DOLBY

Dolby® AC-4: Audio Delivery for
Next-Generation Entertainment Services

June 2015

1 Introduction

Video entertainment is entering a new era, with viewers increasingly seeking flexibility in what they watch, when they watch it, and how they choose to engage with it. To meet the demands of these new consumption patterns and provide flexibility for continued innovation, a new generation of audio delivery technology is required.

This paper describes the technical features and capabilities of the Dolby$^{®}$ AC-4 audio delivery system. Dolby AC-4 is a new audio format that addresses the current and future needs of next-generation video and audio entertainment services, including broadcast and Internet streaming. The core elements of Dolby AC-4 have been standardized with the European Telecommunications Standards Institute (ETSI) as TS 103 190 and adopted by Digital Video Broadcasting (DVB) in TS 101 154 and are ready for implementation in next-generation services and specifications.

2 System Overview

Dolby AC-4 is an audio delivery system designed from a clean sheet that combines high-efficiency audio coding, a flexible transport syntax, and powerful system-level features that optimize delivery and playback on a broad range of devices. The system improves today's use cases by supporting conventional channel-based audio at the low data rates required by next-generation broadcast and streaming services with a number of system-level advancements over existing solutions. For example, these enhancements allow improved accessibility solutions compared with current broadcast systems. Furthermore, Dolby AC-4 enables tomorrow's experiences by fully supporting object-based audio (OBA), creating significant opportunities to enhance the audio experience, including immersive audio and advanced personalization of the user experience.
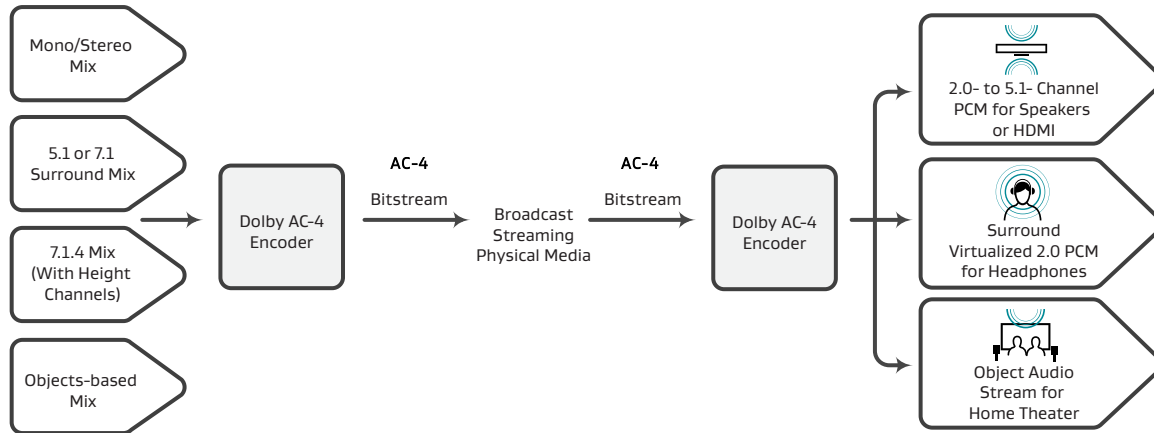
Figure 1: Dolby AC-4 can carry conventional channel-based soundtracks and object-based mixes. Whatever the source type, the decoder renders and optimizes the soundtrack to suit the playback device.

The AC-4 bitstream can carry channel-based audio, audio objects, or a combination of the two. The Dolby AC-4 decoder combines these audio elements as required to output the most appropriate signals for the consumer—for example, stereo pulse-code modulation (PCM) for speakers or headphones or stereo/5.1 PCM over HDMI®. When the decoder is feeding a device with an advanced Dolby renderer—for example, a set-top box feeding a Dolby Atmos® A/V receiver (AVR) in a home theater—the decoded audio objects can be sent to the AVR to perform sophisticated rendering optimized for the listening configuration.

The architecture of the Dolby AC-4 decoder enables powerful audio processing to be performed efficiently in the playback device, reducing the computational and power requirements compared with separate, independent processing stages. This is especially important on portable devices with limited processing and/or battery capability. This is achieved by implementing many of the core coding tools in the quadrature mirror filter bank (QMF) domain, enabling powerful QMF-domain audio processing to be integrated with the decoder without the computational cost of additional transforms to and from the audio-processing filter bank.

The Dolby AC-4 decoder incorporates multiband processing in the QMF domain to tailor the dynamic range and output level to the playback device, which is guided by optional metadata embedded in the bitstream by the service provider or content creator. It also provides the capability to add audio optimization for the type of device—such as spatial enhancement and speaker optimization—directly in the QMF domain without need for additional transforms.

As a result, the overall complexity of a Dolby AC-4 decoder, including integrated loudness and dynamic range control (DRC), is similar to that of a previous-generation Dolby Digital Plus™

decoder with downstream Dolby Volume processing. Similarly, the overall complexity of a Dolby AC-4 decoder with added headphone virtualization is similar to that of a previous-generation Dolby Digital Plus decoder with downstream Dolby Audio™ headphone virtualization.

3 Decoder Overview

3.1 Introduction

The Dolby AC-4 system is a clean-sheet design that builds on state-of-the-art technology and proven know-how to offer high audio quality, rich features, and excellent coding efficiency. This enables high-quality audio to be delivered at around one-quarter of the data rates commonly used in today's HDTV services. To achieve these high compression efficiencies, AC-4 utilizes a number of advanced coding tools. The figure below illustrates a decoder block diagram.
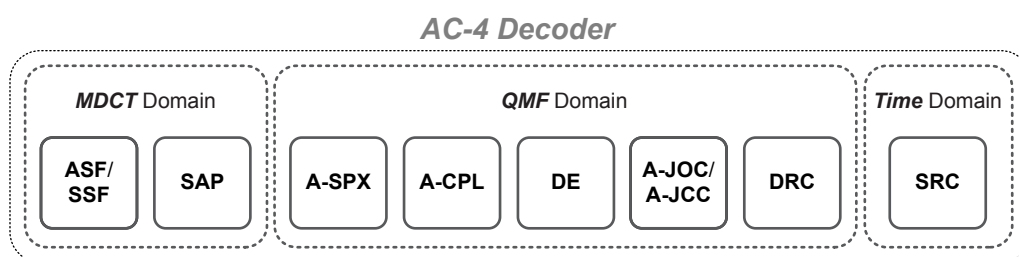
Figure 2: AC-4 Decoder

Key advancements over previous coding systems are discussed in the following sections.

3.2 Dual-Spectral Frontend

In perceptual audio coding, the digital audio is compressed by removal of redundant and irrelevant audio information from the signal. Redundancy is significantly reduced by transforming the audio signal to the frequency domain and applying entropy coding. AC-4 utilizes two different modified discrete cosine transform (MDCT) frontends to code the audio.

For general audio content, the Audio Spectral Frontend (ASF) is used. ASF employs block switching between five transform lengths ranging from 128 to 2,048 samples. The use of multiple block sizes enables the coder to maximize audio quality by using short windows on transient signals such as drums, while using longer windows otherwise to keep the overall data rate low.

AC-4 also contains a dedicated Speech Spectral Frontend (SSF). This prediction-based speech coding tool achieves very low data rates for speech content. Unlike most common speech coders, it operates in the MDCT domain, which enables seamless switching between the ASF and the SSF

as the characteristics of the content change. The SSF is especially important for efficiently delivering multilingual and secondary commentary content where many independent dialogue substreams are encoded and carried in a single AC-4 stream. The figure below shows the top-level structure of the SSF decoder.
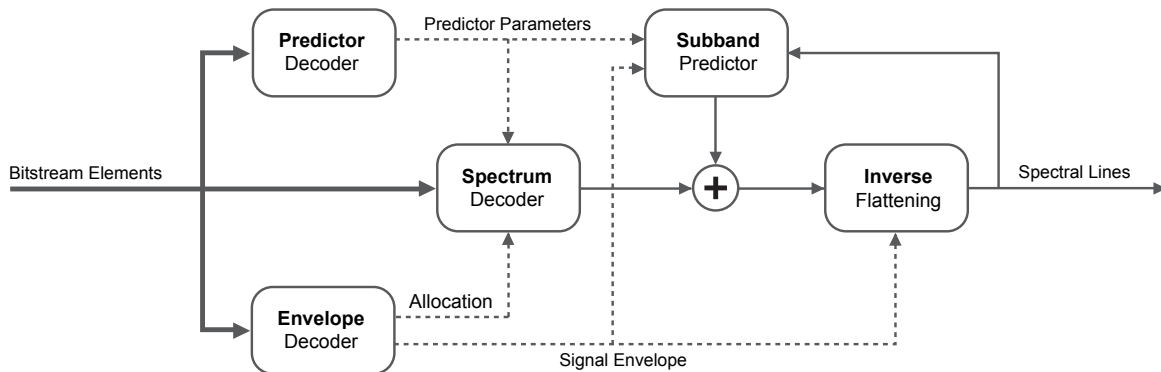

Figure 3: Top-Level Structure of the SSF Decoder

3.3 Stereo Audio Processing

Stereo Audio Processing (SAP) is a waveform-coding tool that improves the coding efficiency of stereo and multichannel signals at all bit rates.

As a superset of existing joint stereo coding techniques, SAP provides Left/Right and Mid/Side coding modes but also offers an additional enhanced Mid/Side coding mode. This mode enables better coding of panned signals than traditional tools (including "intensity stereo") and offers more flexibility to process complex stereo signals.

