Dolby AC-4 and MPEG-DASH

Specification

1 March 2019
Confidential information

Copyright

Dolby Laboratories, Inc.
1275 Market Street
San Francisco, CA 94103-1410 USA
Telephone 415-558-0200
Fax 415-645-4000
http://www.dolby.com

Trademarks

Dolby and the double-D symbol are registered trademarks of Dolby Laboratories.

The following are trademarks of Dolby Laboratories:

Dialogue Intelligence™
Dolby®
Dolby Advanced Audio™
Dolby Atmos®
Dolby Audio™
Dolby Cinema™
Dolby Digital Plus™
Dolby Digital Plus Advanced Audio™
Dolby Digital Plus Home Theater™
Dolby Home Theater®
Dolby Theatre®
Dolby Vision™
Dolby Voice®
Feel Every Dimension™
Feel Every Dimension in Dolby™
Feel Every Dimension in Dolby Atmos™
MLP Lossless™
Pro Logic®
Surround EX™

All other trademarks remain the property of their respective owners.

Confidential information

Contents

1 Introduction to the Dolby AC-4 and MPEG-DASH specification

This documentation specifies the required data formatting and signaling between the server (sender) and the clients (receivers) to enable Dolby AC-4 to be used as an audio format within the MPEG Dynamic Adaptive Streaming over HTTP (MPEG-DASH) standard in conjunction with the ISO base media file format.

1.1 About this documentation

This documentation is written for a system designer to develop a product that can properly multiplex Dolby AC-4 bitstreams into an MPEG-DASH compliant ISO base media file format.

This documentation provides information regarding:

- Signaling of Dolby AC-4 audio streams within the MPEG-DASH Media Presentation Description (MPD) file
- Storage of Dolby AC-4 bitstreams within ISO base media file format files conforming to MPEG-DASH

In addition, the following information is covered:

- The data required to identify a Dolby AC-4 bitstream within an MPEG-DASH compliant ISO base media file format file and an MPD manifest file
- The steps required to properly packetize a Dolby AC-4 bitstream for multiplexing and storage in an MPEG-DASH compliant ISO base media file format file
- The steps required to demultiplex a Dolby AC-4 bitstream from an MPEG-DASH compliant ISO base media file format file

1.2 Conventions used

Modal verbs are used in this documentation to differentiate between mandatory requirements and recommendations.

- Must: indicates that the corresponding requirement or instruction is mandatory.
- Should: indicates that the corresponding statement or instruction is a recommendation and is not mandatory.

1.3 Resources

Standards and Dolby documents provide additional information to assist you in designing your product.

Standards

- ETSI TS 103 190-1 v1.3.1, *Digital Audio Compression (AC-4) Standard, Part 1; Channel-Based Coding*, available from http://www.etsi.org. This documentation describes the channel-based core codec of the AC-4 bitstream syntax.
- ETSI TS 103 190-2 v1.2.1, *Digital Audio Compression (AC-4) Standard, Part 2; Immersive and Personalized Audio*, available from http://www.etsi.org. This documentation extends the AC-4 codec with a number of new use cases relevant for next-generation audio services.
- ISO/IEC 14496-12:2015, *Information Technology—Coding of Audio-Visual Objects, Part 12: ISO Base Media File Format*, available from http://www.iso.org. This documentation is Part 12 of the MPEG-4 specification and describes storage of content in a media file.
- ISO/IEC 23009-1:2014, *Information Technology—Dynamic Adaptive Streaming over HTTP (DASH)—Part 1: Media Presentation Description and Segment Formats*, available from http://www.iso.org.

Other documents

Guidelines forImplementation:DASH-IFInteroperability Points, available from https://dashif.org/guidelines.

1.4 Contacting Dolby

Support services are available to address any questions and to provide advice about integrating Dolby technology into your product.

For product design or testing, contact Dolby at systemsupport@dolby.com. By utilizing Dolby expertise, especially during the design process, many problems that might require design revisions before a product is approved can be prevented.

Dolby is also available to review product plans, including preliminary design information, markings, displays, and control and menu layouts, with the goal of preventing problems early in the product development cycle.

If you have comments or feedback about this documentation, send us an email at documentation@dolby.com.

2 Dolby AC-4 packetization into MPEG-DASH compliant ISO base media filesa fragmented MP4 file

The process of analyzing and preparing a Dolby AC-4 bitstream for storage within an MPEG-DASH compliant ISO base media file a fragmented MP4 file is described. AC-4 sync frames must be converted to raw AC-4 frames before packetization.

2.1 Constraints on Dolby AC-4 elementary streams within MPEG-DASH ISO base media file format segments

A Dolby AC-4 elementary stream must meet certain conditions to be delivered via MPEG-DASH.

Annex E of ETSI TS 103 190-2 describes the constraints that must be applied to the Dolby AC-4 bitstream for storage within the ISO base media file format. Additionally, for MPEG-DASH applications that use the ISO base media file format, these constraints apply:

- The value of the `frame_rate_index` parameter must remain constant.
- The value of the `presentation_config` parameter must remain constant.
- The value of the `fs_index` parameter must remain constant.
- The value of the `channel_mode` parameter must remain constant.
- The value of the `content_classifier` parameter must remain constant.
- The Dolby AC-4 random access points must be encoded correctly to ensure compliance with the requirements described in Parsing and packetizing the Dolby AC-4 bitstream on page 8.
- We strongly recommend that Dolby AC-4 elementary streams be encoded with the same frame rate as the associated video frame rate to ensure constant alignment of video and audio access units in order to utilize the features of A/V alignment.

If a Dolby AC-4 elementary stream does not meet the requirements stated in this specification, the multiplexer may halt processing and inform the user that the bitstream is not compliant.

Related information

Parsing and packetizing the Dolby AC-4 bitstream on page 8
A/V alignment and segmentation on page 16

Random access point on page 33

2.2 Content of the MP4 sample

A Dolby AC-4 MP4 sample corresponds to one raw Dolby AC-4 frame.

For detailed information on the Dolby AC-4 MP4 sample content, refer to Annex E of ETSI TS 103 190-2.

2.3 Reading Dolby AC-4 frames from an .ac4 file

In an .ac4 file, each `raw_ac4_frame` is encapsulated in an `ac4_syncframe`, as defined in ETSI TS 103 190-1. An AC-4 sync frame must first be converted to a raw AC-4 frame by stripping its sync word, frame size, and optional cyclic redundancy check (CRC) words before packetization.

Procedure

1. To locate the start of an AC-4 raw frame, search for the 16-bit sync word. The value of the sync word can be either `0xAC40` or `0xAC41`.
2. Read the 16 bits immediately succeeding the sync word.
   - If the 16-bit value is not equal to `0xFFFF`, the value is the frame size.
   - If the 16-bit value is equal to `0xFFFF`, read the next 24 bits as the frame size.
3. According to the frame size, read the next bits succeeding the frame size as the raw AC-4 frame.
4. Optional: If the sync word is `0xAC41`, the AC-4 simple transport frame includes a 16-bit CRC word based on the `frame_size` element and the `raw_ac4_frame` element. Read the 16 bits immediately succeeding the raw AC-4 frame, and verify that the value matches the CRC calculated using the IBM-CRC-16 polynomial: $x^{16}+x^{15}+x^2+1$.

2.4 Parsing and packetizing the Dolby AC-4 bitstream

A multiplexer packetizes only one raw AC-4 frame as an MP4 sample.

About this task

A seamless switch can be accomplished only at a random access point. The first sample of each MPEG-DASH segment and subsegment must contain a random access point. A Dolby AC-4 elementary stream contains independently decodable frame (I-frame)s to indicate random access points. Therefore, the multiplexer must ensure that the first MP4 sample of each Dolby AC-4 media segment is an I-frame. A Dolby AC-4 media segment may contain one or more random access points.

