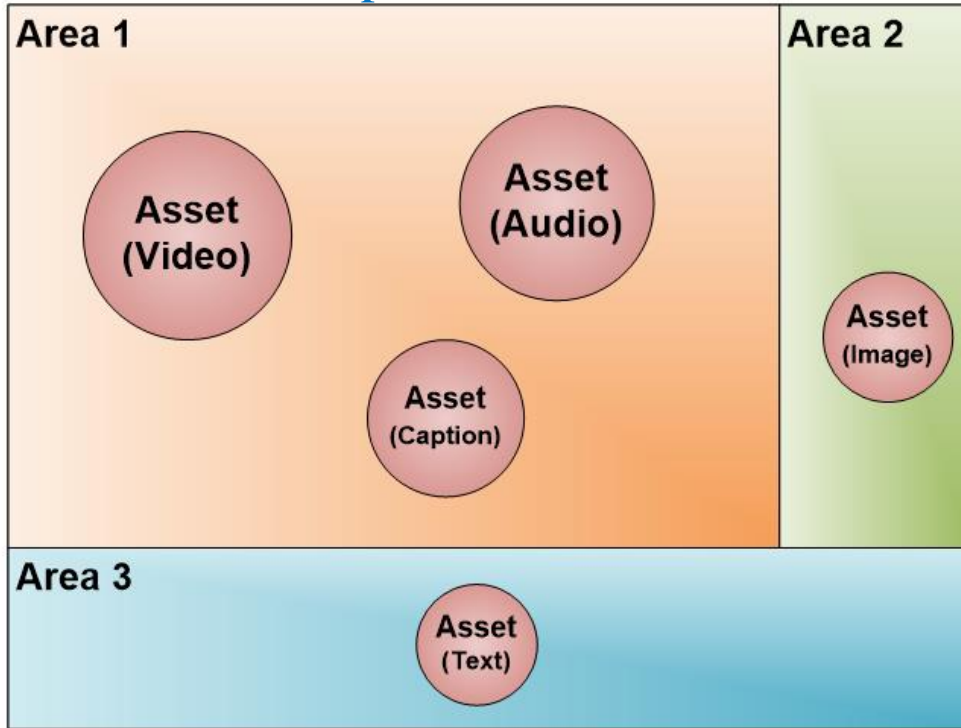


Composition

Composition Information (CI)

Basic Structure of CI

View Example of Asset/Area/View





Composition

EXAMPLE OF MMT CI DOCUMENT

Composition Information (CI)语法示例

```
<!-- MMT-CI:view Type -->
<xsd:complexType name="MMT-CI:viewType">
  <xsd:complexContent>
    <xsd:extension base="HTML5:HTMLElementType">
      <!-- The view element inherits HTML5:HTMLElementType interface which is specified in 3.3.2 of[W3C HTML5]-->
      <xsd:sequence>
        <xsd:element name="divLocation" type="MMT-CI:divLocationType" minOccurs="1"
maxOccurs="unbounded"/>
      </xsd:sequence>
      <xsd:attribute name="begin" type="MMT-CI:beginType" default="0s"/>
      <xsd:attribute name="end" type="MMT-CI:endType" default="indefinite"/>
      <xsd:attribute name="dur" type="MMT-CI:durType"/>
      <xsd:attribute name="type" type="xsd:token" fixed="simple"/>
      <xsd:attribute name="href" type="xlink:hrefType"/>
      <xsd:attribute name="show" type="xsd:token" fixed="embed"/>
      <xsd:attribute name="actuate" type="xlink:actuateType" default="onRequest"/>
      <xsd:attribute name="viewtype" type="MMT-CI:viewtype" default="default"/>
    </xsd:extension>
  </xsd:complexContent>
</xsd:complexType>
```