The output of the SAP tool is either a Left/Right representation of the two channels or, if SAP is combined with Advanced Spectral Extension (A-SPX) and Advanced Coupling (A-CPL), a Mid/Side representation as shown in the following two figures.
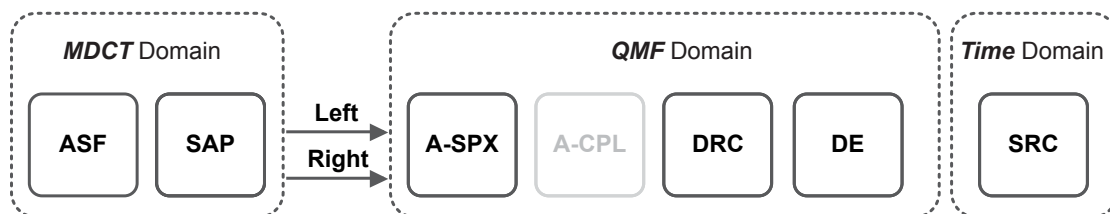
**Figure 4: SAP Output (Left/Right)**
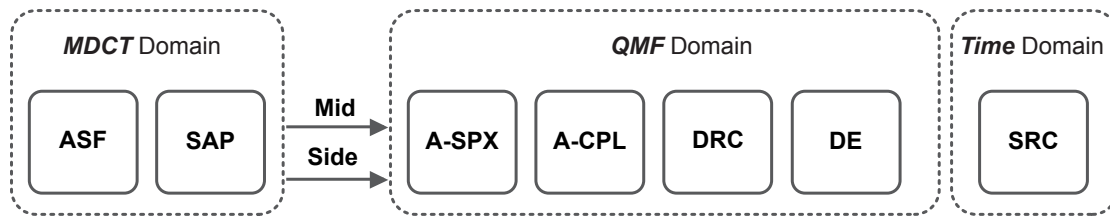
Figure 5: SAP Output (Mid/Side)

3.4 Advanced Spectral Extension

Advanced Spectral Extension (A-SPX) is a coding tool used for efficient coding of high frequencies at low bit rates. This technique improves quality by reconstructing higher frequency sounds, transposing up harmonics from the lower and mid frequencies guided by a side chain of helper data.

A-SPX is similar in concept to the high-frequency reconstruction techniques used in Dolby Digital Plus. However, a key advantage of A-SPX is that it runs on the same sampling frequency as the core coder. This allows waveform coding to be performed selectively in portions of the frequency range where A-SPX is operating to improve performance with critical components. This mixed-mode coding can be frequency interleaved or time interleaved.

The figures below illustrate the mixed modes. W denotes the time/frequency region that is waveform coded, and A-SPX denotes the regions that have been transposed from the low frequency region.
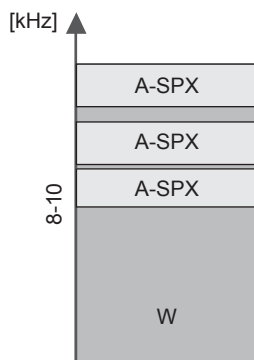
Figure 6: A-SPX Frequency–Interleaved Coding
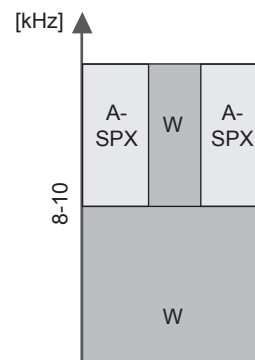
Figure 7: A-SPX Time–Interleaved Coding

In frequency-interleaved coding, a narrow frequency range above the nominal crossover frequency is transmitted as a waveform-coded spectral representation. This provides improved audio quality for content with strong stationary components (for example, tonal signals) in the high-frequency range.

In time-interleaved coding, the content of a small A-SPX portion within a frame is completely replaced with a waveform-coded spectral representation. This improves the quality of high-frequency percussive or transient signals, which are otherwise difficult to reconstruct with current coding systems.

3.5 Advanced Coupling

Advanced Coupling (A-CPL) is a parametric spatial coding tool that enables efficient stereo and multichannel coding at low data rates. It identifies correlations between the channels of stereo or multichannel audio and codes the signal efficiently using waveform coding for the correlated audio and parameters to recreate the perceptually correct spatial relationship between the channels.

A-CPL enables waveform coding to be used at lower bit rates than previous systems, which allows for stereo coding down to 24 kbps and multichannel coding down to 64 kbps.

3.6 Advanced Joint Object Coding

Advanced Joint Object Coding (A-JOC) is a parametric coding tool to code a set of objects efficiently. The technology relies on a parametric model of the object-based content. The tool exploits dependencies among objects and utilizes a perceptually based parametric model to achieve high coding efficiency. The parametric model is constructed on top of a reduced set of objects that is determined by the A-JOC encoder. The reduced set comprises a smaller number of spatial object groups (for example, seven in the example illustrated below), which are coded directly by the core coder. The reduced set of objects may be obtained using a similar approach to spatial coding. The coded objects are accompanied by the object audio metadata (OAMD), which describes the properties of the audio objects. The figure below outlines the basic principle of the A-JOC decoder tool.
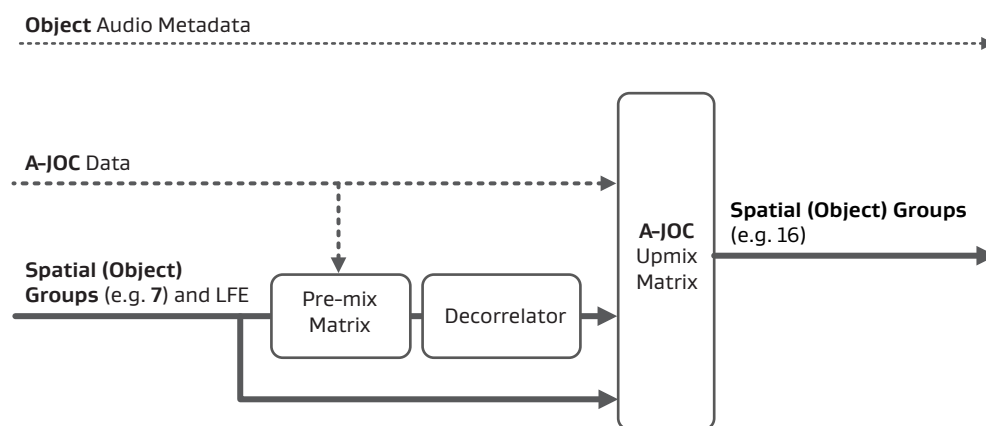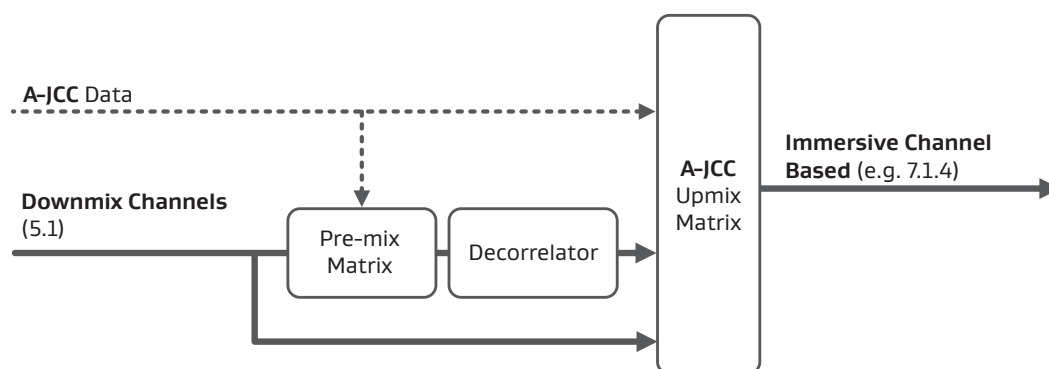
Figure 8: Basic Principle of A-JOC Decoder Tool

3.7 Advanced Joint Channel Coding

Advanced Joint Channel Coding (A-JCC) allows for efficient coding of immersive multichannel signals, including 7.1.4 and 9.1.4 representation by the means of downmix channels and parametric A-JCC data. Key benefits are:

- The optimal downmix is automatically chosen to provide the best audio quality for a given multichannel signal.
- The encoder controls how the height channels are mixed to the horizontal channels.
- The parametric model is scalable in bit rate.

A-JCC Data

Downmix Channels (5.1)

Pre-mix Matrix

Decorrelator

A-JCC Upmix Matrix

Immersive Channel Based (e.g. 7.1.4)

Figure 9: Basic Principle of A-JCC Decoder Tool

3.8 Coding Performance and Coding Tool Use

AC-4 provides a 50% compression efficiency improvement on average over Dolby Digital Plus across content types ranging from mono to immersive audio.

The table below provides an overview of what level of audio quality is achieved given a certain content type and bit rate. The *good* and *excellent* quality statements are intended to match the MUSHRA listening test result scale and are based on both internally and externally conducted listening tests.

	Good Quality	Excellent Quality
Mono	24 kbps	40 kbps
Stereo	40 kbps	64 kbps
5.1	96 kbps	160 kbps
Immersive Audio 7.1.4 Playback	224 kbps	288 kbps

Listening tests at an independent test site were conducted as part of the DVB standardization of AC-4. The table below shows the average results that were achieved:

	MUSHRA Score
5.1 at 192 kbps	90
5.1 at 96 kbps	75
Stereo at 96 kbps	92
Stereo at 64 kbps	89
Mono at 32 kbps	91
Mono at 24 kbps	80

As part of the ATSC 3.0 standardisation effort, precertification listening tests using BS.1116 were conducted with a clear request to show what bit rate was required to achieve above a 4.0 per item score on critical items. AC-4 achieved this with an average score across the items as per the table below.

