

AC-3 编/解码的设计及实现

作者：Steve Vernon

经美国电子电器工程协会允许，发表于 IEEE，消费类电子技术，41 卷，第三篇，1995 年 8 月。

编者按：由于以下文章著于 1995 年，当时数字电视多声道伴音的讨论还仅限于高清 ATSC 系统。但是在随后的发展中，标清数字电视因为要和 DVD 及电影争夺观众，多声道标清电视在 90 年代末蓬勃发展起来。由于 AC-3 能够将单声道，立体声和多声道节目无缝串接在一起播出，并且一种码流可以适应从单声道低端用户到多声道家庭影院的不同需求，因此在世界两大数字电视传输标准 ATSC 和 DVB 中被定为强制的或可选的音频标准，并成为世界上绝大多数国家高清或标清数字电视的音频制式。

摘要

AC-3 是一种供 HDTV 数字电视音频压缩使用的感知编码技术。本文阐述了 AC-3 编/解码的设计及实现，主要是有关其花费最少的方案讨论。

1. 介绍

在最近的几年内，数字音频压缩技术已经成为音响工业的重要的技术。不需要占用很高的带宽就能够实现高质量还音的新型音频格式已经问世，而在传统的技术看来这是必须占用很高带宽。

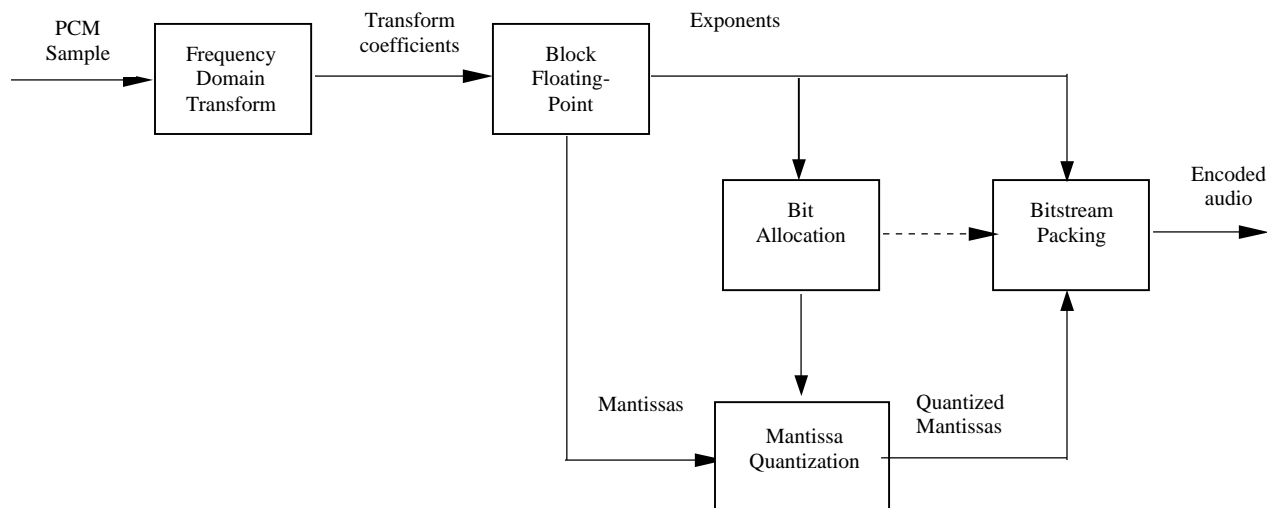
AC-3 编码技术已经被 ATSC 采纳为美国 HDTV（高清数字电视）的音频业务标准。它同时也应用于消费类媒体（LD 光碟、DVD-Video 光碟）以及卫星广播。到目前为止，有超过数十家半导体厂商生产 AC-3 解码芯片。

