

SCTE • ISBE[®]

S T A N D A R D S

Digital Video Subcommittee

SCTE STANDARD

SCTE 243-4 2018

**Next Generation Audio Carriage for Cable Systems:
Part 4 – DTS-UHD Audio Carriage Constraints**

NOTICE

The Society of Cable Telecommunications Engineers (SCTE) Standards and Operational Practices (hereafter called “documents”) are intended to serve the public interest by providing specifications, test methods and procedures that promote uniformity of product, interchangeability, best practices and ultimately the long term reliability of broadband communications facilities. These documents shall not in any way preclude any member or non-member of SCTE from manufacturing or selling products not conforming to such documents, nor shall the existence of such standards preclude their voluntary use by those other than SCTE members.

SCTE assumes no obligations or liability whatsoever to any party who may adopt the documents. Such adopting party assumes all risks associated with adoption of these documents, and accepts full responsibility for any damage and/or claims arising from the adoption of such documents.

Attention is called to the possibility that implementation of this document may require the use of subject matter covered by patent rights. By publication of this document, no position is taken with respect to the existence or validity of any patent rights in connection therewith. SCTE shall not be responsible for identifying patents for which a license may be required or for conducting inquiries into the legal validity or scope of those patents that are brought to its attention.

Patent holders who believe that they hold patents which are essential to the implementation of this document have been requested to provide information about those patents and any related licensing terms and conditions. Any such declarations made before or after publication of this document are available on the SCTE web site at <http://www.scte.org>.

All Rights Reserved

© Society of Cable Telecommunications Engineers, Inc. 2018
140 Philips Road
Exton, PA 19341

Table of Contents

Title	Page Number
NOTICE	2
1 Introduction	4
1.1 Scope	4
2 Normative References	4
2.1 SCTE References	4
2.2 Standards from Other Organizations	4
3 Informative References	4
3.1 SCTE References	5
3.2 Standards from Other Organizations	5
4 Compliance Notation	5
5 Abbreviations and Definitions	5
5.1 Abbreviations	5
5.2 Definitions	5
6 DTS-UHD Compressed MPEG-2 Audio Transport	6
6.1 Overview	6
6.2 MPEG-2 Systems Signaling	6
6.2.1 Buffering Model	6
6.2.2 DTS-UHD Audio Descriptor	6
6.2.3 Semantics for Parameters	7
6.2.4 LongDescriptor Parameters	9
6.3 Audio Preselection Descriptor	12
6.3.1 Descriptor Usage	12
6.3.2 Parameters derived from the BroadcastChunk	12
6.4 DTS-UHD PES Packet Encapsulation	13
6.4.1 Stream Type	13
6.4.2 Stream ID	13
6.4.3 Audio Access Unit Alignment in PES packets	13
6.4.4 Random Access Points	14
6.4.5 Multiple Stream Requirements	14
7 DTS-UHD in ISO Media Files for delivery using MPEG DASH	14
7.1 Introduction	14
7.2 DASH Media Presentation Description (MPD) parameters for DTS-UHD	14
7.3 Design Rules for DTS-UHD Tracks	16
7.3.1 Overview	16
7.3.2 Track Header Box	16
7.3.3 Sync Sample Box	17
7.3.4 Handler Reference Box	17
7.3.5 Sound Media Header Box	17
7.3.6 Sample Description Box	17
7.3.7 AudioSampleEntry Box for DTS-UHD Formats	17
7.3.8 DTSUHDSampleEntry	17
7.3.9 DTSUHDSpecificBox	18
7.3.10 Additional Semantics for DTSUHDSpecificBox	19
7.4 Storage of DTS-UHD elementary streams	19
7.4.1 Overview	19
7.4.2 Random Access Points	19
7.4.3 Restrictions on DTS-UHD Formats	20

1 Introduction

1.1 Scope

This document is part four of a multi-part standard that specifies carriage constraints of Next Generation Audio (NGA) codecs in MPEG-2 Transport Stream and in MPEG ISO-BMFF media segments. In conjunction with ANSI/SCTE 243-1 [1], this document defines the carriage of DTS-UHD audio in MPEG-2 Transport Stream and MPEG DASH using ISO BMFF media segments.

2 Normative References

The following documents contain provisions, which, through reference in this text, constitute provisions of this document. At the time of Subcommittee approval, the editions indicated were valid. All documents are subject to revision; and while parties to any agreement based on this document are encouraged to investigate the possibility of applying the most recent editions of the documents listed below, they are reminded that newer editions of those documents might not be compatible with the referenced version.

