SMPTE STANDARD

Format for Non-PCM Audio and Data in AES3 — MPEG-4 AAC and HE AAC Compressed Digital Audio in ADTS and LATM / LOAS Wrappers

Table of Contents **Page**

3 Barker Avenue, White Plains, NY 10601
(914) 761-1100

Approved
August 2, 2010

Foreword

SMPTE (the Society of Motion Picture and Television Engineers) is an internationally-recognized standards developing organization. Headquartered and incorporated in the United States of America, SMPTE has members in over 80 countries on six continents. SMPTE's Engineering Documents, including Standards, Recommended Practices, and Engineering Guidelines, are prepared by SMPTE's Technology Committees. Participation in these Committees is open to all with a bona fide interest in their work. SMPTE cooperates closely with other standards-developing organizations, including ISO, IEC and ITU.

SMPTE Engineering Documents are drafted in accordance with the rules given in Part XIII of its Administrative Practices.

SMPTE ST 2041-3 was prepared by Technology Committee 32NF.

Intellectual Property

At the time of publication no notice had been received by SMPTE claiming patent rights essential to the implementation of this Standard. However, attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. SMPTE shall not be held responsible for identifying any or all such patent rights.

Introduction

This section is entirely informative and does not form an integral part of this Engineering Document.

The MPEG Committee of ISO/IEC has produced a number of different audio compression technologies. Each of these is nominally an “emission” codec, which means that its design does not include consideration for maintaining audio quality in multiple decode – re-encode cycles. As a result, many operators will desire to keep the original compressed bitstream intact while routing, switching, and when other baseband manipulations are done with associated video signals.

The first of these codecs, MPEG-1 Layer I and II audio (documented in ISO/IEC 11172-3) and MPEG-2 Layers I, II, and III audio (documented in ISO/IEC 13818-3) were intended to be bit-wise compatible within Layers I and II, while MPEG-2 Layer III used “Non-Backwards Compatible” extensions. MPEG-2 Layer III audio has not been widely used in broadcast, however some recent extensions to it have been added via ISO/IEC 14496-3 Subpart 9 which may result in wider use.

The second of these codecs, MPEG-2 AAC (documented in ISO/IEC 13818-7) is intentionally non-compatible with Layer I, II, or III, and is known as “Advanced Audio Coding.” A second set of documents, which extended MPEG-2 AAC were documented as a part of MPEG-4 in ISO/IEC 14496-3. Although MPEG-4 AAC was intended to be backwards compatible with MPEG-2 AAC, no assumption should be made that an MPEG-2 decoder can decode an MPEG-4 AAC bitstream (largely due to the different transport wrappers). When the SBR tool was added to AAC, creating “HE AAC,” it was documented in Amendments to both ISO/IEC 13818-7 and ISO/IEC 14496-3.

It should be noted that there are two alternative “wrappers” standardized for AAC/HE AAC, one called ADTS (widely used in Japanese Digital Broadcasting) and the other known as LATM/LOAS, since the MPEG-4 (ISO/IEC 14496-3) AAC codecs introduced new features and capabilities that require a transport format which can signal their contents. In order to be able to pass the audio bitstream without the necessity to do partial decoding to locate flags the wrapper was devised. LATM/LOAS is specified as the wrapper for MPEG-4 AAC/HE AAC audio streams within DVB transport stream broadcasting applications. As a result, this suite of SMPTE standards will document carriage of both.

The following suite of SMPTE standards defines the carriage of MPEG compressed audio bitstreams within an AES3 carrier bitstream:

SMPTE ST 2041-1, Format for Non-PCM Audio in AES3 – MPEG-1/MPEG-2 Layers I, II, and III Audio

SMPTE ST 2041-2, Format for Non-PCM Audio in AES3 – MPEG-2 AAC/HE AAC Audio in ADTS

SMPTE ST 2041-3, Format for Non-PCM Audio and Data in AES3 – MPEG-4 AAC and HE AAC Compressed Digital Audio in ADTS and LATM/LOAS Wrappers

The bitstreams defined in this standard can be carried independently of video as AES3 bitstreams or embedded into SDI or HD-SDI bitstreams in the normal manner specified by other SMPTE standards.

1 Scope

This Standard specifies the method of packing MPEG-4 AAC and MPEG-4 HE AAC data rate reduced audio streams contained in either an ADTS or an LATM/LOAS wrapper into an AES3 transport stream. It also specifies the content of the data_type_dependent field of the burst_info (Pc) word of the Synchronization preamble used to signal the wrapper type and other necessary information about the payload. The Standard specifies the repetition rate, reference point, decoding latency and use of Pause data bursts in interrupted streams. This Standard supports the MPEG-4 AAC Profile, the MPEG-4 High Efficiency AAC Profile, and the MPEG-4 High Efficiency AAC v2 Profile. This Standard is limited to carriage of a single audio elementary stream.

2 Conformance Notation

Normative text is text that describes elements of the design that are indispensable or contains the conformance language keywords: "shall", "should", or "may". Informative text is text that is potentially helpful to the user, but not indispensable, and can be removed, changed, or added editorially without affecting interoperability. Informative text does not contain any conformance keywords.

All text in this document is, by default, normative, except: the Introduction, any section explicitly labeled as "Informative" or individual paragraphs that start with "Note:”

The keywords "shall" and "shall not" indicate requirements strictly to be followed in order to conform to the document and from which no deviation is permitted.

The keywords, "should" and "should not" indicate that, among several possibilities, one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required; or that (in the negative form) a certain possibility or course of action is deprecated but not prohibited.

The keywords "may" and "need not" indicate courses of action permissible within the limits of the document.

The keyword “reserved” indicates a provision that is not defined at this time, shall not be used, and may be defined in the future. The keyword “forbidden” indicates “reserved” and in addition indicates that the provision will never be defined in the future.

Unless otherwise specified, the order of precedence of the types of normative information in this document shall be as follows: Normative prose shall be the authoritative definition; Tables shall be next; followed by formal languages; then figures; and then any other language forms.