To ensure that each Dolby AC-4 media segment begins with an I-frame, the multiplexer must follow this procedure.

Procedure

1. The multiplexer scans the bitstream to find the first frame with a `b_iframe_global` value of 1. This frame is an I-frame that begins a Dolby AC-4 media segment.
2. As the multiplexer checks subsequent frames for more I-frames, frames that have a `b_iframe_global` value of 1 keep track of the time span.
3. When the multiplexer reaches an I-frame, and the accumulated duration from the first I-frame to the frame immediately preceding the latest found I-frame matches the user's setting, the current Dolby AC-4 media segment is completed.
4. The latest found I-frame becomes the first frame of the next Dolby AC-4 media segment.

Related information

Random access point on page 33

2.5 Signaling Dolby AC-4 bitstreams in an MPEG-DASH–compliant ISO base media file format

Dolby AC-4 uses the basic structures defined within ISO/IEC 14496-12 to signal audio tracks and uses specific extensions within compliance to ISO base media file format to provide detailed information on the characteristics of a Dolby AC-4 bitstream.

The information is included in the `stbl` box of the `moov` box of the MPEG Dynamic Adaptive Streaming over HTTP (MPEG-DASH)–compliant ISO base media file. The locations and hierarchy of the boxes that must be included in `stbl` to identify a Dolby AC-4 stream within an MPEG-DASH–compliant ISO base media file are listed in the table.

Table 1: Sample table box hierarchy for Dolby AC-4 audio tracks

Nesting level				
4	5	6	7	Reference
`stbl`				ISO/IEC 14496-12
	`stsd`			
		`AC4SampleEntry` (header type set to `ac-4` or `enca`)		ETSI TS 103 190-2
			`AC4SpecificBox` (header type set to `dac4`)	ETSI TS 103 190-2
	`stts`			ISO/IEC 14496-12
	`stsc`			
	`stsz`			
	`stz2`			
	`stco`			
	`co64`			

In this table, the value of the nesting level provided for each box is based on the structure of the complete media file, beginning with a nesting value of `0` for the `ftyp` and `moov` boxes. The table also includes a reference to where each box is defined.

The value for the `AC4SampleEntry` box header type depends on whether the file is encrypted. For an unencrypted file, the `AC4SampleEntry` box header type value is `ac-4`.

The value of the `AC4SpecificBox` header type is `dac4`.

The stbl box of a Dolby AC-4 audio track must contain a sync sample box (stss), unless all samples are sync samples (I-frames). The stss box must reference all the sync samples. Random access points are signaled by the sync sample box (stss) in an ISO base media file, or by setting the tr_flags in the track run box in ISO base media segments.

Unless otherwise stated here, all requirements specified in ISO/IEC 14496-12 and Annex E ofETSI TS 103 190-2 must be followed.

2.5.1 Signaling immersive stereo content

Immersive stereo presentations use presentation version 2. To signal immersive stereo content, the ac4_dsi_v1 with changed syntax is used in the AC4SampleEntry.

Dolby AC-4 bitstreams signal immersive stereo content by setting the following fields:

- presentation_version is set to 2.
- b_presentation_group_index is set to 1.
- presentation_group_index is any nonnegative number.
- channel_mode is set to 0b1111000 or 0b11110010, with ch_mode set to 1.
- b_pre_virtualized is set to 0.

To signal immersive stereo content in an MPEG-DASH–compliant ISO base media file, the ac4_dsi_v1 presentations use presentation version 2 and ac4_presentation_v1_dsi elements. To support presentation version 2, syntax changes (as highlighted in the following table) are made to the ac4_dsi_v1.

Syntax	Word size (in bits)
ac4_dsi_v1()	
{	
ac4_dsi_version;..................	3
bitstream_version;................	7
fs_index;........................	1
frame_rate_index;.................	4
n_presentations;..................	9
if (bitstream_version >1)	
{	
b_program_id;.....................	1
if (b_program_id)	
{	
short_program_id;.................	16
b_uuid;...........................	1
ifb_uuid	
program_uuid;.....................	16*8
}	
}	
}	
ac4_bitrate_dsi();	

Syntax	Word size (in bits)
byte_align;.....	0-7
for (i = 0; i < n_presentations; i++)	
{	
presentation_version;.....	8
pres_bytes;.....	8
if (pres_bytes == 255)	
{	
add_pres_bytes;.....	16
pres_bytes += add_pres_bytes;	
}	
if (presentation_version == 0)	
presentation_bytes = ac4_presentation_v0_dsi();	
}	
else	
{	
if (presentation_version == 1 **|| presentation_version == 2)**	
{	
presentation_bytes = ac4_presentation_v1_dsi(pres_bytes);	
}	
else	
{	
presentation_bytes = 0;	
}	
}	
skip_bytes = pres_bytes - presentation_bytes;	
skip_area;.....	skip_bytes*8
}	
}	

Note: The number of bits in byte_align pads the number of bits, counted from the start of ac4_dsi_v1 to a multiple of eight.

For each immersive stereo presentation, the ac4_dsi_v1 must include two ac4_presentation_v1_dsi instances.

1. One instance uses presentation version 2 and signals the immersive stereo content. The corresponding field settings are as follows:
   - presentation_version is set to 2.
   - b_presentation_group_index is set to 1.
   - presentation_group_index is set to the same number as in the corresponding ac4_presentation_info of the Dolby AC-4 bitstream.
   - dsi_presentation_ch_mode is set to 1.
   - b_pre_virtualized is set to 1.

2. A second instance uses presentation version 1 and signals a regular stereo compatible presentation of the immersive stereo content. This ac4_presentation_v1_dsi instance must come after the instance that signals the immersive stereo content. The corresponding field settings are as follows:
   - presentation_version is set to 1.
   - b_presentation_group_index is set to 1.
   - presentation_group_index is set to the same number as in the corresponding ac4_presentation_info of the Dolby AC-4 bitstream.
   - dsi_presentation_ch_mode is set to 1.
   - b_pre_virtualized is set to 0.

Note: Signaling in the MPD file is based on the first instance only.

Signaling immersive stereo with Dolby Atmos

If the ac4_toc of a Dolby AC-4 bitstream contains an immersive stereo presentation (presentation_version = 2) and that presentation contains an AC-4 substream with channel_mode set to 0b1111001, then this immersive stereo presentation is created from Dolby Atmos content and can be signaled as Dolby Atmos.

To signal Dolby Atmos, the ac4_presentation_v1_dsi defined in ETSI TS 103 190-2 is extended, as highlighted in the following table. Dolby Atmos is signaled by setting the dolby_atmos_indicator to 1.

The extension to ac4_presentation_v1_dsi includes additional fields for signalling Dolby AC-4 capabilities.

- The de_indicator field indicates that the audio presentation includes dialogue enhancement metadata and must be set to 1
- The extended_presentation_group_index field is an optional extension to the presentation_group_index to increase its value range beyond 31. If the extension is needed, the b_extended_presentation_group_index field must be set to 1; otherwise, the b_extended_presentation_group_index field must be set to 0. For index values greater than 31, the corresponding value is signalled using only the 9 bits of the extended_presentation_group_index, and the bits of the presentation_group_index must be ignored when the bitstream is read.

