ATSC Standard:
A/342 Part 2, AC-4 System

Doc. A/342-2:2022-03
31 March 2022

Advanced Television Systems Committee
1300 I Street, N.W., Suite 400E
Washington, D.C. 20005
202-872-9160

The Advanced Television Systems Committee, Inc. is an international, non-profit organization developing voluntary standards and recommended practices for broadcast television and multimedia data distribution. ATSC member organizations represent the broadcast, professional equipment, motion picture, consumer electronics, computer, cable, satellite, and semiconductor industries. ATSC also develops implementation strategies and supports educational activities on ATSC standards. ATSC was formed in 1983 by the member organizations of the Joint Committee on Inter-society Coordination (JCIC): the Consumer Technology Association (CTA), the Institute of Electrical and Electronics Engineers (IEEE), the National Association of Broadcasters (NAB), the Internet & Television Association (NCTA), and the Society of Motion Picture and Television Engineers (SMPTE). For more information visit www.atsc.org.

Note: The user's attention is called to the possibility that compliance with this standard may require use of an invention covered by patent rights. By publication of this standard, no position is taken with respect to the validity of this claim or of any patent rights in connection therewith. One or more patent holders have, however, filed a statement regarding the terms on which such patent holder(s) may be willing to grant a license under these rights to individuals or entities desiring to obtain such a license. Details may be obtained from the ATSC Secretary and the patent holder.

Implementers with feedback, comments, or potential bug reports relating to this document may contact ATSC at https://www.atsc.org/feedback/.

Revision History

Version	Date
Candidate Standard approved	3 May 2016
A/242-2:2017 Standard approved	23 February 2017
Document numbering inconsistency fixed on title page and running header, and [14]	7 March 2017
Reference [13] updated to point to the published version of A/341:2017	22 May 2017
Reference [11] updated to point to published version of A/331:2017	6 December 2017
A/342-2:2021 Standard approved	10 March 2021
A/342-2:2022-03 published (references to ATSC documents updated)	31 March 2022

Table of Contents

Index of Figures and Tables

ATSC Standard:
A/342 Part 2, AC-4 System

1. SCOPE

This document standardizes the AC-4 audio system for use in the ATSC 3.0 Digital Television System. It describes the characteristics of the AC-4 audio system and establishes a set of constraints on ETSI TS 103 190-2 [2] for use within ATSC 3.0 broadcast emissions.

1.1 Introduction and Background

The ATSC 3.0 audio system provides an enhanced feature set and improves upon the capabilities of past ATSC audio systems. The system provides listeners with both a personalized and an immersive experience. The ATSC 3.0 audio system establishes a common framework for multiple Next Generation Audio (NGA) systems, both current and future. The AC-4 audio system is one of the NGA systems standardized in ATSC 3.0.

1.2 Organization

This document is organized as follows:

- Section 1 – Outlines the scope of this document and provides a general introduction.
- Section 2 – Lists references and applicable documents.
- Section 3 – Provides definitions of terms, acronyms, and abbreviations for this document.
- Section 4 – Presents an overview of the ATSC 3.0 audio system.
- Section 5 – Specifies the requirements for AC-4 audio coding in the ATSC 3.0 audio system.

2. REFERENCES

All referenced documents are subject to revision. Users of this Standard are cautioned that newer editions might or might not be compatible.

2.1 Normative References

The following documents, in whole or in part, as referenced in this document, contain specific provisions that are to be followed strictly in order to implement a provision of this Standard.

[1] ETSI: "Digital Audio Compression (AC-4) Standard; Part 1: Channel based coding," Doc. TS 103 190-1 V1.3.1 (2018-02), ETSI, Sophia Antipolis Cedex, France.

[2] ETSI: "Digital Audio Compression (AC-4) Standard; Part 2: Immersive and personalized audio," Doc. TS 103 190-2 V1.2.1 (2018-02), ETSI, Sophia Antipolis Cedex, France.

[3] IEEE: "Use of the International Systems of Units (SI): The Modern Metric System," Doc. SI 10, Institute of Electrical and Electronics Engineers, New York, NY.

[4] ISO/IEC: "Information technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats," Doc. 23009-1:2019, International Standards Organization, Geneva, Switzerland.

[5] IETF: BCP 47, "Tags for Identifying Languages," Internet Engineering Task Force, Reston, VA, September 2009.

[6] ATSC: "Techniques for Establishing and Maintaining Audio Loudness for Digital Television," Doc. A/85:2013, Advanced Television Systems Committee, Washington, DC., March 12, 2013. See also A/85:2013 Corrigendum No. 1, "SPL," approved 11 February 2021.

[7] ISO/IEC: "Information technology -- Coding of audio-visual objects -- Part 12: ISO base media file format," Doc. 14496-12:2015, International Standards Organization, Geneva, Switzerland.
[8] ATSC: "ATSC Digital Television Standard, Part 6 – Enhanced AC-3 Audio System Characteristics," Doc. A/53 Part 6:2013, Advanced Television Systems Committee, Washington DC, 7 August 2013.