本文的目的在于介绍 AC-3 编/解码的大致过程，集中讨论有效应用的策略。将要描述 AC-3 编/解码器的基本处理过程以及所需存储空间及其复杂度的相关估计。

2. AC-3 编码系统

AC-3 是一种灵活的音频数据压缩技术，它具有将多种声轨格式编码为一种低码率比特流的能力。支持 8 种不同的声道配置方式，从传统的单声道、立体声到拥有 6 个分离声道的环绕声格式（左声道、中置声道、右声道、左环绕声道、右环绕声道及低音效果声道）。AC-3 的比特流所允许的采样频率可以为 48 kHz、44.1 kHz、或 32 kHz 中的任何一种，并且所支持的码率从 32 kbps（千比特位/秒）到 640 kbps 不等。

图<1>所示的为 AC-3 编码器的方块图。编码过程是在频域内完成的，采用 512 位的 50% 交迭 MDCT（改进型离散余弦变换）。对于有暂态信号，则采用将原先的 512 位变换改为两个 256 位的块切换方式来提高性能。接下来的浮点数变换过程将转换系数集分成指数与尾数对。尾数部分是按照位分配的参数模型进行比特位数可变的量化。



图<1>: AC-3 编码器方框图

AC-3 位分配的模型按照心理声学掩蔽效应来确定给定频带内的尾数位数的多少。根据掩蔽效应的程度，某些尾数只需要极少量的位，甚至无需任何位。这种方式减少了表现原声所需要的比特数，所付出的代价是增加了人耳无法辨别的噪声。

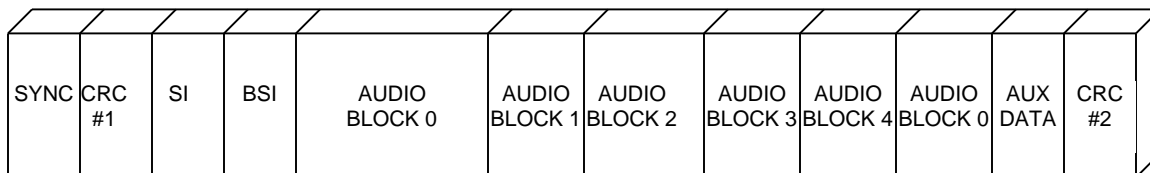
与某些其它的编码系统不同，AC-3 没有将位分配的结果包含在比特流中传输给解码器。而是采用了参量处理的方式，编码器根据变换系数的指数值和某些由信号本身决定的关键的参数来构造其心理声学的掩蔽效应模型。这些参数通过数字码流传输给解码器，与传输原始的位分配数位相比只需极少量的数位。在解码器端，系统根据指数的数值和位分配参数重新构筑位分配功能块。

AC-3 编码的效率随着声源的声道数目增加而上升。这是由两个基本特性所决定的：全局数位集合，和耦合技术。全局数位集合技术可以让位分配器依照需求原则分配音频声道中可使用的数位。如果一个或多个声道在某一特定的瞬间不使用，剩余的声道就可使用比所有的声道对数位的需求都很高时更多的数位假如。

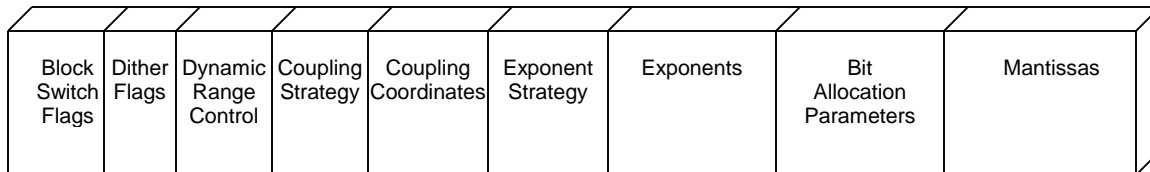
耦合技术(Coupling)属于应用于针对声音频谱中的高频部分实施进一步压缩使之符合需求码率的技术。在高频部分，人耳无法甄别出每个独立的音频声波的周期，而仅对声波的包络产生感应。耦合技术减少了相关声道中的高频组份使之构成一个单一联接声轨，然后产生用以描述各个声道频谱包络特性的辅助数据链。由于耦合技术可能会对原音生成人造的痕迹，因此它通常会被限制在较低码率的情况下使用。

