Dolby

Dolby® AC-4: Audio delivery for next-generation entertainment services

February 2021

Contents

Contents

1 Introduction

Consumer devices are evolving to facilitate increasingly sophisticated entertainment experiences. As devices are paired to new distribution platforms, audiences demand access to their entertainment anywhere, at home or on-the-move. These demands require a comprehensive, flexible audio delivery technology.

Dolby® AC-4 is an audio delivery system designed to exceed expectations. Integrating easily into content workflows and consumer devices, the system supports existing needs and ongoing innovation of broadcast and streaming services. This white paper describes the technical features and capabilities of the system.

Dolby AC-4 has been widely adopted by consumer electronics manufacturers and in next-generation delivery platform specifications. The core technology has been standardized by the European Telecommunications Standards Institute (ETSI) as TS 103 190. It has been adopted by Digital Video Broadcasting (DVB) in TS 101 154, by the Advanced Television Systems Committee (ATSC) for ATSC 3.0, and by numerous regional specification groups.

2 System overview

Dolby AC-4 is an audio delivery system designed from a clean sheet. Delivery and playback are optimized across a broad range of devices through high-efficiency audio coding and powerful system-level features. The system supports conventional channel-based audio and Dolby Atmos immersive audio at low data rates. There are also a number of accessibility advancements over existing solutions, including Dialogue Enhancement for greater clarity. Furthermore, Dolby AC-4 fully supports object-based audio (OBA), allowing for sophisticated personalization.

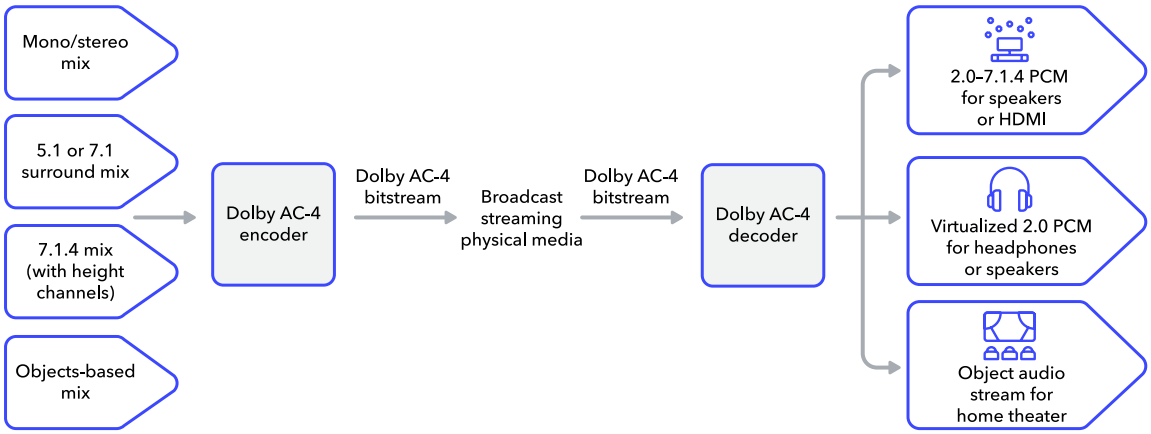

FIGURE 1: Dolby AC-4 can carry conventional channel-based soundtracks and object-based mixes. Whatever the source type, the decoder renders and optimizes the soundtrack to suit the playback device.

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.

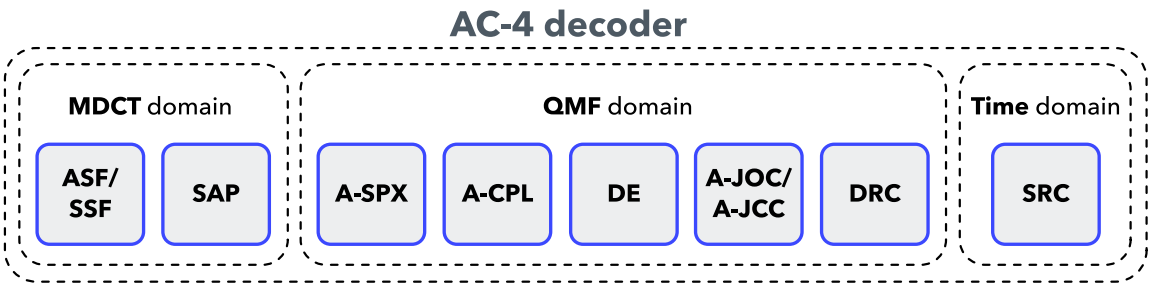

FIGURE 2: AC-4 decoder

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.

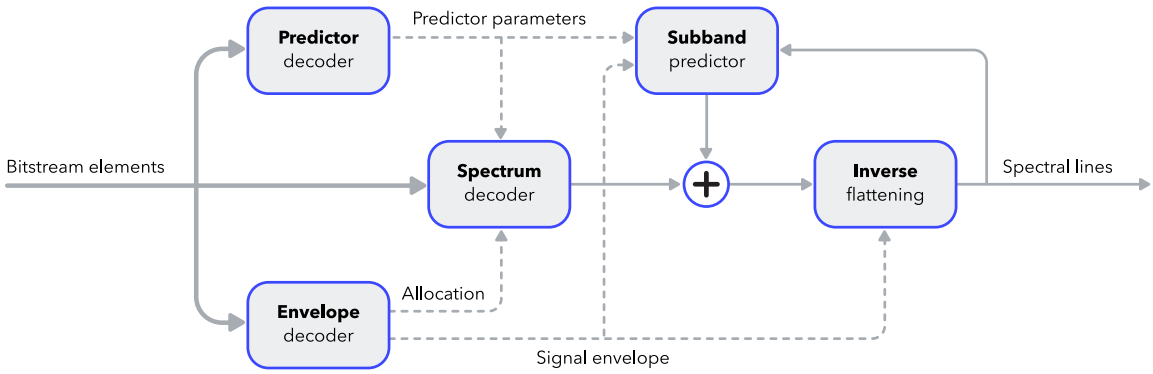

FIGURE 3: Top-level structure of the SSF decoder

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.

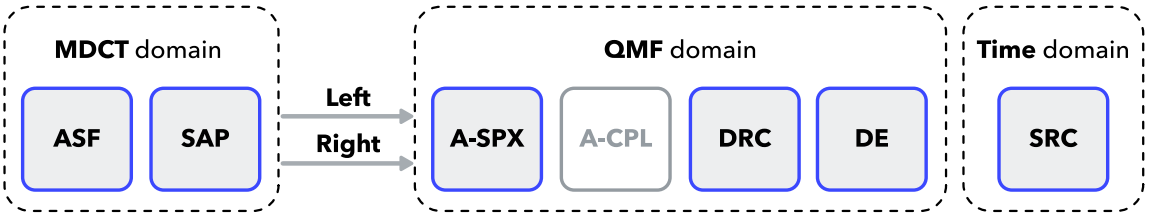

FIGURE 4: SAP output (left/right)

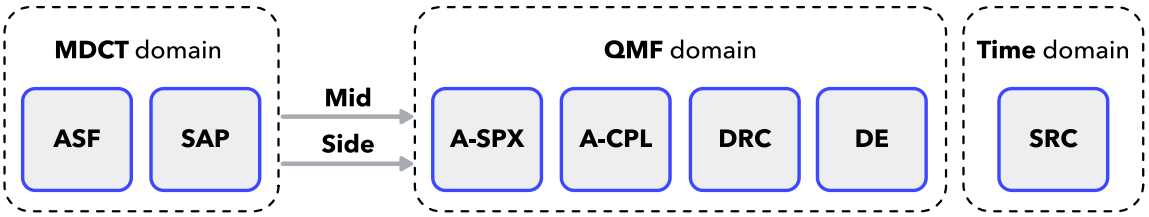

FIGURE 5: SAP output (mid/side)

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 on the next page 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.

