SCTE·ISBE
STANDARDS

Digital Video Subcommittee

AMERICAN NATIONAL STANDARD

ANSI/SCTE 194-2 2018

**DTS-HD AUDIO SYSTEM – Part 2:
Constraints for Carriage over MPEG-2 Transport**

NOTICE

The Society of Cable Telecommunications Engineers (SCTE) / International Society of Broadband Experts (ISBE) 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•ISBE 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•ISBE members.

SCTE•ISBE 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•ISBE 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•ISBE web site at http://www.scte.org.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1. Scope

This document describes the carriage of DTS-HD audio in MPEG-2 systems. The descriptor necessary to signal DTS-HD audio is defined in this document. Multiplexing and transport for cable using MPEG-2 systems are defined in SCTE 54 [2]. Coding constraints for DTS-HD audio elementary streams are defined in SCTE 194-1 [1].

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] SCTE 194-1 201x, "DTS-HD Audio System – Part 1: Coding Constraints for Cable Television"
[2] SCTE 54 2015, "Digital Video Service Multiplex and Transport System Standard for Cable Television"

2.2. Standards from Other Organizations

[3] ETSI TS 102 114 v 1.4.1, "DTS Coherent Acoustics; Core and Extensions with Additional Profiles"
[4] ISO/IEC 13818-1:2007, "Information Technology - Generic coding of moving pictures and associated audio information: Systems".
[5] ISO/IEC 639-2:1998, "Codes for the representation of names of languages - Part 2: Alpha-3 code"

2.3. Published Materials

- No normative references are applicable.

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

- No informative references are applicable.

3.2. Standards from Other Organizations

[6] DTS Document #9302J19200: "DTS-HD PBR API Library Interface", DTS Inc., available at www.dts.com.

3.3. Published Materials

- No informative references are applicable.

4. Compliance Notation

shall	This word or the adjective "***required***" means that the item is an absolute requirement of this document.
shall not	This phrase means that the item is an absolute prohibition of this document.
forbidden	This word means the value specified shall never be used.
should	This word or the adjective "*recommended*" 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 "*optional*" 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

ISBE	International Society of Broadband Experts
SCTE	Society of Cable Telecommunications Engineers

5.2. Definitions

asset	A segregated block of coded audio containing one or more audio channels
audio frame:	A complete logical access unit of an audio stream that corresponds to a defined number of decodable PCM audio samples for a given time segment of the audio presentation
audio stream:	A sequence of synchronized audio frames

core substream	An audio stream component that adheres to the original DTS Coherent Acoustics definition (see Chapter 5 of TS 102 114 [3])
extension	An audio stream component providing a specific enhancement or coding profile
extension substream	An audio stream component utilizing the DTS-HD substream header (see Chapter 7 of TS 102 114 [3])
FSIZE	A core substream header parameter indicating the number of bytes in the core access unit. The actual number of bytes is FSIZE+1
LBR	DTS-HD extension used to implement the low bit rate coding profile
nuExtSSIndex	nuExtSSIndex: A substream header parameter indicating the substream number
PES payload	PES payload: Portion of the PES packet following the PES header
substream	substream: Sequence of synchronized frames comprising one of the logical components of the audio stream
substream_length	substream_length: A substream header parameter that indicates the number of bytes in the substream access unit
X96	Extension that doubles the sampling frequency of the core component
XBR	Extension containing amplitude resolution enhancements to the audio elements stored in the associated core component
XCH	Extension that adds a center surround channel
XLL	Lossless audio coding extension
XXCH	Channel extension, capable of accommodating from 1 to 32 audio channels

6. Transport of DTS-HD audio

6.1. MPEG-2 Systems Signaling

The following sections specify the signaling of DTS-HD audio streams in MPEG programs.

6.1.1. Stream Type

For DTS audio formats, stream_type *shall* be set to 0x88 indicating carriage in PES packets containing DTS-HD elementary stream data.

Note: The stream type value of 0x88 for signaling DTS-HD in North America was chosen to be compatible with the ATSC approach for signaling new codecs. DTS-HD is signaled as stream type 0x06

by DVB. The SCTE standard for DTS-HD differs significantly from the DVB version, for example in the audio descriptor.

6.1.2. Buffering Model

The DTS buffering model and buffer management is designed in accordance with ISO/IEC 13818-1 [4].

For DTS core streams, the main audio buffer size (BSn) *shall* have a fixed value of 9088 bytes, with a drain rate (Rxn) of 2 Mbps. The fixed value above (9088 bytes) was calculated from a double buffer (2*4096 bytes) plus jitter (384 bytes) + packet bursts (512 bytes).

For DTS-HD Lossless formats, the value of BSn *shall* have a fixed value of 66,432 bytes, with an Rxn value of 32 Mbps.

