

### 4.3 打包基本码流及定时信息

音频或视频信号经过编码后得到基本码流，需要经过打包后再送入传输系统或节目系统中。基本码流经过打包后形成的码流称为打包基本码流（PES），其数据包的长度可变，对于视频数据来说，一般取一帧视频图像的基本码流作为数据包的长度，音频数据则以一个音频帧作为数据包的长度。在 PES 数据包的头部包含有关该 PES 数据包的重要信息，如 PES 数据包类型识别，PES 数据包中数据的解码或播放时间，用来再生视频同步并保持视频与音频的同步播放。

#### 4.3.1 基本码流

PES 数据包在理论上属传输层功能的一部分，应该在复用器中产生，但在实际中一般由信源编码器产生。一个 PES 包由包起始码、PES 头标志、PES 包头域和净负荷组成，如图 4-8 所示。

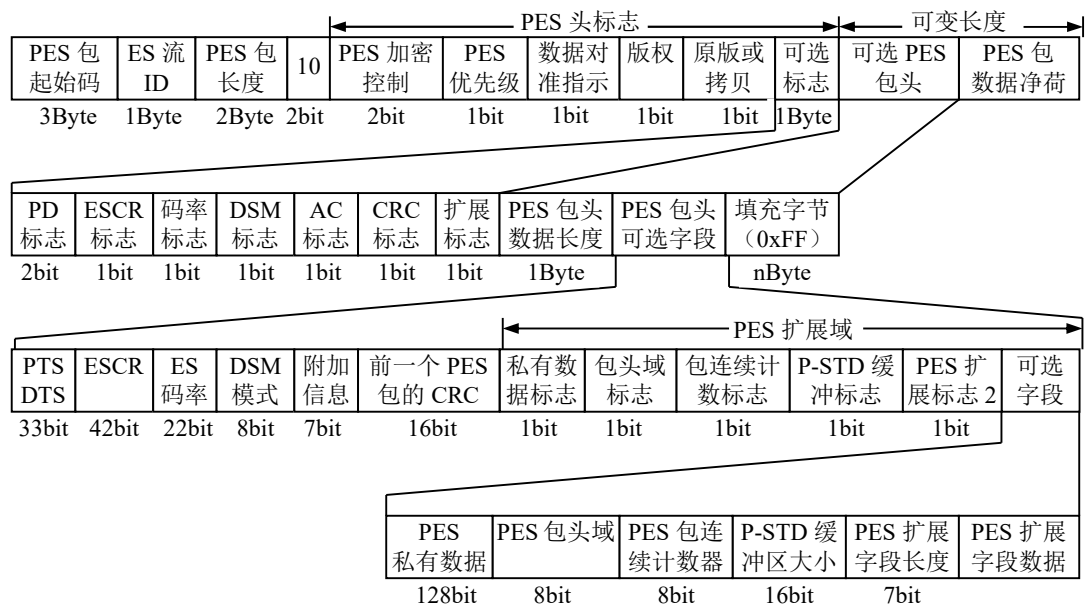


图 4-8 PES 数据流的组成

在 PES 数据的前部有一个 PES 包头数据，它包含了由编码器设置的有关当前 PES 数据的码率、定时及数据描述等基本信息。

PES 包的起始码沿用 ES 流的格式，是一个固定的码字结构，由 23 个“0”和 1 个“1”构成，用于收发两端对 PES 包进行同步。

紧跟在包起始码后面的 8 比特数据为流识别符，用于指示 PES 包承载的 ES 流类型。编码器所生成的每一个 ES 码流均被分配了唯一的识别标志—ID 号，其取值范围在 0xB9 和 0xFF 之间，依据这个 ID 号，可对多路视频 ES 流和多路音频 ES 流进行识别、复用以及解复用。

紧跟在流识别符后面的是 PES 包长度信息，说明其后数据长度。PES 包的长度是可变的。PES 包的包长度标志有两个字节，共 16 个比特，因此 PES 包的最大长度应为  $2^{16}$  字节。但对视频 PES 包而言，这个包长度值设置为“0”，表明对包的大小是没有限制的。视频 PES 包是由一帧编码图像数据构成的，因此视频 PES 包的起始与一帧图像、一个图像序列或一个 GOP 的起始码是对齐的，即视频 PES 包的净数据负荷的第一个字节要么是一帧编码图像的起始码，要么是一个图像序列的起始码，要么是 GOP 的起始码。

在 PES 包长度信息的后面是 16 比特的 PES 包头标志，指示 PES 流的状态以及码流中是否包含有描述比特流特性的参数。其中，前 8 比特分别用来指示码流的加扰、优先级、数据对准、版权以及原版/复制等状况；后 8 比特分别表示码流可选区域中是否有 PTS 或 DTS、基本码流的时钟基准、基本码流码率、数字存储媒体的触发模式、附加的版权信息、前一个 PES 包的循环冗余码、附加的扩展信息等。

如果附加的扩展信息表示存在附加的扩展域，那么后面将有 5 个标志说明扩展域是否包含 PES 私有数据、PES 包头域、PES 包连续性计数器、P-STD 缓冲区大小、PES 扩展域长度等信息。

#### 4.3.2 打包基本码流中的时间标记

在数字电视码流中，在时域上互相独立的视频/音频数据包被时分复用在码流中。解码端在获取视音频数据包后，采用缓冲机制，重新拼装成视频帧和音频帧数据，经解码器解码后输出视频和音频信号。如果在时间上不加以控制，播放的视频图像和声音之间容易出现错位现象，无法达到编码前的播放效果。为了解决此问题，MPEG-2 标准规定要对 ES 流进行打包，在码流中插入视频音频数据间的相关性信息。