3 Normative References

The following standards contain provisions which, through reference in this text, constitute provisions of this recommended practice. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this recommended practice are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

AES3-2009, AES Standard for Digital Audio Engineering — Serial Transmission Format for Two-Channel Linearly Represented Digital Audio Data

ISO/IEC 13818-7:2006, Information Technology — Generic Coding of Moving Pictures and Associated Audio Information — Part 7: Advanced Audio Coding (AAC)

ISO/IEC 14496-3, Information Technology — Coding of Audio-Visual Objects — Part 3: Audio

SMPTE 337-2008, Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface

SMPTE ST 338:2010, Format for Non-PCM Audio and Data in AES3 — Data Types

SMPTE 339-2008, Television — Format for Non-PCM Audio and Data in AES3 — Generic Data Types

SMPTE RP 168-2009, Definition of Vertical Interval Switching Point for Synchronous Video Switching

4 Definitions and Acronyms

4.1 Definitions

4.1.1 AAC LC – AAC Profile

The MPEG-4 AAC Profile is the counterpart to the MPEG-2 AAC Low Complexity Profile, In contrary to the MPEG-2 AAC LC Profile the MPEG-4 AAC Profile enables the usage of an additional tool. This profile is used when there are restrictions on the usage of RAM and processing complexity.

4.1.2 Access Unit

Smallest entity to which timing information can be attributed. An access unit is the smallest individually decodable unit. A decoder consumes access units.

4.1.3 adts_error_check( )

An MPEG-4 ADTS element that carries a CRC value which provides protection for the headers and some of the payload.

4.1.4 AudioMuxElement(1)

An MPEG-4 LATM element that carries payload data for at least one audio elementary stream, related payload length information and multiplex configuration information. This element carries payload data in form of PayloadMux elements. The number in brackets indicates multiplexing configuration (StreamMuxConfig()) is multiplexed into AudioMuxElement(), i.e. in-band transmission.

4.1.5 HE AAC – High Efficiency AAC Profile

This Profile utilizes the Spectral Band Replication (SBR) tool in conjunction with AAC. This Profile is a superset of the AAC Profile. For further information please refer to ISO/IEC 14496-3.

4.1.6 HE AAC v2 – High Efficiency AAC v2 Profile

This Profile utilizes the Spectral Band Replication (SBR) tool and the Parametric Stereo (PS) tool in conjunction with AAC. This Profile is a superset of the High Efficiency AAC Profile. For further information please refer to ISO/IEC 14496-3.

4.1.7 MPEG Surround

A technology used for coding of multichannel signals based on a downmixed signal of the original multichannel signal, and associated spatial parameters. Defined by ISO/IEC 23003-1

4.1.8 PayloadMux

Payload data chunk in an AudioMuxElement that contains potentially multiplexed payload data for multiple audio elementary streams. In general PayloadMux elements can be concatenated inside AudioMuxElements.

4.1.9 StreamMuxConfig

Configuration structure that describes the structure of the LATM payload multiplex.

4.1.10 Sub-data-type

Reference to the type of payload of the data-bursts defined for the use with the specified data-type

4.1.11 Video Sync Point

Signal Alignment Point as defined by Annex A of RP 168.

4.2 Acronyms

ADTS: Audio Data Transport Stream, a wrapper structure defined in ISO/IEC 13818-7 and ISO/IEC 14496-3, consisting of a fixed header, a variable header, an optional error check (CRC_check) and a specified number of raw_data_blocks( ). The fixed header of the ADTS contains the syncword plus all parts of the header which are necessary for decoding and which do not change from frame to frame. The variable header of the ADTS contains header data which changes from frame to frame. The ADTS only supports a raw_data_stream( ) with only one program. The program can have up to 7 channels plus an independently switched coupling channel.

LATM: Low overhead Audio Transport Multiplex. A multiplexing layer defined by ISO/IEC 14496-3, used for multiplexing of audio elementary streams. This Standard is limited to transport of a single audio elementary stream.

LOAS: Low Overhead Audio Stream. A synchronization layer defined by ISO/IEC 14496-3. Provides 3 types of formats which are specifically suited for being used on different type of underlying transmission layers depending on their characteristics.

MDCT: Modified Discrete Cosine Transformation

SBR: Spectral Band Replication

5 MPEG-4 AAC Data in an ADTS Wrapper

When MPEG-4 AAC coded audio is wrapped in ADTS transport syntax it shall be placed into AES3 data streams as specified below.

5.1 Overview

MPEG-4 AAC coded audio shall be transported in an AES3 data stream as a series of Data Bursts. Each Data Burst shall start with a Burst Preamble as defined by SMPTE 337, containing information about the Burst Payload, which shall follow the Burst Preamble. The Burst Payload shall consist of an AAC ADTS Frame. The Burst Payload shall be followed by enough padding words (which shall be PCM zeros, or digital silence) to make the resulting Data Burst duration exactly match that of either 960 or 1024 samples of baseband (PCM) audio that the coded audio represents.

The resulting Data Bursts shall be placed in the audio sample word/aux data fields of AES3 subframes at regular intervals in either the frame or subframe mode (see SMPTE 337, Section 5). Data Bursts shall be placed in the AES3 transport, using either 16, 20, or 24 bits of the available data space. While the 24-bit mode allows more efficient use of the AES3 capacity, the 16- and 20-bit modes allow use with existing equipment limited to 16- or 20-bit operation.

A single AAC ADTS Frame shall form the Burst Payload, as shown by Figure 1. As specified by ISO/IEC 14496-3, each ADTS Frame begins with an Audio Data Transport Stream (ADTS) fixed header, followed by a Variable Header, and an optional ADTS error check (CRC_check) word, followed by the Raw Data Block of AAC coded audio that represents either 960 or 1024 samples of baseband (PCM) audio.