Syntax	Word size (in bits)
ac4_presentation_v1_dsi(pres_bytes)	
{	
presentation_config_v1.....	5
if (presentation_config_v1 == 0x06)	
{	
b_add_emdf_substreams = 1;	
}	
else	
mdcompat;.....	3
b_presentation_group_index;.....	1
if (b_presentation_group_index)	

Syntax	Word size (in bits)
{	
presentation_group_index;.....	5
}	
dsi_frame_rate_multiply_info;.....	2
dsi_frame_rate_fraction_info;.....	2
presentation_emdf_version;.....	5
presentation_key_id;.....	10
b_presentation_channel_coded;.....	1
if (b_presentation_channel_coded)	
{	
dsi_presentation_ch_mode;.....	5
if (dsi_presentation_channel_mode in [11, 12, 13, 14])	
{	
pres_b_4_back_channels_present;.....	1
pres_top_channel_pairs;.....	2
}	
presentation_channel_mask_v1;.....	24
}	
b_presentation_core_differs;.....	1
if (b_presentation_core_differs)	
{	
add_pres_bytes;.....	16
pres_bytes += add_pres_bytes;	
}	
b_presentation_core_channel_coded;.....	1
if (b_presentation_core_channel_coded)	
{	
dsi_presentation_channel_mode_core;....	2
}	
}	
b_presentation_filter;.....	1
if (b_presentation_filter;)	
{	
b_enable_presentation;.....	1
n_filter_bytes;.....	8
for (i = 0; i < n_filter_bytes; i++)	
{	
filter_data;.....	8
}	
}	

Syntax	Word size (in bits)
if (presentation_config_v1 == 0x1f)	
{	
ac4_substream_group_dsi();	
}	
else	
{	
b_multi_pid;.....	1
if (presentation_config_v1 in [0, 1, 2])	
{	
ac4_substream_group_dsi();	
ac4_substream_group_dsi();	
}	
if (presentation_config_v1 in [3, 4])	
{	
ac4_substream_group_dsi();	
ac4_substream_group_dsi();	
ac4_substream_group_dsi();	
}	
if (presentation_config_v1 == 5)	
n_substream_groups_minus2;.....	3
n_substream_groups = n_substream_groups_minus2 + 2;	
for (sg = 0; sg < n_substream_groups; sg++)	
{	
ac4_substream_group_dsi();	
}	
}	
if (presentation_config_v1 > 5)	
n_skip_bytes;.....	7
for (i = 0; i < n_skip_bytes; i++)	
skip_data;.....	8
}	
}	
}	
b_pre_virtualized;.....	1
b_add_emdf_substreams;.....	1
}	
if (b_add_emdf_substreams)	
{	
n_add_emdf_substreams;.....	7
for (j = 0; j < n_add_emdf_substreams; j++)	

Syntax	Word size (in bits)
{	
substream_emdf_version;.....	5
substream_key_id;.....	10
}	
}	
b_presentation_bitrate_info;.....	1
if (b_presentation_bitrate_info)	
{	
ac4_bitrate_dsi();	
}	
b_alternative;.....	1
if (b_alternative)	
{	
byte_align;.....	0...7
alternative_info();	
}	
byte_align;.....	0...7
if (bits_read() <= (pres_bytes - 1) * 8)	
{	
de_indicator;.....	1
dolby_atmos_indicator;.....	1
reserved;.....	4
b_extended_presentation_group_index;..	1
if (b_extended_presentation_group_index)	
{	
extended_presentation_group_index;.....	9
}	
else	
{	
reserved;.....	4
}	
}	
}	

Note: The number of bits in byte_align pads the number of bits, counted from the start of ac4_presentation_v1_dsi to a multiple of eight.

2.6 Deriving the contents of AC4SpecificBox

The AC4SpecificBox contains the AC-4 decoder specific information (ac4_dsi).

The contents of the `AC4SpecificBox` are derived from the AC-4 table of contents (`ac4_toc`) included in the Dolby AC-4 frame. The process of deriving the contents of the `ac4_dsi` from the `ac4_dsi` is described in section E.6 of ETSI TS 103 190-2.

2.7 A/V alignment and segmentation

Dolby AC-4 can adapt its frame rate to match commonly used video frame rates (for example, 23.976, 25, and 29.97 fps). Therefore, Dolby AC-4 frames and corresponding video access units can maintain temporal alignment so long as the same frame rate is used in both the audio and video encoders.

Recommended Dolby AC-4 frame rates are listed in the following table.

Highest video frame rate used (in fps)	Recommended audio frame rates (in fps, in the order of preference)
120	30, 24[1], 25[1]
119.88	29.97
100	25, 24[1]
60	30, 24[1], 25[1]
59.94	29.97
50	25, 24[1]
48	24, 25[1]
30	30, 24[1], 25[1]
29.97	29.97
25	25, 24[1]
24	24, 25[1],
23.976	23.44 (native)

Aligned audio and video are especially important at the beginning and particularly at the end of segments when trying to achieve seamless transition for certain use cases (such as ad insertion) without introducing artifacts through audio gaps or additional audio.

Dolby AC-4 I-frames should be placed temporally aligned with the I-frames of the video to enable seamless switching. Most importantly, the first I-frames in the video and audio segments should be temporally aligned. It is acceptable that the succeeding I-frames in the corresponding segments are not aligned.

In the following three figures, the segment length is two seconds. For the first audio and video pair, the I-frame intervals of both the video and audio streams are one second. The I-frames, frame rates, and segment sizes are all aligned. For the second and third audio and video pairs, the I-frame intervals for the video streams are two seconds, while the I-frame intervals for the audio streams are one second. The second I-frame in an audio segment is not aligned with any video I-frame. All of the three examples are suitable as all the first I-frames in video and audio segments are temporally aligned.

[1] Recommended only if a perfect segment alignment of audio and video segments can be achieved using the frame rate, but not otherwise.