MC 小结: MMT

◆ MMT: MPEG Media Transport

□ 替代MPEG-2 TS

◆ Encapsulation

◆ Delivery

□ NTP Timestamp

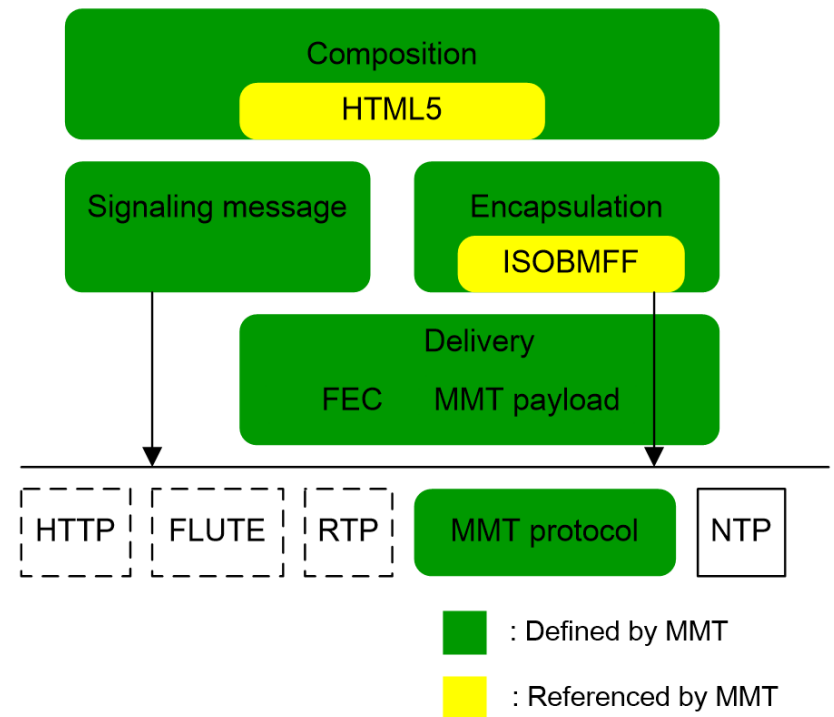
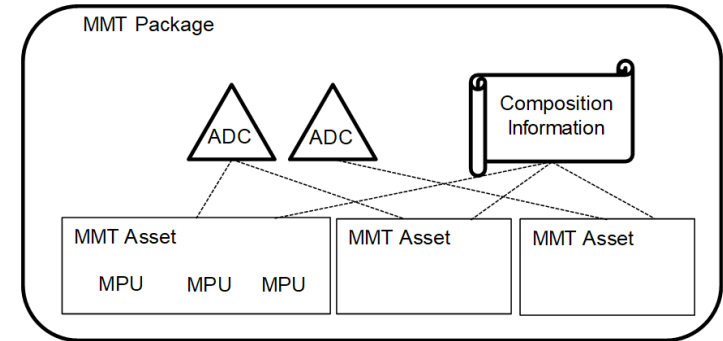
□ AL-FEC

◆ Signaling

□ Clock Relation Information

◆ Composition

□ HTML5





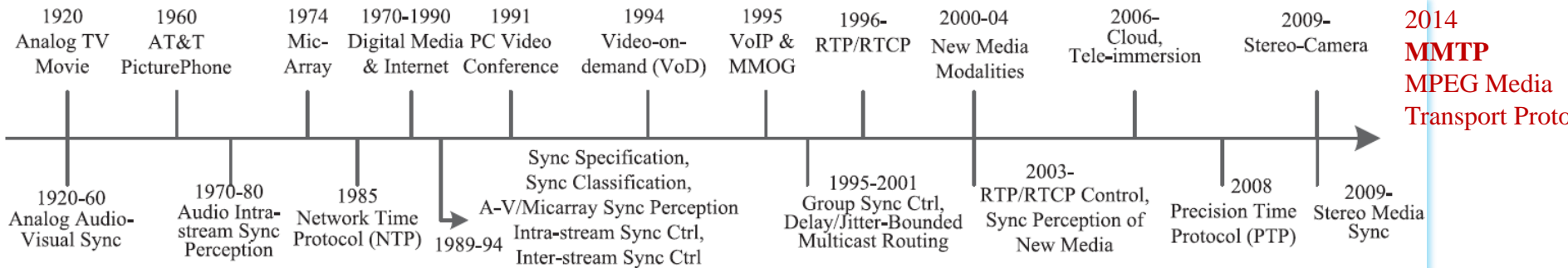
小结：典型多媒体系统媒体的同步

- ◆ MP3如何在播放时进行时间控制？
 - Samples Per Frame
 - Sampling Rate
- ◆ DVD/HDTV中不同媒体流如何实现同步？
 - ES→PES→TS
 - PCR/DTS/PTS
- ◆ RTP/RTCP的流同步如何实现？
 - SSRC Identifier
 - NTP timestamp
 - RTP timestamp
- ◆ MMT: MPEG Media Transport
 - NTP timestamp
 - Clock Relation Information