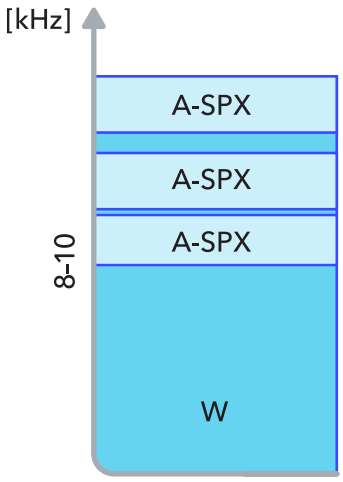

FIGURE 6: A-SPX frequency-interleaved coding

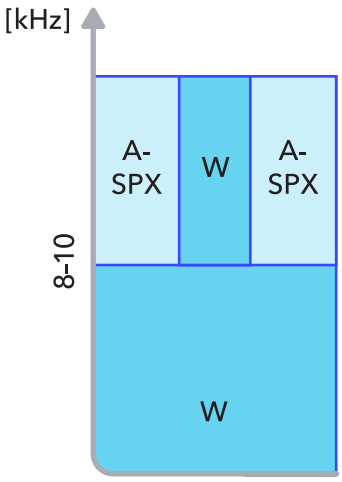

FIGURE 7: A-SPX time-interleaved coding

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 efficiently code a set of objects and beds. 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.

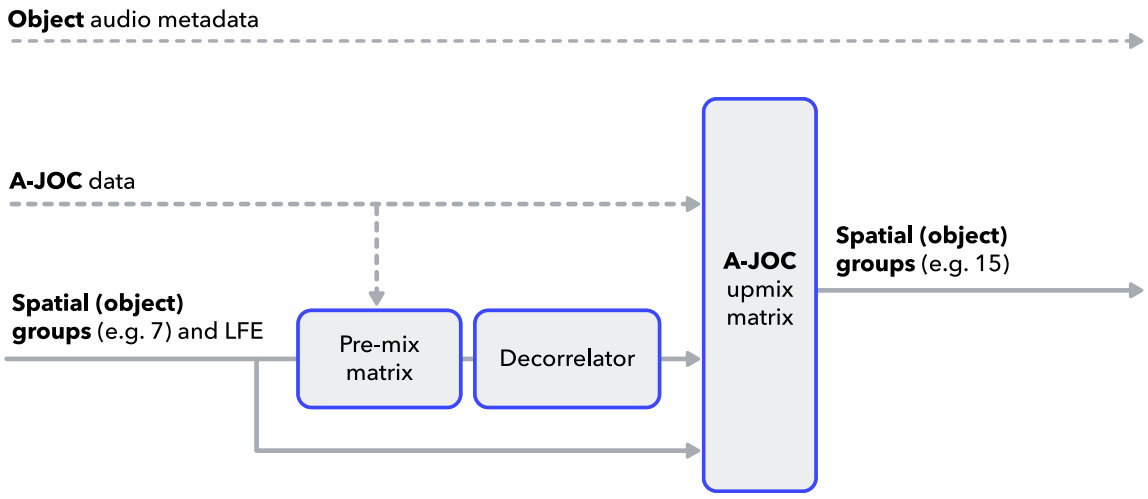

FIGURE 8: Basic principle of A-JOC decoder tool

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.

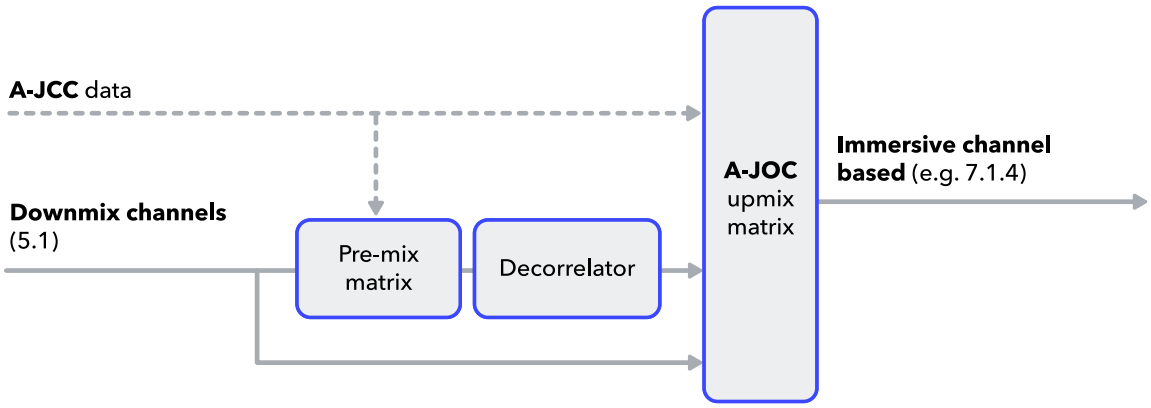

FIGURE 9: Basic principle of A-JCC decoder tool

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	32 kbps	48 kbps
Stereo	48 kbps	64 kbps
5.1	96 kbps	144 kbps
Immersive audio 7.1.4 playback	192 kbps	320 kbps

TABLE 1: Overview of audio quality that is achieved given a certain content type and bit rate

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 Table 2 below.

Format	High quality emission acc. BS.1548 consistently above 4 on BS.1116
Stereo	96 kbps
5.1	192 kbps
Immersive audio 7.1.4 playback	288 kbps

TABLE 2: AC-4 average score

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.

		MDCT Domain		QMF domain			Time domain	
Codec mode		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	•	•			•	•	•

TABLE 3: Tools applied for coding channel-based audio across a broad range of channel configurations and data rates

4 Overview of bitstream syntax

An AC-4 elementary stream 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.