2.1 SCTE References

- [1] ANSI/SCTE 243-1 2017, Next Generation Audio Carriage Constraints for Cable Systems: Part 1 – Common Transport Signaling
- [2] ANSI/SCTE 242-1 2017, Next Generation Audio Coding Constraints for Cable Systems: Part 1 – Introduction and Common Constraints
- [3] SCTE 242-4 2018, Next Generation Audio Coding Constraints for Cable Systems: Part 4 – DTS-UHD Audio Coding Constraints

2.2 Standards from Other Organizations

- [4] ISO/IEC 13818-1:2015, Information Technology – Generic coding of moving pictures and associated audio information: Systems
- [5] ATSC Standard A/342-1:2017, A/342 Part 1, Audio Common Elements
- [6] ISO/IEC 14496-12:2015, Information technology - Coding of audio-visual objects - Part 12: ISO Base Media File Format
- [7] ETSI TS 103 491 V1.1.1 (2017-04), DTS-UHD Audio Format; Delivery of Channels, Objects and Ambisonic Sound Fields
- [8] ETSI TS 103 584 V1.1.1 (2018-01), DTS-UHD Point Source Renderer
- [9] DVB BlueBook A038 (2017-12), Digital Video Broadcasting (DVB); Specification for Service Information (SI) in DVB systems (Final draft of ETSI EN 300 468 v 1.16.1)
- [10] IETF RFC 6381 The 'Codecs' and 'Profiles' Parameters for "Bucket" Media Types
- [11] DASH IF: “Guidelines for Implementation: DASH-IF Interoperability Points for ATSC 3.0, Version 1.0,”, January 31, 2016. <http://dashif.org/wp-content/uploads/2017/02/DASH-IF-IOP-for-ATSC3-0-v1.0.pdf>

3 Informative References

The following documents might provide valuable information to the reader but are not required when complying with this document.

3.1 SCTE References

[12] ANSI/SCTE 54:2015, Digital Video Service Multiplex and Transport Subsystem Standard for Cable Television.

3.2 Standards from Other Organizations

[13] IETF RFC4151 The ‘tag’ URI Scheme

4 Compliance Notation

shall	This word or the adjective “ <i>required</i> ” means that the item is an absolute requirement of this document.
shall not forbidden	This phrase means that the item is an absolute prohibition of this document. This word means the value specified shall never be used.
should	This word or the adjective “ <i>recommended</i> ” means that there may exist valid reasons in particular circumstances to ignore this item, but the full implications should be understood and the case carefully weighted before choosing a different course.
should not	This phrase means that there may exist valid reasons in particular circumstances when the listed behavior is acceptable or even useful, but the full implications should be understood and the case carefully weighed before implementing any behavior described with this label.
may	This word or the adjective “ <i>optional</i> ” means that this item is truly optional. One vendor may choose to include the item because a particular marketplace requires it or because it enhances the product, for example; another vendor may omit the same item.
deprecated	Use is permissible for legacy purposes only. Deprecated features may be removed from future versions of this document. Implementations should avoid use of deprecated features.

5 Abbreviations and Definitions

5.1 Abbreviations

SCTE	Society of Cable Telecommunications Engineers
LFE	Low Frequency Effects

5.2 Definitions

This specification uses the definitions defined in ANSI/SCTE 242-1[2] and, by incorporation, ATSC A/342-1[5]. The following terms have definitions specific to DTS-UHD and shall apply to all clauses in this document.

Audio Frame – unit of coded audio that, when decoded, will generate defined number of uncompressed Linear PCM audio samples for each wave form

Audio Stream – sequence of synchronized Audio Frames.

BroadcastChunk - block of data within an audio stream containing data that maps audio components to preselections as defined in SCTE 242-4[3]

Frame Duration –time represented by one decoded Audio Frame

non-sync frame – Audio Frame that may require prior frames for correct decode (refer to SCTE 242-4 [3])

Object – Audio Element as defined in ATSC A/342-1[2] and referenced in ANSI/SCTE 242-1[1]

Object Group - selected collection of audio objects to be played together

Presentation - selected collection of Channels, or Objects and Object Groups used together to generate the rendered output

sync frame – Audio Frame that is independently decodable (refer to SCTE 242-4 [3]).

6 DTS-UHD Compressed MPEG-2 Audio Transport

6.1 Overview

The following sections specify the signaling of DTS-UHD audio component in MPEG programs.

6.2 MPEG-2 Systems Signaling

6.2.1 Buffering Model

It is recommended that the main audio buffer size (BS_n , as defined in ISO/IEC 13818-1 [4] clause 2.4.2) has a fixed value of 66,434 bytes.

6.2.2 DTS-UHD Audio Descriptor

The DTS-UHD Audio Descriptor shall be included in the program map section following the relevant ES_info_length field for any DTS-UHD audio stream component coded in accordance with TS 103 491 [7] that is included in the MPEG program.