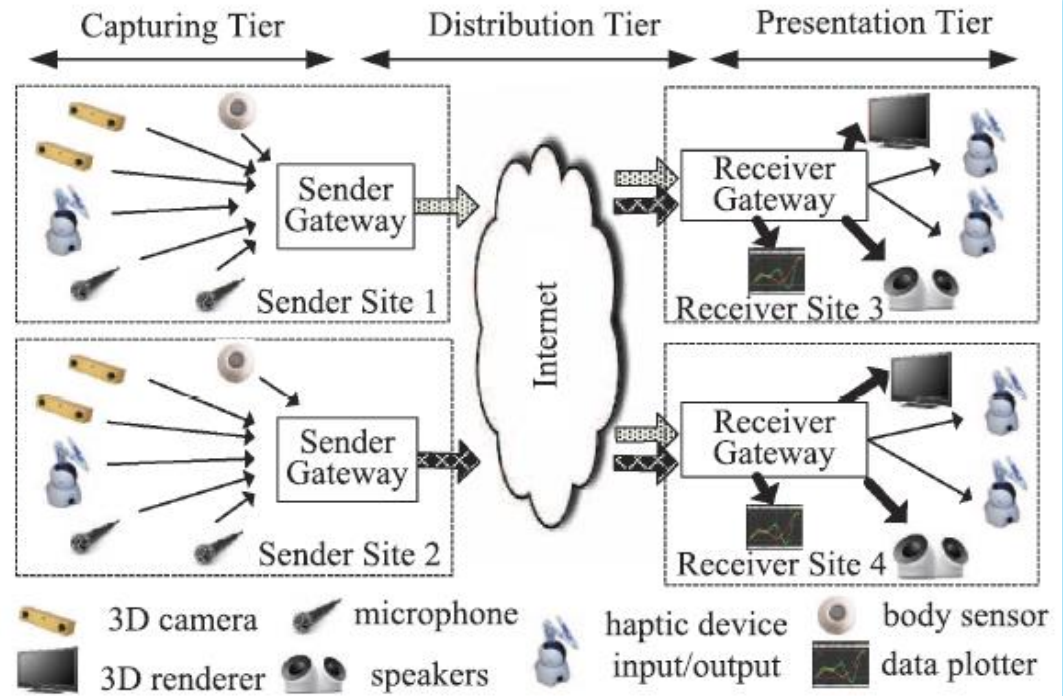


总结：过去的多媒体同步研究

timeline of multimedia and synchronization technologies



- Intra-Stream Synchronization**
- Intra-Media Synchronization**
- Multiview-Videos**
- Spatial Audios**
- Haptic or Body Sensory Streams**
- Intra-Bundle Synchronization**
- Intra-Session Synchronization**



MC 主要内容

- ◆ 什么是多媒体同步
- ◆ 如何描述多媒体同步
- ◆ 典型多媒体设备接口信号的同步机制
 - CD-DA 音频的流内同步
 - 模拟电视信号的行同步与场同步
 - ITU-R BT. 601 数字视频图像帧的同步
 - HDMI 数字多媒体接口的同步
- ◆ 典型多媒体系统的媒体同步方法
 - MP3流同步
 - DVD/HDTV的流同步
 - 基于RTP的流媒体同步
- ◆ 展望未来



将音频数据同步隐藏到帧间编码

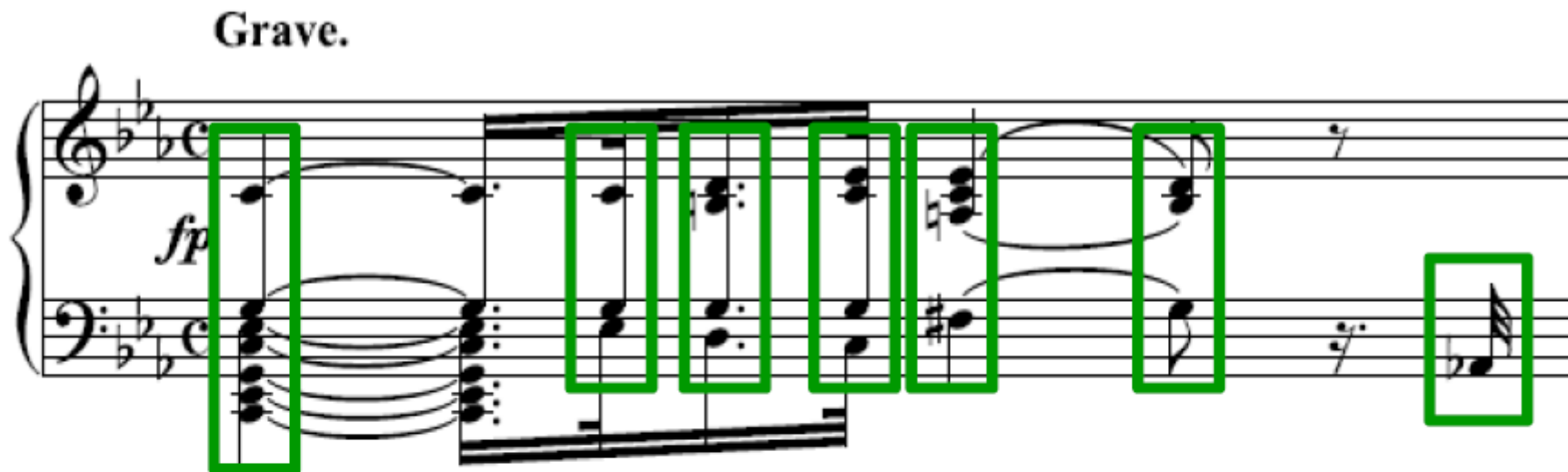
◆ 帧内编码主要应用于整个视频序列的第一帧、I 帧以及 P 帧中的部分宏块。如果使用帧内编码宏块进行音频信号的嵌入，既会严重影响视频图像的质量，而且嵌入量也相对较小，不容易满足音视频同步编码的要求。