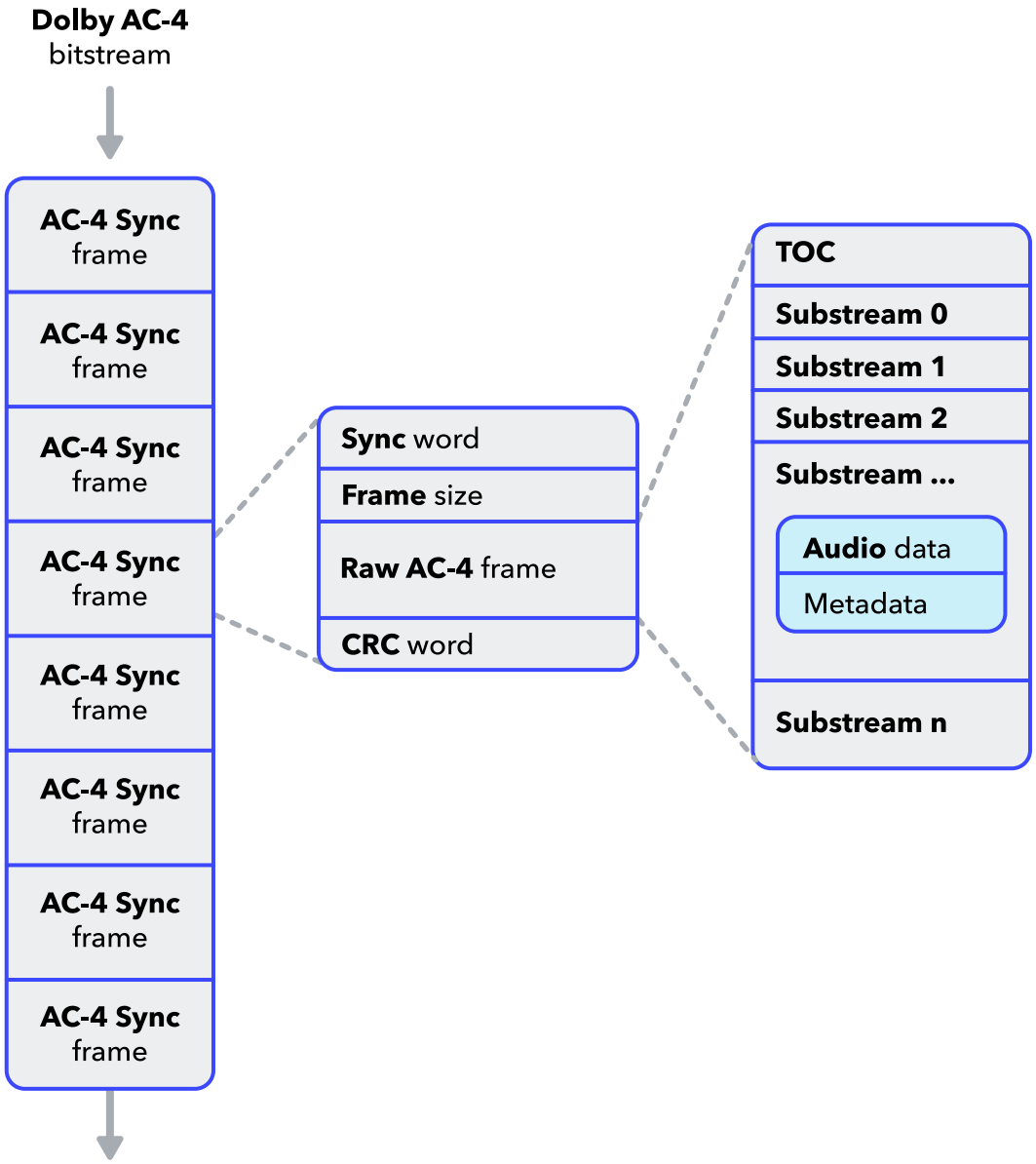

FIGURE 10: High-level bitstream syntax (AC-4 part 2)

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.

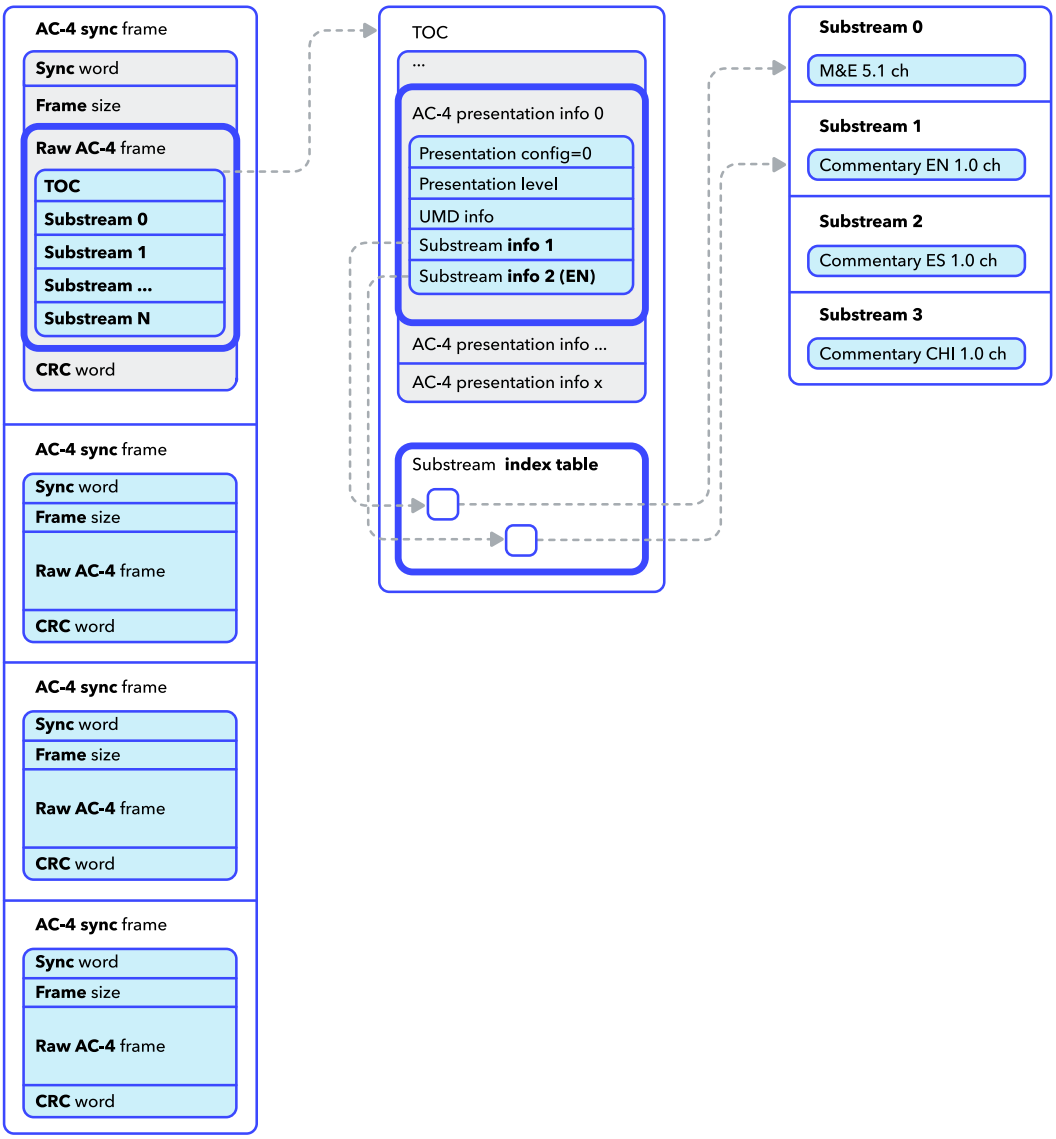

FIGURE 11: Live 5.1 sports broadcast with four presentations

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 of 50% or more compared with using the same format to deliver multiple 5.1 streams.