Format	Bit Rate	BS.1116 Score
Stereo	96 kbps	4.61
5.1	192 kbps	4.63
Immersive Audio 7.1.4 Playback	288 kbps	4.64

The various AC-4 coding tools are employed selectively based on the content, channel or object configuration, and target data rate for optimal performance across a wide range of data rates and audio types—from mono dialogue tracks through complex OBA. This gives AC-4 tremendous flexibility, making it an excellent choice for all types of broadcast and Internet delivery services.
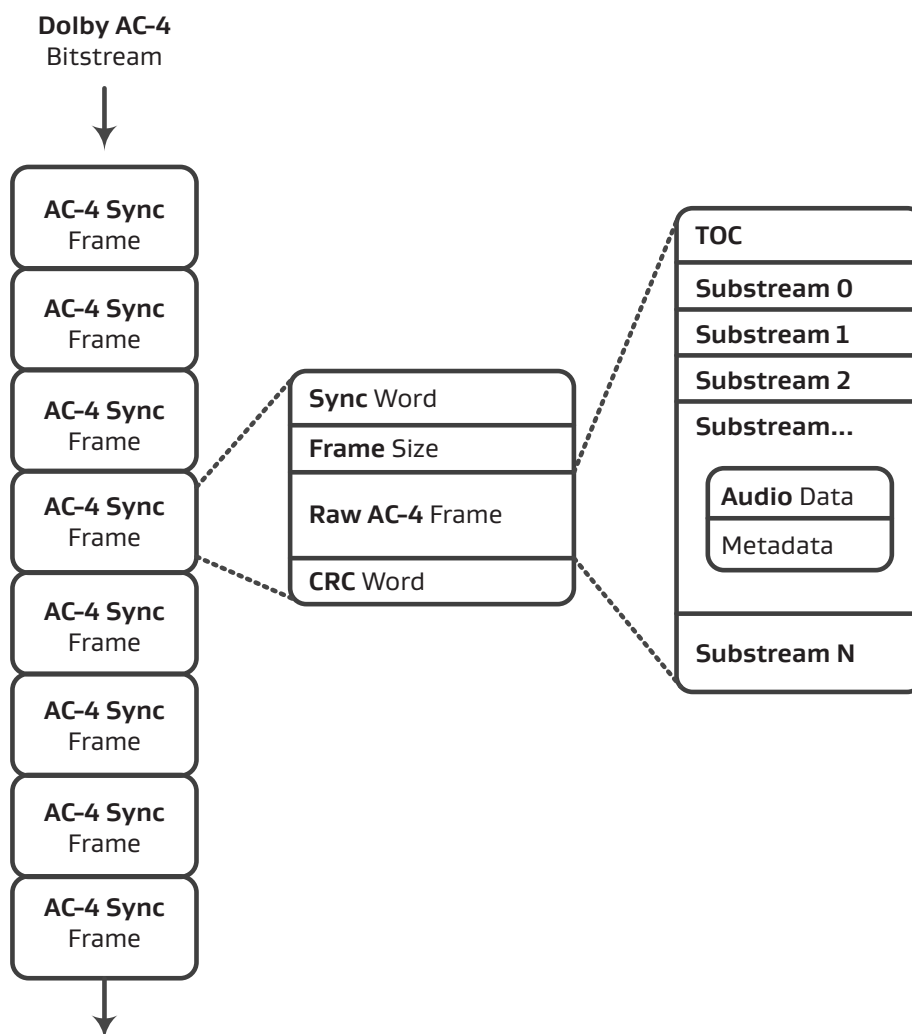
The table below shows how the various tools are applied for coding channel-based audio across a broad range of channel configurations and data rates.

Codec Mode		MDCT Domain		QMF Domain			Time Domain	
		Spectral Frontend	SAP	A-SPX	Advanced Coupling	DRC DE	SRC	Limiter
Mono	16–64 kbps	✓		✓		✓	✓	✓
	>64 kbps	✓				✓	✓	✓
Stereo	24–48 kbps	✓	✓	✓	✓	✓	✓	✓
	48–96 kbps	✓	✓	✓		✓	✓	✓
	>96 kbps	✓	✓			✓	✓	✓
5.1	(Up to 5.1 Ch) 64–128 kbps	✓	✓	✓	✓	✓	✓	✓
	(Up to 5.1 Ch) 128–320 kbps	✓	✓	✓		✓	✓	✓
Immersive Audio 7.1.4 Playback	(Up to 7.1 Ch) 160–256 kbps	✓	✓	✓	✓	✓	✓	✓
	(Up to 7.1 Ch) 256–320 kbps	✓	✓	✓		✓	✓	✓
	(Up to 7.1 Ch) >320 kbps	✓	✓			✓	✓	✓

4 Overview of Bitstream Syntax

An AC-4 bitstream consists of synchronization frames, each beginning with a sync word and optionally ending with a cyclic redundancy check (CRC) word. The sync word allows a decoder to easily identify frame boundaries and begin decoding. The CRC word allows a decoder to detect the occurrence of bitstream errors and perform error concealment when it detects an error.

The data carried within each synchronization frame is referred to as the raw AC-4 frame. Each raw frame contains a Table of Contents (TOC) and at least one substream containing audio and related metadata. The following figure shows the high-level bitstream structure.

Dolby AC–4
Bitstream

AC–4 Sync
Frame

AC–4 Sync
Frame

AC–4 Sync
Frame

AC–4 Sync
Frame

AC–4 Sync
Frame

AC–4 Sync
Frame

AC–4 Sync
Frame

AC–4 Sync
Frame

Sync Word

Frame Size

Raw AC–4 Frame

CRC Word

TOC

Substream 0

Substream 1

Substream 2

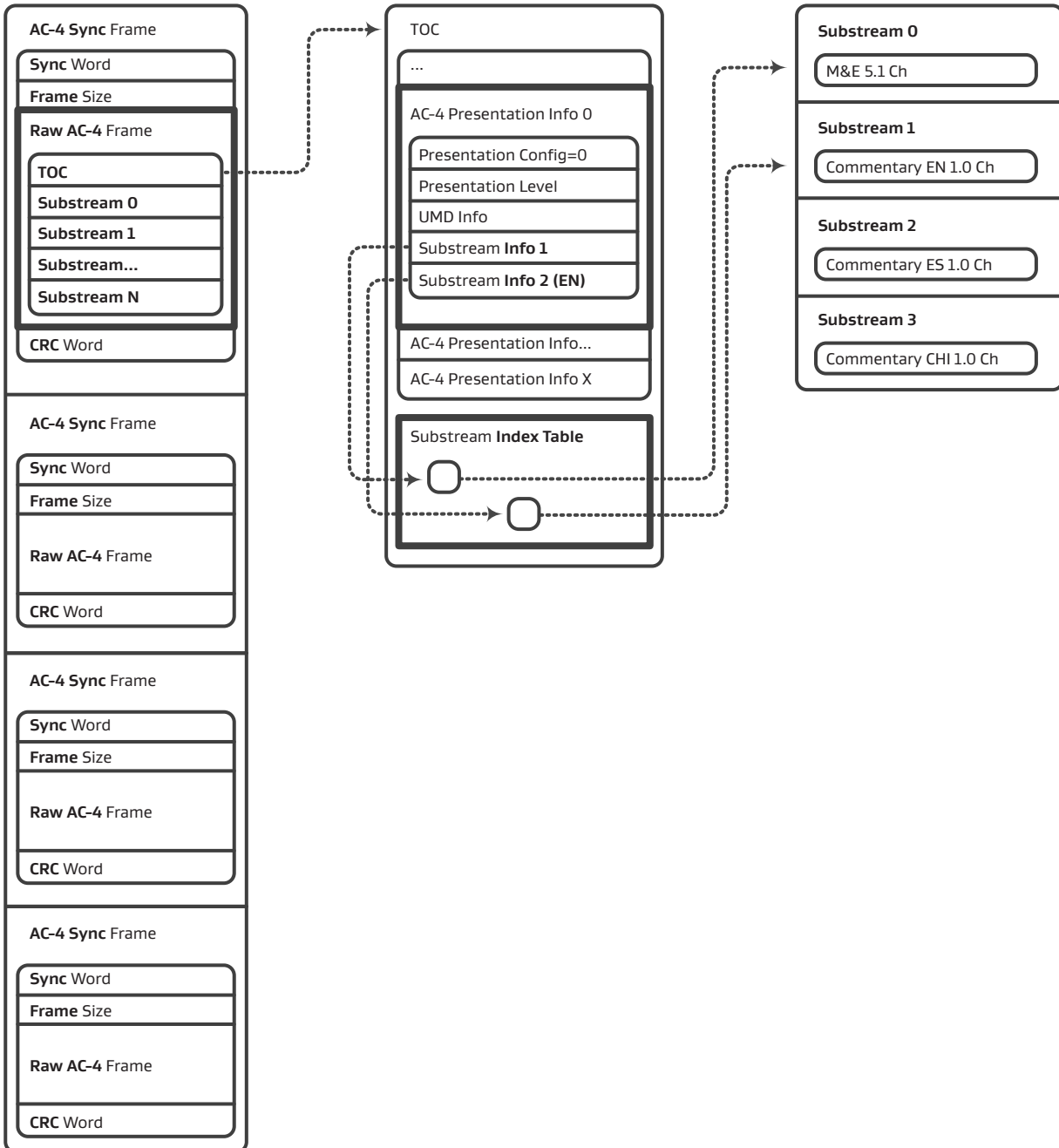**Substream...**

Audio Data

Metadata

Substream N

Figure 10: High-Level Bitstream Syntax (AC-4 Part 2)

The TOC contains the inventory of the bitstream. Each audio substream can carry either one or more audio channels or an individual audio object. This structure provides flexibility and extensibility that allows the AC-4 format to meet future requirements.