2.1 AC-3 比特流的特征

AC-3 的比特流是由帧(Frame)构成的（详见图 2），在恒定的时间间隔内，其所有编码的声道所包含的信息就能体现在 1536 个 PCM 采样值的信息。每一个 AC-3 的帧(Frame)都具有固定的尺寸，只由采样频率及编码数据率决定。同时，每个帧 (Frame)都是独立的实体而且并不与前一个帧 (Frame)分享数据，除了在 MDCT 所固有的去交迭变换。



图<2>: AC-3 帧结构



图<3>: AC-3 声音块结构

在每个 AC-3 帧(Frame)的开头是 SI 域（同步信息 Sync Information）及 BSI 域（比特流信息）。SI 域及 BSI 域描述了比特流的结构，包括采样频率、数据码率、编码声道的数目及其他一些系统描述的元素。每个帧用两个 CRC（循环冗余码校验字）字来提供侦错手段，一个位于帧的起始处，另一个位于帧的结尾。

每个帧有 6 个声音块，每个块表示为每个编码声道包含 256 个 PCM 取样 (sample)（详见图 3）。声音块中所含内容包括块转换标志，耦合坐标、指数、位分配参数、尾数。允许在帧的内部进行数据分享，比如在声音块 0 中的数据可以被同一帧内的后续块所使用。

在每个帧的结尾处有一个可选的辅助数据域。在这个区域内允许系统设计者在 AC-3 比特流中嵌入可在整个系统内传递的、自有的控制字及状态字信息。

AC-3 的编/解码器被设计成一个完整的音频子系统的解决方案，它拥有普通的低码率编/解码所没有的许多特性。这些特性包括适用于消费类音频回放系统的动态范围压缩特性(Dynamic Range Compression)、对话归一(Dialog Normalization)以及缩混特性(Downmixing)，缩混特性可以将多声道音频进行转换为特定数目的声道输出。动态范围控制(Dynamic Range Control)的控制字是嵌入在 AC-3 比特流内，并被解码器应用，可以使同一个比特流源在不同模式下进行还音。

3. 解码器的实现策略

图 4 表明了 AC-3 基本的解码过程。为了使解码器的资源及反应时间的需求达到最低，每个 AC-3 的帧都是通过一系列的循环嵌套结构来解码。

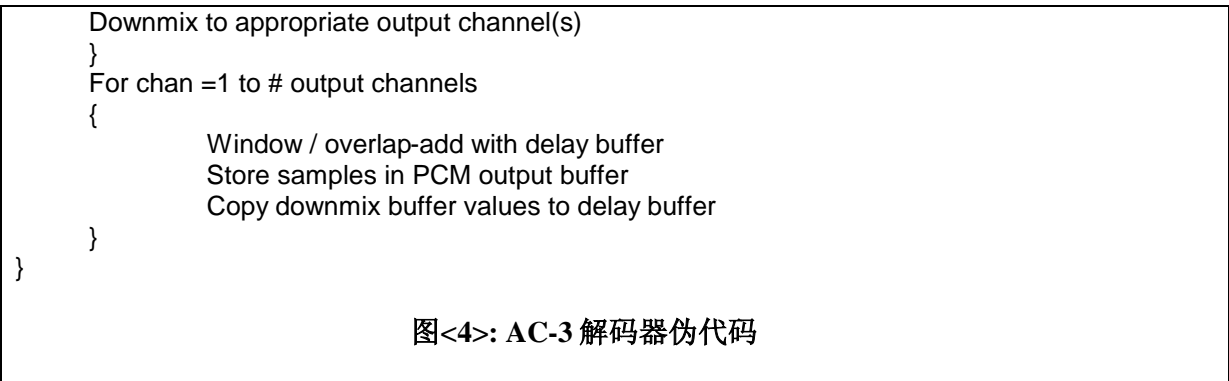
第一个步骤是建立帧的排列结构。这阶段包括查找 AC-3 的同步字，其次是确认侦错的 CRC 校验字表示无错误发生。一旦发现帧的同步字，解码器就会将 BSI 域的数据包进行解压缩释放，也就确认了诸如编码的声道数目之类的重要信息。