2.2 Informative References

The following documents contain information that may be helpful in applying this Standard.

[9] ITU: "Algorithms to measure audio programme loudness and true-peak audio level," Recommendation ITU-R BS.1770-4, International Telecommunications Union, Geneva, Switzerland, 2015.
[10] ITU: "Requirements for Loudness and True-Peak Indicating Meters," Recommendation ITU-R BS.1771-1, International Telecommunications Union, Geneva, Switzerland.
[11] ATSC: "ATSC Standard: Signaling, Delivery, Synchronization, and Error Protection," Doc. A/331:2022-03, Advanced Television Systems Committee, Washington DC, 31 March 2022.
[12] ATSC: "ATSC Standard: Video – HEVC," Doc. A/341:2022-03, Advanced Television Systems Committee, Washington DC, 31 March 2022.
[13] ATSC: "ATSC Standard: A/342 Part 1, Audio Common Elements," Doc. A/342-1:2022-03, Advanced Television Systems Committee, Washington DC, 31 March 2022.
[14] IANA: language subtag registry, http://www.iana.org/numbers.html under "Language Tags."

3. DEFINITION OF TERMS

With respect to definition of terms, abbreviations, and units, the practice of the Institute of Electrical and Electronics Engineers (IEEE) as outlined in the Institute's published standards [3] shall be used. Where an abbreviation is not covered by IEEE practice or industry practice differs from IEEE practice, the abbreviation in question will be described in Section 3.3 of this document.

3.1 Compliance Notation

This section defines compliance terms for use by this document:

shall – This word indicates specific provisions that are to be followed strictly (no deviation is permitted).

shall not – This phrase indicates specific provisions that are absolutely prohibited.

should – This word indicates that a certain course of action is preferred but not necessarily required.

should not – This phrase means a certain possibility or course of action is undesirable but not prohibited.

3.2 Treatment of Syntactic Elements

This document contains symbolic references to syntactic elements used in the audio, video, and transport coding subsystems. These references are typographically distinguished by the use of a different font (e.g., restricted), may contain the underscore character (e.g., sequence_end_code) and may consist of character strings that are not English words (e.g., dynrng).

3.2.1 Reserved Elements

One or more reserved bits, symbols, fields, or ranges of values (i.e., elements) may be present in this document. These are used primarily to enable adding new values to a syntactical structure without altering its syntax or causing a problem with backwards compatibility, but they also can be used for other reasons.

The ATSC default value for reserved bits is '1'. There is no default value for other reserved elements. Use of reserved elements except as defined in ATSC Standards or by an industry standards setting body is not permitted. See individual element semantics for mandatory settings and any additional use constraints. As currently reserved elements may be assigned values and meanings in future versions of this Standard, receiving devices built to this version are expected to ignore all values appearing in currently reserved elements to avoid possible future failure to function as intended.

3.3 Acronyms and Abbreviations

The following acronyms and abbreviations are used within this document.

A/V	Audio/Video
AC-3	Audio Codec 3
AC-4	Audio Codec 4
AD	Audio Description
ATSC	Advanced Television Systems Committee
BCP	Best Current Practice
CRC	cyclic redundancy check
DASH	Dynamic Adaptive Streaming over HTTP
DRC	Dynamic Range Control
E-AC-3	Enhanced AC-3
EHFR	Efficient High Frame Rate
EMDF	Extensible Metadata Delivery Format
ETSI	European Telecommunications Standards Institute
fps	frames per second
HD-SDI	High Definition Serial Digital Interface
HEVC	High efficiency video coding
HTTP	Hypertext Transfer Protocol
IANA	Internet Assigned Numbers Authority
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IPTV	Internet Protocol Television
ISO	International Organization for Standardization
ISOBMFF	ISO Base Media File Format
ITU	International Telecommunication Union
kbps	kilobits per second
kHz	kiloHertz
LKFS	Loudness, K-weighted, relative to Full Scale

M&E	Music and Effects
MPD	Media Presentation Description
NGA	Next Generation Audio
PCM	Pulse-code modulation
RAP	Random-access point
TOC	Table of Contents
TS	transport stream
TV	Television
URL	Uniform Resource Locator

3.4 Terms

The following terms are used within this document.

Adaptation Set – A set of interchangeable encoded versions of one or several media content components.

Alternative Presentation – A Presentation that supports the personalized audio use-case by providing alternative object properties to be used in combination with common object essences.

Audio Description – As defined in A/342 Part 1 [13], "Audio Description" is the insertion of audio narrated descriptions of a television program's key visual elements into natural pauses between the program's dialogue.

Audio Element – The smallest addressable unit of an audio Program.

Audio Emergency Information – As defined in A/342 Part 1 [13], "Emergency Information" is data to be presented aurally, such as the reading of a text crawl, and is distinct from Emergency Alert System (EAS) data and audio.