For all other DTS-HD formats, the value of BSn *shall* have a fixed value of 17,814 bytes, with an Rxn value of 8 Mbps.

6.1.3. Registration Descriptor

When DTS-HD audio is present in the transport stream, a registration_descriptor with the format_identifier set to "SCTE", as described in SCTE 54 [2], *shall* be included.

6.1.4. DTS-HD Audio Descriptor

The DTS-HD Audio Descriptor *shall* be included in the program map section following the relevant ES_info_length field for any DTS-HD audio stream component coded in accordance with SCTE 194-1 that is included in the MPEG program.

The syntax for the DTS-HD Audio Descriptor is shown in Table 1, Table 2, and Table 3. Semantics for these syntax elements are defined in Sections 6.1.4.1, Section 6.1.4.2 with Table 4, and Section 6.1.4.3 with Table 5 to Table 9.

Table 1 - DTS-HD Audio Descriptor

Syntax	No. of bits	mnemonic
DTS-HD_audio_descriptor(){		
descriptor_tag	8	uimsbf
descriptor_length	8	uimsbf
substream_core_flag	1	bslbf
substream_0_flag	1	bslbf
substream_1_flag	1	bslbf
substream_2_flag	1	bslbf
substream_3_flag	1	bslbf
reserved;	3	bslbf
if (substream_core_flag)		
substream.(substream_core)	See Table 2	
if (substream_0_flag)		
substream.(substream_0)	See Table 2	
if (substream_1_flag)		
substream.(substream_1)	See Table 2	
if (substream_2_flag)		
substream.(substream_2)	See Table 2	
if (substream_3_flag)		
substream.(substream_3)	See Table 2	
for (i=0; i<N; i++)		
additional_info_byte[i]	8	bslbf
}		

Table 2 - Substream

Syntax	No. of bits	mnemonic
substream.(substream name) {		
substream_length	8	uimsbf
num_assets	3	bslbf
channel_count	5	bslbf
LFE_flag	1	bslbf
sampling_frequency	4	bslbf
sample_resolution	1	bslbf
reserved	2	bslbf
for (i=0; i ≤ num_assets; i++)		
asset [i]	See Table 3	
}		

Table 3 - Asset

Syntax	No. of bits	Mnemonic
asset {		
asset_construction	5	bslbf
vbr_flag	1	bslbf
post_encode_br_scaling_flag	1	bslbf
component_type_flag	1	bslbf
language_code_flag	1	bslbf
if (post_encode_br_scaling_flag)		
bit_rate_scaled	13	
else		bslbf
bit_rate	13	
reserved	2	bslbf
if (component_type_flag)		
component_type	8	bslbf
if (language_code_flag)		
ISO_639_language_code	24	bslbf
}		

6.1.4.1. Semantics for Table 1

All syntax elements in the DTS-HD Audio Descriptor *shall* be set consistent with these elements in the audio elementary stream. In the event of a conflict between the descriptor and the bitstream, the bitstream *shall* be the preferred resource.

descriptor_tag: this *shall* be set to 0x7B.

descriptor_length: this 8-bit field specifies the total number of bytes following the descriptor_length field. It *shall* also be used to determine the number of additional_info_bytes contained in the audio descriptor.

substream_core_flag: *shall* be set to 1 if a core substream exists in the audio stream.

substream_0_flag: *shall* be set to 1 if an extension substream with nuExtSSIndex =0 exists in the audio stream.

substream_1_flag: *shall* be set to 1 if an independent extension substream with nuExtSSIndex =1 exists in the audio stream.

substream_2_flag: *shall* be set to 1 if an independent extension substream with nuExtSSIndex =2 exists in the audio stream.

substream_3_flag: *shall* be set to 1 if an independent extension substream with nuExtSSIndex =3 exists in the audio stream.

reserved: the reserved bits throughout the various sub-structures serve to force byte alignment of several major subcomponents in the DTS-HD descriptor. These bits *shall* be set to '0' and are reserved for future

definition. The reserved field ***shall*** be ignored by receivers built to this version of specification *should* they become defined in the future.

additional_info_byte: this byte array of from 0 to N bytes is reserved for future use. The number of bytes in this array is determined by comparing descriptor_length to the number of bytes parsed up to the end of the last substream element.

6.1.4.2. Semantics for Table 2

substream_length: this 8-bit field specifies the total number of bytes following the substream_length field in the substream structure, including the embedded asset structures.

num_assets: represents the number of audio assets stored in the substream. The number of audio assets stored in the substream is equal to num_assets+1. For substream_core num_assets is always 0. For all independent extension substreams, this value is identical to nuNumAssets in the extension substream header.

channel_count: the maximum number of channels that *may* be presented including LFE (if present).

Note 1: For multi asset presentations, the sum of the channels in all assets *may* be greater than channel_count.