其次的步骤就是将 6 个声音块分别解压缩释放。为了使 PCM 输出的缓冲区的容量要求达到最小值，所有的声音块都遵循每个时间段仅对一个声音块解码的原则。当每个声音块的解码结束语时，所得到的 PCM 结果将会拷贝至输出缓冲区。通常情况下，它拥有双倍的容量为系统的数/模（D/A）转换器提供直接中断存取手段。

```

AC-3 frame alignment / CRC check
Unpack BSI data
For Block =1 to 6
{
    Unpack fixed data
    For chan =1 to # coded channels
    {
        Unpack exponents
        For band =1 to # bands
        {
            Compute bit allocation
            Unpack mantissas
            Scale mantissas / undo coupling
            Denormalize mantissas by exponents
        }
    }
    Compute partial inverse transform
}

```

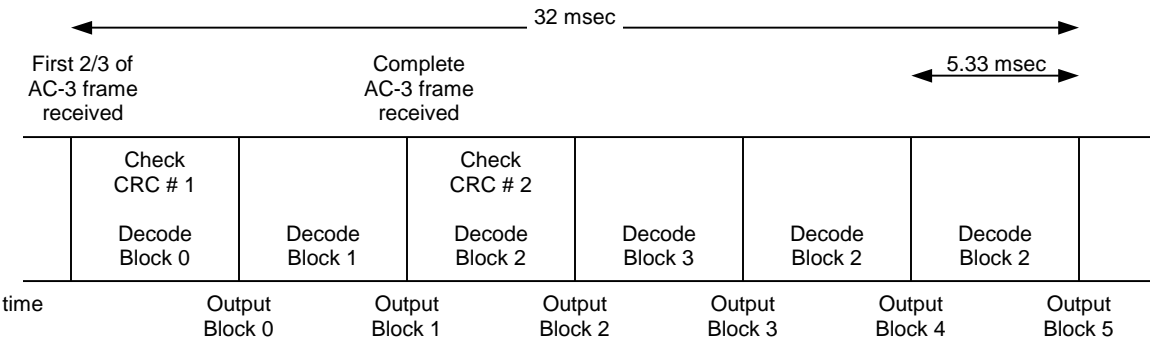


3.1 解码器的同步时间

图 5 表示了 AC-3 解码器同步的时间过程。按照设计要求，所有的 AC-3 编码器应确保最初两个声音块包含在 AC-3 帧的第一个 5/8 长度段内。同样，第一个 CRC 校验字仅针对 AC-3 帧的第一个 5/8 段的内容进行纠错校验。这个措施有助于减少解码器所需的工作时间。因为解码器可以在整个 AC-3 帧被完全接收之前就能开始对最初的两个声音块进行解压缩及处理。

一旦最初 2/3 长度的 AC-3 帧被接收，解码器就开始检查第一个 CRC 校验字。如果没有错误发生，解码器将对声音块 0 进行解码并且将重构的 PCM 样本拷贝至输出缓冲区。在每个声音块处理的结束阶段（对于 48kHz 采样频率的系统为 5.33 毫秒），输出中断操作将开始向数/模转换器（DAC）传输声音块 0 的样本，并着手处理声音块 1 的解码工作。

在声音块 1 解码的结束阶段，解码器将完成接收整个 AC-3 帧的过程并且检验第二个 CRC 校验字。可以看到，这种方案使得整个解码器所需的工作时间相当于处理一个声音块的时间加上接收 2/3 长度的 AC-3 帧所需的时间（在连续输入的 48kHz 采样系统中为 27 毫秒）。



图<5>: AC-3 解码器同步周期

3.2 解码器的输入处理

声音块处理的过程可以明显的分为两个阶段，这里所谈及的是输入和输出阶段。输入阶段包括所有的比特流的解压缩过程和编码声道的处理。输出阶段主要是指 MDCT 逆变换的窗口和正交叠加阶段。