Audio Program Component – Logical group of one or more Audio Elements that is used to define an audio Presentation, e.g., complete main, Music and Effects, dialog, etc.

Codec – A system consisting of an encoder and decoder.

I-Frame – An independently decodable frame.

Initialization Segment – Segment containing metadata that is necessary to present the media streams encapsulated in Media Segments.

Media Presentation – Collection of essences that establishes a bounded or unbounded presentation of media content.

Media Presentation Description – Formalized description for a Media Presentation for the purpose of providing a streaming service.

Media Segment – Segment that complies with media format in use and enables playback when combined with zero or more preceding Segments and an Initialization Segment (if any).

Personalization – A feature that provides consumers control over the listening experience.

Presentation – A set of one or more AC-4 Substreams intended to be decoded and presented to the user simultaneously. In the context of this document, the term Presentation and the term Audio Presentation, as defined in Part 1 of this standard [13], refer to identical concepts.

Primary Bit Stream – The elementary stream that serves as the entry-point. It carries the default Presentation in a multi-stream Program.

Program – A single content segment.

Representation – A collection and encapsulation of one or more media streams in a delivery format and associated with descriptive metadata.

Segment – Unit of data associated with an HTTP-URL and optionally a byte range that are specified by an MPD.

Segment Index – Compact index of the time range to byte range mapping within a Media Segment separately from the MPD.

Stream Access Point – Position in a Representation enabling playback of a media stream to be started using only the information contained in Representation data starting from that position onwards (preceded by initializing data in the Initialization Segment, if any).

Subsegment – Unit within Media Segments that is indexed by a Segment Index.

Substream – A part of an AC-4 elementary stream that contains audio data and/or corresponding metadata. A Substream is referenced by one or more Presentations.

Track – A collection of related samples in an ISO base media file. For media data, a Track corresponds to a sequence of images or sampled audio.

Reserved – Set aside for future use.

4. SYSTEM OVERVIEW

AC-4 is a highly efficient audio Codec that supports legacy channel-based content, object-based and channel-based immersive content, personalized content, and advanced metadata as defined in ETSI TS 103 190-2 [2].

4.1 AC-4 System Features

The AC-4 audio system has several features specific to AC-4 that are utilized in the ATSC 3.0 audio system.

4.1.1 A/V Sync and Frame Alignment

The A/V frame alignment feature of AC-4 avoids complex problems that can occur when trying to keep content in sync at segment boundaries, without compromising the audio at switching points. When enabled, this feature simplifies splicing workflows. It also simplifies transcoding from or to formats that use video-based data frame alignment, such as HD-SDI.

AC-4 supports the generation of coded audio frames that represent the same time interval as the associated video frame. With this frame alignment, audio can be passed transparently through a cable, satellite, or Internet Protocol Television (IPTV) turnaround facility, eliminating the need to decode and re-encode audio to correct time-base differences or to perform switching/splicing. All common integer and fractional video frame rates are supported. To select the correct frame rate and time-align the audio frame boundaries to video, the AC-4 encoder is provided with reference timing information. There is a direct whole-frame relationship between video and audio frames from a reference time, such as at the start of a coded video sequence or at a random access point (RAP).

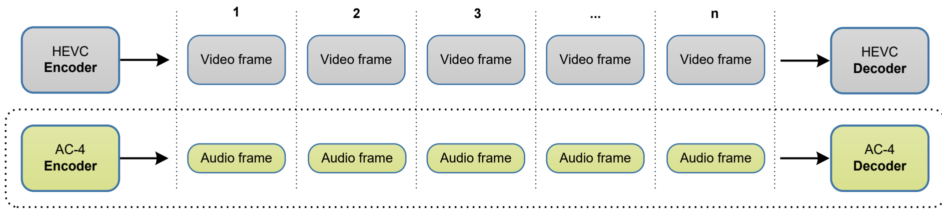

Figure 4.1 Timing Relationship When Using Video Frame Alignment in AC-4.

4.1.2 Dialog Enhancement

AC-4 provides user-controlled enhancement of dialog intelligibility during decoding. The AC-4 system generates dialog enhancement information in the bit stream for all content, including content that has the dialog pre-mixed with other elements. User control of dialog enhancement may be configured at or signaled to the AC-4 encoder based on broadcaster or programmer needs and requirements.

4.1.3 Extensible Metadata Delivery Format Support

The AC-4 bit stream includes support for the carriage of Extensible Metadata Delivery Format (EMDF) data defined in ETSI TS 103 190-1 [1]. EMDF provides a structured and extensible container for additional user data (for example, third-party metadata and third-party application data) to be carried in AC-4 bit streams.

4.1.4 Loudness and Dynamic Range Control

The AC-4 bit stream carries an extensive set of loudness and Dynamic Range Control metadata that is compliant with worldwide regulations including ATSC A/85 [6].