The syntax for the DTS-UHD Audio Descriptor shall be as defined in Table 1 below. All values presented in the DTS-UHD descriptor are interpreted as unsigned integer values, most significant bit first.

Table 1: DTS-UHD descriptor

Syntax	size (bits)	mnemonic	Reference
dts-uhd_descriptor {	8	bslbf	
descriptor_tag	8	uimsbf	
descriptor_length	8	uimsbf	
if (descriptor_tag = 0x7F)	8	uimsbf	6.2.3.1
descriptor_tag_extension (8)			
DecoderProfileCode	6	uimsbf	6.2.3.2
FrameDurationCode	2	bslbf	6.2.3.3
MaxPayloadCode	3	bslbf	6.2.3.4
ExtendedDescriptor	1	bslbf	
LongDescriptor	1	bslbf	6.2.3.6
StreamIndex	3	bslbf	6.2.3.7
if (LongDescriptor) {			

Syntax	size (bits)	mnemonic	Reference
NumPresentationsCode	5	uimsbf	6.2.4.1
ChannelMask	32	bslbf	6.2.4.2
BaseSamplingFreqCode	1	bslbf	6.2.4.3
SampleRateMod	2	bslbf	6.2.4.4
RepresentationType	3	bslbf	6.2.4.5
for (np=0; np<NumPresentations; np++) { IDTagPresent[np] }	NumPresentations	uimsbf	6.2.4.6
ByteAlign	0-7	bslbf	6.2.4.7
for (np=0; np<NumPresentations; np++) { if (IDTagPresent[np]) { PresentationIDTag[np] } // end of If IDTagPresent } // end of ID tag loop } // end of if LongDescriptor	128		6.2.4.8
if (ExtendedDescriptor) { ByteCount	6	uimsbf	
reserved_bits	2	bslbf	reserved_bits
ExtendedPayloadBytes[ByteCount] }	ByteCount * 8	bslbf	6.2.3.9
}			

6.2.3 Semantics for Parameters

6.2.3.1 descriptor_tag

The value of `descriptor_tag` shall be 0x7F and the value of `descriptor_tag_extension` shall be set to 0x21, as assigned for use in DVB, ATSC and SCTE.

6.2.3.2 DecoderProfileCode

This field indicates the DTS-UHD decoder profile required to decode this stream. Valid range of *DecoderProfileCode* shall be from 0 to 63. The relationship between the *DecoderProfile* and *DecoderProfileCode* shall be as follows:

$$DecoderProfile = DecoderProfileCode + 2$$

where *DecoderProfile* indicates the DTS-UHD decoder profile as defined in TS 103 491 [7]

For DTS-UHD NGA streams, *DecoderProfile* shall be 3 or greater.

Note: When *DecoderProfile* = 2, the `audio_preselection_descriptor` shall not be present.

6.2.3.3 FrameDurationCode

FrameDurationCode indicates frame duration according to Table 2. The units for Frame Duration are sample periods relative to Base Sampling Frequency defined below.

Table 2: Frame Duration

FrameDurationCode	Frame Duration (samples)
0	512
1	1024
2	2048
3	4096

6.2.3.4 MaxPayloadCode

MaxPayloadCode indicates the maximum size of the audio payload, as indicated in Table 3. Note that *Maxpayload* is not the size of the largest audio frame in the presentation, but rather a "not to exceed" value for buffer configuration and digital audio interface purposes, and is inclusive of all required preambles, headers, burst spacing, etc.

Table 3: MaxPayload

MaxPayloadCode	MaxPayload (bytes)
0	2048
1	4096
2	8192
3	16384
4	32768
5	65536
6	131072
7	reserved

Note that *MaxPayloadCode* can be used to determine the minimum IEC sample clock frequency required to transfer the DTS-UHD stream on a digital audio interface.

6.2.3.5 ExtendedDescriptor

ExtendedDescriptor is a 1-bit flag to be interpreted as a Boolean. If *ExtendedDescriptor* == 1, then additional private information bytes beyond this descriptor definition shall exist at the end of this definition. *ExtendedPayloadBytes* bytes shall be included in *descriptor_length*.

6.2.3.6 LongDescriptor

LongDescriptor is a 1-bit flag to be interpreted as a Boolean. For SCTE systems carrying NGA audio, when the *audio_preselection_descriptor()* is present, *LongDescriptor* shall be 0. For legacy channel based audio and when *audio_preselection_descriptor()* is not used, *LongDescriptor* may be set to 1. When *LongDescriptor* is set to 1 the *LongDescriptor* parameters shall be correctly populated as described in 6.2.4 below.