Conventional approach		Presentation approach	
Stream	**Data rates/kbps**	**Stream**	**Data rates/kbps**
English 5.1	144	English Commentary	48
Spanish 5.1	144	Spanish Commentary	48
Chinese 5.1	144	Chinese Commentary	48
Commentary-free 5.1	144	Commentary-free 5.1	144
TOTAL	576	TOTAL	288
		Savings	**50%**

TABLE 4: A data-rate savings in excess of 50% can be offered by a presentation-based approach

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.

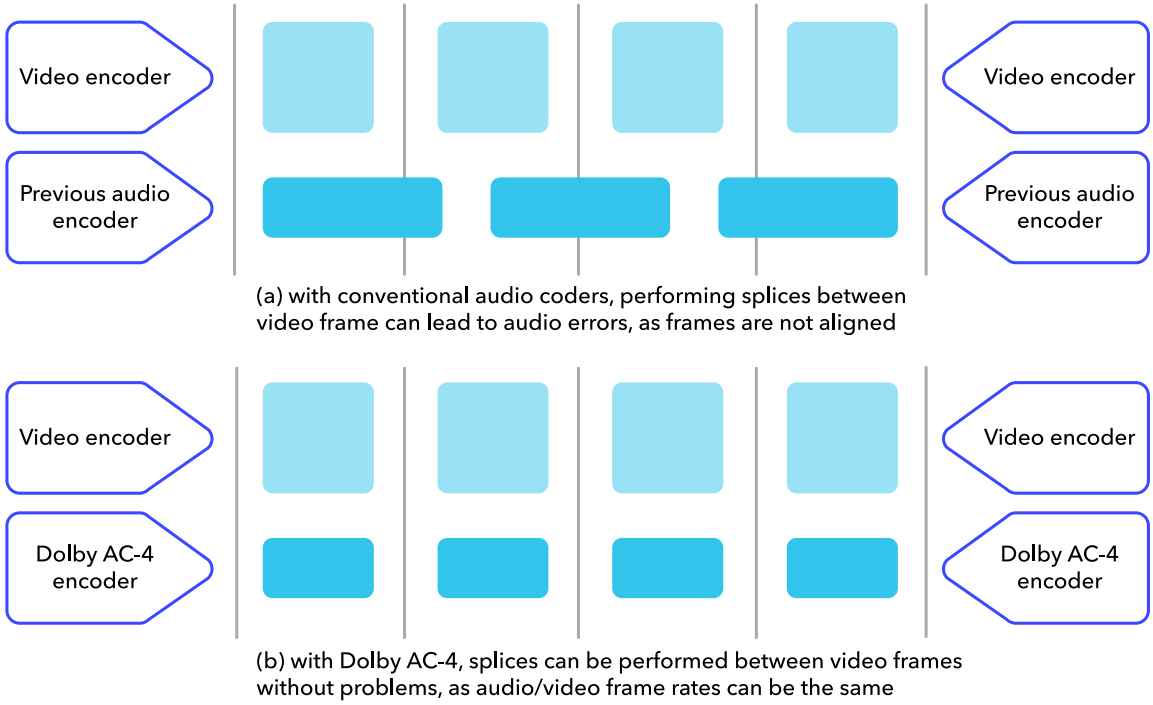

FIGURE 12: Audio/video frame alignment in AC-4

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.

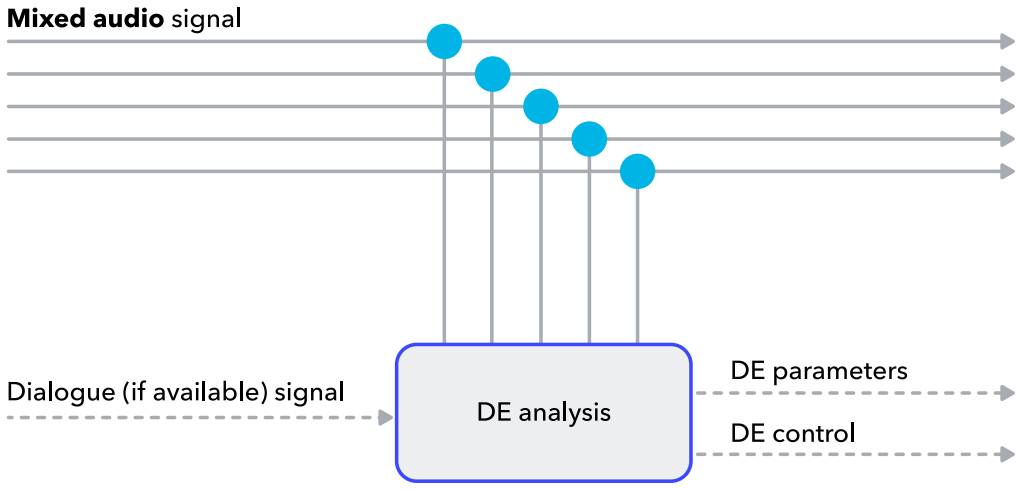

FIGURE 13: Dialogue enhancement for channel-based content (encoder functionality)

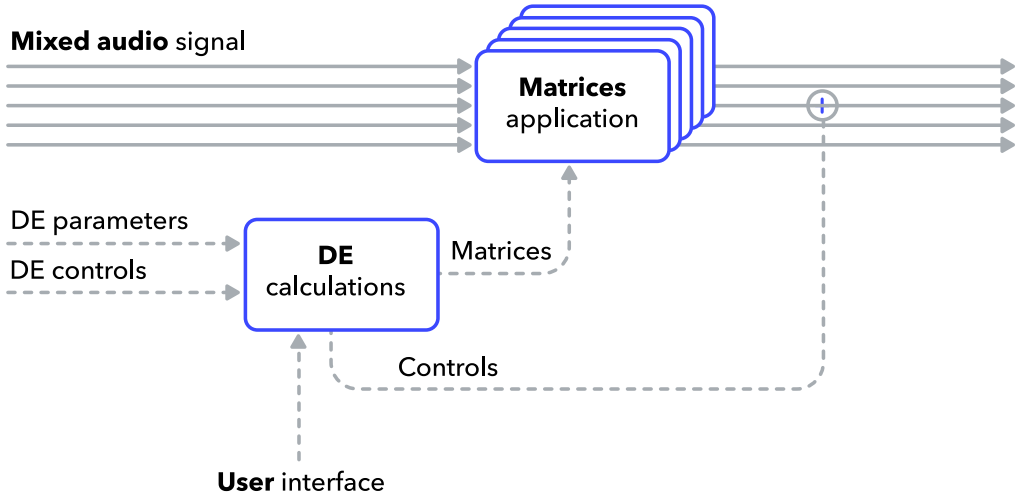

FIGURE 14: Dialogue enhancement for channel-based content (decoder functionality)

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 AC-4 decoder levels

In order to manage decoder complexity, AC-4 provides a concept of levels. Levels describe the complexity of the input to the decoder per Presentation in a stream. Levels do not limit how the content is rendered, and hence presented to the consumer. The details of Levels are defined in ETSI 103 190-1 and –2. Levels are signaled in the mdcompat field for each presentation info in the AC-4 TOC. The Decoder implemented in all consumer devices to date conforms to Level-3. That decoder is able to decode channel-based presentations based on a Main Audio component up to 7.1.4 or presentations based on a Music and Effects audio component up to 5.1 and a Dialogue audio component up to 3.0. Both can be combined with an Associated Audio component up to 2.0. A Level-3 Decoder is also able to decode object audio or mixed Presentations consisting of up to 17.1 audio objects, based on 10 independent, or core audio objects. A stream can combine presentations with different Levels. A Level-3 Decoder can decode presentations indicating Level-3 or lower and is able to ignore presentations with higher levels. Overall a level-3 decoder supports streams with up to 64 presentations.