4.1.4.1 Loudness

The loudness metadata included in the AC-4 elementary stream allows for a full range of parameters to be described, including:

- True peak and maximum true peak, as specified in Recommendation ITU-R BS.1770 [9]
- Relative gated loudness values, as specified in Recommendation ITU-R BS.1770 [9]
- Speech gated loudness values, as specified in ATSC A/85 [6]
- Dialog gating type, as specified in ETSI TS 103 190-1 [1]
- Momentary and maximum momentary loudness, as specified in Recommendation ITU-R BS.1771 [10]
- Short term and maximum short term loudness, as specified in Recommendation ITU-R BS.1771 [10]

4.1.4.2 Dynamic Range Control (DRC)

The Dynamic Range Control (DRC) elements of AC-4 provide a flexible set of DRC options to address a wide variety of device profiles and user applications. This includes legacy E-AC-3 (ATSC A/53 Part 6 [8]) DRC profiles as specified in ATSC A/85 [6] as well as custom DRC profiles that can be defined for each output mode (e.g. home theater, flat panel TV, portable speaker, portable headphone).

4.1.5 Intelligent Loudness Management

In addition to the carriage of a richer set of loudness metadata beyond dialnorm, AC-4 incorporates a means to verify that the loudness information carried in the AC-4 bit stream is correct. The system can use this means to signal to devices after decoding, that the loudness metadata is correct and no further loudness processing is needed. This behavior protects the audio from further processing that could degrade audio quality. The AC-4 encoder incorporates a real-time loudness leveler that can be dynamically enabled if the incoming loudness metadata cannot be validated.

4.1.6 Target Devices

AC-4 supports device-specific metadata to optimize rendering based on the output-device characteristics. Target-device metadata can optionally enable conditional authoring and rendering based on output speaker configuration. This feature gives content creators artistic flexibility in creating an optimal sounding mix for all output speaker configurations without the compromises of downmixing.

4.1.7 Alternative Metadata

Alternative metadata supplements existing object metadata to allow different renditions of the same object to be created for each Presentation. Alternative metadata can also be defined for each target device.

4.1.8 Advanced Single-Stream and Multi-Stream (Hybrid) Presentations

AC-4 enables advanced single-stream Presentations by carrying multiple Audio Program Components in a single AC-4 bit stream. This allows all Audio Program Components of a single Presentation, as well as components of multiple Presentations, to be carried within a single AC-4 bit stream.

Hybrid delivery involves transport of one Audio Program Component over a first path, such as a broadcast path, and one or more Audio Elements over a second path, such as broadband (Internet) or an alternate physical layer pipe. AC-4 supports advanced multi-stream Presentations to enable hybrid-delivery use cases as defined in ATSC A/331 [11].

4.1.9 Core and Full Decoding

The AC-4 decoder has two decoding modes: core decoding and full decoding. The core decoding mode enables a low-complexity decode of a complete audio Presentation for devices with limited output capabilities (e.g., mobile devices, tablets, televisions, etc.). The full decoding mode enables decoding of a complete audio Presentation for devices with expanded output capabilities (e.g., Audio/Video Receiver). The choice of decoding modes enables a single bit stream to be compatible with a wide range of device types and applications.

4.1.10 High Sampling Frequencies

AC-4 supports high sampling frequencies of 96 kHz and 192 kHz. However, Part 1 of this standard [13] constrains sampling frequency to 48 kHz. The AC-4 bit stream is structured such that bit streams with high sampling frequencies can be decoded to PCM at 48 kHz without any penalties. This feature minimizes the complexity of decoders that do not need to support high sampling frequencies.

4.1.11 Seamless Audio Switching

DASH allows transitions between Representations within the same Adaptation Set in order to optimize playback quality for changing network conditions. AC-4 enables seamless switching between AC-4 streams of the same media content with the following types of configuration changes:

- Bit rate changes
- Channel mode changes

- 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)

4.2 High-Level AC-4 Structure

AC-4 supports the carriage of multiple audio Presentations in a single bit stream. Presentation information for each Presentation includes instructions for selecting and mixing Audio Program Components.

The AC-4 TOC (table of contents) contains a list of one or more Presentations that are carried in the stream. Presentations consist of Substream groups where each Substream group has a specific role in the user experience: Music and Effects (M&E), Dialog, Associated Audio etc. Substream groups can be shared between Presentations so that parts common to several Presentations do not need to be transmitted twice.