◆ 在 H.264 帧间编码中，除了 SKIP 模式，还有 7 种不同的编码模式：mode1 (Inter 16×16)、mode2 (Inter 16×8)、mode3 (Inter 8×16)、mode4 (Inter 8×8)、mode5 (Inter 8×4)、mode6 (Inter 4×8)以及 mode7 (Inter 4×4)。作者利用帧间编码模式的多样性，使得帧间编码模式和要嵌入的音频信号之间产生映射关系，使得不同的帧间编码模式携带不同的音频信息。实验结果显示不会对视频产生太大影响。

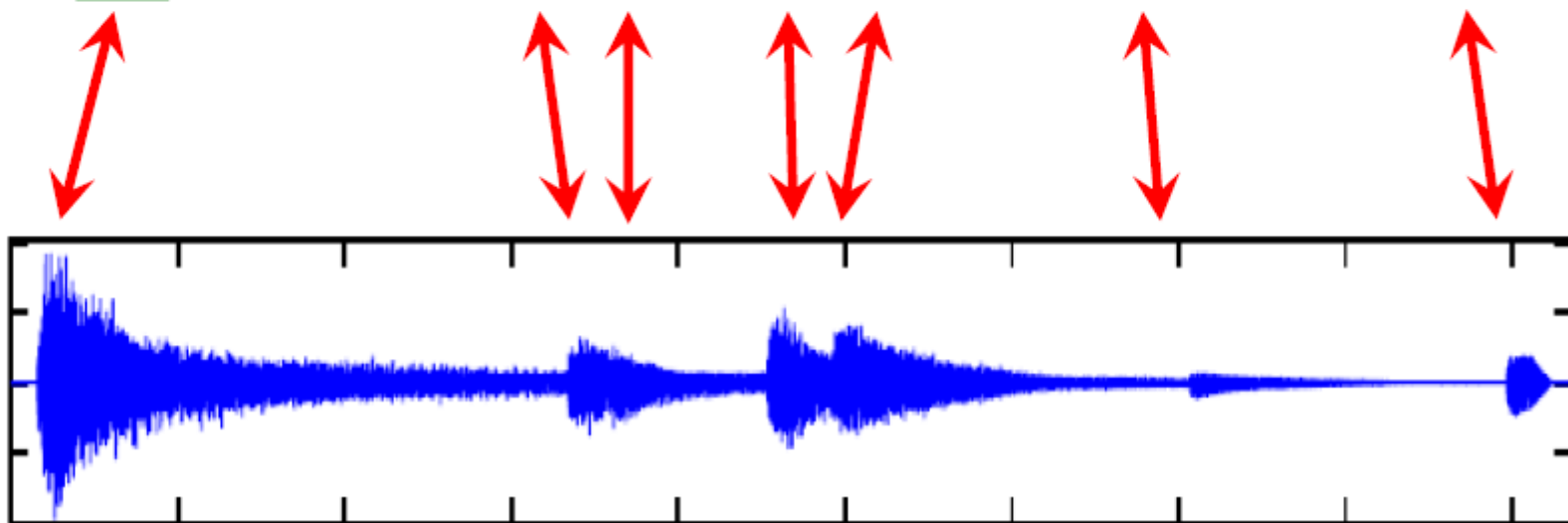


自动读谱演奏?