AC-4 also allows multiple Presentations to be carried in a single bitstream. Each Presentation defines a way of mixing a set of audio substreams to create a unique rendering of the program. Instructions for which substreams to use and how to combine them for each Presentation are specified in a Presentation info element carried in the TOC.

Presentations enable multiple versions of the audio experience, such as different languages or commentary, to be delivered in a single bitstream in a convenient, bandwidth-efficient manner. An example is shown in the figure below, where four versions of a live 5.1 sports broadcast—the original English version, two alternate languages (Spanish and Mandarin Chinese), and a commentary-free version—are combined into a single AC-4 bitstream.

Figure 11: Live 5.1 Sports Broadcast with Four Presentations

Rather than transmitting four separate 5.1-channel streams, as would be done with current technologies, the AC-4 bitstream contains four substreams:

- Music/effects mix without commentary (5.1 channels)
- English commentary (mono)
- Spanish commentary (mono)
- Mandarin Chinese commentary (mono)

The TOC contains four Presentation info elements, one for each playback experience. The figure above shows the English Presentation (Presentation 0) selected, which instructs the decoder to combine the music/effects and English commentary substreams to create the English output. Similarly, the Spanish and Mandarin Chinese Presentation info elements instruct the decoder to render the common music/effects substream with the relevant language commentary, while the commentary-free Presentation info element renders only the music/effects substream. All necessary mixing is done in the decoder, eliminating the need to implement this with surrounding system components.

The table below demonstrates how this Presentation-based approach can offer data-rate savings in excess of 50% compared with using the same format to deliver multiple 5.1 streams.

Conventional Approach		Presentation Approach	
Stream	**Data Rate/kbps**	**Stream**	**Data Rate/kbps**
English 5.1	144	English Commentary 1.0	40
Spanish 5.1	144	Spanish Commentary 1.0	40
Chinese 5.1	144	Chinese Commentary 1.0	40
Commentary-free 5.1	144	Commentary-free 5.1	144
TOTAL	576	TOTAL	264
		Savings	**54%**

Comparing the Presentation approach of AC-4 with delivery of multiple 5.1 streams using Dolby Digital (as used by many HDTV services today), the data-rate savings exceed 80%.

Use of Presentations also provides a way to deliver optimal audio experiences to devices with very different capabilities using a single audio bitstream. For example, decoders in some devices might be able to decode only up to 5.1 channels, where the additional channels are unnecessary for that device and would impose an unrealistic processing load.

A service provider wishing to offer, say, a 7.1.4-channel service could ensure compatibility with both simple and advanced devices by incorporating a 7.1.4-channel and a 5.1-channel Presentation within a single AC-4 bitstream. The formal interoperability test program for Dolby AC-4 ensures that all decoders conform to clearly defined functional levels so that advanced services can easily be configured to suit the target device base.
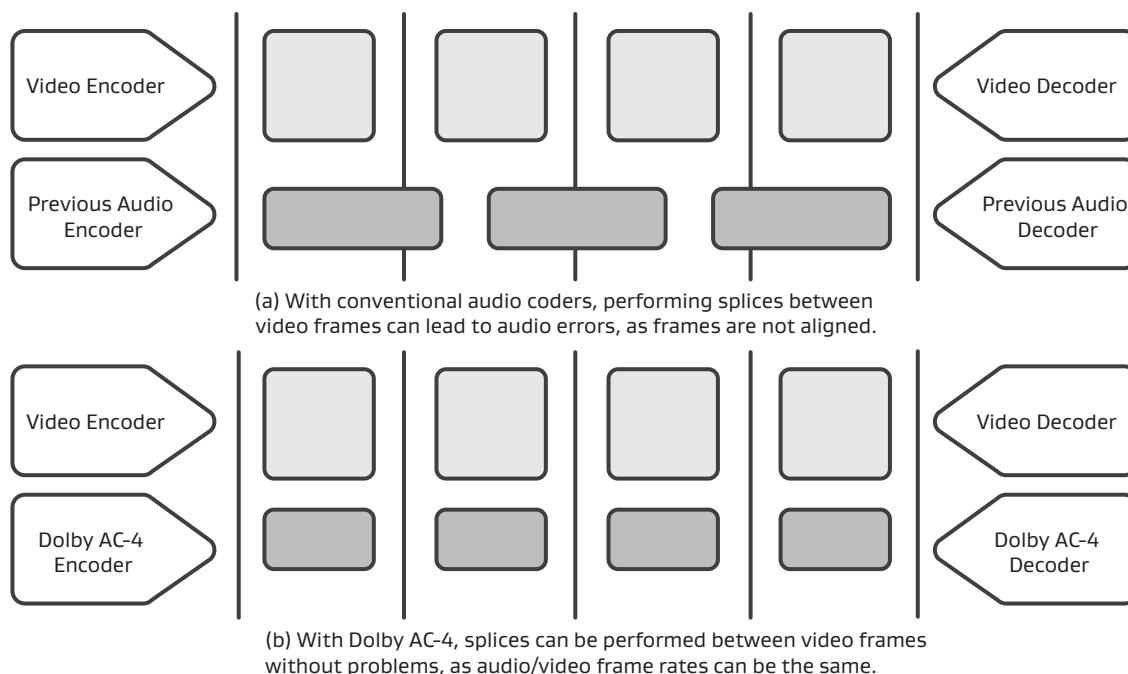
5 System Features

5.1 Audio/Video Frame Alignment

In current digital broadcast systems, encoded audio and video utilize different frame rates. Although in isolation these rates make sense, the combination of two different rates in a final delivery stream or package makes further manipulation of the program in the transport domain complex. Applications such as editing, ad insertion, and international turnarounds become challenging to implement, as the switching points at the end of video frames do not align with the ends of audio frames.

If not implemented carefully, this can result in sync errors between video and audio or audible audio errors. Current solutions to this involve decoding and re-encoding the audio, which introduces potential sync errors, quality loss, and, in the case of OBA, misalignment of audio and time-critical positional metadata.

In Dolby AC-4, a new approach is taken. The Dolby AC-4 encoder features an optional video reference input to align the audio and video frames. The encoded audio frame rate can therefore be set to match the video frame rate, and as a result, the boundaries of the audio frames can be precisely aligned with the boundaries of the video frames. The Dolby AC-4 system accommodates current broadcast standards, which specify video rates from 23.97 to 60 Hz as well as support for rates up to 120 Hz for new ultra high-definition specifications.

(a) With conventional audio coders, performing splices between video frames can lead to audio errors, as frames are not aligned.

(b) With Dolby AC-4, splices can be performed between video frames without problems, as audio/video frame rates can be the same.

Figure 12: Audio/Video Frame Alignment in AC-4

This approach simplifies implementation for developers of downstream systems and reduces the risk of sync errors and other artifacts caused by trimming or cutting a stream. It does not require that the cut point be known when encoding the material, which provides flexibility for downstream manipulation within headends, the delivery network, and consumer devices.
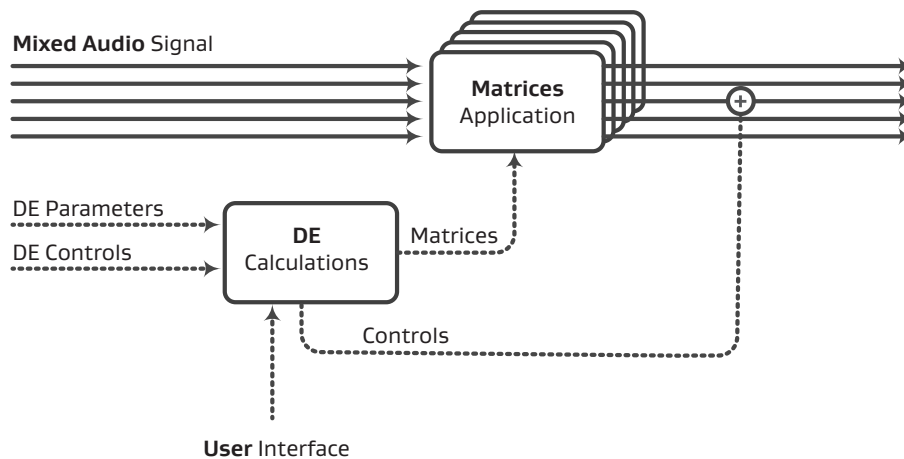
5.2 Dialogue Enhancement

In practical tests, viewers have been found to have widely differing preferences for dialogue or commentary levels. Dolby AC-4 addresses this challenge by providing mechanisms for viewers to tailor the dialogue level to suit their individual preferences. These flexible mechanisms work with both legacy content that contains dialogue already mixed into the main audio and new content where a separate dialogue track is available to the Dolby AC-4 encoder.

Conventional dialogue enhancement techniques in products such as TVs and tablets have applied single-ended processing in the playback device to attempt to detect and adjust dialogue elements of the mixed audio. While these techniques have the advantage that they do not require specially produced content, they have the disadvantage of not being wholly predictable, and their effectiveness is limited by the processing power available in the playback device.