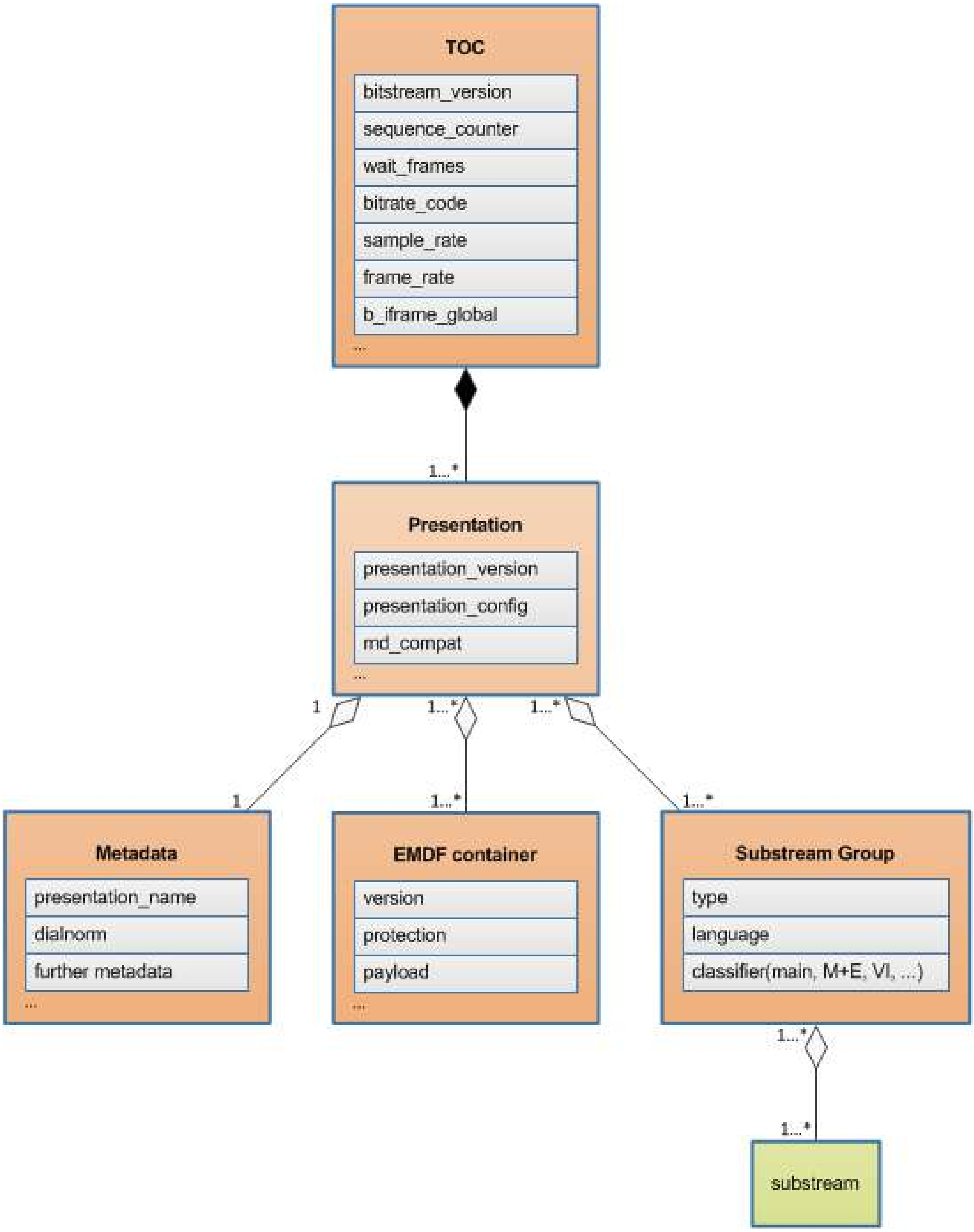

Figure 4.2 High level bit stream structure.

A Presentation informs the decoder which parts of an AC-4 stream are intended to be decoded simultaneously at a given point in time. As such, Presentations describe available user experiences. Features such as loudness and dialog enhancement are therefore managed by the Presentation.

Figure 4.3 shows an example TOC with several Presentations for M and E (shown as "Music & Effects" in the drawing below) with different language dialog Substreams. The selected Presentation contains the 5.1 M and E Substream, as well as an English dialog Substream.