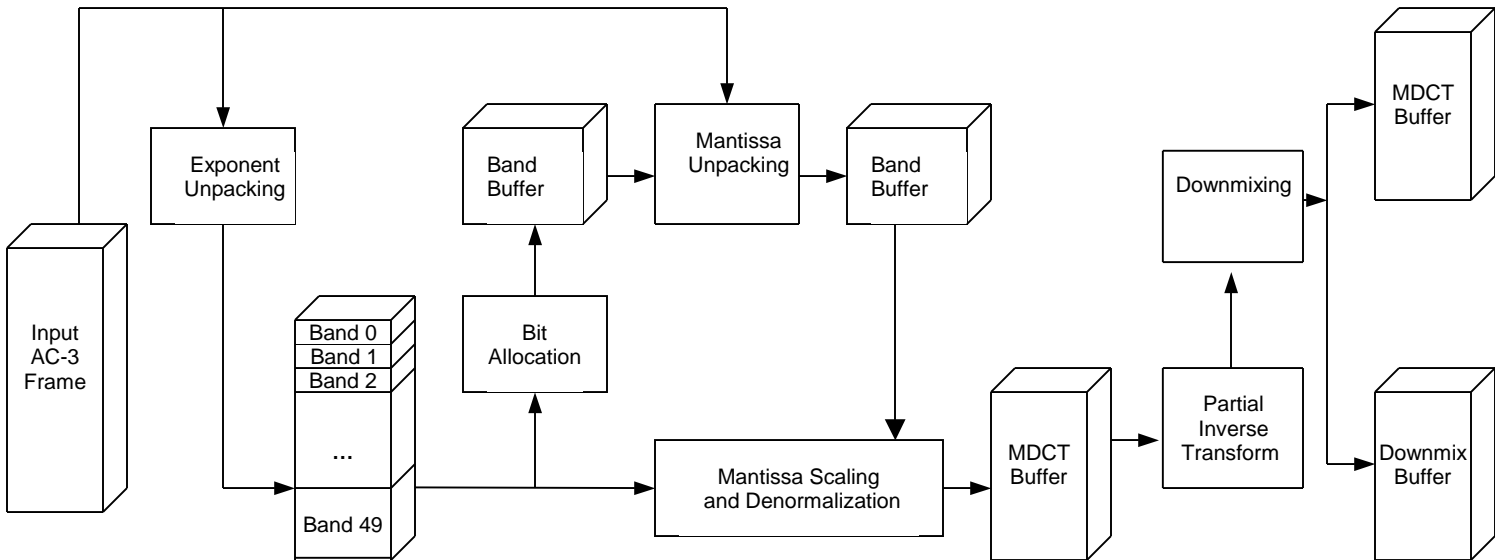
有这种区别是因为 AC-3 解码器的输出声道的数目并不需要与比特流的已编码的声道数目相吻合。通过使用一种叫缩混(Downmixing)的技术，解码器就可以接收任意的编码声道数目的比特流并且产

生任意数目的声道输出。输入过程是基于每个已编码声道进行处理，然而输出过程是基于每个输出声道进行处理。

输入过程由解码器对固定的声音块数据解压缩的阶段开始，其功能是对位于声音块起始处的各个参数及标志进行数据采集。固定数据包括诸如块切换标志、耦合标志、指数及位分配参数等信息。术语“固定数据”(Fixed Data)指的是这些比特流元素实际的字长是预先知道的，并且无需进行位分配的工作。

指数区域在固定数据区域内占有最大的单体空间，这是由于它们包括由所有编码声道而来的信息。因为编码模式的不同会导致指数对于尾数个数的不同，可能是一个指数对应一个尾数，直到每个声道对应 253 个尾数。实际上其并未将所有解压缩后的指数存入相应本地存储单元，而是在普遍的情况下仅存储指示相应指数区域的指示器，并且仅在需要的时候进行解压缩——每次一个声道。