6.2.3.7 StreamIndex

When a given Audio Program is delivered using multiple streams, StreamIndex is the index used to indicate stream priority for prioritizing mixing metadata. The main stream shall have StreamIndex=0, while auxiliary streams shall have StreamIndex values in the range of 1-7.

If the Audio Program is delivered in a single stream, StreamIndex shall be 0.

The value of StreamIndex shall correspond to the value of StreamID carried in the DTS-UHD BroadcastChunk as defined in SCTE DVS 242-4[3].

6.2.3.8 reserved_bits

These bits serve to byte align some parameters in the descriptor and may be defined for use in future specifications. For the purposes of this specification, all *reserved_bits* shall be set to 0 and ignored by the receiver.

6.2.3.9 ExtendedPayloadBytes

The number of bytes, *ByteCount*, is determined by parsing the parameters defined here and comparing the number of bytes remaining to be parsed based on *descriptor_length*. If *ExtendedPayloadBytes* exist, receivers built to this this specification shall ignore this data.

6.2.4 LongDescriptor Parameters

The parameters described in this section are conditionally present if *LongDescriptor* [Table 1] is set to 1.

6.2.4.1 NumPresentationsCode

This field represents the number of audio presentations encoded within DTS-UHD elementary stream. The valid range of the bitstream parameter *m_ucNumAudioPres* defined in TS 103 491 [7] is from 1 to 32. The relationship between *NumPresentationsCode* and *NumPresentations* shall be as follows:

$$NumPresentations = m_ucNumAudioPres = NumPresentationsCode + 1$$

6.2.4.2 ChannelMask

A bit mask that indicates the channel layout encoded in the default presentation of the DTS-UHD bitstream according to the bit-to-channel-label mapping described in Table 4. Note that the ChannelMask table is derived from Channel Bitmask defined in TS 103 584 [8].

Table 4: Speaker Labels for ChannelMask

Label	Notation	ChannelMask
C	CENTER	0x00000001
L	LEFT	0x00000002
R	RIGHT	0x00000004
Ls	SRRD_LEFT	0x00000008
Rs	SRRD_RIGHT	0x00000010
LFE1	LFE_1	0x00000020
Cs	SRRD_CENTER	0x00000040
Lsr	REAR_SRRD_LEFT	0x00000080
Rsr	REAR_SRRD_RIGHT	0x00000100

Label	Notation	ChannelMask
Lss	SIDE_SRRD_LEFT	0x00000200
Rss	SIDE_SRRD_RIGHT	0x00000400
Lc	LEFT_CENTER	0x00000800
Rc	RIGHT_CENTER	0x00001000
Lh	HIGH_LEFT	0x00002000
Ch	HIGH_CENTER	0x00004000
Rh	HIGH_RIGHT	0x00008000
LFE2	LFE_2	0x00010000
Lw	LEFT_WIDE	0x00020000
Rw	RIGHT_WIDE	0x00040000
Oh	TOP_CENTER_SRRD	0x00080000
Lhs	HIGH_SIDE_LEFT	0x00100000
Rhs	HIGH_SIDE_RIGHT	0x00200000
Chr	HIGH_REAR_CENTER	0x00400000
Lhr	HIGH_REAR_LEFT	0x00800000
Rhr	HIGH_REAR_RIGHT	0x01000000
Cb	LOW_FRONT_CENTER	0x02000000
Lb	LOW_FRONT_LEFT	0x04000000
Rb	LOW_FRONT_RIGHT	0x08000000
Ltf	TOP_FRONT_LEFT	0x10000000
Rtf	TOP_FRONT_RIGHT	0x20000000
Ltr	TOP_REAR_LEFT	0x40000000
Rtr	TOP_REAR_RIGHT	0x80000000

6.2.4.3 BaseSamplingFrequencyCode

BaseSamplingFrequencyCode is derived from the parameter *m_unClockRateInHz* defined in TS 103 491 [7]. This field shall indicate the base sampling frequency, according to Table 5.

Table 5: Base Sampling Frequency

BaseSamplingFrequencyCode	Clock Rate (Hz) ¹
0	44100
1	48000

Consistent with the constraints in ANSI/SCTE 242-1 [2], *BaseSamplingFrequencyCode* shall equal 1.

6.2.4.4 SampleRateMod

This field indicates the sampling frequency of the audio samples stored in the bitstream. It modifies the *BaseSamplingFrequency* by multiplying by the corresponding value indicated in Table 6.

Table 6: SampleRateMod

SampleRateMod	Multiplier
0	1
1	2
2	4
3	8

Consistent with the constraints in ANSI/SCTE 242-1 [2], *SampleRateMod* shall equal 0.