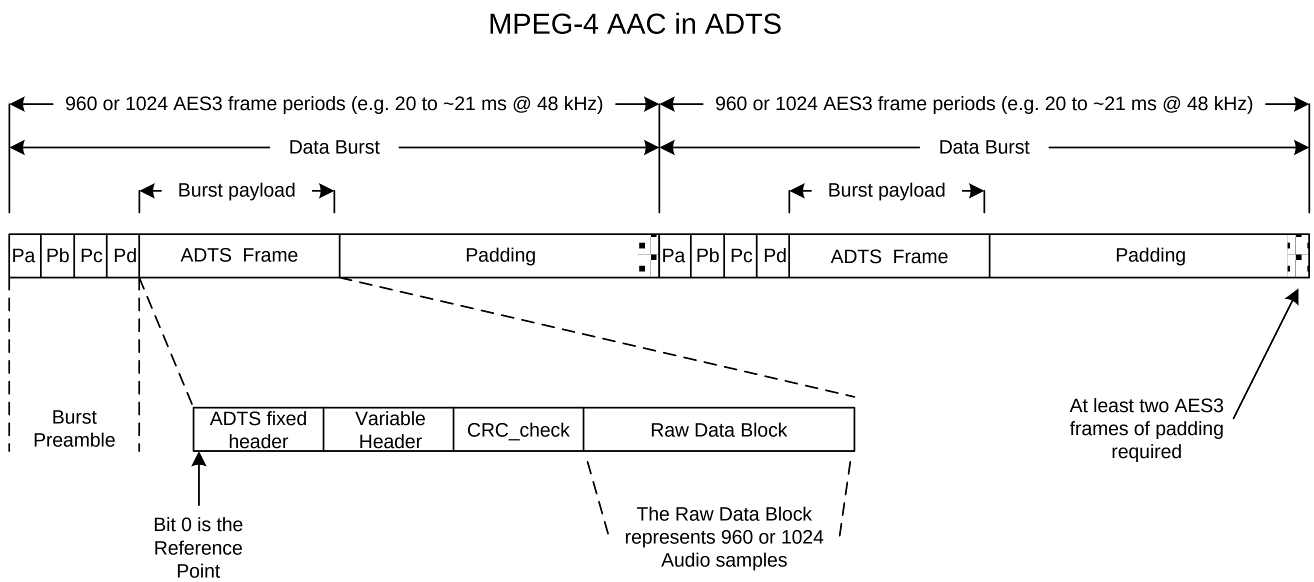

Figure 1 – MPEG-4 AAC audio data in an ADTS wrapper, transported in an AES3 stream

5.2 Burst_preamble

The Pc word (burst_info value) of the burst_preamble carries the data_type identifier, the data_type_dependent and the data_stream_number information (see SMPTE 337, Table 7)

5.2.1 data_type identifier

The data_type identifier shall be set to 10, indicating that the audio is coded as an MPEG-4 AAC stream.

5.2.2 data_type_dependent

The values of the data_type_dependent bits shall be as shown in Table 1, and as described below.

Table 1 – Values of data_type_dependent field for MPEG-4 AAC in an ADTS wrapper

data_type_dependent bit number	Meaning
0 - 1	Reserved, must be set to '01'
2	0 => Data Burst length of 1024 AES3 frames 1 => Data Burst length of 960 AES3 frames
3	0 => Parametric Stereo is not used 1 => Parametric Stereo is used
4	0 => MPEG Surround is not used 1 => MPEG Surround is used

data_burst_length – Bit 2 shall either be set to 0 to indicate that the Data Burst is 1024 AES3 frames long or set to 1 to indicate that the Data Burst 960 AES3 frames long.

parametric_stereo – Bit 3 shall be set to 0 to indicate that Parametric Stereo is not present, as use of the PS tool implies presence of SBR which is covered by Section 6.

mpeg_surround - Bit 4 shall be set to 0 to indicate that MPEG Surround is not present. As the ADTS header does not provide any mechanism to explicitly signal the presence of MPEG Surround data, such signaling may be either done outside the scope of this standard or by partially decoding the MPEG-4 AAC raw data to see if MPEG Surround data is present.

5.2.3 data_stream_number

The data_stream_number may be set to any number from 0h to 6h. . The value should be 0 (zero) indicating only one program is carried in the AES3 stream. 7h is a reserved value. See SMPTE 337, clause 7.1.3.5.

5.3 AAC Burst Payload

The MPEG-4 AAC encoder produces a stream of Raw Data Blocks, as defined by ISO/IEC 14496-3. ADTS fixed and variable headers and an optional ADTS error check (CRC_check) word as defined by ISO/IEC 14496-3 shall be prepended to the Raw Data Block. Each Raw Data Block shall contain the data that represents either 1024 or 960 audio samples of all the audio channels in a single program. The length of the Raw Data Block depends on the encoded bit rate.

5.4 AES3 Frame Rate (Sampling Frequency)

The frame rate of the AES3 stream used to transport the AAC ADTS Frames shall be the same as the rate at which the encoded audio was sampled.

5.5 AAC Reference Point

The Reference Point of an AAC Burst Payload shall be bit 0 of the Burst Payload, as shown in Figure 1.

5.6 AAC Standard Repetition Rate

AAC Burst Payloads occur at the standard Repetition Rate if the Reference points for consecutive data bursts (in the same data stream number) occur either 1024 or 960 (the indicated data_block_length) AES3 frames apart.

5.7 AAC Decoding Latency (Professional)

A reference decoder shall output the first PCM sample of the decoded audio exactly two Data Burst periods after the first bit of the first Data Burst is received by the decoder.

Note: The decoding latency of two Data Burst periods does not include the encoding latency. The encoding latency needs to be added to the decoding latency when calculating the total delay of the audio system.

5.8 AAC Reference Position

The Reference Position of a Burst Payload is defined by the relationship of the decoded audio to an associated video signal. A Burst Payload is in the Reference Position when the decoded audio from that Burst Payload is in sync with the associated video.

The Reference Point of the Burst Payload carried in an AES3 stream whose sampling frequency is locked to the associated video signal shall therefore precede the Video Sync Point by two Data Burst periods.