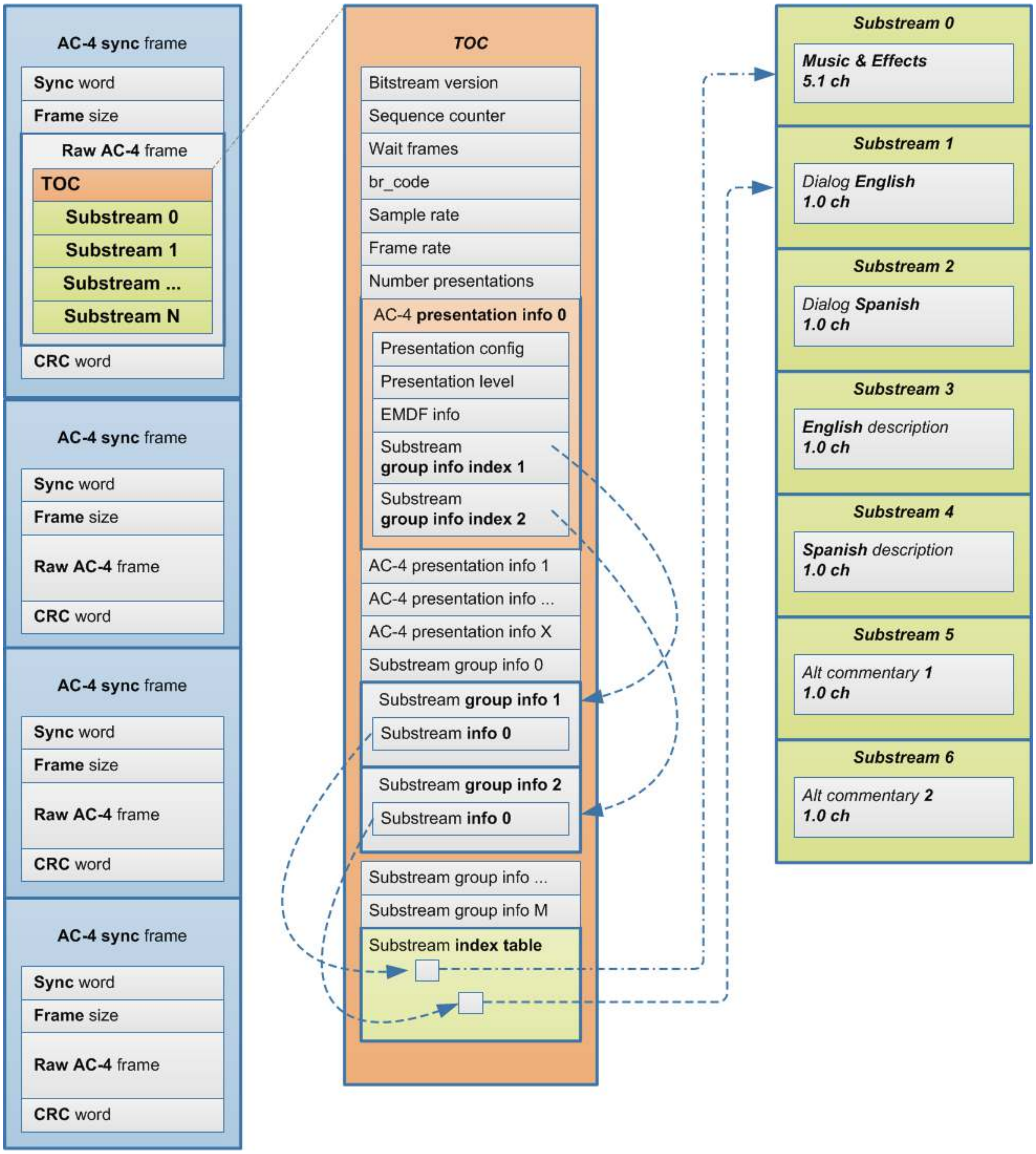

Figure 4.3 Example of complex frame with several Presentations and Substream Groups.

Refer to ETSI TS 103 190-2 [2] Section "4.5.1 Bitstream Structure" for details.

5. AC-4 SPECIFICATION

This section comprises the constraints for Elementary Stream components carrying audio encoded with the AC-4 codec.

5.1 Bit Stream Syntax

Refer to ETSI TS 103 190-2 [2] Section "6.0 Bitstream Syntax" for the full definition of the AC-4 bit stream syntax.

5.2 General Constraints on AC-4 coding

AC-4 audio elementary streams shall comply with the syntax and semantics contained in ETSI TS 103 190-2 [2], as well as the parts of ETSI TS 103 190-1 [1] referenced by ETSI TS 103 190-2 [2], as constrained in the following sections.

5.2.1 Bit Stream Constraints

AC-4 bit streams shall be constrained as follows:

- The value of bitstream_version shall be set to 2.
- The value of fs_index shall be set to 1, which indicates a base sampling frequency of 48 kHz.
- The value of b_sf_multiplier shall be set to 0, which indicates a sampling frequency equal to the base sampling frequency of 48 kHz.
- The stream frame rate shall be from the following set: 48000/2048, 24000/1001, 24, 25, 30000/1001, 30, 50, 60000/1001, 60, 100, 120/1.001, 120 Hz. This set corresponds to all video frame rates specified in ATSC A/341 [12] and additionally includes the native AC-4 frame rate that does not correspond to a video frame rate.
- When using 100, 120/1.001 or 120Hz stream frame rate, the *efficient high frame rate* (EHFR) mode shall be used, except for temporary switches to high frame rates for I-Frame insertion. Note: Using the *efficient high frame rate* (EHFR) mode circumvents the inherent bitrate penalty when increasing the audio frame rate.
- Primary Bit Streams, distributed over broadcast, shall contain at least one Presentation with an md_compat element value that is less than or equal to 3. This Presentation, along with all the Substreams it references, shall be contained in the Primary Bit Stream.
- The value of all reserved elements shall be 0. Note that this is different from the ATSC default value for reserved bits as stated in Section 3.2.1 of this document.
- The maximum AC-4 frame size shall be limited to the maximum overall frame size, as shown in Table 5.1.