6.2.4.5 RepresentationType

This value conveys general information about the type of spatial audio signals encoded in the default presentation of the DTS-UHD bit-stream according to the description in Table 7. This parameter shall match the bitstream metadata parameter *ucObjRepresTypeIndex* defined in TS 103 491 [7].

Allowed values for *RepresentationType* are affected by and may affect *ChannelMask*, as indicated.

Table 7: Representation Type

RepresentationType	Description
0	Multi-channel representation in layout described by <i>ChannelMask</i>
1	Multi-channel representation in layout described by <i>ChannelMask</i> obtained by rendering 2D content, (i.e. no height channels), with spatial resolution higher than indicated by the encoded layout, (e.g. LtRt, 5.1ES)
2	Multi-channel representation in layout described by <i>ChannelMask</i> obtained by rendering 3D content, (i.e. includes height), with spatial resolution higher than indicated by the encoded layout, (e.g. 5.1 NeoX),
3	Binaurally processed audio (2 waveforms, L, R for headphones)
4	Ambisonic representation
5	Audio tracks with or without an associated mixing matrix to a particular channel mask based output layout
6	3D Objects based representation
7	Combination of objects with different representation types

When *RepresentationType* has a value of 0, *ChannelMask* shall be set to a value that directly corresponds to the encoded channels.

When *RepresentationType* has a value of 1 or 2, *ChannelMask* shall be set to a value that corresponds to the layout of the encoded channels into which higher spatial resolution components have been rendered (i.e. the corresponding channels that have been delivered to the audio compression engine). For example, LtRt material would have a *ChannelMask* of 0x00000006, representing the two channels of encoded audio content that may be presented at a higher spatial resolution on, for example, a 5.1-channel system.

When *RepresentationType* is equal to 3, *ChannelMask* shall be set to 0x00000006, indicating Left and Right channels are active.

When *RepresentationType* is greater than 3, *ChannelMask* shall be set to 0x00000000, indicating that the audio presentation is channel layout agnostic.

6.2.4.6 IDTagPresent

One *IDTagPresent* flag exists for each presentation in the audio track, with the total number indicated by *NumPresentations*. The *IDTagPresent* flags appear in order of the increasing value of the Audio Presentation Index (*ucAudPresIndex*) defined in ETSI TS 103 491 [7]. If *IDTagPresent* is equal to 1 for a given presentation, then a 16 byte ID for that presentation (*PresentationIDTag*) shall be present in the descriptor.

6.2.4.7 ByteAlign

A variable number of bits (from 0 to 7) shall be inserted to pad to the end of the current byte boundary.

6.2.4.8 PresentationIDTag

PresentationIDTag for a specific audio presentation shall be present only when the corresponding *IDTagPresent* flag is set to 1. This 16-byte field can be used to identify the corresponding audio presentation.

6.3 Audio Preselection Descriptor

6.3.1 Descriptor Usage

The audio preselection descriptor, defined in ETSI EN 300 468 [9], may be present as indicated in SCTE 243-1 [1]. The following table shall define the mapping of the DTS-UHD bitstream properties to parameters in the descriptor.

Table 8: DTS-UHD Audio Preselection Parameters

Field in the audio preselection descriptor	Mapping
num_preselections	The number of preselections presented in the bitstream, and constrained as indicated in section 6.2.4.1
preselection_id	points to the preselection as described in section 6.3.2.2f or the associated language
ISO_639_language_code	Corresponds to the language for the current preselection, as described in section 6.3.2.3
audio_description	Indicates whether a scene description track is available, as described in section 6.3.2.4
spoken_subtitles	Indicates whether a dialog track for the visually impaired is present, as described in section 6.3.2.5
dialogue_enhancement	Indicates whether a program with dialogue enhancement metadata is present, as described in section 6.3.2.6

6.3.2 Parameters derived from the BroadcastChunk

When the DTS-UHD BroadcastChunk (defined in SCTE 242-4 [3]) is present in the main DTS-UHD stream, the corresponding parameters in the audio_preselection_descriptor (as defined in Table 8) shall be consistent with the values in the BroadcastChunk, as described below.

6.3.2.1 num_preselections

This is the number of Preselections being defined in this descriptor. At least 1 preselection is required. If all Preselections defined in the BroadcastChunk are defined in the audio_preselection_descriptor, then the total number of Preselections can be calculated as:

$$num_preselections = \sum_{i=0}^{numLanguages} (numSelectionSets[i] + 1)$$

6.3.2.2 preselection_id

preselection_id identifies a particular DTS-UHD Program as follows:

if *LanguageIndex* = 0

$$Preselection_id = ProgramIndex + 1;$$

If *LanguageIndex* > 0

$$Preselection_id = ProgramIndex + 1 + \sum_{i=0}^{LanguageIndex-1} (numSelectionSets[i] + 1)$$

where *LanguageIndex* and *ProgramIndex* are defined in Table 2.