6.2 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.

The AC-4 Decoder selects either core decoding or full decoding depending on the output channel configuration and the complexity of the selected presentation for decoding.

6.2.1 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.

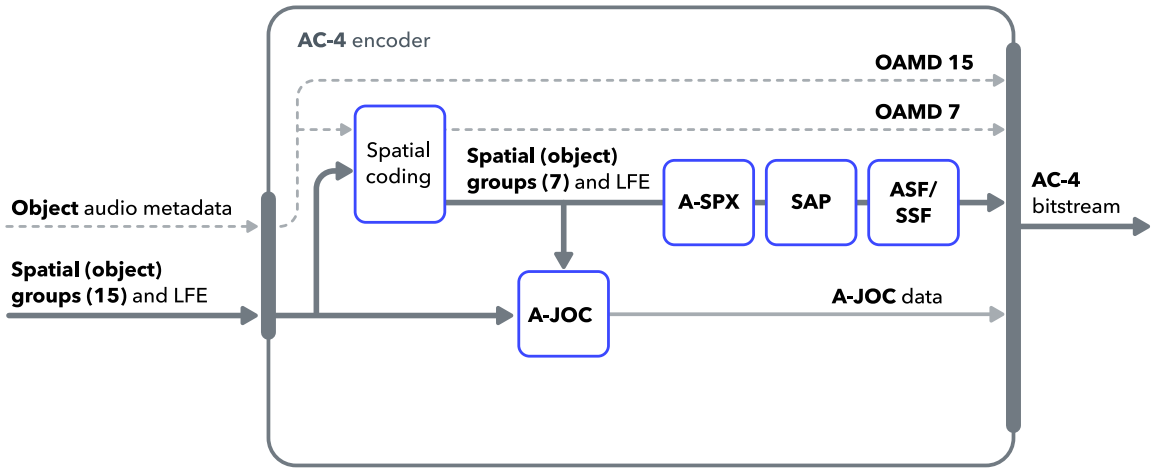

FIGURE 15: Encoding object-based audio (adaptive downmix)

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.

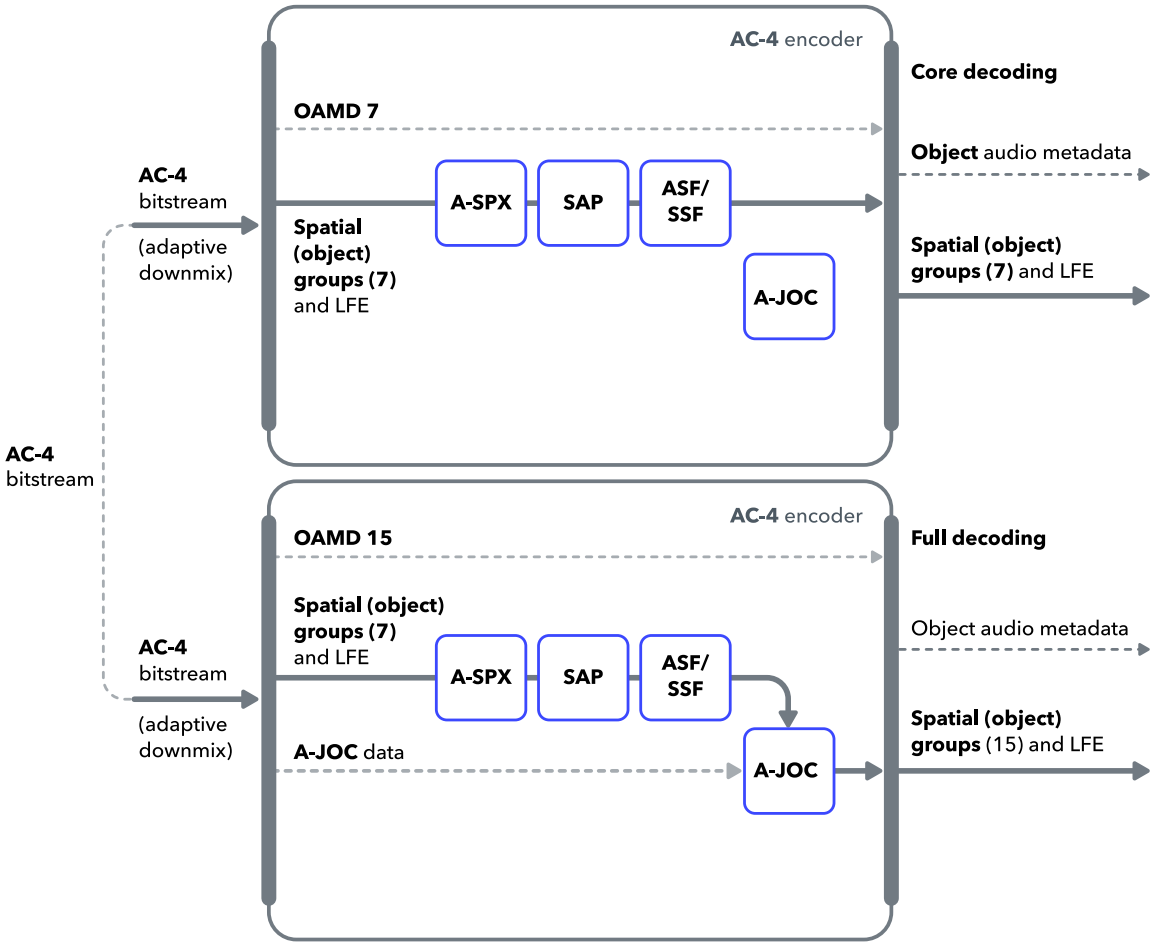

FIGURE 16: A-JOC core and full decoding

6.2.2 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 S-CPL is used, "side signals" are also created; these are conceptually similar to side signals in traditional mid-side coding. A-CPL can optionally work with bandlimited side signals.

- 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.
- Simple Coupling (S-CPL) provides the highest audio quality by coding the side signals up to the full audio bandwidth.

The figure below illustrates the channel-based immersive encoder.

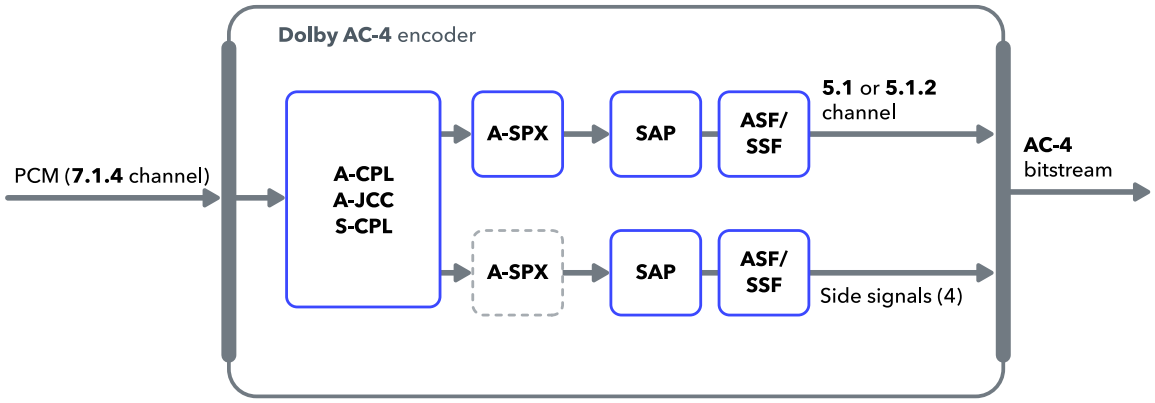

FIGURE 17: Dolby AC-4 encoder (7.1.4 input example)

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). A-JCC coded signals are further processed by the A-JCC core decoding module to create a 5.1.2 intermediate representation.