Figure 1: Alignment of I-frame intervals, frame rates, and segment sizes

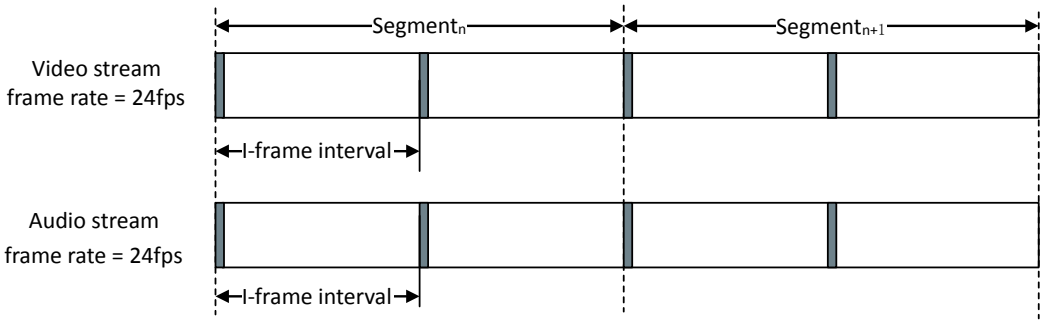


Figure 2: Alignment of frame rates and segment sizes

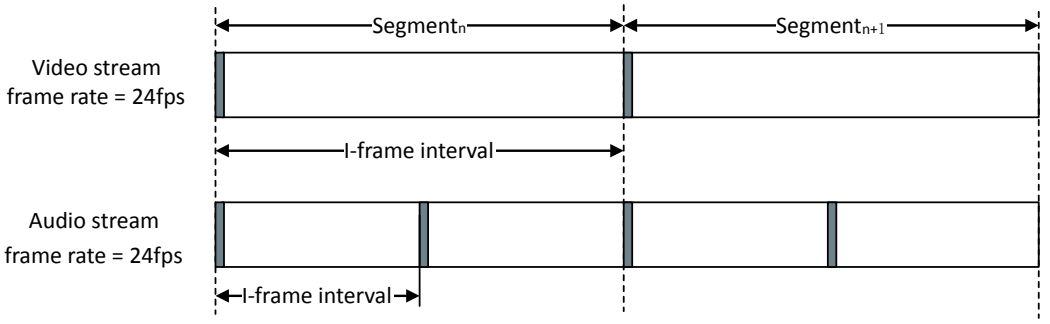
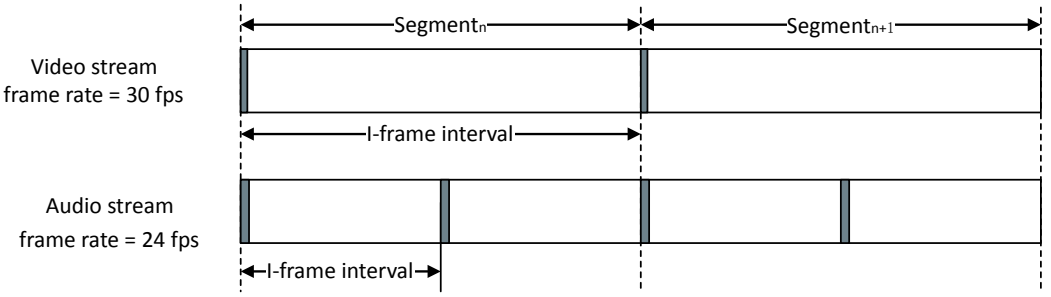
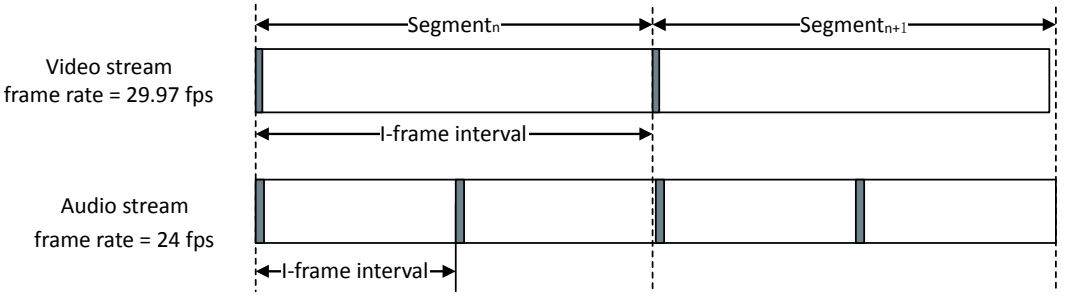


Figure 3: Alignment of segment sizes



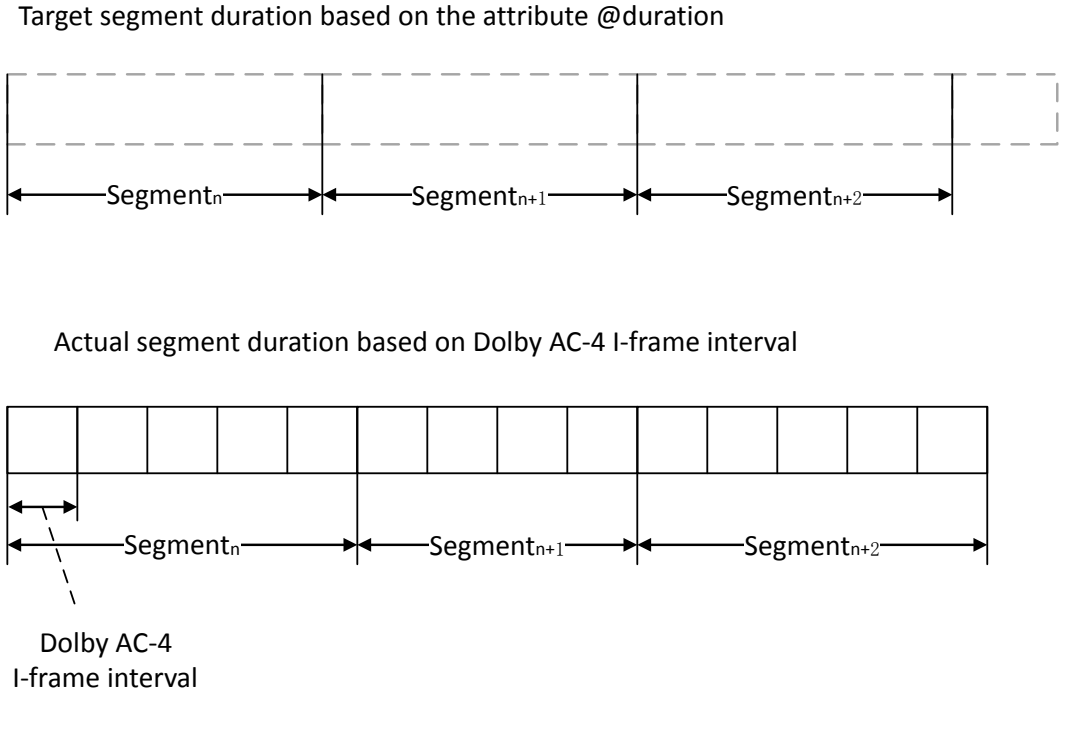
The following figure shows an example where the first I-frames of the video segment$_{n+1}$ and audio segment$_{n+1}$ are not aligned, and thus this user case is not recommended.

Figure 4: I-frames in segment$_{n+1}$ are not aligned



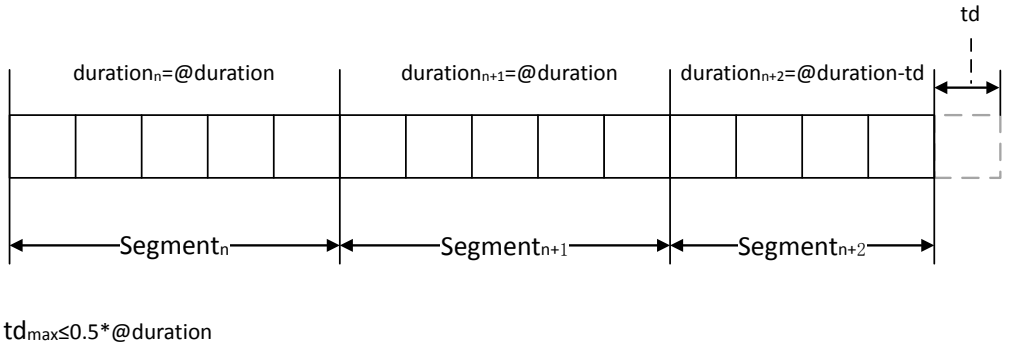
If I-frame alignment cannot be achieved, or if the target segment duration is not an integer multiple of the Dolby AC-4 I-frame interval, Dolby AC-4 segment durations are allowed to fluctuate to maintain close alignment with video segments or the target segment duration timeline. The following figure shows an example where an Dolby AC-4 segment (segment$_{(n+1)}$ in the figure) is shorter by one Dolby AC-4 I-frame interval to maintain close segment alignment.