Image



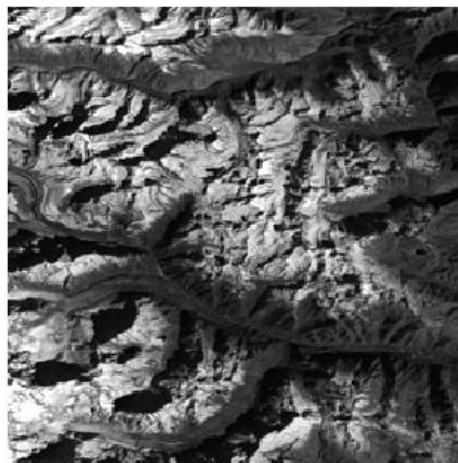
Audio



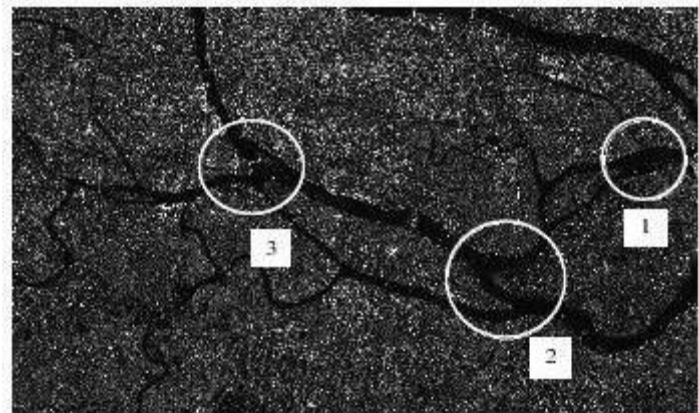
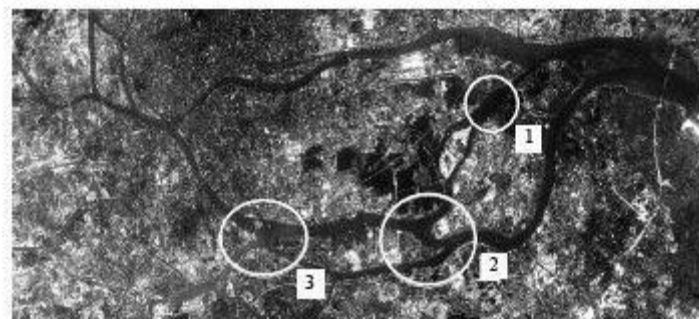
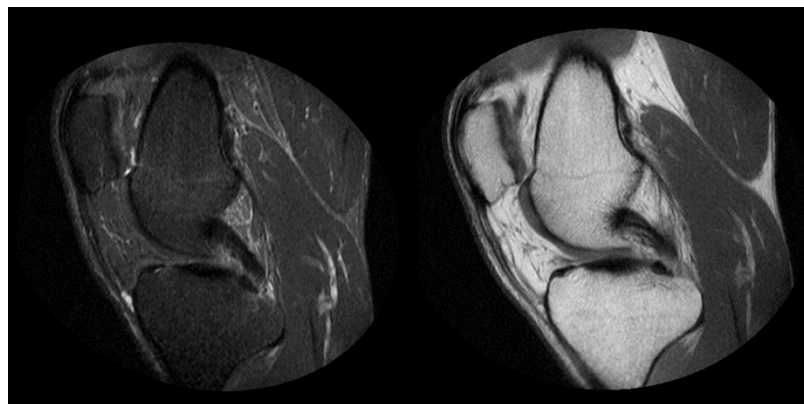
MC 多光谱图像配准



(a) 参考图像



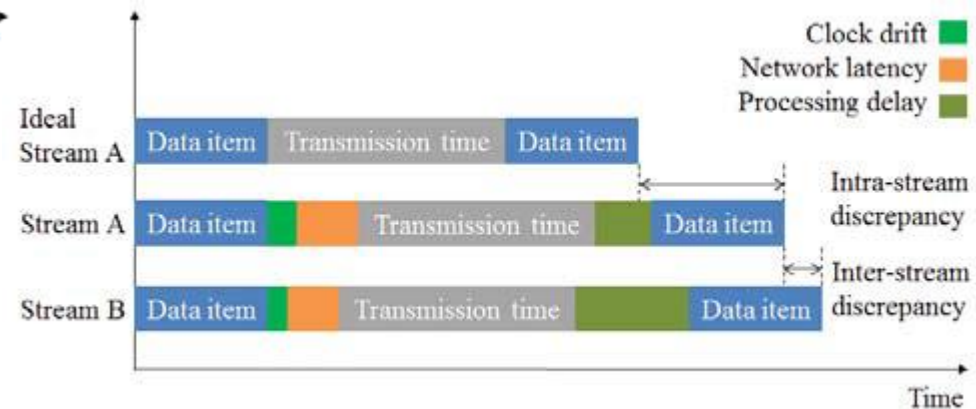
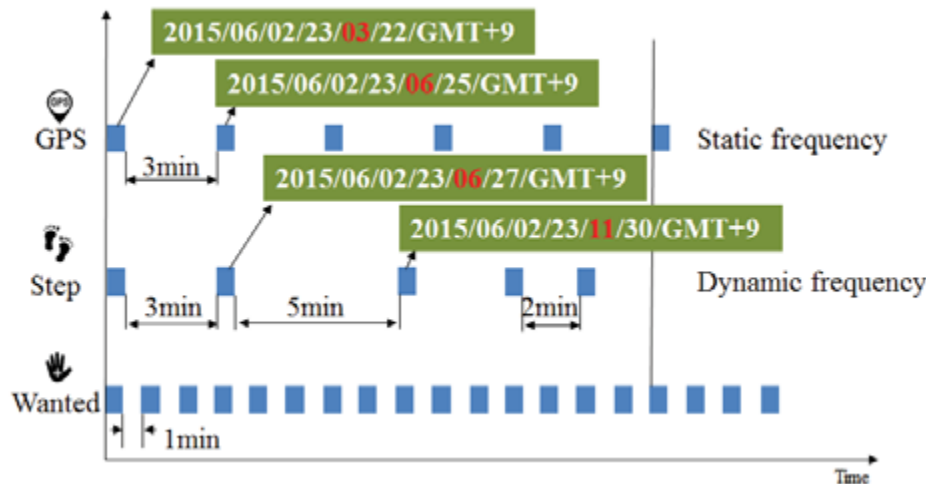
(b) 待配准图像