一旦“固定数据”(Fixed Data)被解压缩后，解码器开始处理每个编码声道（详见图 6）。首先，输入帧(Input Frame)的给定声道的指数部分进行解压缩。然后是位分配的计算，其作用是将指数和位分配参数分离出来并且计算每一个压缩的尾数的字长。然后，尾数从输入帧(Input Frame)中被解压缩释放。尾数部分按照比例提供合适的动态范围控制(如有需要，会忽略耦合操作)，并由指数来指导去规格化。最终，进行局部逆变换，其结果被缩混(Downmix)到适当的缩混缓冲区用于接下来的输出处理。



图<6>: 解码器输入处理

首先在实施这些步骤之前，每个声道各自的指数部分被解压缩到具有 256 字符称为“MDCT Buffer”的缓冲器。这些指数被重新构成为多达 50 个频带以便进行位分配。在每个频带中指数数值会随着声音频率的升高而增长，大致上符合模仿心理声学的重要频带的对数算法（logarithmic Division）。

对于每个位分配频带来说，指数部分与位分配参数组合起来产生此频带内的每个尾数的尾数字长。这些字长值存放在一个 24 位字符的缓冲区内(最大的位分配的频带可由 24 位两进制频率组成)。一旦计算出字符长度的数值，相对应的尾数就从输入帧(Input Frame)中被解压缩并且存入频带缓冲区的适当位置。最后这些尾数被标定后，由相对应的指数执行去规格化，并且将其重新写入 MDCT 缓冲区的相应区域。当所有的频带处理完成，所有的尾数完成解压缩，应该对在 MDCT 缓冲区内所有剩余的区域进行清零。

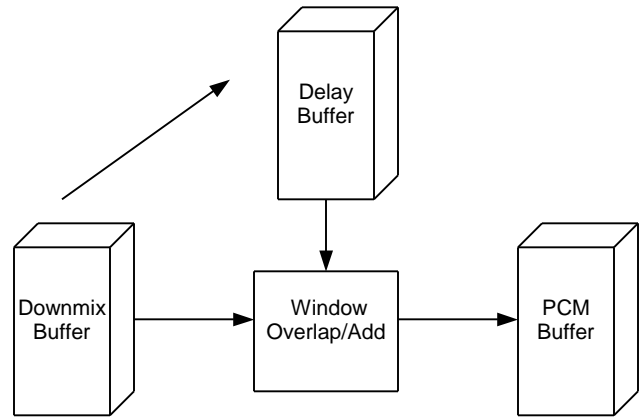
这时，在 MDCT 缓冲区内进行局部的逆变换。局部逆变换的实现包括一个 128 位的复杂的预乘过程，一个 128 位的复杂快速傅里叶（FFT）逆变换阶段，以及一个 128 位复杂的后乘过程。此过程的结果输出结果将缩混到相应的缩混缓冲区。

请注意如果不采用块切换，缩混处理完全可以先于局部逆变换在频域中完成。在采用块切换的变换中，使用的是不同的局部逆变换进程，并且这两种不同变换的结果不能直接联合直到进行到窗口函数转换阶段。

如同本文后面所叙述的，输入处理阶段使用相对较少的系统内存，因为它具有在同一时间段内处理一个声道的输入码流的能力，甚至于处理一个频带。在另一方面，此阶段是在整个解码阶段信息处理量 MIPS（每秒处理百万条指令）最大的部分，因为位分配及解压缩进程的计算量通常为整个解码复杂度的一半。

3.3 解码器的输出处理

一旦输入处理阶段完成并且缩混缓冲区完全产生，解码器就可以进入输出处理阶段（详见图 7）。对于每个输出声道，缩混缓冲区及其相应的 128 位字符长，占半个数据块时间的延时缓冲区被调入窗口一起产生一个 256 位的 PCM 输出取样。然后，这些取样根据精度的大小排列并且拷贝到输出缓冲区。当这个阶段完成后，缩混缓冲区的一半数据被拷贝至相应的延时缓冲区，为下一个声音块的重新整合提供必要的 50% 迭交信息。