Figure 5: Variable Dolby AC-4 segment durations to maintain segment alignment



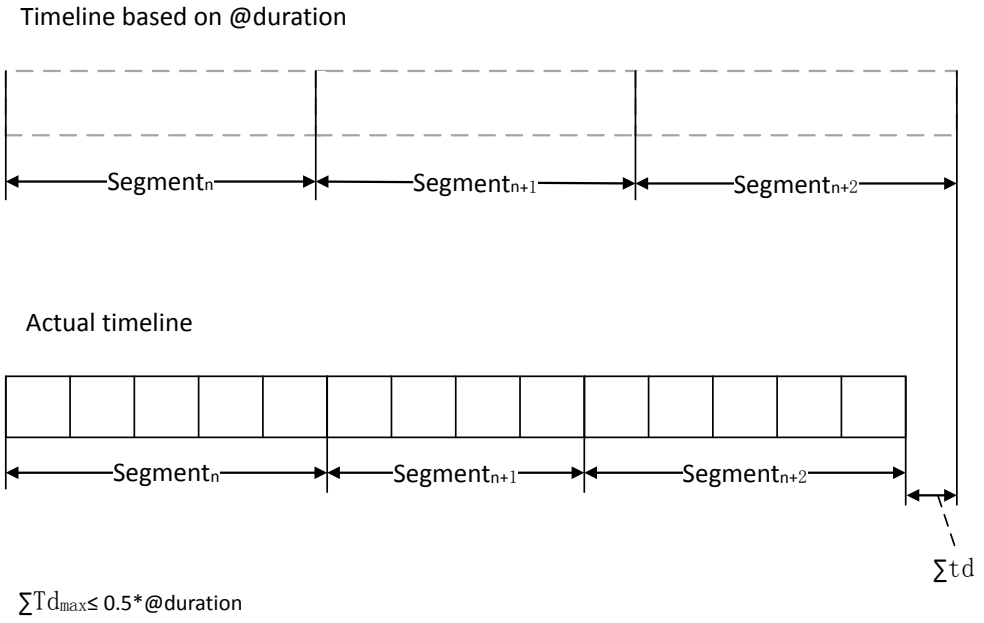
If the `SegmentTimeline` element is used to reference segments, the segment timeline must signal accurate segment durations. Otherwise, you must ensure that AC-4 segments still have almost equal durations. The maximum duration deviation for a signal segment must be within ±50% of the signaled segment duration (for example, as indicated by the `@duration` attribute). The maximum accumulated duration deviation over multiple segments must be within ±50% of the signaled segment, as constrained in DASH-IF interoperability points. The following two figures illustrate the maximum duration deviation of a single segment and the maximum accumulated duration deviation, respectively.

Figure 6: Maximum duration deviation of a single segment



$td_{max} \leq 0.5 * @duration$

Figure 7: Maximum accumulated deviation



To minimize segment duration fluctuation and meet the preceding constraint, the AC-4 I-frame interval must be at most one fourth of the target segment duration.

3 Media Presentation Description with Dolby AC-4

Use proper values in an MPD to describe a Dolby AC-4 bitstream. Examples of MPD files for different scenarios are provided.

3.1 Media Presentation Description

An MPD can be used to describe media presentations with one or multiple Dolby AC-4 bitstreams.

Although the syntax of the MPD is capable of using common XML elements to describe almost any media format, the encoding type and the configuration of a Dolby AC-4 bitstream that is part of a content presentation constrain the parameter values of some of these elements.

The MPD supports these scenarios:

- Media presentations that consist of a single Dolby AC-4 bitstream
- Media presentations that consist of multiple full Dolby AC-4 bitstreams, with each bitstream stored in a separate MP4 file or segment file

  Note: The scenario where two audio services (one main and one associated) that are stored in separate files and are intended to be decoded and then mixed together is not supported.

3.2 Adaptation sets

An adaptation set describes the overall media presentation.

The adaptation set typically consists of multiple representation of the same audio, video, or audio/video content, with each instance encoded at a different data rate. A representation describes the parameters of each individual encoding of an adaptation set:

- The codecs attribute specifies the codecs used to encode all representations within the adaptation set. For Dolby AC-4 bitstreams, the value element of the codecs attribute is created according to the syntax described in IETF RFC 6381. The value consists of a dot-separated list of these four parts, of which the latter three are represented by two-digit hexadecimal numbers:
  - the fourCC ac-4
  - the bitstream_version as indicated in the ac4_dsi()

- the presentation_version as indicated for the first or default presentation in the ac4_dsi()
- the mdcompat parameter as indicated for the first or default presentation in the ac4_dsi()

- The mimeType attribute describes the encapsulation format used to store the Dolby AC-4 bitstreams present in the adaptation set. For adaptation sets that conform to ISO/IEC 14496-12, the mimeType attribute is set to audio/mp4.
- The lang attribute corresponds to the language of the first or default presentation conveyed in the language_tag_bytes of the ac4_substream_group_dsi structure (within the ac4_dsi_v1 structure), which is tagged as dialog or complete main in the corresponding content_classifier.

3.3 Representations

Each adaptation set carries one or more representations. Each representation is associated with a Representation element.

All representations in an adaptation set are perceptually identical, meaning that one or more major parameters (such as the bit rate or the channel configuration) may differ across the Dolby AC-4 bitstreams in the same adaption set.

3.3.1 AudioChannelConfiguration descriptor

The channel configuration for a Dolby AC-4 stream is specified in the AudioChannelConfiguration descriptor.

MPEG channel configuration scheme

For all the Dolby AC-4 channel configurations that can be mappable to a value with the MPEG channel configuration scheme, the AudioChannelConfiguration descriptor uses the MPEG channel configuration scheme as follows:

schemeIdUri: urn:mpeg:mpegB:cicp:ChannelConfiguration

A Dolby AC-4 channel configuration is represented by the presentation_channel_mask_v1 parameter contained in the first presentation_v1_dsi (by order of appearance in the presentation loop) within the ac4_dsi_v1 structure. The following table provides a mapping from ac4_dsi_v1/presentation_channel_mask_v1 to a value with the MPEG channel configuration scheme. The mapping is based on the speaker configurations specified in ISO/IEC 23001-8.

presentation_channel_mask_v1 in hexadecimal representation	Value with the MPEG channel configuration scheme
000002	1
000001	2
000003	3
008003	4
000007	5
000047	6
020047	7
008001	9
000005	10

presentation_channel_mask_v1 in hexadecimal representation	Value with the MPEG channel configuration scheme
008047	11
00004F	12
02FF7F	13
06FF6F	13
000057	14
040047	14
00145F	15
04144F	15
000077	16
040067	16
000A77	17
040A67	17
000A7F	18
040A6F	18
00007F	19
04006F	19
01007F	20
05006F	20

Alternative AudioChannelConfiguration scheme

Note: If applicable, the scheme described by the `schemeIdUri: urn:mpeg:mpegB:cicp:ChannelConfiguration` is preferred.

In some cases, the MPEG channel configuration scheme is not applicable. For example, for a 5.1.2 channel configuration, there is no corresponding value in the MPEG channel configuration scheme. Alternatively, the `AudioChannelConfiguration` descriptor can use the `AudioChannelConfiguration` scheme described in the `schemeIdUri: tag:dolby.com, 2015:dash:audio_channel_configuration:2015`.

The `value` element contains a six-digit hexadecimal representation of the 24-bit speaker group index bit field, which describes the channel assignment of the referenced Dolby AC-4 bitstream according to the table.

Speaker group index	Location
0	L
	R
1	C
2	Ls
	Rs
3	Lb
	Rb
4	Tfl
	Tfr

Speaker group index	Location
5	Tbl
	Tbr
6	LFE
7	TL
	TR
8	Tsl
	Tsr
9	Tfc
10	Tbc
11	Tc
12	LFE2
13	Bfl
	Bfr
14	Bfc
15	Cb
16	Lscr
	Rscr
17	Lw
	Rw
18	Vhl
	Vhr
19	Reserved
20	Reserved
21	Reserved
22	Reserved
23	Reserved

Bit 0, which indicates the presence of the L and R channel, is the most significant bit of the `AudioChannelConfiguration` descriptor. Bits 19 to 23 must be set to `0`. For example, for a stream with 5.1.2 channel configuration of L, C, R, Ls, Rs, TL, TR, LFE, the `value` element must contain the value `0x0000C7` (the hexadecimal equivalent of the binary value `0000 0000 0000 0000 1100 0111`).

The following table shows some examples of common `AudioChannelConfugration` values.