With Dolby AC-4, dialogue enhancement is instead implemented by utilizing the dramatically higher processing power of the audio encoder to analyze the audio stream and generate a highly reliable parametric description of the dialogue, whether or not a separate dialogue track is available. These parameters are sent with the audio in the AC-4 stream and used by the playback device to adjust the dialogue level under user control.
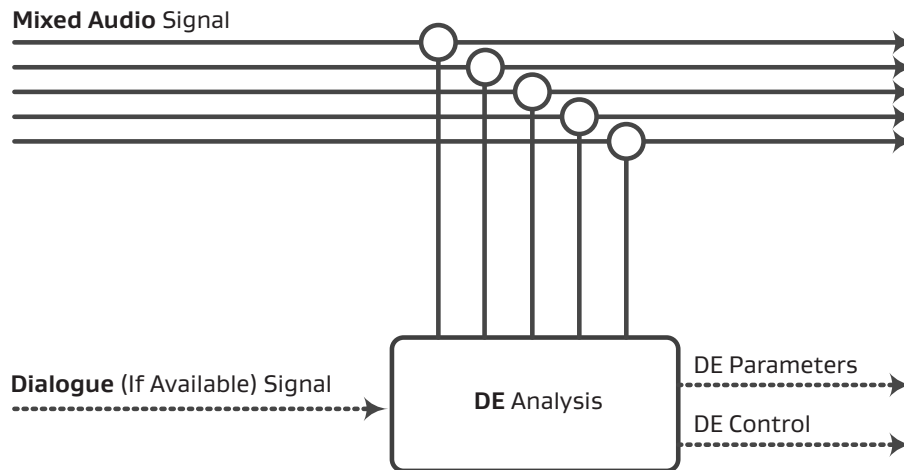
Figure 13: Dialogue Enhancement for Channel-Based Content (Encoder Functionality)

Figure 14: Dialogue Enhancement for Channel-Based Content (Decoder Functionality)

If the dialogue is available as a separate audio track, the encoder creates the parameters based on the joint analysis of the mixed audio signal and the separate dialogue signal. These parameters are more precise than those extracted from the mixed audio signals as described previously and allow more precise dialogue adjustments in the decoder.

Alternatively, if desired (for example, to perform language substitution), the dialogue and

music/effects tracks can be sent in the AC-4 bitstream as separate substreams for optimum performance and maximum flexibility.

The combination of guided and automatic modes in the Dolby AC-4 system means that dialogue enhancement functionality may be implemented by the service provider in a broad, predictable, and effective manner. The automatic analysis method (at the encoder) provides a solution that is easy to deploy with legacy content and workflows as the industry transitions to interchange of separate dialogue tracks in the long term.

5.3 Advanced Loudness Management

Over the last decade, approaches to managing the loudness of broadcast and streaming services have changed considerably. The broadcast industry, for example, has made significant steps towards using a long-term loudness measure, rather than just a short-term peak measure, to align the levels of programming and provide a more consistent and pleasant experience for viewers. This has resulted in recommendations from the International Telecommunication Union on international program interchange levels, from the European Broadcasting Union on broadcast levels, and from several other national and international groups on local requirements.

However, the need to achieve loudness consistency and meet regional loudness mandates has often led to the introduction of loudness processing at multiple points in the chain—for example, in content creation, in the broadcast station, and at the operator. In many cases, this redundant processing results in compromised sound quality.

To help services ensure loudness consistency and compliance with regulations, the Dolby AC-4 encoder incorporates integrated intelligent loudness management. The encoder assesses the loudness of incoming audio and can, if desired, update the loudness metadata (dialnorm) to the correct value or signal the multiband processing required to bring the program to the target loudness level. Rather than processing the audio in the encoder, this information is added to the bitstream in the DRC metadata so processing can be applied downstream in the consumer device appropriately for the playback scenario. The process is therefore non-destructive; the original audio is carried in the bitstream and available for future applications.

To avoid the problems associated with cascaded leveling processes, Dolby AC-4 makes use of the extended metadata framework standardized in the European Telecommunications Standards Institute (ETSI) 102 366 Annex H. This framework carries information about the loudness processing history of the content so that downstream devices can intelligently

disable or adjust their processing accordingly, maximizing quality while maintaining consistency. Annex H metadata can be carried throughout the program chain, either with the baseband audio prior to final encoding or inserted into transmission bitstreams including Dolby Digital Plus and AC-4.

If the incoming audio presented to the Dolby AC-4 encoder has previously been produced or adjusted to a target loudness level by a trusted device, this can be signalled to the encoder using Annex H metadata. In this case, the integrated loudness leveling processing of the encoder will be automatically disabled, so that the audio is delivered without further adjustment, maximizing quality and preserving the original creative decisions.

Because the Annex H metadata in the AC-4 bitstream also indicates any loudness processing that has been applied, this can be used to automatically disable unnecessary loudness processing that might be in place downstream—for example, in cable or Internet Protocol Television (IPTV) headends. The extended Annex H metadata also includes additional loudness measures such as short-term loudness, which can assist compliance in regions that regulate based on these characteristics.

6 Decoder System Features

6.1 Encoding and Decoding (Core and Full) Scenarios for Immersive Audio

This section highlights the availability of two decoding modes—core decoding and full decoding—by giving examples of encoding and decoding scenarios for channel-based and spatial object groups–based immersive audio.

The core decoding mode enables a lower-complexity, reduced-channel-count or reduced-object-count output from the decoder, while the full decoding mode does the full decoding of the stream. These two decoding modes are made possible by the structure of the bitstream and the coding methods employed. They are an essential feature of AC-4, allowing a stream to play on lower-cost, lower-complexity devices, typically outputting stereo or 5.1, as well as on full-capability devices with high-channel-count output or other needs for full fidelity.

6.1.1 Object-Based (Spatial Object Groups) Immersive

In the case of object-based audio using spatial object groups, the input to the encoder consists of spatial object groups (of which there are 15 in the example below) and the LFE channel, as

well as their corresponding object audio metadata (OAMD). An Advanced Joint Object Coding (A-JOC) module on the encoder side is used to provide the A-JOC data to the bitstream.

In the example below, the spatial coding module further reduces the spatial object groups to seven while creating associated OAMD. The seven spatial groups are then encoded using the core encoder modules.
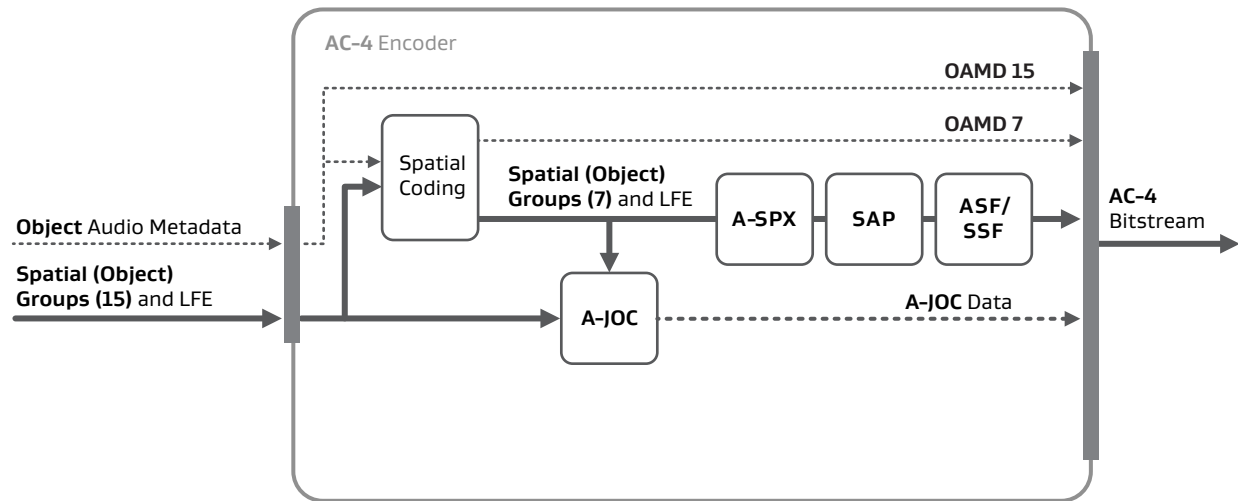
Figure 15: Encoding Object-Based Audio (Adaptive Downmix)

As outlined in the figure below, the same bitstream can then be decoded by a playback device running in core decoding or full decoding. The difference is that for full decoding the A-JOC decoding module is used, resulting in 15 spatial object groups being output by the decoder. In both decoder modes there is the need for a renderer, which is further described in section 6.2.
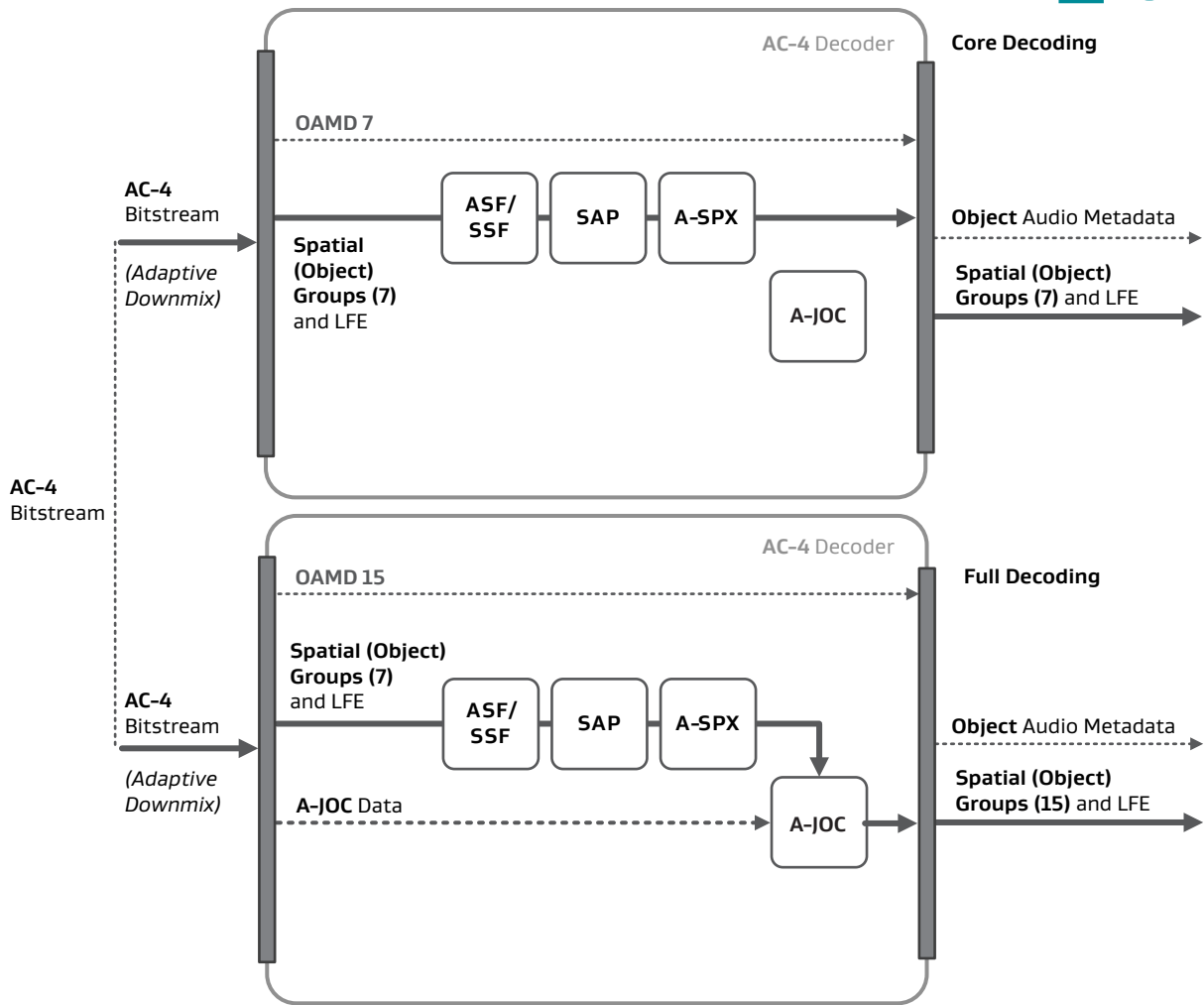
**Figure 16: A-JOC Core and Full Decoding**