Table 5.1 Maximum Frame Sizes

Frame Rate Index	Frame Rate [1/s]	Max Single Presentation Frame Size (for md_compat=3) [byte]	Max Overall Frame Size [byte]
0	24000/1001	7994	127904
1	24	7986	127776
2	25	7666	122656
3	30000/1001	6390	102240
4	30	6386	102176
5	48000/1001	3990	63840
6	48	3986	63776
7	50	3826	61216
8	60000/1001	3186	50976
9	60	3186	50976
10	100	1906	30496
11	120000/1001	1586	25376
12	120	1586	25376
13	48000/2048	8178	130848

5.2.2 Presentation Constraints

AC-4 Presentations shall be constrained as follows:

- The value of the presentation_version element shall be greater than or equal to 1.
- The value of the presentation_config element shall remain constant within a single program.
- For every presentation, presentation_id shall be present in every AC-4 frame; each presentation shall have a presentation_id unique across all presentations contained in the ac4_toc( ).
- The presentations shall be sorted in the ac4_toc() by their preference, i.e., the presentation with highest priority shall be the first presentation.
- The presentation_id values shall be in ascending (but not necessarily consecutive) order.
- The maximum bitrate shall be limited to 1521 kbps for a single Presentation with md_compat less than or equal to 3.
- When the presentation_name element is serialized into multiple chunks, the length of the full Presentation name shall be less than or equal to 255 bytes.
- If a Presentation Substream has b_pres_ndot value set to 1, then the following metadata elements shall be included and set in the Presentation Substream:
    - loro_centre_mixgain and loro_surround_mixgain, unless default values as specified in ETSI TS 103 190-1 [1] Table 144 are used.
    - ltrt_centre_mixgain and ltrt_surround_mixgain, unless default values as specified in ETSI TS 103 190-1 [1] Section 4.3.12.2.12 are used.
    - cdmx_parameters, unless defaults based on ETSI TS 103 190-2 [2] Table 153 are used.
    - all applicable downmix loudness correction gains, unless default values as specified in ETSI TS 103 190-2 [2] Section 4.8.5.3 are used.
    - preferred_dmx_method, unless the default value of "not indicated" is used.
    - if b_alternative is set to 1, then all target specific metadata.
    - dialog panning metadata, unless default loudspeaker position is used.
    - associate audio panning metadata, unless default loudspeaker position is used.

5.2.3 Substream Group Constraints

AC-4 Substream groups shall be constrained as follows:

- The value of content_classifier shall remain constant within a single program.
- If Program language is signaled:
  - The b_language_indicator flag shall be set to 1 in each AC-4 audio frame.
  - The b_serialized_language_tag flag shall be set to 0.
- The sequence of language_tag_bytes shall contain a language tag that conforms to the syntax and semantics defined in IETF BCP 47 [5].
- Note: The IANA language subtag registry can be found here [14].

5.2.4 Substream Constraints

AC-4 Substreams shall be constrained as follows:

- The value of sus_ver shall be greater than or equal to 1.
- For channel-based Substreams, the value of channel_mode shall remain constant within a single Program.
- For channel-based Substreams carrying associated audio, the value of channel_mode shall be set to '00' (Mono) or '10' (Stereo).
- For channel-based Substreams carrying dialog, the value of channel_mode shall be set to '00' (Mono), '10' (Stereo), or '1100' (3.0).
- If b_audio_ndot is 1, then the following metadata shall be included and set in the Substream:
  - dialog panning metadata, unless default loudspeaker position is used
  - pre-encoding downmix or upmix info if applicable
  - phase90_info_2ch or phase90_info_mc, unless the default value of "not indicated" is used
  - channels classifier if applicable

5.3 Loudness and DRC Constraints

In order to comply with ATSC A/85 [6], the following encoding constraints apply[1]:

- The dialnorm_bits parameter shall be used to indicate the measured loudness of the audio Program in LKFS[2] units, as recommended in ATSC A/85 [6].
- The value of the loud_prac_type element should be set to '0001'. If the loud_prac_type element is set to '0001':
  - The value of the b_loudcorr_type flag shall be set to 0, if the audio Program has been corrected with an infinite look-ahead (file-based). If the loudness correction was based on a combination of real-time loudness measurement and dynamic range compression, the flag shall be set to 1.
  - The value of the b_loudcorr_dialgate flag shall be set to 0 when the dialnorm_bits element represents the relative-level gated loudness of the Program, and shall be set to 1 when the dialnorm_bits element represents the speech/dialog Program loudness.

[1] For regions with loudness regulations that use a loudness practice type other than ATSC A/85 [6], these parameters can be set according to local loudness regulations.

[2] The LKFS unit is equivalent to a decibel and is defined in ITU-R BS.1770 [9].