Table 2: Example AudioChannelConfiguration Values

Channel layout	Location	**`AudioChannelConfiguration`** value
1.0	C	`0x000002`
2.0	L, R	`0x000001`
5.1	L, R, C, LFE, Ls, Rs	`0x000047`
5.1.2	L, R, C, LFE, Ls, Rs, TL, TR	`0x0000C7`
5.1.4	L, R, C, LFE, Ls, Rs, Tfl, Tfr, Tbl, Tbr	`0x000077`

Table 2: Example AudioChannelConfiguration Values (continued)

Channel layout	Location	**AudioChannelConfiguration** value
7.1.2	L, R, C, LFE, Ls, Rs, Lb, Rb, TL, TR	0x0000CF
7.1.4	L, R, C, LFE, Ls, Rs, Lb, Rb, Tfl, Tfr, Tbl, Tbr	0x00007F

The b_presentation_channel_coded parameter in the ac4_dsi_v1 structure is false if the audio contains objects.

For content that conveys audio objects that are rendered to positions/coordinates independent from speaker configurations, the hexadecimal value must be set to 800000.

3.4 Accessibility descriptor

If the AdaptationSet includes an accessibility descriptor that describes the type of accessible audio service being provided, the AdaptationSet provides for enhanced accessibility. The required attribute schemeIdUri is set to urn:tva:metadata:cs:AudioPurposeCS:2007, as defined in section B.1 of ETSI TS 102 822-3-1, signaling the namespace for the accessibility descriptor.

The audio purpose classification scheme (AudioPurposeCS), which is used to describe the type of accessible audio service that is being delivered, is defined in section A.15 of ETSI TS 102 822-3-1. The value of the termID attribute shows the type of accessible audio service carried in the Dolby AC-4 bitstream, which is indicated by the value of the content_classifier parameter in the default presentation of the Dolby AC-4 bitstream, or in the AC4SpecificBox of the Dolby AC-4 audio track. The corresponding values of the termID attribute and content_classifier parameter are listed in the table.

Table 3: Corresponding termID attribute and content_classifier parameter values

termID attribute value	AudioPurposeCS name	content_classifier parameter value
1	Audio description for the visually impaired	010
2	Audio description for the hearing impaired	011
3	Supplemental commentary	101
4	Director's commentary	101
5	Educational notes	101
6	Main program audio	000
7	Clean feed (no effects mix)	100

3.5 Media Presentation Description for immersive stereo content

In an immersive stereo representation, an additional Supplementalproperty descriptor must be included.

For the immersive stereo representation in the MPD file, the value attribute of the AudioChannelConfiguration descriptor must be set to 2. The representation must also include the following Supplementalproperty descriptor.

```
<SupplementalProperty schemeIdUri="tag:dolby.com,2016:dash:virtualized_content:2016"
value="1"/>
```

3.6 ATSC 3.0 specific constrains on Media Presentation Description

MPD files for ATSC 3.0 services must adhere to the constraints defined in DASH-IF Interoperability for ATSC 3.0 and the Broadcast TV Profile constraints from ISO/IEC 23009-1.

3.7 Compatibility with ATSC 3.0 and DVB-DASH

To be compatible with ATSC 3.0 and DVB-DASH, required settings must be used for the AudioChannelConfiguration descriptor and the Accessibility descriptor.

In DASH-IF Interoperability for ATSC 3.0, it is mandatory to set the value of the schemeIdUri attribute in the AudioChannelConfiguration descriptor to tag:dolby.com, 2015:dash:audio_channel_configuration:2015. For DVB-DASH, it is recommended to use a different schemeIdURI, but the use of the same schemeIdURI as defined for ATSC 3.0 is also allowed. To be compatible with both ATSC 3.0 and DVB-DASH, tag:dolby.com, 2015:dash:audio_channel_configuration:2015 must be used for the schemeIdUri attribute in the AudioChannelConfiguration descriptor.

In DASH-IF Interoperability for ATSC 3.0, the default Role scheme is used for the Accessibility element. DVB-DASH uses a dedicated DVB scheme for indication of accessibility features. To be compatible with DVB-DASH, both Accessibility descriptors must be used. For example:

```
<Role schemeIdUri="urn:mpeg:dash:role:2011" value="commentary"/>
<Accessibility schemeIdUri="urn:mpeg:dash:role:2011" value="descriptions"/>
<Accessibility schemeIdUri="urn:tva:metadata:cs:AudioPurposeCS:2007" value="1"/>
```

3.8 Media Presentation Description file examples

This section contains example MPD files for different media presentations.

3.8.1 Media Presentation Description for a single video component and single audio component

This MPD example describes a simple media presentation that consists of a single video component with a single 5.1-channel (L, C, R, Ls, Rs, LFE) Dolby AC-4 audio component. Three representations of the video content and three representations of the audio content are provided, each at a different data rate.

The media presentation complies with the ISO base media file format live profile, as defined in ISO/IEC 23009-1:2012.

```
<?xml version="1.0" encoding="utf-8"?>
```

```
<MPD xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dolby="http://www.dolby.com/ns/online/DASH" xmlns="urn:mpeg:dash:schema:mpd:2011"
xsi:schemaLocation="urn:mpeg:dash:schema:mpd:2011"
type="static" minimumUpdatePeriod="PT2S" timeShiftBufferDepth="PT30M"
availabilityStartTime="2011-12-25T12:30:01" minBufferTime="PT4S"
profiles="urn:mpeg:dash:profile:isoff-live:2011">
   <BaseURL>http://cdn1.example.com/</BaseURL>
   <BaseURL>http://cdn2.example.com/</BaseURL>
   <Period>
      <!-- Video -->
      <AdaptationSet mimeType="video/mp4" codecs="avc1.4D401F"
       frameRate="30000/1001" segmentAlignment="true" startWithSAP="1">
         <BaseURL>video/</BaseURL>
         <SegmentTemplate timescale="90000" media="$Bandwidth$/$Index$.m4s"
initialization="$Bandwidth$/0.mp4">
            <SegmentTimeline>
               <S t="0" d="180180" r="10"/>
            </SegmentTimeline>
         </SegmentTemplate>
         <Representation id="v0" width="320" height="240" bandwidth="250000" />
         <Representation id="v1" width="640" height="480" bandwidth="500000" />
         <Representation id="v2" width="960" height="720" bandwidth="1000000" />
      </AdaptationSet>
      <!-- 5.1 channel English Audio -->
      <AdaptationSet mimeType="audio/mp4" codecs="ac-4.02.01.01" lang="en"
segmentAlignment="true" startWithSAP="1">
         <SegmentTemplate timescale="48000" media="audio/en/$Bandwidth$/$Index$.m4s"
          initialization="audio/en/$Bandwidth$/0.mp4">
            <SegmentTimeline>
               <S t="0" d="96768" r="10"/>
            </SegmentTimeline>
         </SegmentTemplate>
         <AudioChannelConfiguration
schemeIdUri="urn:mpeg:mpegB:cicp:ChannelConfiguration" value="6"/>
         <Representation id="a0" bandwidth="192000" />
         <Representation id="a1" bandwidth="256000" />
         <Representation id="a2" bandwidth="384000" />
      </AdaptationSet>
   </Period>
</MPD>
```