6.3.2.3 ISO_639_language_code

If this value is present, it shall equal *ISO639_code* in the BroadcastChunk for the corresponding *preselection_id*.

6.3.2.4 audio_description

The value of *audio_description* shall be equal to that of *AudioDescription* in the corresponding SelectionSet indicated by *preselection_id*.

6.3.2.5 spoken_subtitles

The value of *spoken_subtitles* shall be equal to that of *SpokenSubtitles* in the corresponding SelectionSet indicated by *preselection_id*.

6.3.2.6 dialogue_enhancement

The value of *dialogue_enhancement* shall be equal to that of *DialogueEnhancement* in the corresponding SelectionSet indicated by *preselection_id*.

6.4 DTS-UHD PES Packet Encapsulation

6.4.1 Stream Type

For DTS-UHD transport streams used in SCTE systems, *stream_type* shall be set to 0x06 indicating stream is carrying a private data type in PES packets. The mapping of values of PID to *stream_type* is indicated in the transport stream PMT.

6.4.2 Stream ID

All DTS-UHD elementary streams shall use a *stream_id* of 0xBD, indicating private stream 1, in accordance with ISO/IEC 13818-1 [4]. Multiple DTS streams may share the same value of *stream_id* since each stream is carried with a unique PID value.

6.4.3 Audio Access Unit Alignment in PES packets

DTS-UHD access units shall be delivered in the MPEG-TS in playback order. Multiple access units are permitted in a PES packet. Audio access units shall begin with a valid DTS sync word. Valid DTS sync words are listed in Table 9.

Table 9: DTS-UHD Syncwords

Name	Syncword	Description
DTSUHD_SYNC	0x40411BF2	DTS-UHD Sync Frame
DTSUHD_NOSYNC	0x71C442E8	DTS-UHD Non-sync Frame
DTSUHD_BCHUNK	0x2A3E2523	DTS-UHD BroadcastChunk

6.4.4 Random Access Points

If a PES packet is to be used for random access to a DTS-UHD stream the following conditions shall be satisfied according to 13818-1 [4]:

Optional PES header shall be present and shall contain PTS

Data_Alignment_Indicator within the PES header shall be set to 1

The first access unit in the PES packet shall be a DTS-UHD sync frame, as indicated by the presence of DTSUHD_SYNC.

The transport packet that contains the header of a random access PES packet may have the *random_access_indicator*, if present, set to 1. Transport packets that do not meet this requirement shall have the *random_access_indicator*, if present, set to 0.

6.4.5 Multiple Stream Requirements

When a given Audio Program is delivered using multiple streams, the Random Access Points of auxiliary streams shall be aligned with the Random Access Points of the main stream.

When an Audio Program containing multiple streams is played back, the synchronized audio frames from each stream shall be delivered to the DTS-UHD playback system in sequence from lowest StreamIndex value (which contains the main DTS-UHD stream), in increasing order. This will coincide with the sequentially increasing values of StreamID as indicated in the DTS-UHD BroadcastChunk.

7 DTS-UHD in ISO Media Files for delivery using MPEG DASH

7.1 Introduction

This chapter describes the DTS-UHD audio track in relation to the ISO file and the constraints on the DTS-UHD audio formats.

7.2 DASH Media Presentation Description (MPD) parameters for DTS-UHD

This section details the DASH MPD attributes and elements used with DTS-UHD.

Table 10: DTS-UHD DASH MPD Elements and Attributes

Element or Attribute Name	Description
@codecs	<p>The value of the codecs attribute shall be created according to the syntax described in RFC 6381 [10].</p> <p>For DTS-UHD, @codecs is the associated 4cc with no additional suffix.</p> <p style="padding-left: 40px;">if DecoderProfile = 2 then @codecs = 'dtsx'</p> <p style="padding-left: 40px;">if DecoderProfile > 2 then @codecs = 'dtsy'</p>
Preselection@tag	<p>This field shall map to a selectable Presentation specified within a DTS-UHD stream by value of ucAudPresIndex, as defined in TS 103 491 [7]</p>
AdaptationSet@tag	<p>This field shall list selectable Presentations specified within a DTS-UHD stream by values of ucAudPresIndex, as defined in TS 103 491 [7]</p>
ContentComponent@tag	<p>This field shall correspond to a value of ucAudPresIndex as defined in TS 103 491 [7] for the stream indicated in the tag.</p>
AudioChannelConfiguration	<p>DTS-UHD shall use the following scheme URI following the 'tag' URI scheme defined in RFC 4151[13]:</p> <p style="padding-left: 40px;">tag:dts.com,2018:uhd:audio_channel_configuration</p> <p>and the value shall be the ChannelMask as defined in 6.2.4.2 of this specification.</p>
@audioSamplingRate	<p>The value shall equal the maximum audio sampling frequency of the decoded audio presentation.</p> <p>For SCTE this is "48000" for 48 kHz per ANSI/SCTE 243-1[1].</p>
@mimeType	<p>The MIME type to be used with DTS-UHD shall be "audio/mp4".</p>
RandomAccess	<p>The type to be used with DTS-UHD shall be "closed", i.e. the SAP type is 1.</p>
Language	<p>The language indicated should equal the language of the dialogue element as described in ANSI/SCTE 242-4, section 7.2.4.3.</p>
Role	<p>The Role@value should be set by the content author.</p>