5.9 Use of Pause Data Bursts to Fill Gaps in AAC Data Streams

Pause data-bursts (ref SMPTE 339, clause 6) should be used to fill gaps in the AAC data stream. The Pa word of the first Pause data burst transmitted after the last valid AAC data burst shall occur one (1) Data Burst period after the Pa word of the preceding AAC data burst. The Pause data-bursts should be transmitted with a repetition period of 64 AES3 frames, unless other repetition periods are necessary to precisely fill the stream gap (whose length might not be a integer multiple of 64 AES3 frames), subject to the zero padding requirements of SMPTE 337, clause 7.3.

6 MPEG-4 HE AAC Data in an ADTS Wrapper

When MPEG-4 HE AAC coded audio is wrapped in ADTS transport syntax it shall be placed into AES3 data streams as specified below.

6.1 Overview

MPEG-4 HE AAC or HE AAC v2 (HE AAC with the addition of the Parametric Stereo coding tool) coded audio shall be transported in an AES3 data stream as a series of Data Bursts. Each Data Burst shall start with a Burst Preamble as defined by SMPTE 337, containing information about the Burst Payload, which shall follow the Burst Preamble. The Burst Payload shall consist of an HE AAC ADTS Frame. The Burst Payload shall be followed by enough padding words (which shall be PCM zeros, or digital silence) to make the resulting Data Burst duration exactly match that of either 1920 or 2048 samples of baseband (PCM) audio that the coded audio represents.

The resulting Data Bursts shall be placed in the audio sample word/aux data fields of AES3 subframes at regular intervals in either the frame or subframe mode (see SMPTE 337, Section 5). Data Bursts shall be placed in the AES3 transport, using either 16, 20, or 24 bits of the available data space. While the 24-bit mode allows more efficient use of the AES3 capacity, the 16- and 20-bit modes allow use with existing equipment limited to 16- or 20-bit operation.

A single HE AAC ADTS Frame shall form the Burst Payload, as shown by Figure 2. As specified by ISO/IEC 14496-3, each ADTS Frame begins with an Audio Data Transport Stream (ADTS) fixed header, followed by a Variable Header, and an optional ADTS error check (CRC_check) word, followed by the Raw Data Block of AAC coded audio that represents either 1920 or 2048 samples of baseband (PCM) audio.

The ADTS header does not have any provision for explicitly signaling the presence of SBR data in the Raw Data block. The only way of setting the data_type field of the Pc word in the Burst Preamble correctly is to decode enough of the Raw Data block to determine if there is any SBR data present, and thus to differentiate between AAC and HE AAC coded audio. The same considerations apply to determining the correct setting for the data_type_dependent field parameter values.

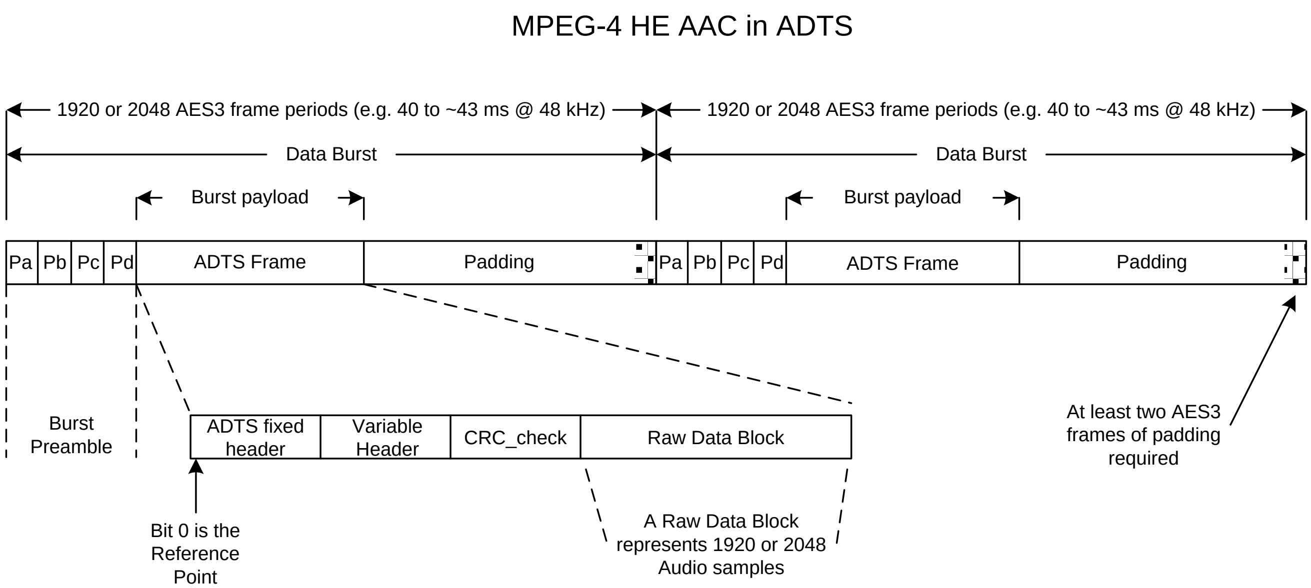

Figure 2 – MPEG-4 HE AAC audio data in an ADTS wrapper, transported in an AES3 stream

6.2 Burst_preamble

The Pc word (burst_info value) of the burst_preamble carries the data_type identifier, the data_type_dependent and the data_stream_number information (see SMPTE 337, Table 7)

6.2.1 data_type identifier

The data_type identifier shall be set to 11, indicating that the audio is coded as an MPEG-4 HE AAC stream.

6.2.2 data_type_dependent

The values of the data_type_dependent bits shall be as shown in Table 2, and as described below.