6.1.2 Channel-Based Immersive

For channel-based immersive audio (which in the example below is 7.1.4), different tools are used depending on the bit rate.

These tools, which code the spatial properties of the audio signal, aim to reduce the number of waveforms to be coded by the subsequent tools (A-SPX, SAP, and ASF/SSF), and in doing so create a parametric representation. When A-CPL is used, "side signals" are also created; these are conceptually similar to side signals in traditional mid-side coding.

- Advanced Joint Channel Coding (A-JCC) is used for the lowest possible bit rates; a core of 5.1 channels are coded.

- Advanced Coupling (A-CPL) is used for intermediate to high rates; a core of 5.1.2 audio channels are coded, and the audio side signal can optionally be coded to further increase audio quality.

The figure below illustrates the channel-based immersive encoder.
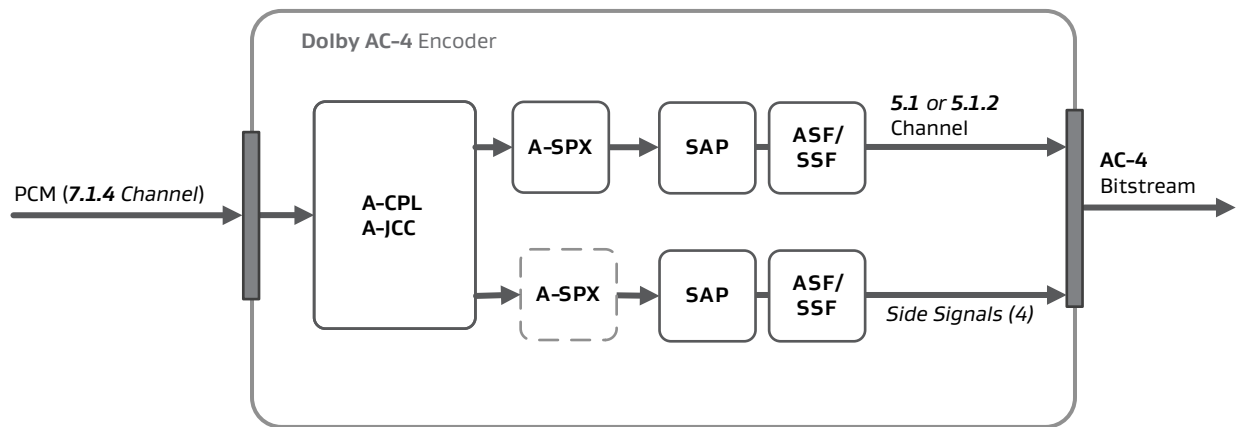
Figure 17: Dolby AC-4 Encoder (7.1.4 Input Example)

If the decoder is configured to do core decoding, the 5.1 or 5.1.2 waveform-coded channels are decoded by the core decoding tools—Audio Spectral Frontend (ASF), SAP, Advanced Spectral Extension (A-SPX), and, optionally, the core part of A-JCC, depending on coding configuration.

If the decoder is configured to do full decoding, the 5.1 or 5.1.2 waveform-coded channels, along with optional side signals, are decoded by the core decoding tools ASF, SAP, A-SPX, and the A-CPL or A-JCC, depending on coding configuration.

The figure below illustrates the differences in the PCM output decoding scenario when doing core decoding and full decoding.
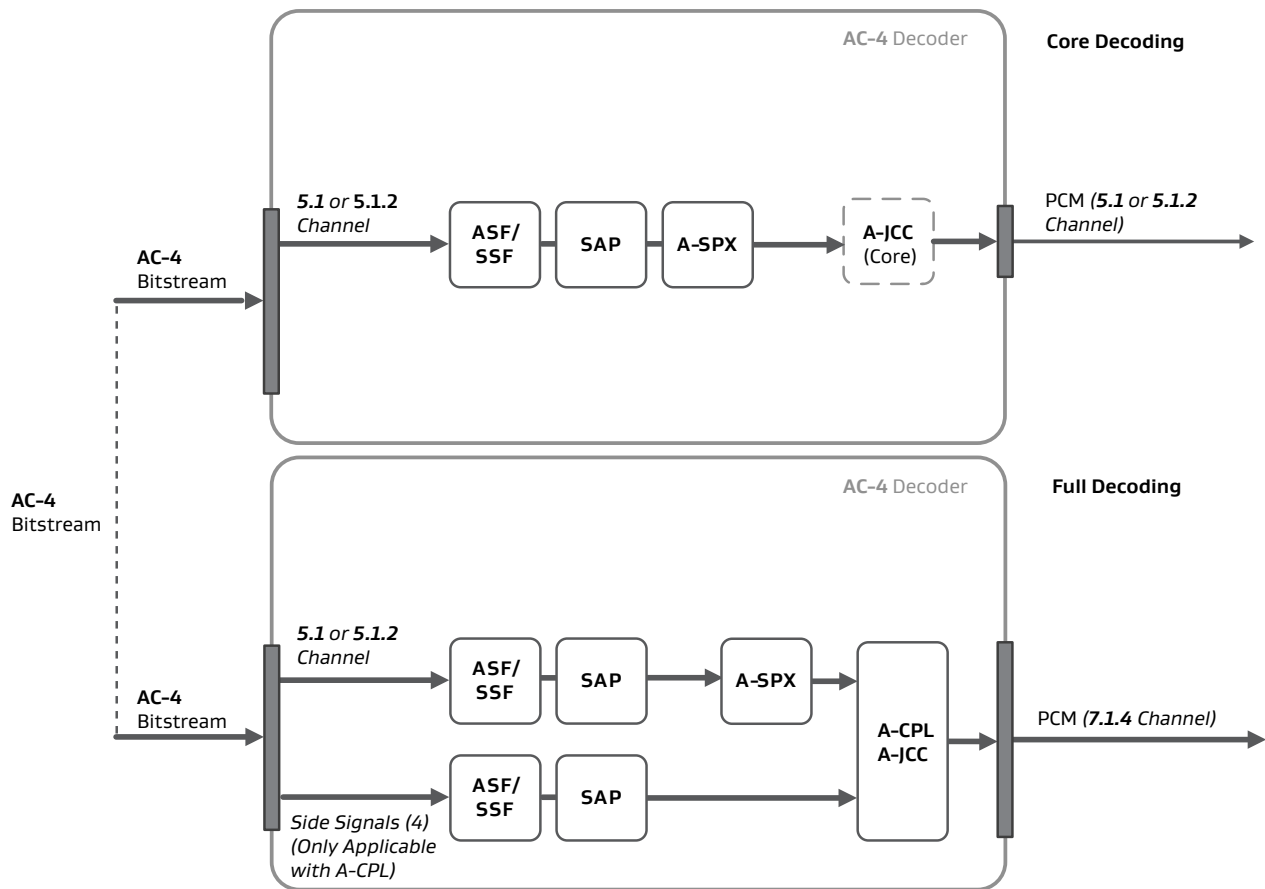
Figure 18: PCM Output Decoding

6.2 Audio Renderer

The audio renderer mixes the audio substreams into the required number of output channels, most commonly stereo or 5.1 surround. For applications containing multiple languages or multiple commentaries the renderer can combine channel-based substreams such as a 5.1-channel music/effects mix and separate dialogue tracks. The appropriate substreams are identified using the Presentation info element of the TOC, as discussed in Section 4.

For object-based audio (OBA) the renderer accepts audio objects and accompanying metadata, such as object type, position, width, divergence, and so on. For each object, the audio renderer determines the best way to recreate its position with the speakers available, creating the best representation of the original experience in the listening environment.

Because object rendering is a critical process for faithful reproduction, the Dolby AC-4 decoder uses a renderer based on that used on film mix stages for Dolby Atmos content mastering and in other Dolby object audio monitoring tools.