图<7>: 解码器输出处理

与输入处理阶段相比，解码器的输出处理阶段所需要的 MIPS 相对少得多，然而却使用了相对大得多的内存。对于每个输出声道，解码器需要高达 256 位字符的缩混缓冲区以及一个 128 位字符的延时缓冲区，另外还有一个 512 位字符的输出缓冲区（假定属于双缓冲输出）。另外，这些缓冲区需要足够的字符宽度来保持优良的宽带信噪比使之符合专业音频的标准。比如 16 位精度的 PCM 输出，相应的延时缓冲区至少应为 18 位，而且缩混缓冲区及 MDCT 缓冲区至少应为 20 位。对于精度更高的 PCM 输出，缓冲区的字符宽度相应的增加。

3.4 解码器内存及 MIPS（每秒处理百万条指令的能力）总结

在图 8 的表格中总结了一个 AC-3 解码器内存及 MIPS 的要求。内存分为 RAM 和 ROM。在这些表格中，参数 K 指的是输入比特流所包含的编码声道数目，参数 N 为输出的声道数目。这些表格同时假定输入帧为单缓冲器，而 PCM 输出缓冲器是双缓冲器，按照 HDTV 的规范配置，AC-3 比特流是采样频率为 48kHz、码率为 384kbps 的六个声道的编码节目。

RAM 数据结构	字	位
Input AC-3 Frame (AC-3 输入帧)	768	16
Audio block fixed data (声音块固定数据)	=384	16
MDCT Buffer (MDCT 缓冲区)	256	≥20

Downmix Buffers (缩混缓冲区)	N*256	≥20
Delay Buffers (延时缓冲区)	N*128	≥18
PCM Buffers (PCM 缓冲区)	N*512	≥16

ROM 数据结构	字	位
Transform lookup tables (变换查询表)	768	≥20
Other lookup tables (其它查询表)	=1250	16
Program code (程序代码)	=3500	32

进程任务	MIPS	% of Total
Frame alignment / CRC check (帧排列/CRC 校验)	1.3	5%
Unpack BSI (BSI 解压缩)	0.2	1%
Unpack fixed data (固定数据解压缩)	0.9	3%
Unpack exponents (指数解压缩)	K*0.5	9%
Compute bit allocation (位分配计算)	K*1.6	30%
Unpack mantissas (尾数解压缩)	K*1.1	20%
Scale mantissas & denormalize (尾数衡量&去规整)	K*0.3	6%
Partial inverse transform (局部逆变换)	K*0.9	20%
Downmixing (缩混)	K*0.1	2%
Window / overlap-add (窗口 / 正交叠加)	N*0.2	4%

6 (5.1)声道解码器: 6.6K RAM, 5.4K ROM, 27.3 MIPS
2 声道解码器: 3.1K RAM, 5.4K ROM, 26.5 MIPS

图 8: 解码器内存/MIPS 总结

从有关内存表格中, 我们可以清楚地看到两声道解码器 (将任何数量的输入声道缩混为两声道输出) 所需的 RAM 资源比 6 个声道解码器要少很多。由于 RAM 的费用占有整个芯片相当可观的份额, 这指出了两声道解码器的生产成本比全性能解码器的成本要低得多。两声道解码器提供了跨越到多声道音频的低成本解决方案, 因为用户开始可以欣赏 6 声道节目源的缩混信号, 将来可以再升级到完全分离的多声道回放。

对于 MIPS, 两声道的解码器与 6 声道的解码器在计算的复杂度上大致相等。这就表明了在处理器内核不做很大修改的情况下, 两声道的解码器芯片设计就可以扩展到完全的 6 声道的产品。所有有关 MIPS 的数目及固化源代码 ROM 的大小是参照 Zoran ZR38001 DSP 处理器的指令集定义的, 此款芯片是通用型处理器, 除了通用的 DSP 特性之外, 它还包括了一个单循环的柱状移位寄存器 (**barrel shifter**) 和一个四重循环的蝶型 FFT 处理。

4. 关于编码器设计的几点思考

与解码不同, AC-3 的编码并不是一个固定的过程, 它更要求精细的方案分析以期达到最理想的性能。几乎所有的通过比特流方式传递给解码器的参数必须是由编码器根据信号的特征有选择地决定。