Table 2 – Values of data_type_dependent field for MPEG-4 HE AAC in an ADTS wrapper

data_type_dependent bit number	Meaning
0 - 1	Reserved, must be set to ‘01’
2	0 => Data Burst length of 2048 AES3 frames 1 => Data Burst length of 1920 AES3 frames
3	0 => Parametric Stereo is not used 1 => Parametric Stereo is used
4	0 => MPEG Surround is not used 1 => MPEG Surround is used

data_burst_length – Bit 2 shall either be set to 0 to indicate that the Data Burst is 2048 AES3 frames long or set to 1 to indicate that the Data Burst 1920 AES3 frames long.

parametric_stereo – Bit 3 shall be set to 0 to indicate that Parametric Stereo is not present, or to 1 to indicate that Parametric Stereo is used. As the ADTS header does not provide any mechanism to explicitly signal the presence of PS data, such signaling may be either done outside the scope of this standard or by partially decoding the MPEG-4 AAC raw data to see if PS data is present.

mpeg_surround - Bit 4 shall be set to 0 to indicate that MPEG Surround is not present or set to 1 to indicate that MPEG Surround is present. As the ADTS header does not provide any mechanism to explicitly signal the presence of MPEG Surround data, such signaling may be either done outside the scope of this standard or by partially decoding the MPEG-4 AAC raw data to see if MPEG Surround data is present.

6.2.3 data_stream_number

The data_stream_number may be set to any number from 0h to 6h. . The value should be 0 (zero) indicating only one program is carried in the AES3 stream. 7h is a reserved value. See SMPTE 337, clause 7.1.3.5.

6.3 HE AAC Burst Payload

The MPEG-4 HE AAC encoder produces a stream of Raw Data Blocks, as defined by ISO/IEC 14496-3. ADTS fixed and variable headers and an optional ADTS error check (CRC_check) word as defined by ISO/IEC 14496-3 shall be prepended to the Raw Data Block. Each Raw Data Block shall contain the data that represents either 2048 or 1920 audio samples of all the audio channels in a single program. The length of the Raw Data Block depends on the encoded bit rate.

6.4 AES3 Frame Rate (Sampling Frequency)

The frame rate of the AES3 stream used to transport the HE AAC ADTS Frames shall be the same as the rate at which the encoded audio was sampled.

6.5 HE AAC Reference Point

The Reference Point of an HE AAC Burst Payload shall be bit 0 of the Burst Payload, as shown in Figure 2.

6.6 HE AAC Standard Repetition Rate

HE AAC Burst Payloads occur at the standard Repetition Rate if the Reference points for consecutive data bursts (in the same data stream number) occur either 2048 or 1920 (the indicated data_block_length) AES3 frames apart.

6.7 HE AAC Decoding Latency (Professional)

A reference decoder shall output the first PCM sample of the decoded audio exactly two Data Burst periods after the first bit of the first Data Burst is received by the decoder.

Note: The decoding latency of two Data Burst periods does not include the encoding latency. The encoding latency needs to be added to the decoding latency when calculating the total delay of the audio system.

6.8 HE AAC Reference Position

The Reference Position of a Burst Payload is defined by the relationship of the decoded audio to an associated video signal. A Burst Payload is in the Reference Position when the decoded audio from that Burst Payload is in sync with the associated video. The Reference Point of the Burst Payload carried in an AES3 stream whose sampling frequency is locked to the associated video signal shall therefore precede the Video Sync Point by two Data Burst periods.

6.9 Use of Pause Data Bursts to Fill Gaps in HE AAC Data Streams

Pause data-bursts (ref SMPTE 339, clause 6) should be used to fill gaps in the HE AAC data stream. The Pa word of the first Pause data burst transmitted after the last valid HE AAC data burst shall occur one (1) Data Burst period after the Pa word of the preceding HE AAC data burst. The Pause data-bursts shall be transmitted with a repetition period of 64 AES3 frames, unless other repetition periods are necessary to precisely fill the stream gap (whose length might not be an integer multiple of 64 AES3 frames), subject to the zero padding requirements of SMPTE 337, clause 7.3.

7 MPEG-4 AAC Data in an LATM/LOAS Wrapper

When MPEG-4 AAC coded audio is wrapped in LATM/LOAS transport syntax it shall be placed into AES3 data streams as specified below.

7.1 Overview

MPEG-4 AAC-coded audio shall be transported in an AES3 data stream as a series of Data Bursts. Each Data Burst shall start with a Burst Preamble as defined by SMPTE 337, containing information about the Burst Payload, which shall follow the Burst Preamble. The Burst Payload shall be followed by enough padding words (which shall be PCM zeros, or digital silence) to make the resulting Data Burst duration exactly match that of either 960 or 1024 samples of baseband (PCM) audio that the coded audio represents, as shown by Figure 3.

The resulting Data Bursts shall be placed in the audio sample word/aux data fields of AES3 subframes at regular intervals in either the frame or subframe mode (see SMPTE 337, Section 5). Data Bursts shall be placed in the AES3 transport, using either 16, 20, or 24 bits of the available data space. While the 24-bit mode allows more efficient use of the AES3 capacity, the 16- and 20-bit modes allow use with existing equipment limited to 16- or 20-bit operation.

The Burst Payload shall consist of a 3 byte LOAS header, followed by an AudioMuxElement. As specified by ISO/IEC 14496-3, an AudioMuxElement begins with a flag, followed by a Stream Configuration element which might be present in all or just some of the AudioMuxElements, depending on data rate requirements. The Mux Payload is the next element. It begins with a length code followed by an Audio Access Unit carrying AAC coded audio that represents either 960 or 1024 samples of baseband (PCM) audio. Padding bits are appended to the Mux Payload to ensure that the AudioMuxElement is byte aligned.

Only the following LATM/LOAS configurations are supported:

The numSubFrames parameter of the LATM StreamMuxConfig shall be 0.

The numProgram parameter of the LATM StreamMuxConfig shall be 0.

The numLayer parameter of the LATM StreamMuxConfig shall be 0 except for audio streams where MPEG Surround (MPS) data is embedded within the payload of the first LATM layer, and this MPS data is explicitly signaled in the second LATM layer. In such cases the presence of a second layer in LATM frames is allowed and therefore numLayer shall be 1, indicating 2 layers. In this configuration there exists no payload associated with the second LATM layer and therefore the payload size indication for the second layer in LATM shall be set to zero.