Different products require different rendering capabilities in addition to the core and full decoding capabilities. The picture below shows that a simple renderer only capable of outputting stereo and 5.1 is sufficient in many cases, while an advanced render is needed to render to higher numbers of speakers with a high degrees of flexibility.
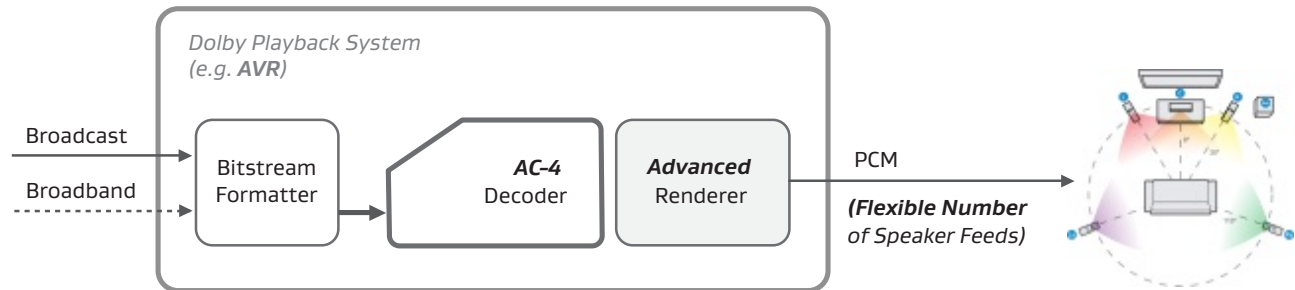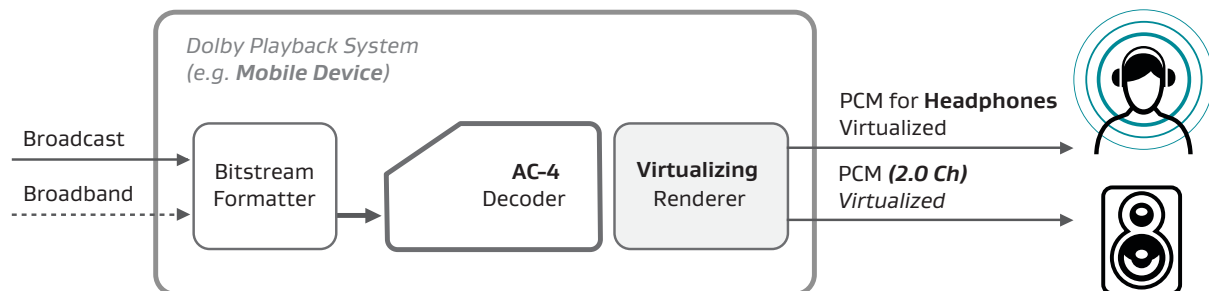

Figure 19: AVR Playback System
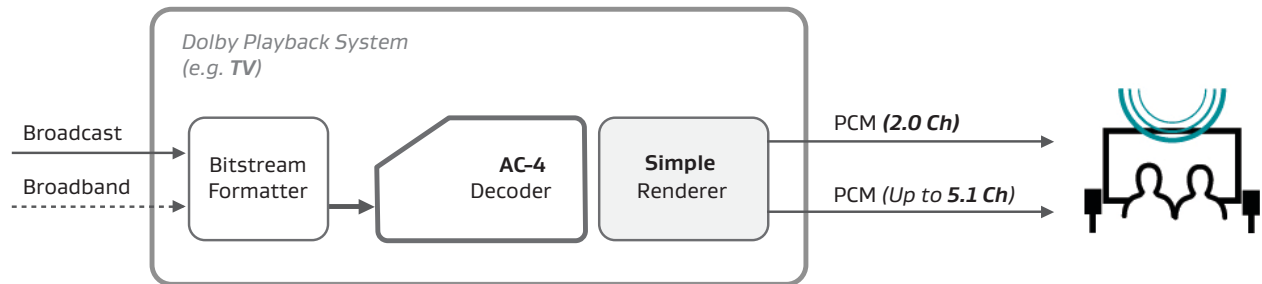

Figure 20: TV Playback System

In other playback conditions the renderer can be combined with a virtualizer to efficiently provide immersive audio experiences over headphones or stereo loudspeakers, which is well suited to portable use cases.

Figure 21: Mobile Device Playback System

With the successful launch of Dolby Atmos in the home theatre world, there is an installed base of immersive audio playback systems that can be reached by sending decoded audio channels and objects via Metadata-enhanced Audio Transmission (MAT) over HDMI. This

means that even though there is currently no AC-4–capable AVR, it is possible to provide an immersive experience building on the success of Dolby Atmos.
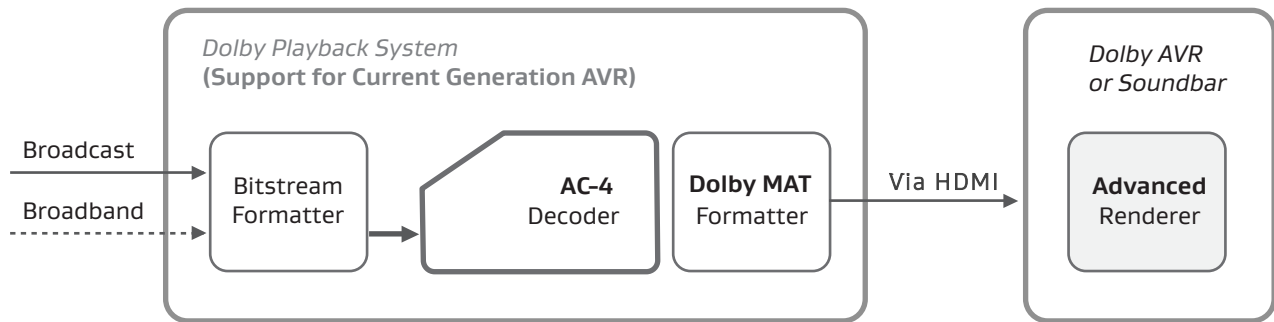
Figure 22: Connectivity to Current Dolby Atmos AVRs

6.3 Dynamic Range Control (DRC)

The Dolby AC-4 decoder applies DRC to tailor the dynamic range and the typical output level to suit the listening scenario. As outlined in Section 2, implementing DRC in the QMF domain enables powerful multiband and multichannel processing which improves quality over previous wideband approaches.

Dolby AC-4 supports a number of DRC modes to adapt the content to different listening environments and playback scenarios. Each mode is associated with a type of playback device and has guidelines for decoder-defined playback reference levels.

Four standard DRC decoder modes have been defined, each with a corresponding output level range. In addition to the standard modes, it is possible to add up to four user-definable modes to support future or proprietary device types.

| DRC Decoder Mode	Output Level Range
AVR and Home Theater	–30 to –27
Flat-panel TV	–16 to –17
Portable Device Speakers	–16 to 0
Portable Device Headphones	–16 to 0
Four More Reserved	
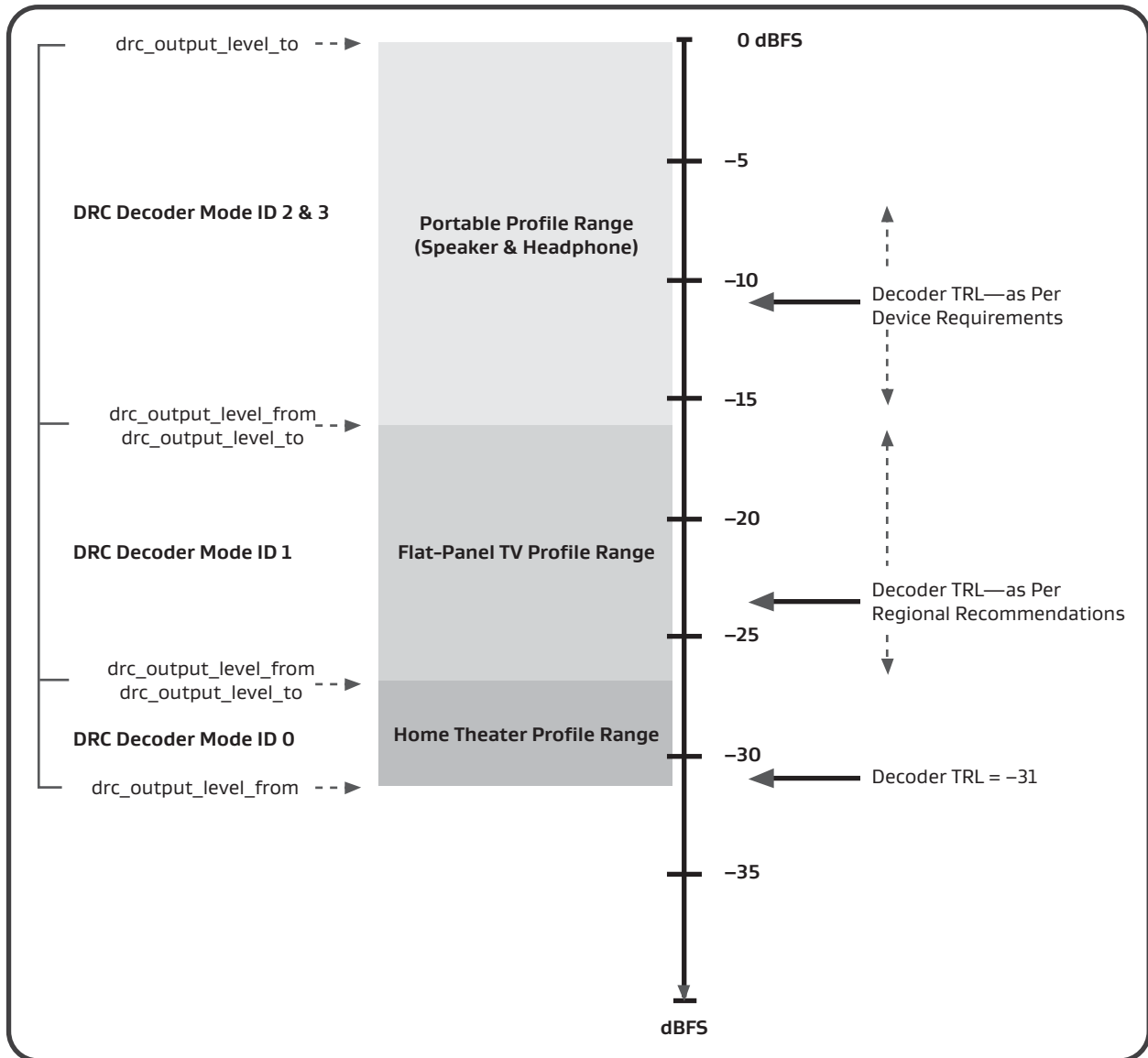
Figure 23: Playback Device Target Reference Level and DRC Profile Mapping