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 S-CPL, 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.

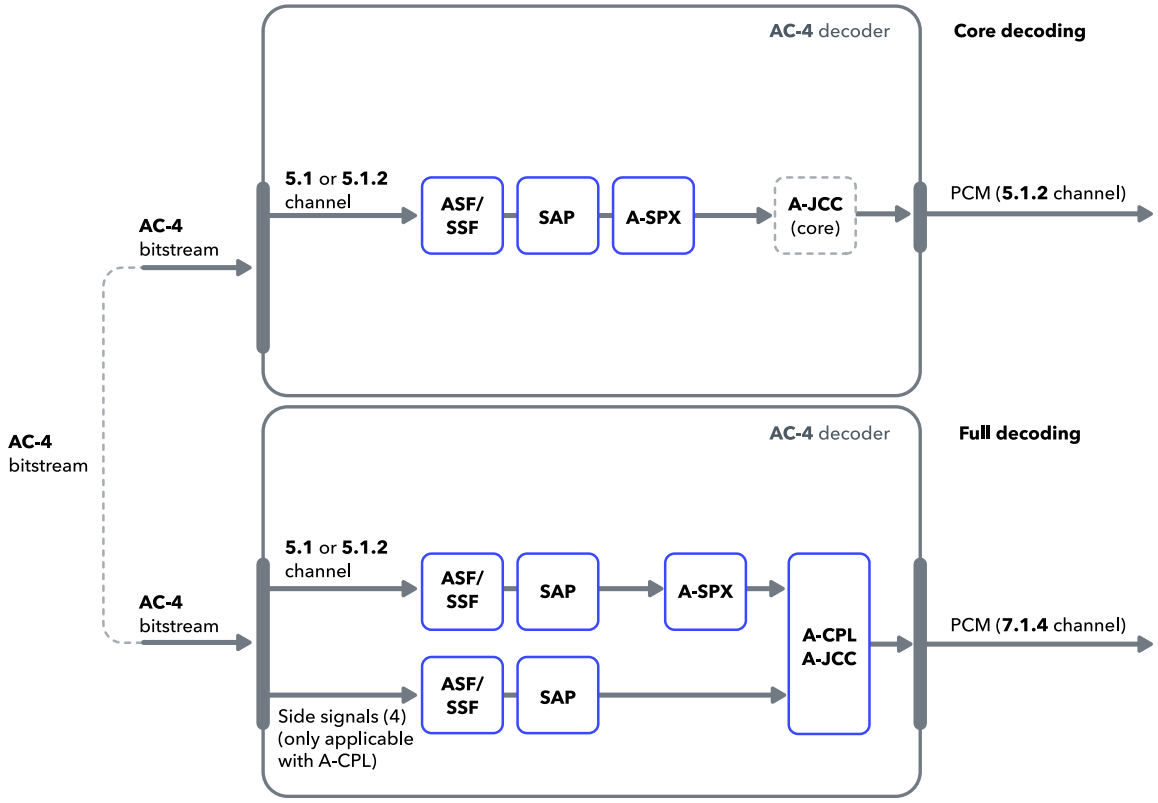

FIGURE 18: PCM output decoding

6.3 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 can render to higher numbers of speakers with a high degree of flexibility.

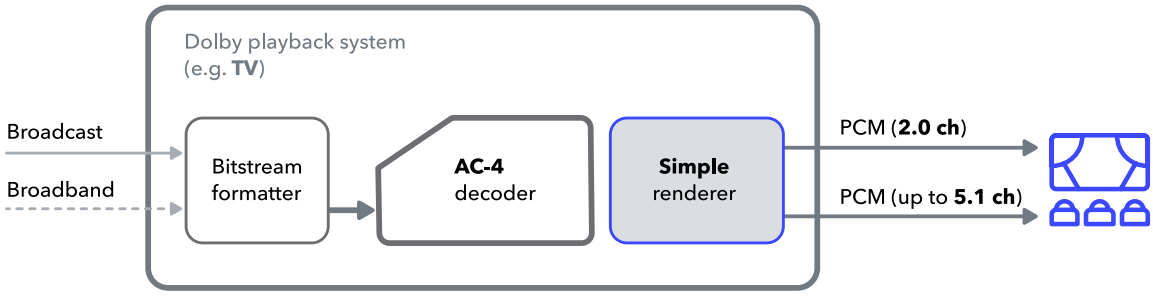

FIGURE 19: TV playback system

In other playback conditions the renderer can be combined with a virtualizer to efficiently provide immersive audio experiences over headphones or loudspeakers.

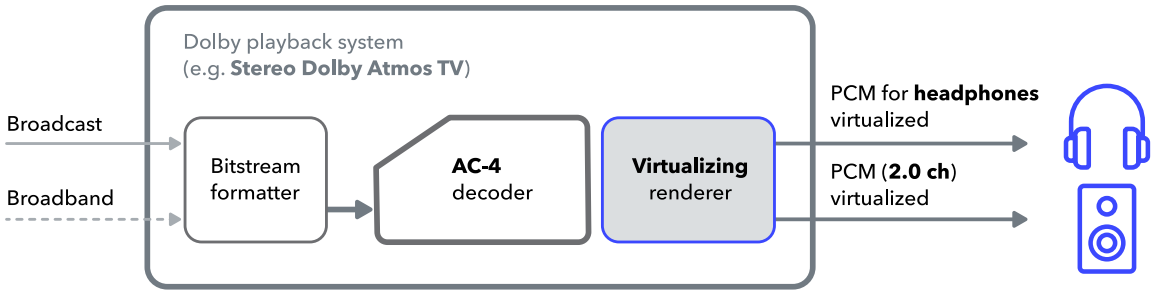

FIGURE 20: Stereo living room device playback system

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 and eARC. 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 using off-the-shelf Set-top Boxes or TVs as source devices, and Dolby Atmos capable Soundbars and AVRs as sink devices.