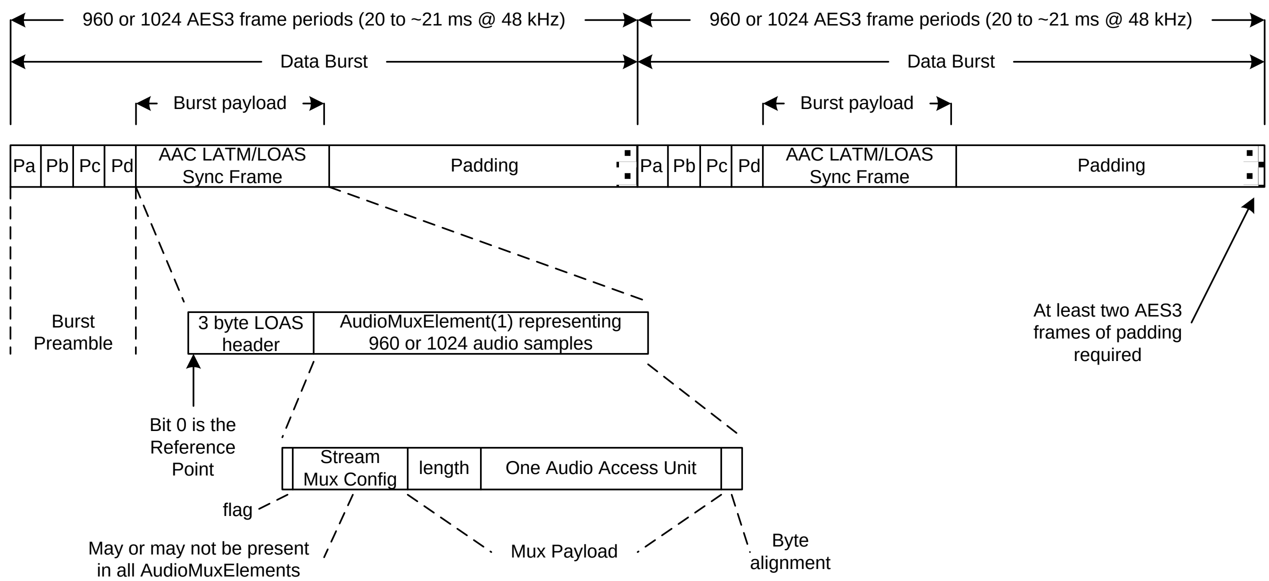

Figure 3 – MPEG-4 AAC audio data in an LATM/LOAS wrapper, transported in an AES3 stream

7.2 Burst_preamble

The Pc word (burst_info value) of the burst_preamble carries the data_type identifier, the data_type_dependent and the data_stream_number information (see SMPTE 337, Table 7)

7.2.1 data_type identifier

The data_type identifier shall be set to 10, indicating that the audio is coded as an MPEG-4 AAC stream.

7.2.2 data_type_dependent

The values of the data_type_dependent bits shall be as shown in Table 3, and as described below.

Table 3 – Values of data_type_dependent field for MPEG-4 AAC in an LATM/LOAS wrapper

data_type_dependent bit number	Meaning
0 - 1	Reserved, must be set to ‘00’
2	0 => Data Burst length of 2048 AES3 frames 1 => Data Burst length of 1920 AES3 frames
3	0 => Parametric Stereo is not used 1 => Parametric Stereo is used
4	0 => MPEG Surround is not used 1 => MPEG Surround is used

data_burst_length – Bit 2 shall either be set to 0 to indicate that a 1024 word long transform is used or set to 1 to indicate that a 960 word long transform is used.

parametric_stereo – Bit 3 shall be set to 0 to indicate that Parametric Stereo is not present.

mpeg_surround - Bit 4 shall be set to 0 to indicate that MPEG Surround is not present, or to 1 to indicate that MPEG Surround is used.

7.2.3 data_stream_number

The data_stream_number shall be set to any number from 0h to 6h. . The value should be 0 (zero) indicating only one program is carried in the AES3 stream. 7h is a reserved value. See SMPTE 337, clause 7.1.3.5.

7.3 AAC Burst Payload

The Burst Payload is made up of a 3 byte LOAS header, followed by an AudioMuxElement. The AudioMuxElement carries a Stream Configuration element which may be present in all or just some of the AudioMuxElements, depending on data rate requirements. A length code and an Audio Access Unit carrying AAC coded audio that represents either 960 or 1024 samples of the baseband (PCM) audio for a single program, make up the Mux Payload. The Mux Payload is followed by padding bits to ensure that the AudioMuxElement is byte aligned. The length of the Sync Frame depends on the encoded bit rate.

7.4 AES3 Frame Rate (Sampling Frequency)

The frame rate of the AES3 stream used to transport the AAC Sync Frames shall be the same as the rate at which the encoded audio was sampled.

7.5 AAC Reference Point

The Reference Point of an AAC Burst Payload shall be bit 0 of the Burst Payload, as shown in Figure 3.

7.6 AAC Standard Repetition Rate

AAC Burst Payloads occur at the standard Repetition Rate if the Reference points for consecutive data bursts (in the same data stream number) occur either 1024 or 960 (the indicated data_block_length) AES3 frames apart.

7.7 AAC Decoding Latency (Professional)

A reference decoder shall output the first PCM sample of the decoded audio exactly two Data Burst periods after the first bit of the first Data Burst is received by the decoder.

Note: The decoding latency of two Data Burst periods does not include the encoding latency. The encoding latency needs to be added to the decoding latency when calculating the total delay of the audio system.

7.8 AAC Reference Position

The Reference Position of a Burst Payload is defined by the relationship of the decoded audio to an associated video signal. A Burst Payload is in the Reference Position when the decoded audio from that Burst Payload is in sync with the associated video. The Reference Point of the Burst Payload carried in an AES3 stream whose sampling frequency is locked to the associated video signal shall therefore precede the Video Sync Point by two Data Burst periods.