An AC-4 decoder may be provisioned to align with a number of device categories and applications along with how the target reference level (TRL) parameter maps to the 4 decoder dynamic range control modes as shown in Figure 23. For example, if −23 dBFS is selected for the target reference level (TRL), the AC-4 DRC Decoder Mode ID (representing the parameterized compression curve for that playback mode) overlapping with -23 will be used to generate the DRC information in the decoder, which in this example is the Flat Panel TV profile. For Home Theatre AVR use cases, the decoder user/system would provision the decoder target reference loudness to -31.

Selecting -31 TRL will place the DRC Profile selection into the DRC Decoder Mode ID 0 range, and the system will use the Home Theatre Profile to generate the DRC info in the decoder. Portable/Mobile devices have flexibility to set the TRL value to best align with their internal gain structure needs.

DRC parameters for each output mode are generated by the Dolby AC-4 encoder (Figure 24) or by an external third-party processor (Figure 25) and transported in the AC-4 bitstream as DRC gain values (wideband or multiband). Alternatively, the desired DRC characteristic can be expressed in the bitstream as a parameterized compression curve.
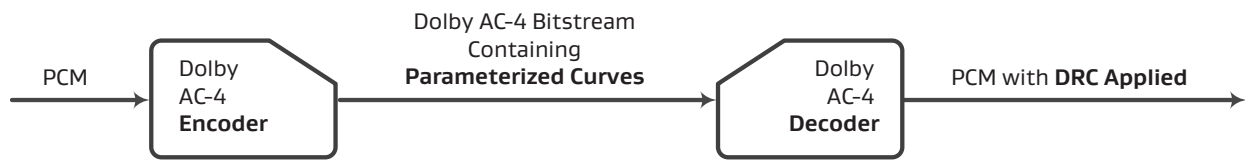
Figure 24: Parameterized Curves

This curve can be created by the service provider or content creator to suit the content and their house style. These curves may be selected from a number of presets in the encoder or may be customized if desired. Parameterized compression curves provide benefits such as lower bit-rate overhead and higher audio quality for traditional channel audio content, with even larger gains for OBA and immersive audio content, where a DRC gain per channel or object, as used in other systems, becomes costly.
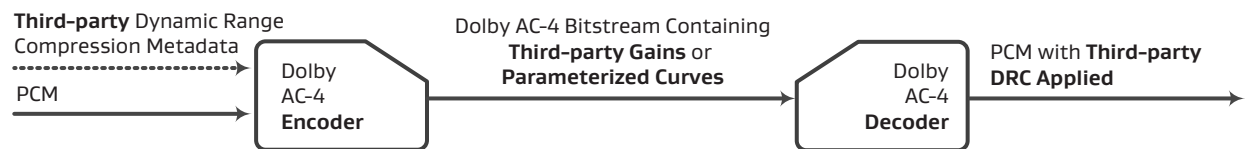
Figure 25: Third-Party Gains

The Dolby AC-4 decoder calculates gains based on the compression curve transmitted for the selected mode and the target playback reference level of the device. The target playback reference level for each mode is not fixed but instead can be defined in the decoder. This enables flexibility to match the loudness of other content sources depending on the listening scenario.

6.4 Simultaneous Outputs

To enable a Dolby AC-4 decoder to support multiple outputs—for example, to feed several different devices simultaneously—the decoding process is split into two stages: the input stage and the output stage.

The input stage decodes the selected channels and objects, provides the configuration of the current frame, and performs error detection.

Each output stage adapts the audio to the particular playback scenario:

- Output mapping—for example, stereo, multichannel, or objects
- User preferences—for example, the selection of dialogue enhancement or the main and/or associated audio
- Environmental conditions—for example, late-night mode
- Other device-specific playback optimization

An arbitrary number of output stages can be connected to a single input stage, which allows rendering different output configurations from one input bitstream simultaneously. Each output stage can be freely configured—for example, one for headphone output at a high output level and another for 5.1-channel mode at line level. This saves computational load compared with running an input decode stage for each output.

The diagram below shows examples of a decoder with a single output and a decoder with three independent outputs, each of which is optimized for a different kind of device.
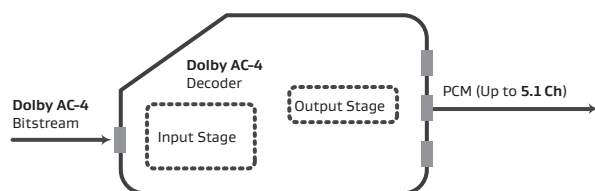
Figure 26: Single Output
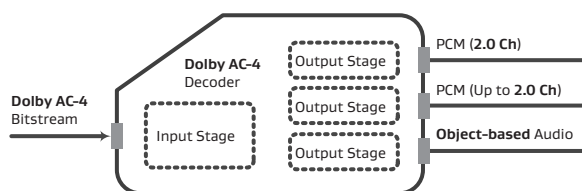
Figure 27: Three Independent Outputs

6.5 Seamless Switching

The Dolby AC-4 decoder supports seamless switching between data rates and configurations. This is achieved by a decoder that reconfigures itself according to the input frame. A smooth transition is ensured because all relevant metadata is carried in precise alignment with the audio to which it refers.

Seamless switching also enables glitch-free transitions between different audio streams. To facilitate detection, the sequence number of each frame is checked in the decoder; if the frame numbers are nonconsecutive, a crossfade is triggered. Because the codec operates on the overlap-add principle, there is sufficient audio available to perform a short crossfade so as to avoid any artifacts that would otherwise arise from a hard switch. If the content is the same before and after a switch—as in, for example, a switch of the bit rate in an adaptive streaming scenario—that transition becomes unnoticeable.

This allows a service provider using adaptive streaming to utilize a wide range of bit rates / operation points with one audio coding system, as the Dolby AC-4 decoder will smoothly switch at transition points between rates. Furthermore, to ease the integration in playback devices, the Dolby AC-4 decoder has a fixed latency across all configurations. This simplifies integration for developers, as it is not necessary to add code to reconfigure the system depending on the input type. It also removes the risk of sync errors and audible glitches when switching configurations.

7 System Extensibility

In Section 4 above, it was noted that extensibility of the AC-4 format is enabled through the use of a TOC, which specifies the number of audio substreams and Presentations that are carried by a stream. This provides the flexibility to configure streams containing multiple versions of the audio to suit different users or devices. It also enables services to be offered in the future that leverage completely new features, without compromising compatibility with deployed legacy receivers.

Dolby AC-4 also provides the potential for insertion of additional streams of third-party data, which will travel transparently in the encoded audio bitstream, all in perfect synchronization. If desired, the data may be secured so that it cannot be modified without detection.

Applications of this functionality for service providers are potentially broad and powerful. One example is in audience measurement applications, where the data stream is used for securely carrying a program identification code that could be extracted in the client of any suitably equipped playback device. Other applications include sending signaling to other home devices—for example, to trigger advertising on other screens, to control other entertainment-related devices such as seat rumblers, or to trigger events on the increasingly diverse array of intelligent devices in the home or car.

8 Standardization and Deployment

In readiness for next-generation service deployments, the core of the Dolby AC-4 technology has been standardized by the ETSI as TS 103 190 in April 2014, which may be downloaded freely via the ETSI website.

In February 2015, AC-4 was added to the latest version of the DVB audio video toolbox (TS 101 154 v13) for use in new deployments such as DVB-T2 and ultra high-definition broadcast systems. It is also proposed for forthcoming specifications such as ATSC 3.0 and for next-generation streaming services. Part 2 of the ETSI AC-4 standard process, which adds OBA enhancements, is underway in 2015.

Dolby AC-4 is being made available via approved chip and software implementations to ensure consistency of operation. An interoperability-testing program ensures that all products incorporating Dolby AC-4 perform consistently and fully support the relevant feature sets.

9 Conclusion

This paper has discussed the features and capabilities of a new audio delivery system, Dolby AC-4. The format offers several performance and system-level advancements that will benefit a broad range of next-generation services, including lower data rates, multidevice optimizations, audio/video frame alignment, and extended loudness management. It capitalizes on state-of-the-art technology developed from both the Dolby Digital and HE-AAC families and on practical experience gained from deploying digital broadcast and streaming services in all regions of the world.

Dolby AC-4 provides mechanisms that will enhance access and engagement for a broader audience, such as dialogue enhancement. It also offers a progression path for the implementation of new user experiences, from simple multi-commentary applications to object-based personalization and fully immersive Dolby Atmos cinematic experiences.

The capabilities of Dolby AC-4 highlight the fact that it is a single format to be utilized across a full range of services, from basic to premium and traditional broadcast to streaming. And, now, with AC-4's standardization by the industry, planning for deployment can begin.