不同便携设备采集数据的同步

Traditional applications directly deal with raw data, regardless of whether there are discrepancies, thus lead to inaccurate analysis results. In this context, to improve the quality of the data and make the results more reliable, it is necessary to consider the time discrepancy among multiple wearable devices.





云存储：文件的同步/备份



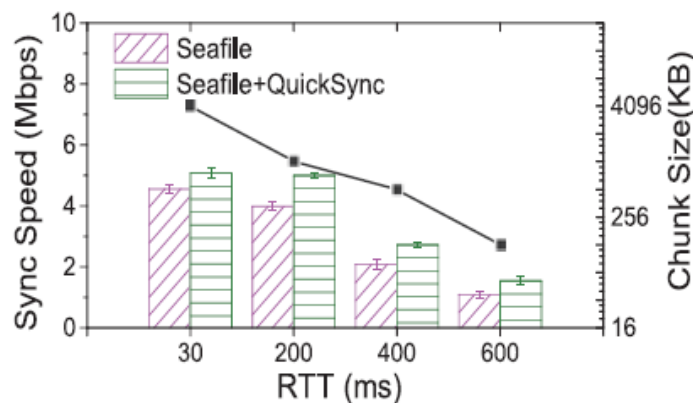


云存储同步性能

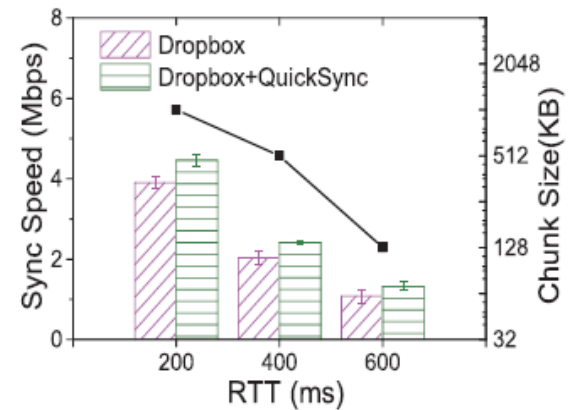
Capability Implementation of Four Popular Cloud Storage Services

Capabilities	Windows				Android			
	Dropbox	Google Drive	OneDrive	Seafile	Dropbox	Google Drive	OneDrive	Seafile
Chunking	4 MB	8 MB	var.	var.	4 MB	260 KB	1 MB	×
Bundling	✓	×	×	×	×	×	×	×
Deduplication	✓	×	×	✓	✓	×	×	×
Delta encoding	✓	×	×	✓	×	×	×	×
Data compression	✓	✓	×	×	×	×	×	×

In this paper, we identify, analyze, and address the **synchronization (sync) inefficiency problem** of modern mobile cloud storage services. Our measurement results demonstrate that existing commercial sync services fail to make full use of available bandwidth, and generate a large amount of unnecessary sync traffic in certain circumstances even though the incremental sync is implemented.