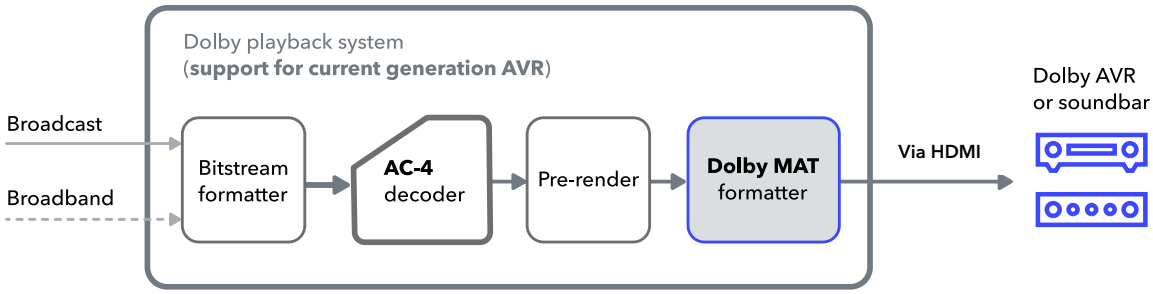

FIGURE 21: Connectivity to current Dolby Atmos AVRs

6.4 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	-31 to -27
Flat panel TV	-26 to -17
Portable device speakers	-16 to 0
Portable device headphones	-16 to 0
Four user-definable modes	

TABLE 5: Output level ranges for DRC decoder modes

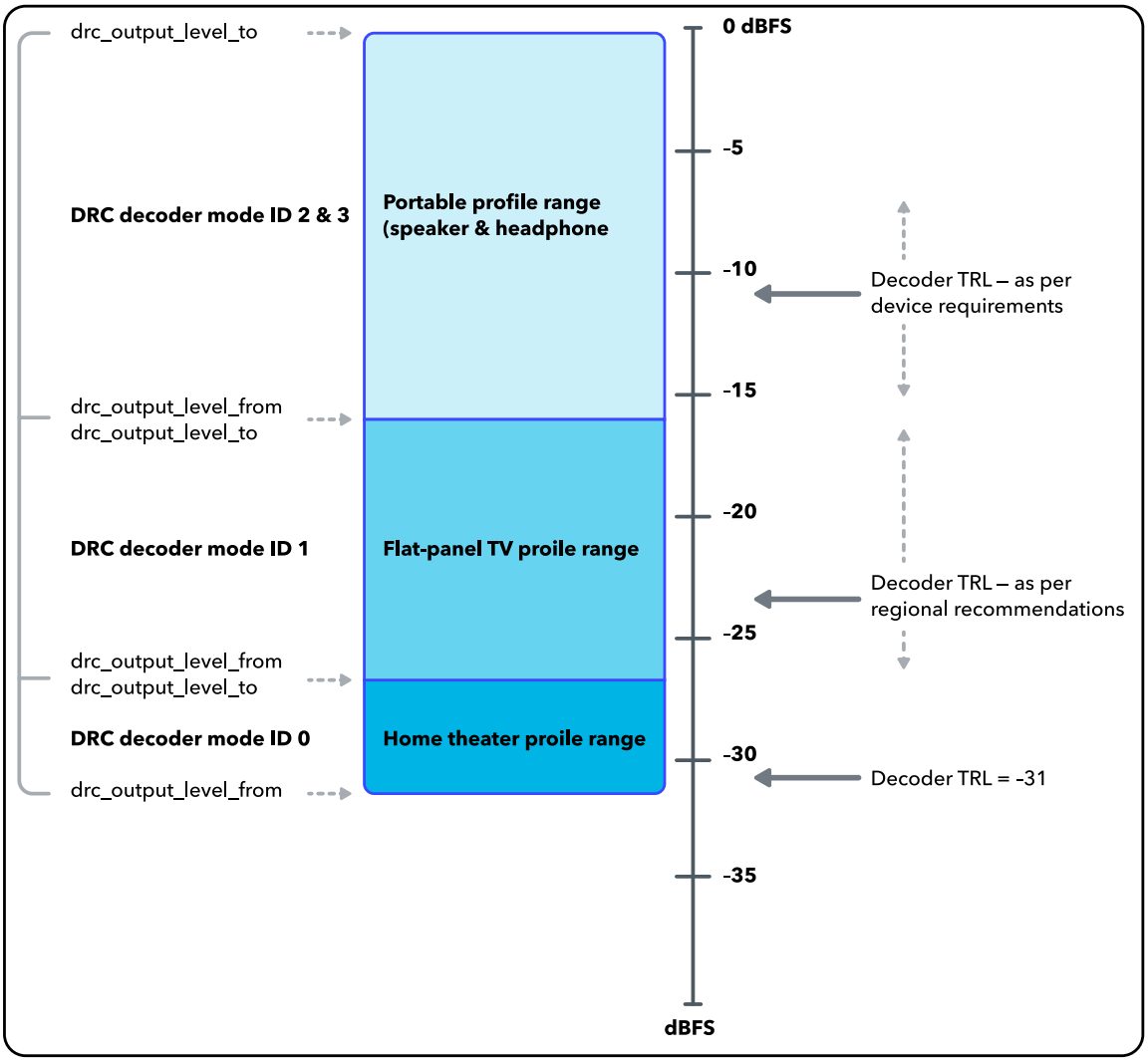

FIGURE 22: Playback device target reference level and DRC profile mapping

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

decoder dynamic range control modes as shown in Figure 22. For example, if –23 dBFS is selected for the target reference level (TRL), the highest AC-4 DRC Decoder Mode ID (representing the parameterized compression curve for that playback mode) present in the selected presentation that overlaps 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) and transported in the AC-4 bitstream as a parameterized compression curve.

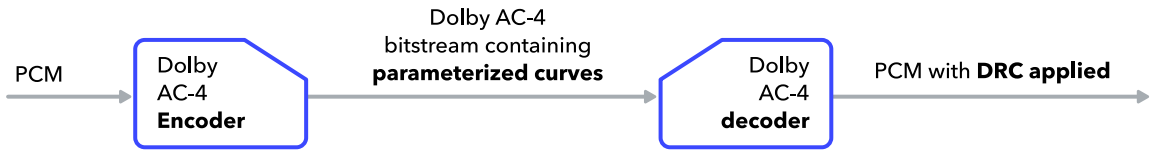

FIGURE 23: Parameterized curves

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 over traditional DRC gains 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.

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.

6.5 Seamless switching

The Dolby AC-4 decoder supports seamless switching between data rates for sensible 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 on I-frame boundaries. 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.

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 Immersive Stereo (IMS) for mobile

At Dolby we have worked for many years in close collaboration with mixers to define the best Atmos experience for playback on loudspeakers and headphones. IMS is a codec tool that was developed specifically to efficiently transmit the Atmos experience to mobile devices and ensures that the device renders the Atmos experience correctly. Using IMS, immersive audio (both object-based and channel-based immersive) is coded as two channels and associated control data. At the playback side a low-complexity decoding process based on the two channels and the additional control data is applied to create the Atmos experience for playback on headphones and stereo speakers integrated in mobile devices, such as mobile phones and tablets, or the two channels are decoded to LoRo for non-virtualized stereo. As such, single IMS stream can be decoded to three different experiences in the AC-4 decoder:

- Headphone virtualized
- Speaker virtualized (on integrated speakers of mobile devices)
- LoRo (non virtualized stereo)

For mobile delivery the transmission rate and the playback complexity of IMS is significantly lower compared to a generic transmission of channel-based immersive or object-based immersive (with AC-4 or Dolby Digital Plus). The recommended data rates for IMS for a given quality according to the MUSHRA scale are shown in the table below. Near transparent quality is reached at a bitrate of 256 kbps.