Accessibility	<p>If the presentation includes a visual description as described in Table 8, or spoken subtitle as described in Table 8, then an Accessibility element should indicate “audio-description/visually impaired”.</p> <p>If the presentation supports dialog enhancement either through metadata presents or user interactivity, as described in Table 8, then an Accessibility element should indicate “enhanced audio intelligibility”.</p> <p>If the audio track is an emergency alert message, then the Accessibility element shall indicate “emergency”.</p>
Label	The Label for a Representation should be set by the content author.

Preselections may be signaled in the MPD via Preselection Elements or Preselection Descriptors as specified in clause 5.4.3 of DASH-IF IOP for ATSC3.0 [11]

The value of the Preselection Descriptor provides two fields, separated by a comma:

- the tag of the Preselection, which shall map to a selectable Presentation specified within a DTS-UHD stream by value of ucAudPresIndex, as defined in TS 103 491 [7]
- a white space separated list of the id(s) of the contained elements/content components of this Preselection, in processing order. The first id shall define the main element.

7.3 Design Rules for DTS-UHD Tracks

7.3.1 Overview

In this section, operational rules for boxes defined in 14496-12 [6] as well as definitions of private extensions to those ISO file format standards are described.

An ISO media file may contain one or more audio tracks. The tracks are composed in conformity to ISO base media file format and described in 14496-12 [6], for an audio track structure. The sub-sections that follow describe the construction of an audio track containing DTS-UHD audio.

7.3.2 Track Header Box

The syntax and values for the Track Box ('tkhd') and its sub-boxes shall conform to 14496-12 [6], and the following fields of each box shall be set to the following specified values. There are some “template” fields declared to use; see the syntax of 'tkhd' in 14496-12 [6].

flags = 000007h, except for the case where the track belongs to an alternate group;

layer = 0;

volume = 0100h;

matrix = {00010000h,0,0,0, 00010000h,0,0,0, 40000000h};

width = 0;

height = 0;

7.3.3 Sync Sample Box

All DTS-UHD sync frames are samples that are Sync Samples (random access points). All DTS-UHD non-sync frames are samples that are not random-access points.

Per the requirements of 14496-12 [6] the Sync Sample box shall be present if the DTS-UHD track contains samples that are not Sync Samples. More information about DTS-UHD random access points is provided in section 7.4.2.

7.3.4 Handler Reference Box

The syntax and values for the Handler Reference Box ('hdlr') shall conform 14496-12 [6], and the fields of this box shall be set to the following specified values.

handler_type = 'soun'

7.3.5 Sound Media Header Box

The syntax and values for the Sound Media Header box ('smhd') shall conform to 14496-12 [6], and the fields of this box shall be set to the following specified values.

balance = 0;

7.3.6 Sample Description Box

The SampleEntry box for DTS-UHD is derived from the AudioSampleEntry box defined in 14496-12 [6]. The DTS-UHD specific SampleEntry box is identified by a unique codingname value (see Table 11). It indicates the audio format used to encode the audio track and describes the configuration of the audio elementary stream.

Table 11: Defined Audio Formats

codingname	Description
dtsx	DTS-UHD bitstream compatible with Decoder Profile = 2
dtsy	DTS-UHD bitstream compatible with Decoder Profile > 2

7.3.7 AudioSampleEntry Box for DTS-UHD Formats

The syntax and values of the AudioSampleEntry Box shall conform to DTSUHDSampleEntry.

The configuration of the DTS-UHD elementary stream is described in the DTSUHDSpecificBox ('udts'), within DTSUHDSampleEntry. The syntax and semantics of the DTSUHDSpecificBox are defined in section 7.3.9.