7.9 Use of Pause Data Bursts to Fill Gaps in AAC Data Streams

Pause data-bursts (ref SMPTE 339) should be used to fill gaps in the AAC data stream. The Pa word of the first Pause data burst transmitted after the last valid AAC Data Burst shall occur one (1) Data Burst period after the Pa word of the preceding AAC data burst. The Pause data-bursts shall be transmitted with a repetition period of 64 AES3 frames, unless other repetition periods are necessary to precisely fill the stream gap (whose length may not be an integer multiple of either 3 or 64 AES3 frames), subject to the zero padding requirements of SMPTE 337, clause 7.3.

8 MPEG-4 HE AAC Data in an LATM/LOAS Wrapper

When MPEG-4 AAC or HE AAC v2 (HE AAC with the addition of the Parametric Stereo tools) coded audio is wrapped in LATM/LOAS transport syntax it shall be placed into AES3 data streams as specified below.

8.1 Overview

MPEG-4 HE AAC or HE AAC v2 coded audio shall be transported in an AES3 data stream as a series of Data Bursts. Each Data Burst shall start with a Burst Preamble defined by SMPTE 337, containing information about the Burst Payload, which shall follow the Burst Preamble. The Burst Payload shall be followed by enough padding words (which shall be PCM zeros, or digital silence) to make the resulting Data Burst duration exactly match that of either 1920 or 2048 samples of baseband (PCM) audio that the coded audio represents, as shown by Figure 4.

The resulting Data Bursts shall be placed in the audio sample word/aux data fields of AES3 subframes at regular intervals in either the frame or subframe mode (see SMPTE 337, Section 5). Data Bursts shall be placed in the AES3 transport, using either 16, 20, or 24 bits of the available data space. While the 24-bit mode allows more efficient use of the AES3 capacity, the 16- and 20-bit modes allow use with existing equipment limited to 16- or 20-bit operation.

The Burst Payload shall consist of a 3 byte LOAS header, followed by an AudioMuxElement. As specified by ISO/IEC 14496-3, an AudioMuxElement begins with a flag, followed by a Stream Configuration element which might be present in all or just some of the AudioMuxElements, depending on data rate requirements. The Mux Payload is the next element. It begins with a length code followed by an Audio Access Unit carrying HE AAC coded audio that represents either 1920 or 2048 samples of baseband (PCM) audio. Padding bits are appended to the Mux Payload to ensure that the AudioMuxElement is byte aligned.

Only the following LATM/LOAS configurations are supported:

The numSubFrames parameter of the LATM StreamMuxConfig shall be 0.

The numProgram parameter of the LATM StreamMuxConfig shall be 0.

The numLayer parameter of the LATM StreamMuxConfig shall be 0 except for audio streams where MPEG Surround (MPS) data is embedded within the payload of the first LATM layer, and this MPS data is explicitly signaled in the second LATM layer. In such cases the presence of a second layer in LATM frames is allowed and therefore numLayer shall be 1, indicating 2 layers. In this configuration there exists no payload associated with the second LATM layer and therefore the payload size indication for the second layer in LATM shall be set to zero.

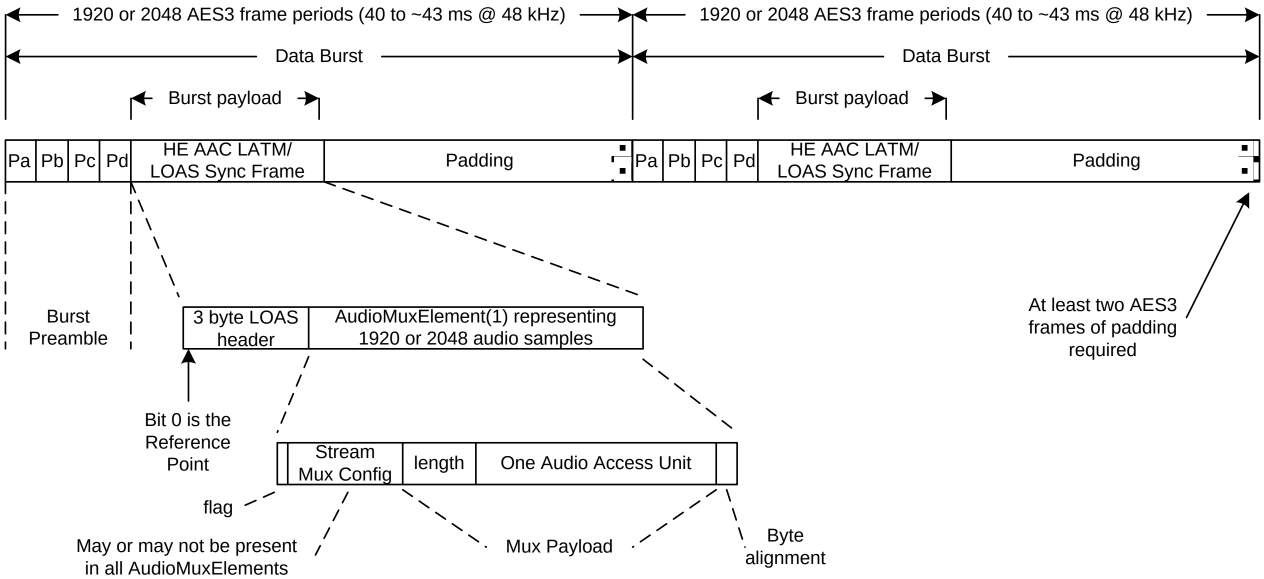

Figure 4 – MPEG-4 HE AAC audio data in an LATM/LOAS wrapper, transported in an AES3 stream

8.2 Burst_preamble

The Pc word (burst_info value) of the burst_preamble carries the data_type identifier, the data_type_dependent and the data_stream_number information (see SMPTE 337, Table 7)

8.2.1 data_type identifier

The data_type identifier shall be set to 11, indicating that the audio is coded as an MPEG-4 HE AAC stream.

8.2.2 data_type_dependent

The values of the data_type_dependent bits shall be as shown in Table 4, and as described below.