	Good quality (MUSHRA)	Excellent quality (MUSHRA)
IMS	64 kbps	112 kbps

TABLE 6: Data rates for IMS

Comparing computational playback complexity, playback of IMS is around 3-4 times lower than playback of channel-based immersive or object-based immersive (with AC-4 or Dolby Digital Plus) including playback side virtualization.

IMS can be used in combination with all AC-4 system features such as dual-ended dialog enhancement, and is supported for accessibility and audio personalization use-cases.

IMS can also be used for 5.1 content. Similarly as for Atmos content, bitrate and playback complexity for 5.1 transmission and playback is lower compared to channel-based immersive or object-based immersive transmission and playback (with AC-4 or Dolby Digital Plus).

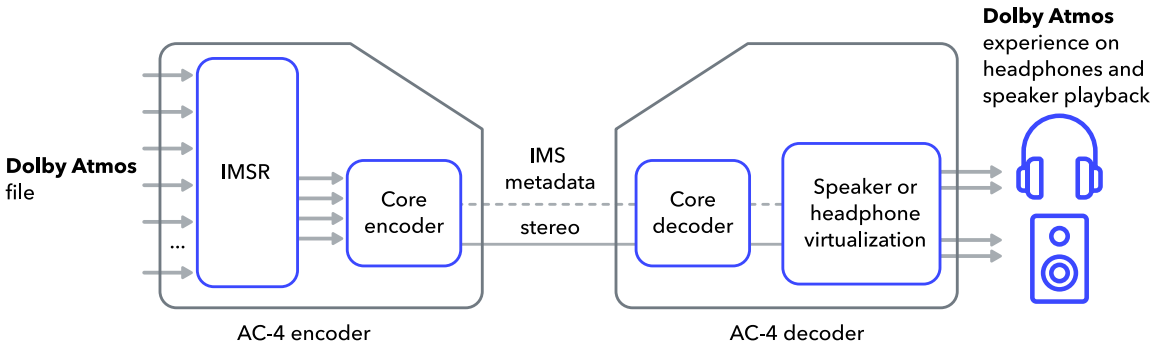

FIGURE 24: The basic principle of IMS

The basic principle of IMS is illustrated in the figure above. In the AC-4 Encoder the immersive input signal, which may consist of a full Atmos printmaster or a channel-based immersive representation such as 7.1.4, is first rendered and analysed in the IMS Renderer (IMSR) into signals that are used in the second stage of IMS encoding to generate the IMS control data that are transmitted in the bitstream alongside the two-channel signal. On the decoder side, the core decoder decodes the signals and transforms them into headphone or speaker virtualized signals based on the control data to create an Atmos experience, or decodes the two channel signal to LoRo for non-virtualized stereo.

9 Pro partner additions

Support for AC-4 in the content distribution ecosystem is provided by products from Professional Partners Dolby works closely with. Following are different categories of products available in support of deploying AC-4.

Manufacturer	Product(s)	Supports
ATEME	Titan Live	2.0, 5.1, Atmos
DS broadcast	BGE9000	2.0, 5.1
Harmonic	Electra X2	2.0, 5.1
Harmonic	Electra XOS	2.0, 5.1
Kai Media	KME-U4K	2.0, 5.1
Synamedia	vDCM	2.0, 5.1

TABLE 7: Video transcoders/encoders

In many cases, the packager is integrated with the video encoder. Those are listed here along with standalone packagers.

Manufacturer	Product(s)	Supports
Ateme	Titan Live	ATSC-DASH
Bento4	Bento4	DASH
DS Broadcast	BGE9000	ATSC-DASH
Harmonic	Electra X2	ATSC-DASH
Harmonic	Electra XOS	ATSC-DASH
Kai Media	KME-U4K	ATSC-DASH
Shaka	Shaka	DASH
Synamedia	vDCM	ATSC-DASH

TABLE 8: Packagers

Manufacturer	Product(s)	Supports
DekTec	StreamXpert	2.0, 5.1, Atmos, DASH
DS Broadcast	BGD-4100	2.0, 5.1
Kai media	KMR-U4K	2.0, 5.1

TABLE 9: Monitoring decoders

Some video encoders can support the pass-through of an encoded AC-4 stream. A standalone audio encoder allows the addition of AC-4 audio without requiring a replacement of the video encoder.

Manufacturer	Product(s)	Supports
Telos Alliance	LA-5300	2.0, 5.1, Atmos

TABLE 10: Standalone audio encoder

10 Standardization and deployment

In readiness for next-generation service deployments, the core of the Dolby AC-4 technology was standardized by the ETSI as TS 103 190 Part 1 in April 2014, adding object-based audio (OBA) via TS 103 190 Part 2 in September of 2015. The most recent versions of these documents 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.

On February 23, 2017 the ATSC A/342 part 2 "AC-4 System" standard was approved. On October 19, 2017 A/300 the "ATSC 3.0 System" standard was approved, specifying that all terrestrial and hybrid broadcast services in a given region shall use one audio system for that region and that broadcasters in North America selected the A/342 Part 2 (AC-4) for Mexico, Canada, and the U.S. The most recent versions of these standards may be freely downloaded via the ATSC website.

In 2017, the Digital Video Subcommittee of the Society of Cable Telecommunications Engineers (SCTE) / International Society of Broadband Experts (ISBE) adopted ANSI/SCTE 242-2 (2017) Next Generation Audio Coding Constraints for Cable Systems: Part 2– AC-4 Audio Coding Constraints, and ANSI/SCTE 243-2 2017 Next Generation Audio Carriage Constraints for Cable Systems: Part 2 – AC-4 Audio Carriage Constraints, and these are referenced within SCTE 54 Digital Video Service Multiplex and Transport System Standard for Cable Television.

In December of 2018, CTA WAVE published their Content Specification CTA-5001-A (later updated to CTA-5001-B), to include AC-4 as a media profile. The WAVE project, from the Consumer Technology Association (CTA), focuses on commercial internet video and web applications, and developing interoperability tools for global compatibility, and publishes content specifications, test content, and HTML interfaces for media based on MPEG CMAF format.

AC-4 has been included as the sole next generation audio codec in several European regional broadcast specifications, including NorDig v3.1.1 (unified specification for

DVB-T2 covering Denmark, Norway, Finland, Iceland, Sweden and Ireland), Italy UHD-Book v1.0, and Poland DVB-T2.

In September 2019, AC-4 was added in the NorDig Unified Test Plan v3.1.1 and updated NorDig Unified IRD Specification v3.1.1 was released.

On May 2020, Dolby AC-4 was included in the Brazilian Digital Terrestrial Television Standard ABNT NBR 15602-2:2020.

AC-4 broadcasts are on the air for ATSC 3.0 stations in the US and broadcast trials have been completed in several other countries, including Spain, France, Italy and Poland. Dolby AC-4 has been used for high profile live event broadcasts including the French Open Tennis and the European Athletics Championships.

European broadcasters are actively exploring how HbbTV streaming enables a fast track to launch of Dolby AC-4 enhanced services including immersive audio, extended accessibility provision and audio personalisation. AC-4 is supported in the latest version of the HbbTV specification (ETSI TS 102 796 v1.5.1) and official test suite, and has regularly featured in HbbTV Association plugfest events to validate receiver interoperability.

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.

11 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, services are on air, multiple new deployments are underway and planning for new feature rollouts can begin.