例如, 使用暂态检测程序来决定是否对 MDCT 变换的尺寸进行块切换。其下一步是检查频域指数及其他固定数据来决定是否在每个声音块中重复传递不同的数据, 还是在声音块之间分享通用数据。最终, 采纳某个精益求精的位分配模型来确认位分配参数以求达到最佳的主观听音质量。

由于上述原因, 我们无法详细说明 AC-3 编码器的复杂度上限。虽然如此, 仍可以在几乎所有的 AC-3 编码器设计中采纳以下几点意见。

在通常情况下，编码器（设计）必须考虑到同一个帧(Frame)内在所有声音块中包含所有声道极端情况，来决定最佳的位分配参数以及固定数据重新利用的策略。尽管解码器能够在同一时间段内每次一个声道的实施逆变换（以及在同一时间段内每次一个声音块），编码器通常会在任何编码开始之前将 6 个声音块中的所有声道完成所有变换。其结果是，编码器的系统内存要求将比解码器多出好几倍。

另外，编码器的位分配的程序通常是一个交互式的过程，这是因为编码器需要在不超出允许码率的情况下尽可能多得（对信息）分配数位。最终，编码器的任务还包括产生某些能被解码器简单应用的控制字，比如耦合参数和动态范围(Dynamic range)压缩控制字。

要提供编码的策略分析是一个复杂的过程，编码器所要求的 MIPS 的性能比解码器相关要求高出很多。现今的编码器设计根据其精细程度的不同，要求其 MIPS 的性能是它们相应解码器的 2 至 6 倍。值得注意的是，如果 2 声道的码率设置超过 192kbps，或者将 6 声道的码率设置超过 384kbps，为达到几近透传音频性能，仅需要中等复杂度的编码。

最后，由于编码器的字符长度影响会到解码器的动态范围，在进行进一步变换前应仔细计算。为保证 PCM 的精确度至少达到 18 位以上，当今的编码器采用浮点计算或者是 24 位固定数位处理器。

5. 推论

以上介绍了 AC-3 音频数据压缩系统，及介绍了编码及解码方案的大致情况。同时也在如何设计高效率解码器应用上介绍了一些技术细节。而且 AC-3 解码器可以通过一系列的紧密的嵌套循环来实现。

这里所介绍的解码器芯片有益的特性是其所需 RAM 的数量大致与其输出的声道数目成正比。由于解码器所具备的缩混功能可以产生任意数目的声道输出，这就意味着可以生产能够兼容多声道音频比特流的低成本解码器。

6. 感谢

本文的作者感谢杜比实验室工程项目组的其他成员对 AC-3 编/解码技术的贡献，他们中包括 Craig Todd, Louis Fielder, Grant Davidson, Mark Davis, Matt Fellers 以及 Marina Bosi。

7. 参考书目

- [1] G. Davidson, L. Fielder, and M. Antill, "Low-Complexity Transform Coder for Satellite Link Applications," AES 89th Convention, Los Angeles, Sept. 1990, Preprint 2966.
- [2] L. Fielder and G. Davidson, "AC-2: A Family of Low-Complexity Transform-Based Music Coders," 1991 AES Workshop on Digital Audio, London, Oct. 1991.
- [3] G. Stoll and Y. F. Dephy, "High-Quality Audio Bit-Rate Reduction System Family for Different Applications," Proc. of IEEE Intl. Conf. On Comm., Atlanta, pp. 937-941, April 1990.
- [4] ATSC A/52, "Digital Audio Compression (AC-3) Standard," United States Advanced Television Systems Committee.
- [5] C. Todd, G. Davidson, M. Davis, L. Fielder, B. Link. And S. Vernon, "AC_3 Flexible Perceptual Coding for Audio Transmission and Storage," AES 96th Convention, Amsterdam, Feb. 1994, Preprint 3796.
- [6] J. Princen and A. Bradley, "Analysis / Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation," IEEE Trans. ASSP, vol. ASSP=34, no.5, pp. 1153-1161, Oct. 1986.