时间标记在数字电视码流中起着十分重要的作用。在模拟电视信号中，通过加入复合同步信号解决了视频图像显示的同步问题。同时，视频和音频信号通过频分复用进行同时传输，音频和视频的同步问题几乎可以忽略。但是，压缩的数字电视码流则不同：压缩使得电视信号失去了行场同步信息；由于双向预测编码引起的帧重排，导致视频图像的发送顺序和显示顺序不一致；音频基本码流和视频基本码流时分复用。

在 PES 包中有两个时间标记：显示时间标记 PTS（Presentation Time Stamp）和解码时间标记 DTS（Decoding Time Stamp）。这两个时间标记对数字电视的解码和显示是非常重要的，PTS 用于通知解码器何时显示一帧已解码的图像，而 DTS 指示何时对接收到的一帧图像的编码码流进行解码。

时间标记能够指明特定的进入单元属于哪一个时间段，是一个 33 比特数字，由 90KHz 时钟驱动的计数器计数得到。这个时钟是对 27MHz 节目时钟基准进行 300 分频后获得。时间标记出现的次数是平均分布的，不必在每个显示单元中插入时间标记。MPEG 标准规定数字电视节目流和传输流中的两个时间标记的间隔不能超过 700ms。对于 PES 包内没有指定 DTS/PTS 的图像帧或声音帧，可以在已知的 DTS/PTS 基础上添加一个计算的偏移量来得到。偏移量的大小等于当前帧与参考帧之家解码/播放的次序间隔和解码播放的帧率的乘积。

通过在视频或音频数据包的包头插入时间标记来获得同步。当解码器收到所选择的 PES 数据包时，它便对每个进入单元进行解码，并缓存到 RAM 中。当时间计数到达时间标记的值时，便读出 RAM 中的数据。时间标记有两个作用，首先，每个基本数据流都能获得有效的时基校正。其次，视频和音频基本数据流可以同步到一块形成节目。

由于一个 PES 包对应一帧图像，因此在每个 PES 包中均应设定与该帧图像对应的 PTS 值。至于 DTS，它不能独立出现，必须与 PTS 一起发生。DTS 的值可由 PTS 的值得到，除非对解码过程有特殊要求，一般不设定也不传送 DTS。当编码图像帧是 B 帧时，也就是说不需要对编码帧顺序进行重排时，DTS 值与 PTS 值是相同的。PTS 是 PES 头中最重要的功能，PTS 的差错将导致图像与伴音对不齐之类的错误。

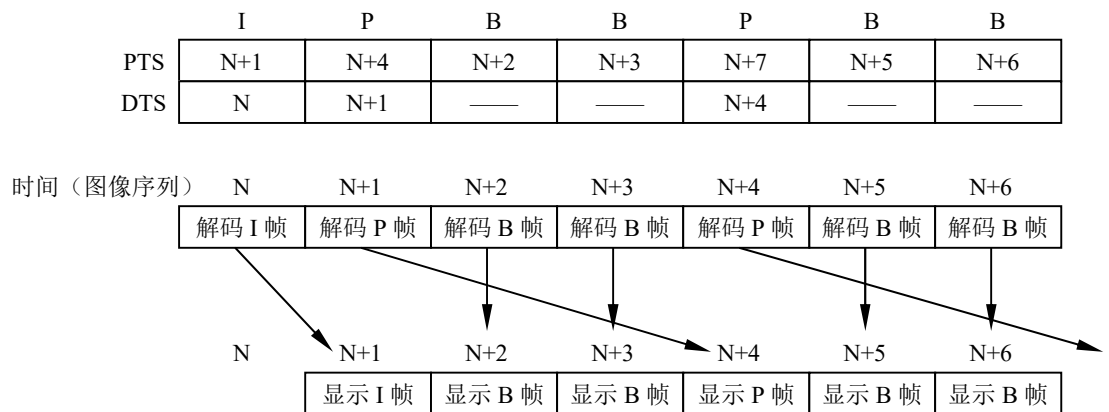


图 4-9 MPEG-2 码流的解码和显示顺序

由于在 MPEG-2 中使用双向预测编码，一个画面（如 I 帧和 P 帧）需要在显示前的一段时间被解码，作为解码 B 帧画面的数据源。例如，GOP 结构为 IBBPBBP 的视频序列，传输则按照 IPBBPBB 的顺序，如图 4-9 所示。由于 B 画面是同时被解码和显示的，所以只含有 PTS。当接收到 IPBB 序列时，I 和 P 帧画面必须在第一个 B 帧画面之前被解码。解码器一次只能解码一幅画面，所以 I 帧画面先被解码并储存。当 P 画面被解码时，已被解码的 I 画面就是输出，所以后面可以接 B 帧画面。如图 4-9 所示，当接收到含有 I 帧画面的进入单元时，在包头会同时有 DTS 和 PTS，这些时间标记由一个画面的时间作间隔。如果使用双向编码，后面必须跟 P 画面，并且该画面也有 DTS 和 PTS 时间标记，这两个时间标记的间隔为三个帧时间，可以容纳插入两个 B 帧画面。因此，收到 IPBB 序列，I 帧画面被延时一个帧时间显示，P 帧画面被延时三帧时间显示，两个 B 帧画面则没有延时，这样，显示序列则变为 IBBP。可见，如果 GOP 结构发生改变，在 I 帧画面和 P 帧画面之间有更多的 B 画面，那么在 P 帧画面中 DTS 与 PTS 之间的差异就会增大。在数据包的包头设置 PTS/DTS 标记是为了指明只有 PTS 存在还是 PTS 和 DTS 时间标记同时存在。

音频数据包可能含有多个进入单元，数据包头部含有一个 PTS。由于音频数据包总是按顺序发送的，所以音频数据包中没有 DTS。