3.8.2 Media Presentation Description for a single video component and an immersive stereo audio component

The MPD example consists of two adaptation sets. One adaptation set contains an H.264 video representation, and the other one contains a 2.0-channel immersive stereo audio representation.

```
<?xml version="1.0"?>
<MPD
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xmlns:dolby="http://www.dolby.com/ns/online/DASH"
 xmlns="urn:mpeg:dash:schema:mpd:2011"
```

```
 xsi:schemaLocation="urn:mpeg:dash:schema:mpd:2011"
 type="static"
 mediaPresentationDuration="PT0H0M30.040S"
 minBufferTime="PT4S"
 profiles="urn:mpeg:dash:profile:isoff-live:2011">

 <Period start="PT0S">
    <!-- Video -->
    <AdaptationSet mimeType="video/mp4" codecs="avc1.4d4028" segmentAlignment="true"
startWithSAP="1">
      <SegmentTemplate timescale="25000"
        media="Bitrate_1280x720p_25fps_h264_ims_72kbps_$Number$.mp4"
        initialization="Bitrate_1280x720p_25fps_h264_ims_72kbps.mp4"
        startNumber="1">
        <SegmentTimeline>
          <S t="0" d="50000" r="15"/>
        </SegmentTimeline>
      </SegmentTemplate>
      <Representation id="1" bandwidth="5375000" width="1280" height="720"
frameRate="25"/>
    </AdaptationSet>
    <!-- Audio -->
    <AdaptationSet mimeType="audio/mp4" codecs="ac-4.02.02.00" lang="en"
segmentAlignment="true" startWithSAP="1">
      <SegmentTemplate timescale="48000"
        media="Bitrate_ims_72kbps_25fps_ac4_$Number$.mp4"
        initialization="Bitrate_ims_72kbps_25fps_ac4.mp4"
        startNumber="1">
        <SegmentTimeline>
          <S t="0" d="96000" r="15"/>
        </SegmentTimeline>
      </SegmentTemplate>
      <Representation id="2" bandwidth="76776" audioSamplingRate="48000">
          <AudioChannelConfiguration
schemeIdUri="urn:mpeg:mpegB:cicp:ChannelConfiguration" value="2"/>
          <SupplementalProperty schemeIdUri="tag:dolby.com,2016:dash:virtualized_content:
2016" value="1"/>
      </Representation>
    </AdaptationSet>
 </Period>
</MPD>
```

3.8.3 Media Presentation Description with multiple adaptation sets

The MPD example consists of three adaptation sets. The first one is a video adaptation set containing an H.264 video representation. The other two are audio adaptation sets, containing a 2.0-channel Dolby AC-4 audio representation and a 5.1-channel Dolby AC-4 audio representation respectively.

```
<?xml version="1.0"?>
<MPD
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xmlns:dolby="http://www.dolby.com/ns/online/DASH"
 xmlns="urn:mpeg:dash:schema:mpd:2011"
```

```
 xsi:schemaLocation="urn:mpeg:dash:schema:mpd:2011"
 type="static"
 mediaPresentationDuration="PT0H2M39.080S"
 minBufferTime="PT4S"
 profiles="urn:mpeg:dash:profile:isoff-live:2011">

 <Period start="PT0S">
    <!-- Video -->
    <AdaptationSet mimeType="video/mp4" codecs="avc1.4d4028" segmentAlignment="true"
startWithSAP="1">
      <SegmentTemplate timescale="25000"
        media="../Video/Silent_1280x720p_25fps_h264_$Number$.mp4"
        initialization="../Video/Silent_1280x720p_25fps_h264.mp4">
        <SegmentTimeline>
          <S t="0" d="50000" r="79"/>
        </SegmentTimeline>
      </SegmentTemplate>
      <Representation id="1" bandwidth="3912332" width="1280" height="720"
frameRate="25"/>
    </AdaptationSet>
    <!-- Audio -->
    <AdaptationSet mimeType="audio/mp4" codecs="ac-4.02.01.00" lang="en"
segmentAlignment="true" startWithSAP="1">
      <SegmentTemplate timescale="48000"
        media="../Audio/Silent_2ch_64kbps_25fps_ac4_$Number$.mp4"
        initialization="../Audio/Silent_2ch_64kbps_25fps_ac4.mp4">
        <SegmentTimeline>
          <S t="0" d="96000" r="79"/>
        </SegmentTimeline>
      </SegmentTemplate>
      <Representation id="2" bandwidth="82964" audioSamplingRate="48000">
        <AudioChannelConfiguration schemeIdUri="urn:mpeg:mpegB:cicp:ChannelConfiguration"
value="2"/>
      </Representation>
    </AdaptationSet>
    <!-- Audio -->
    <AdaptationSet mimeType="audio/mp4" codecs="ac-4.02.01.01" lang="en"
segmentAlignment="true" startWithSAP="1">
      <SegmentTemplate timescale="48000"
        media="../Audio/Silent_6ch_128kbps_25fps_ac4_$Number$.mp4"
        initialization="../Audio/Silent_6ch_128kbps_25fps_ac4.mp4">
        <SegmentTimeline>
          <S t="0" d="96000" r="79"/>
        </SegmentTimeline>
      </SegmentTemplate>
      <Representation id="3" bandwidth="150312" audioSamplingRate="48000">
        <AudioChannelConfiguration schemeIdUri="urn:mpeg:mpegB:cicp:ChannelConfiguration"
value="6"/>
      </Representation>
    </AdaptationSet>
 </Period>
</MPD>
```

3.8.4 Media Presentation Description with a single audio component

The MPD example consists of one audio adaptation set with a 5.1.2 channel based immersive Dolby AC-4 representation.

```
<?xml version="1.0"?>
<MPD
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:dolby="http://www.dolby.com/ns/online/DASH"
  xmlns="urn:mpeg:dash:schema:mpd:2011"
  xsi:schemaLocation="urn:mpeg:dash:schema:mpd:2011 DASH-MPD.xsd"
  profiles="urn:mpeg:dash:profile:isoff-broadcast:2015"
  type="dynamic">
  <Period>
    <AdaptationSet id="1" mimeType="audio/mp4" codecs="ac-4.02.01.02" lang="en"
segmentAlignment="true" startWithSAP="1">
      <Role schemeIdUri="urn:mpeg:dash:role:2011" value="main"/>
       <SegmentTemplate timescale="48000" media="audio_singlepres_512cm/$Time$.m4s"
initialization="audio_singlepres_512cm/init.mp4">
        <SegmentTimeline>
          <S t="0" d="96000" r="36"/>
        </SegmentTimeline>
      </SegmentTemplate>
      <Representation id="r1" audioSamplingRate="48000" bandwidth="160000">
        <AudioChannelConfiguration schemeIdUri="tag:dolby.com,
2015:dash:audio_channel_configuration:2015" value="0000C7"/>
      </Representation>
   </AdaptationSet>
    </Period>
</MPD>
```

4 Demultiplexing a Dolby AC-4 bitstream from an MPEG-DASH compliant ISO base media file segment

Buffering and parameter conflicts must be considered when product designers implement an ISO base media file format file demultiplexer that supports Dolby AC-4 bitstreams.

4.1 Buffering considerations

A Dolby AC-4 decoder must buffer at least six frame units until it starts decoding (according to the buffer signaling model, indicated by `wait_frames` in the table of contents).

To allow for possible jitter, we recommend a buffer size of seven frame units. Therefore, the minimum buffer size between a demultiplexer and a decoder can be calculated using this equation:

$N = 7 \times R_{stream} \div F_{frame}$

In this equation, R_{stream} and F_{frame} are the total bit rate and the frame rate of a Dolby AC-4 stream.

For example, the data rate is 1,536 kbps and the frame rate is 23.976 fps, which makes the minimum buffer size 56.056 KB.

4.2 AC4SampleEntry and Dolby AC-4 bitstream parameter conflicts

The `AC4SampleEntry` contains only basic parameters describing the audio bitstream and is intended to be used by the system for information purposes only. It is possible that differences may occur between the `AC4SampleEntry` and the parameters of the Dolby AC-4 bitstream (in which case, the information from the `AC4SpecificBox` must be ignored and the parameters of the Dolby AC-4 bitstream must be used for decoding purposes).

4.3 AC4SpecificBox and Dolby AC-4 bitstream parameter conflicts

The `AC4SpecificBox` is intended to be used by the system for information purposes only.