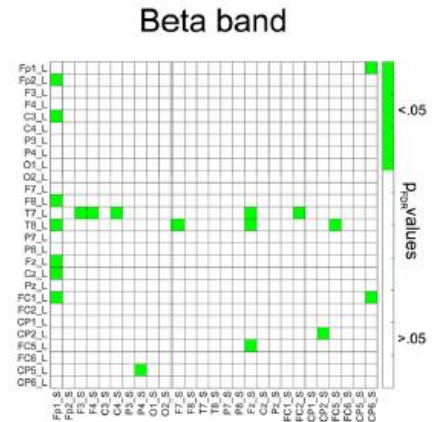
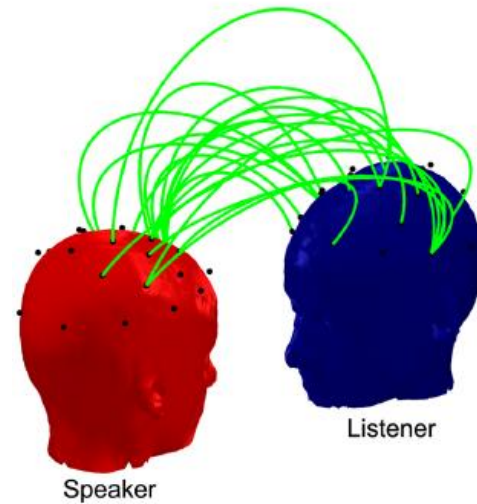
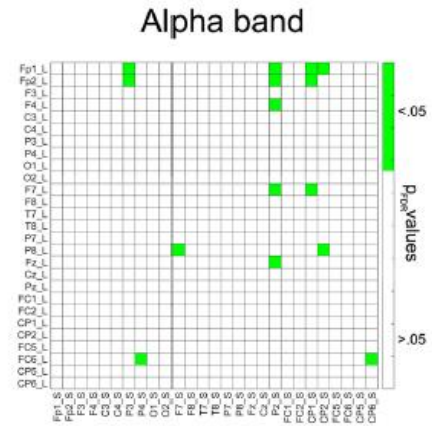
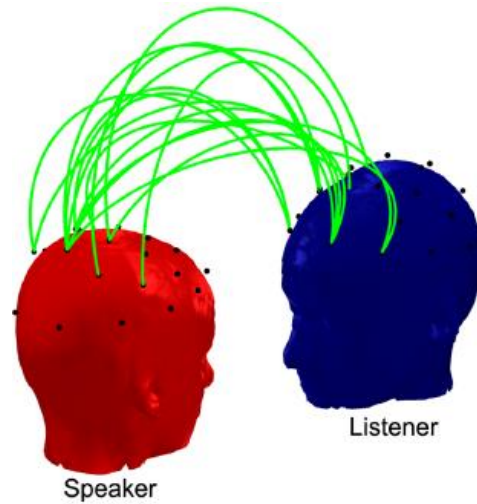
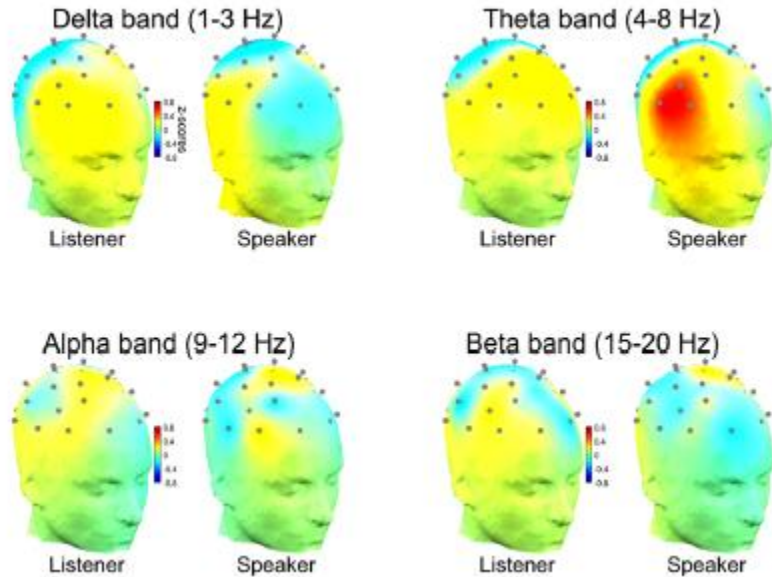
(a) Seafile platform



(b) Dropbox platform

MC 说者、听者脑电的同步

Results show that brain oscillations are synchronized between listener and speaker during oral narratives. This interpersonal synchronization is mediated in part by a lower-level sensory mechanism of speech-to-brain synchronization, but also by the interactive process that takes place in the situation per se.





从空间区域看用户群体行为

- ◆ 通过WiFi热点的流量将具有类似行为的用户进行聚类，即刻画用户的群体行为
 - track the attention of the crowds during the 2006 FIFA World Cup final
 - F. Calabrese et al., “Eigenplaces: Segmenting Space through Digital Signatures,” IEEE Pervasive Computing, vol. 9, no. 1, 2010, pp. 78–84.
- ◆ 通过分析社交媒体（如Twitter）的footprint（如关键字标注）对聚类后的人群所讨论的话题内容进行分析
 - CitySensing: Fusing City Data for Visual Storytelling, IEEE Multimedia, July–September 2015