7.3.8 DTSUHDSampleEntry

DTSUHDSampleEntry extends the AudioSampleEntry box defined in 14496-12 [6]:

```
class DTSUHDSampleEntry (codingname) extends AudioSampleEntry (codingname) {
    DTSUHDSpecificBox()    // 'udts' box
```

}

For DTSUHDSampleEntry(), the following values inherited from AudioSampleEntry are set as follows:

codingname shall be set to one of the values listed in Table 11.

channelcount shall be set to the number of decodable output channels, which are described by the *ChannelMask* field of the DTSUHDSpecificBox defined in section 7.3.9 of this document.

samplesize shall be set to 16.

samplerate shall be set to the base sampling frequency, as indicated by the *BaseSamplingFreqCode* field of the DTSUHDSpecificBox.

7.3.9 DTSUHDSpecificBox

The syntax and semantics of the DTSUHDSpecificBox ('udts') are shown below.

Note, many of the fields of the DTSUHDSpecificBox are identical to fields of the same name in the DTS-UHD Audio Descriptor for MPEG-2 systems and reference those sections of this document for definition.

Table 12: DTS-UHD Specific Box

Syntax	Reference
class DTSUHDSpecificBox extends Box ('udts') {	
bit (6) DecoderProfileCode;	6.2.3.2
bit (2) FrameDurationCode;	6.2.3.3
bit (3) MaxPayloadCode;	6.2.3.4
bit (5) NumPresentationsCode;	6.2.4.1
unsigned int(32) ChannelMask;	6.2.4.2
bit (1) BaseSamplingFrequencyCode	6.2.4.3
bit (2) SampleRateMod	6.2.4.4
bit (3) RepresentationType;	6.2.4.5
bit (3) StreamIndex	6.2.3.7
bit (1) ExpansionBoxPresent;	7.3.10.1
bit (NumPresentations) IDTagPresent[NumPresentations];	6.2.4.6
ByteAlign [0..7]	7.3.10.2
for (i = 0; i < NumPresentations; i++) {	
if (IDTagPresent[i] == TRUE)	
unsigned char PresentationIDTag[16];	6.2.4.8
}	
if (ExpansionBoxPresent) {	
box DTSExpansionBox [];	7.3.10.3
}	
}	

7.3.10 Additional Semantics for DTSUHDSpecificBox

7.3.10.1 ExpansionBoxPresent

This flag, if set to 1, indicates that at least one private box is present at the end of the DTSUHDSpecificBox.

7.3.10.2 ByteAlign

A variable number of bits (from 0 to 7) shall be inserted to pad to the end of the current byte boundary.

7.3.10.3 StreamIndex

When a given Audio Program is delivered as streams in multiple tracks, StreamIndex is the index used to indicate stream priority for prioritizing mixing metadata. The track containing the main stream shall have StreamIndex=0, while tracks containing auxiliary streams shall have StreamIndex values in the range of 1-7.

If the Audio Program is limited to a single track, StreamIndex shall be 0.

7.3.10.4 DTSExpansionBox

If *ExpansionBoxPresent* is set to 1, then one or more private boxes referred to as *DTSExpansionBoxes* can be found at the end of the ‘udts’ box, filling the remaining space of the box..

Playback devices not equipped to support one or more of these private boxes shall rely on the defined parameters in ‘udts’ to play the audio track using the default presentation parameters.

7.4 Storage of DTS-UHD elementary streams

7.4.1 Overview

Storage of DTS-UHD elementary streams within an ISO media file shall be according to this section.

7.4.2 Random Access Points

A sample, as defined in ISO/IEC 14496-12 [6] shall consist of a single DTS-UHD audio frame, including the leading syncword, as defined in TS 103 491 [7].

DTS-UHD audio samples may be either sync frames or non-sync frames, as identified by the first four bytes of the sample, which constitute the SYNCWORD, shown in Table 13.

Table 13 : DTS-UHD SYNCWORD

SYNCWORD	Designation
0x40411BF2	DTS-UHD Sync Frame
0x71C442E8	DTS-UHD Non-sync Frame

For all DTS-UHD tracks containing both sync frames and non-sync frames, the first sample in the track shall be a sync frame. For fragmented track files, the first sample in each fragment shall be a sync frame.

For DTS-UHD tracks containing non-sync frames, the methods provided in 14496-12 [6] for locating random access points such as the sync sample box ('stss'), the track fragment random access box ('tfra'), and random access information in 'trun', 'traf' and 'trex', shall be applied as required.

7.4.3 Restrictions on DTS-UHD Formats

This section describes the restrictions that apply to the DTS-UHD streams packaged in an ISO media file.

The following conditions shall not change in any DTS-UHD audio stream:

- Duration of Synchronized Frame

- Audio Channel Arrangement