- If the b_loudcorr_dialgate flag value is set to 1:
  - The loudrelgat element shall be used to signal the measurement of the integrated loudness of the audio Program, as specified in Recommendation ITU-R BS.1770 [9], without loudness or DRC metadata applied.
  - The loudspchgat element shall be used to signal the measurement of the dialog/speech loudness of the audio Program, as specified in ITU-R Recommendation BS.1770 [9], without loudness or DRC metadata applied.

All four default DRC modes, indicated by drc_decoder_mode_id values 0 to 3 as defined in ETSI TS 103 190-1 [1] Section 4.3.13.3.1, shall be included in the AC-4 elementary stream.

5.4 Personalized Audio Constraints

The following parameters shall be present in the bit stream:

- Substream groups that contain dialog shall include content_type information, with b_language_indicator set to 1, indicating the presence of Program language.
- If Alternative Presentations exist (b_alternative flag is equal to 1):
  - The first target, by order of appearance, shall have a target level that is equal to or less than the Presentation decoder compatibility level (target_level <= md_compat).
  - The first target, by order of appearance shall include the stereo speaker target device category.
- Audio Emergency Information shall be inserted in the AC-4 elementary stream by setting the content_classifier of the associated audio substream group to '010' ("visually impaired") and placing this presentation before the corresponding AD presentation.

5.5 Constraints for Advanced Multi-Stream Presentations

The following constraints apply to bit streams that include advanced multi-stream Presentations:

- The b_multi_pid element shall be set to 1, if Presentations include additional Audio Elements that are contained in a separate elementary stream (such as for "hybrid" delivery).
- For multi-stream transmission, Presentations in the primary and in secondary elementary streams shall be marked as matching, by signaling the same presentation_id element value in the respective ac4_presentation_v1_info element in the TOCs of the elementary streams.

5.6 ISO Base Media File Format Packaging Rules

This section describes how AC-4 bit streams shall be packetized and signaled in the ISO Base Media File Format (ISOBMFF).

5.6.1 Signaling an AC-4 Bit Stream in ISO Base Media File Format

The basic structures defined within ISO/IEC 14496-12 [7] to identify audio Tracks shall be used with specific extensions to provide detailed information on the characteristics of an AC-4 stream. The AC4SampleEntry Box shall be included in the Sample Description Box according to ETSI TS 103 190-2 [2], Annex E. The box type of the AC4SampleEntry Box shall be 'ac-4'.

5.6.2 Additional Constraints on AC-4 Elementary Streams

ETSI TS 103 190-2 [2], Annex E, describes the constraints that shall be applied to the AC-4 bit stream for use within the ISOBMFF. For the use of DASH with the ISOBMFF, the following additional constraints shall apply when packaging AC-4 audio into Representations:

- The value of the frame_rate_index parameter shall remain constant within each Representation.
- The value of the presentation_config parameter shall remain constant within each Representation.
- The values of the channel_mode parameters shall remain constant within each Representation.
- The values of the content_classifier parameters shall remain constant within each Representation.
- AC-4 access units should be encoded with the same frame rate as the associated video frame rate.
- AC-4 access units should be encoded temporally aligned with the video access units from the corresponding video to ensure continuous alignment of video and audio access units in order to utilize the features of A/V alignment.
- AC-4 I-Frames should be placed temporally aligned with the I-Frames of the video to enable synchronous switching.

5.6.3 Packaging of AC-4 Bit Streams into ISO Base Media File Format Segments

Each AC-4 raw frame shall be packaged as an ISOBMFF sample. For more information refer to ETSI TS 103 190-2 [2], Annex E.3. AC-4 sync frames as described in ETSI TS 103 190-2 [2], Annex C, shall not be packetized directly into ISOBMFF.

5.6.4 Random Access Point and Stream Access Point

A random access point shall be signaled by means of the file format sync sample box 'stss' in ISOBMFF or by means of the respective flag in the Track run box in case of ISO base media segments.

An AC-4 elementary stream contains I-Frames (Stream Access Point type = 1) to indicate random access points. A seamless switch can be accomplished only at a random access point.

- The first sample in an ISOBMFF file shall be a random access point.
- In cases of fragmented ISOBMFF files, the first sample of each Segment and Subsegment shall be a random access point.
- A Segment or Subsegment with AC-4 may contain more than one random access point.

5.6.5 Packaging of Individual Audio Elements

In scenarios where AC-4 Representations from different Adaptation Sets form an AC-4 Presentation, the following constraints shall apply:

- The duration of corresponding Segments and Subsegments of those Representations shall be identical.
- The Segments and Subsegments of Representations from different Adaptation Sets shall be temporally aligned, i.e., the beginning and the end of Segments from Representations in different Adaptation Sets forming an AC-4 Presentation shall have the same time stamps.
- The value of the frame_rate_index parameter of each Representation shall be identical.

– End of Document –