For example, the system might use the `AC4SpecificBox` for any of these scenarios:

- Informing the onscreen display of the channel configuration of the audio stream
- Offering the user the ability to select between the different audio services that are being carried (in the case where multiple independent substreams are present)

- Recovering errors (if there is an interruption in audio data delivery and the system cannot ascertain the configuration of the complete Dolby AC-4 bitstream from the substream structure in the MP4 sample due to CRC errors)

Because it is possible that differences may occur between the AC4SpecificBox and the parameters of the Dolby AC-4 bitstream, the information from the AC4SpecificBox must not be used to configure the audio decoder or the audio subsystem of the device. If the device encounters a conflict, the Dolby AC-4 bitstream parameter value must always take precedence over the value of the corresponding parameter in the AC4SpecificBox.

5 MPEG-DASH and Dolby AC-4 overview

Introduction about MPEG-DASH and Dolby AC-4

5.1 MPEG Dynamic Adaptive Streaming over HTTP with ISO containers

MPEG-DASH is an HTTP-based adaptive streaming technology that enables high-quality streaming of media content over the Internet.

MPEG-DASH works by breaking the content into a sequence of small HTTP-based file segments, each of which can be downloaded independently. Alternative segments at different data rates are available, allowing streaming to adapt to different network conditions.

5.1.1 MPEG-DASH Media Presentation Description file

A Media Presentation Description (MPD) file is a hierarchical XML document that provides information for an MPEG-DASH client about the available content for a media presentation.

The media presentation contains the encoded audio and video streams that are segmented into chunks for online delivery. The MPD is a manifest of the available content, which describes the configuration of the content, potentially available content alternatives, the location of the content (URLs), and other characteristics (including timing information and media characteristics such as video resolution and bit rates).

The MPEG-DASH client uses the information in the MPD for constructing the HTTP URLs that then allow it to access segments containing the actual audio and video content.

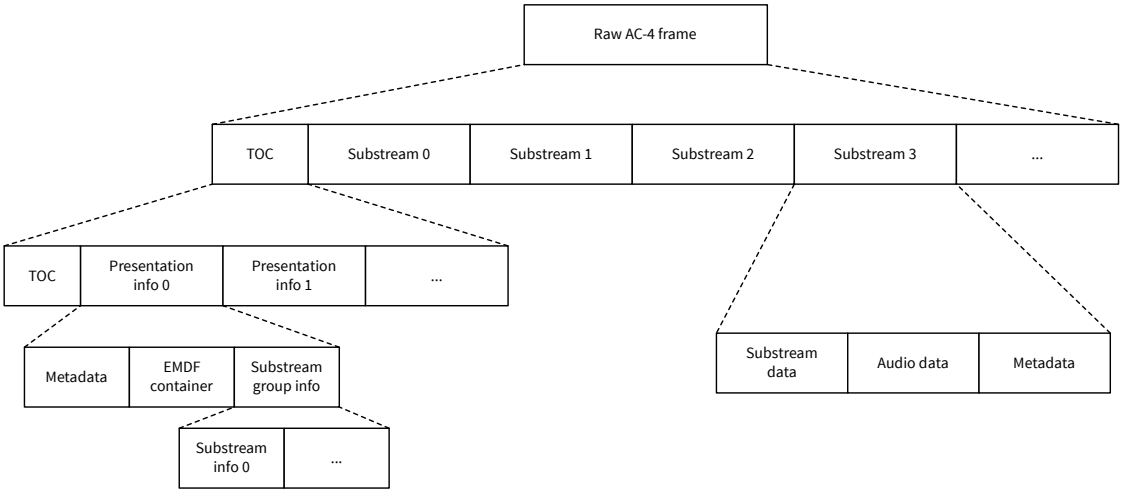
5.2 Raw AC-4 frame

Each raw AC-4 frame is composed of a table of contents and at least one substream.

The table of contents can be considered as the bitstream inventory where all important information for the overlaying system resides. Each table of contents contains the table of contents data and at least one presentation. An audio presentation informs the decoder which parts of an AC-4 stream are intended to be played back simultaneously at a given point in time. Presentations consist of substream groups. Substream groups consist of one or more individual substreams. The substream group carries properties common to all substreams contained in the substream group. Substreams in one substream group are either all channel coded or all object coded. Substreams are decodable units that represent a specific channel configuration (mono, stereo, 5.1, or 7.1). All of the payload information (such as audio data and metadata) is stored in substreams.

The raw AC-4 frame structure is shown in the following figure, as described in ETSI TS 103 190-2.

Figure 8: Raw AC-4 frame structure



5.3 Random access point

The Dolby AC-4 decoder features a seamless switch between Dolby AC-4 elementary streams of the same media content and different configuration options.

Configuration changes can include:

- Bit-rate changes
- Channel mode changes
- Sampling-rate changes where the higher sampling rate is an integer multiple of the lower sampling rate (for example, from 96 to 48 kHz)
- Frame-rate changes where the higher frame rate is a factor of two or four times the lower frame rate (for example, from 48 to 24 fps and vice versa)
- Any combinations of these

A seamless switch means that the audio output has no audible artifacts or distortion introduced by audio gaps or additional audio during and after switching.

A seamless switch can be accomplished only at a random access point. Comparable to random access frames in H.264 video streams, the Dolby AC-4 elementary stream contains I-frames to indicate random access points. The `b_iframe_global` flag inside the table of contents indicates an I-frame. Unlike video I-frames, audio I-frames are for signaling only. Audio can still be decoded if an audio I-frame is not present at a random access point. When a switch is not immediately followed by an I-frame, the Dolby AC-4 decoder performs partial decoding until it receives the next I-frame. Partial decoding may lead to degraded audio quality, including short periods of silence, limited audio bandwidth, or a downmix of the content.

Glossary

CRC
Cyclic redundancy check.

DASH
Dynamic Adaptive Streaming over HTTP. An adaptive bit-rate streaming protocol that enables high-quality streaming of media content over the Internet delivered from HTTP.

Dolby AC-4
An audio coding system that enables high audio quality at very low bit rates. It consists of a channel-based audio codec (lossy, low bit-rate audio codec), object-based audio extensions, and system integration components.

HTTP
Hypertext Transfer Protocol. An application protocol for hypermedia information systems, and the foundation for data communication for the World Wide Web.

immersive stereo
A technology that delivers a virtualized immersive experience to headphones or stereo speakers through a Dolby AC-4 bitstream with appropriate stereo content and metadata that converts the stereo signal into the virtualized experience.

I-frame
Independently decodable frame. An I-frame is a single frame that contains all the data needed to decode the frame. Dolby AC-4 uses the concept of I-frames to achieve greater coding efficiency. An I-frame is indicated by the value of the `b_iframe_global` bitstream parameter.

LSB
Least-significant bit.

MP4 sample
A single ISO base media file track sample, as defined in section 3.1.10 of *ISO/IEC 14496-12*.

MPD
Media Presentation Description. A manifest used in MPEG Dynamic Adaptive Streaming over HTTP (MPEG-DASH) to describe the available streaming content, its various alternatives, URL addresses, and other characteristics, as well as segments that contain the actual multimedia bitstreams in the form of chunks, in single or multiple files.

MPEG
Moving Picture Experts Group. An ISO/IEC working group that develops video and audio encoding standards. Also the name of a family of digital video and audio coding standards.

MPEG-4
An MPEG standard (ISO/IEC 14496) for a group of audio and video coding formats and related technologies.

MPEG-DASH
MPEG Dynamic Adaptive Streaming over HTTP. An adaptive bit-rate streaming protocol that enables high-quality streaming of media content over the Internet delivered from HTTP.

MSB
Most-significant bit.

presentation
References to AC-4 substreams to be decoded and presented simultaneously.

presentation configuration
Set of metadata to describe how a presentation must be decoded.

raw AC-4 frame
The actual codec frame that consists of a table of contents plus several byte-aligned substreams.

substream
A decodable unit that represents a specific channel configuration (mono, stereo, or 5.1) and contains audio data and corresponding metadata.

URI
Uniform Resource Identifier. A group of characters identifying a resource on a network (typically, the Internet).