Table 4 – Values of data_type_dependent field for MPEG-4 HE AAC in an LATM/LOAS wrapper

data_type_dependent bit number	Meaning
0 - 1	Reserved, must be set to '00'
2	0 => Data Burst length of 2048 AES3 frames 1 => Data Burst length of 1920 AES3 frames
3	0 => Parametric Stereo is not used 1 => Parametric Stereo is used
4	0 => MPEG Surround is not used 1 => MPEG Surround is used

data_burst_length – Bit 2 shall either be set to 0 to indicate that a 2048 word long transform is used or set to 1 to indicate that a 1920 word long transform is used.

parametric_stereo – Bit 3 shall be set to 0 to indicate that Parametric Stereo is not present, or to 1 to indicate that Parametric Stereo is used.

mpeg_surround - Bit 4 shall be set to 0 to indicate that MPEG Surround is not present, or to 1 to indicate that MPEG Surround is used.

8.2.3 data_stream_number

The data_stream_number shall be set to any number from 0h to 6h. . The value should be 0 (zero) indicating only one program is carried in the AES3 stream. 7h is a reserved value. See SMPTE 337

8.3 HE AAC Burst Payload

The MPEG-4 HE AAC encoder produces a stream of Audio Access Units for a single audio program, as defined by ISO/IEC 14496-3. Each Access Unit is combined with a length code to form the Payload Mux of an AudioMuxElement. Stream Mux Configuration data is prepended to at least a fraction of the Payload Mux data blocks to form (with a flag bit and byte alignment bits) an AudioMuxElement, as defined by ISO/IEC 14496-3. A LOAS header, as defined by ISO/IEC 14496-3, is prepended to the AudioMuxElement to form the Burst Payload, or HE AAC LATM/LOAS Sync Frame. Each HE AAC Sync Frame shall contain the data that represents either 2048 or 1920 audio samples of all the audio channels in a single program. The length of the Sync Frame depends on the encoded bit rate.

8.4 AES3 Frame Rate (Sampling Frequency)

The frame rate of the AES3 stream used to transport the HE AAC Sync Frames shall be the same as the rate at which the encoded audio was sampled.

8.5 HE AAC Reference Point

The Reference Point of an HE AAC Burst Payload shall be bit 0 of the Burst Payload, as shown in Figure 4.

8.6 HE AAC Standard Repetition Rate

HE AAC Burst Payloads occur at the standard Repetition Rate if the Reference points for consecutive data bursts (in the same data stream number) occur either 2048 or 1920 (the indicated data_block_length) AES3 frames apart.

8.7 HE AAC Decoding Latency (Professional)

A reference decoder shall output the first PCM sample of the decoded audio exactly two Data Burst periods after the first bit of the first Data Burst is received by the decoder.

Note: The decoding latency of two Data Burst periods does not include the encoding latency. The encoding latency needs to be added to the decoding latency when calculating the total delay of the audio system.

8.8 HE AAC Reference Position

The Reference Position of a Burst Payload is defined by the relationship of the decoded audio to an associated video signal. A Burst Payload is in the Reference Position when the decoded audio from that Burst Payload is in sync with the associated video. The Reference Point of the Burst Payload carried in an AES3 stream whose sampling frequency is locked to the associated video signal shall therefore precede the Video Sync Point by two Data Burst periods.

8.9 Use of Pause Data Bursts to Fill Gaps in HE AAC Data Streams

Pause data-bursts (ref SMPTE 339) should be used to fill gaps in the HE AAC data stream. The Pa word of the first Pause data burst transmitted after the last valid HE AAC Data Burst shall occur one (1) Data Burst period after the Pa word of the preceding HE AAC data burst. The Pause data-bursts shall be transmitted with a repetition period of 64 AES3 frames, unless other repetition periods are necessary to precisely fill the stream gap (whose length may not be an integer multiple of either 3 or 64 AES3 frames), subject to the zero padding requirements of SMPTE 337, clause 7.3.

Annex A Bibliography (Informative)

IEC 60958, Digital Audio Interface

IEC 61937-1 (2007-01), Digital Audio — Interface for Non-Linear PCM Encoded Audio Bitstreams Applying IEC 60958 — Part 1: General

IEC 61937-2 (2007-05), Digital Audio — Interface for Non-Linear PCM Encoded Audio Bitstreams Applying IEC 60958 — Part 2: Burst-info

IEC 61937-4 (2003-05), Digital Audio — Interface for Non-Linear PCM Encoded Audio Bitstreams Applying IEC 60958 — Part 4: Non-Linear PCM Bitstreams According to the MPEG Audio Format

IEC 61937-6 (2006-01), Digital Audio — Interface for Non-Linear PCM Encoded Audio Bitstreams Applying IEC 60958 — Part 6: Non-Linear PCM Bitstreams According to the MPEG-2 AAC and MPEG-4 AAC Audio Formats

IEC 61937-11 (2010-05), Digital Audio — Interface for Non-Linear PCM Encoded Audio Bitstreams Applying IEC 60958 — Part 11: MPEG-4 AAC and its extensions in LATM/LOAS

ISO/IEC 11172-3:1993, Information Technology — Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1,5 Mbit/s - Part 3: Audio

ISO/IEC 13818-1 | ITU-T H.222.0, Information Technology — Generic Coding of Moving Pictures and Associated Audio Information: Systems

ISO/IEC 13818-3:1998, Information Technology — Generic Coding of Moving Pictures and Associated Audio Information — Part 3: Audio

ISO/IEC 23003-1:2007, Information Technology — MPEG Audio Technologies — Part 1: MPEG Surround

SMPTE 272M-2004 Television – Formatting AES/EBU Audio and Auxiliary Data into Digital Video Ancillary Data Space

SMPTE 276M-1995, Television — Transmission of AES-EBU Digital Audio Signals Over Coaxial Cable

SMPTE 299-1-2009, 24-Bit Digital Audio Format for SMPTE 292 Bit-Serial Interface (Note: Document previously numbered SMPTE 299-2009 – Content unchanged)