通过social media的footprint挖掘时空关系

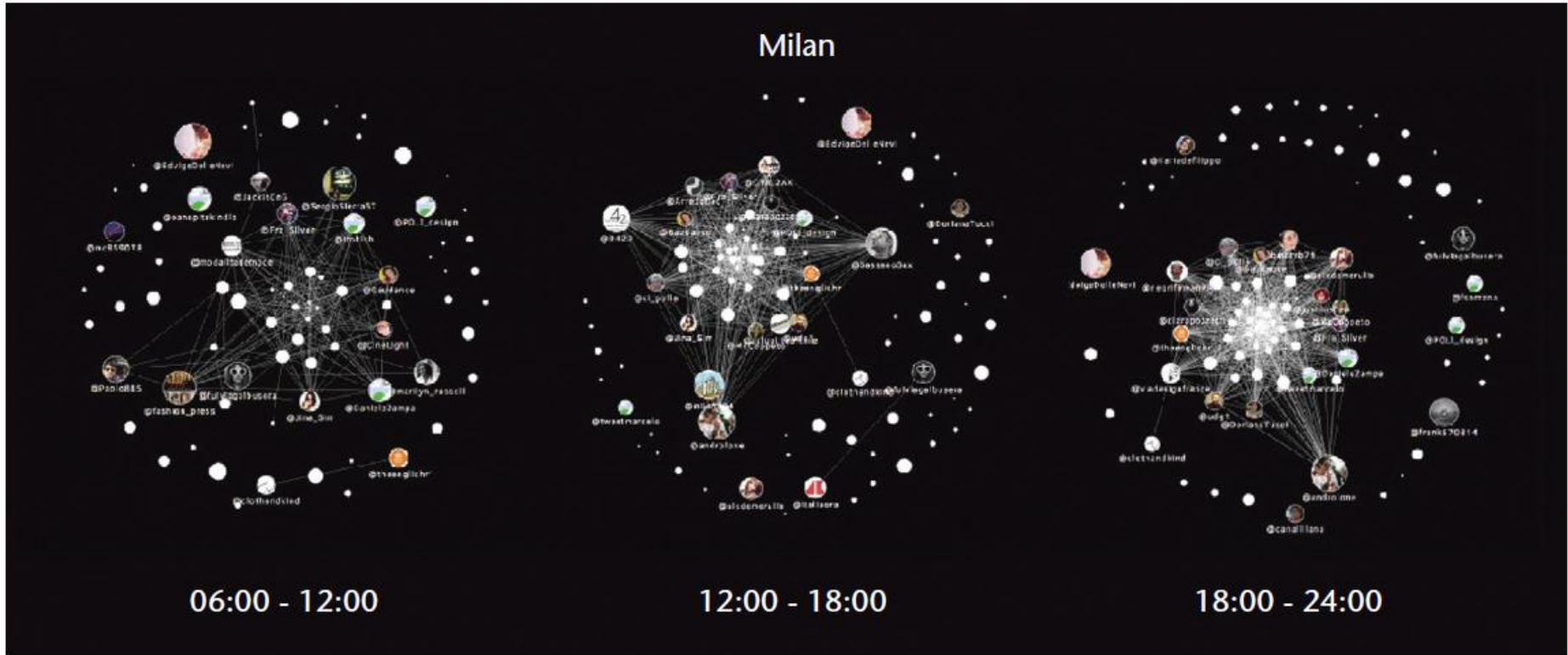


Figure 7. This graph view highlights the scale-free nature of the graph of people when connected by the places, events, and hashtags they discuss on social media.



分布式多媒体同步的新问题

◆ 点到点系统 ← 时间

- 流内同步（Intra-stream Synchronization）是指在单一连接上的单一媒体流内各个媒体单元之间存在的时间关系。
- 流间同步（Inter-stream Synchronization）是指在多个连接上的相关媒体流之间存在的时间关系。

◆ 多点系统：用户间的同步 ← 空间

- 一个发送者，多个接收者之间的同步
- 多个发送者，一个接收者同步不同来源的信息
- 多个发送者，多个接收者情形

◆ 内容同步：在内容理解的基础上 ← 语义

- 多源信息融合？

多点参与的系统约始于2000，内容同步约始于2010……

MC 总结：多媒体同步

- ◆ 什么是多媒体同步
 - 时间、空间；同步容限
- ◆ 如何描述多媒体同步
 - SMIL、基于Petri Net的并发系统描述
- ◆ 典型多媒体设备接口信号的同步机制
 - 行消隐、场消隐
 - 消隐信号持续时间传送音频等信息
- ◆ 典型多媒体系统的媒体同步方法
 - 各种时间戳的产生机制和时间计量方法
 - MPEG2-TS→RTP(网络友好的H.264)→MMT
- ◆ 展望未来：时间/空间/内容/设备...多维度的同步？