Note 2: The number of channels actually presented in a particular implementation *may* be less than the maximum number available to be presented. For example a 7.1 presentation *may* be presented as 5.1, or 2.0, depending on the configuration of the receiver.

LFE_flag: If LFE_flag is set to 1, then this substream contains an LFE channel.

sampling_frequency: This parameter is interpreted according to Table 4 and indicates the maximum sampling frequency stored in the elementary stream. Note that not all values in the table are valid for the substreams of type substream_core.

Table 4 - sampling_frequency

substream_core.sampling_frequency substream_*n*.sampling_frequency	Sampling Frequency
2	32 kHz*
12	48 kHz
13	96 kHz
14	192 kHz*
Sampling frequencies indicated with (*) ***shall not*** be used with a core substream	

sample_resolution: this parameter indicates whether the decoded audio *should* be treated as 16-bit or 24-bit samples. If any resolution ≤ 16 bits are indicated, sample_resolution = 0, indicating 16-bit audio. Otherwise, sample_resolution = 1, indicating 24-bit audio.

6.1.4.3. Semantics for Table 3

asset_construction: this parameter provides details about the internal construction of the audio stream and is interpreted according to Table 5.

Table 5 - asset_construction

asset_construction	Core substream				Extension substream					
	Core	XCH	X96	XXCH	Core	XXCH	X96	XBR	XLL	LBR
1	✓									
2	✓	✓								
3	✓			✓						
4	✓		✓							
5	✓					✓				
6	✓							✓		
7	✓	✓						✓		
8	✓			✓				✓		
9	✓					✓		✓		
10	✓						✓			
11	✓	✓					✓			
12	✓			✓			✓			
13	✓					✓	✓			
14	✓								✓	
15	✓	✓							✓	
16	✓		✓						✓	
17									✓	
18										✓
19					✓					
20					✓	✓				
21					✓				✓	

vbr_flag: this flag is set to 1 if the audio asset has a variable bit rate, otherwise this flag is 0.

post_encode_br_scaling_flag: this flag is set to 1 if scaling of the bit stream has occurred after it was encoded, otherwise this flag is 0.

component_type_flag: *shall* be set to 1 if the field component_type is present.

language_code_flag: *shall* be set to 1 when ISO_639_language_code field is present.

bit_rate_scaled: represents the scaled bit rate of the coded elementary stream as a 10.3 unsigned fractional fixed point value, (i.e. a 13 bit integer value divided by 8). This field is exists in the descriptor when post_encode_br_scaling_flag = 1.

If the stream is variable bit rate, and dynamically updating the bit rate field is not possible or practical due to system limitations, or exceeds 1023.875 kbits/sec, then bit_rate_scaled *shall* be set to 0.

bit_rate: is a 13-bit unsigned integer representing the bit rate of the coded elementary stream in kbits/s. This field exists in the descriptor when post_encode_br_scaling_flag = 0.

If the stream is variable bit rate, and dynamically updating the bit rate field is not possible or practical due to system limitations, then bit_rate *shall* be set to 0.

component_type: This 8-bit field specifies the type of the audio component as described in Table 6.

Table 6 - component_type

component_type bits	Description
b7 (MSB)	Reserved
b6	full service flag (see **Error! Reference source not found.**)
b5 to b3	service type flags (see **Error! Reference source not found.**)
b2 to b0	number of channels flags (see **Error! Reference source not found.**)

Table 7 - Full Service Flag

full service flag (b6)	Description
0	Decoded audio stream is intended to be combined with another decoded audio stream before presentation
1	Decoded audio stream is a full service (suitable for decoding and presentation to the listener)

Table 8 - Service Type Flags

service type flags			Description	Restrictions	
b5	b4	b3		full service flag (b6)	number of channel flags (b2 to b0)
0	0	0	Complete Main (CM)	**shall** be set to 1	
0	0	1	Music and Effects (ME)	**shall** be set to 0	
0	1	0	Visually Impaired (VI)		
0	1	1	Hearing Impaired (HI)		
1	0	0	Dialogue (D)	**shall** be set to 0	
1	0	1	Commentary (C)		**shall** be set to 000
1	1	0	Emergency (E)	**shall** be set to 1	**shall** be set to 000
1	1	1	Voiceover (VO)	**shall** be set to 0	**shall** be set to 000
1	1	1	Reserved	**shall** be set to 1	
The values of the service type flags **shall** only be considered valid if the conditions identified in the restrictions column are satisfied					

Table 9 - Number of Channels Flags

number of channels flags			Description
b2	b1	b0	
0	0	0	Mono
0	0	1	Reserved
0	1	0	2 channel (stereo, LoRo)
0	1	1	2 channel matrix encoded (stereo, LtRt)
1	0	0	Multichannel audio (>2 channels)
1	0	1	Reserved
1	1	0	Reserved
1	1	1	Reserved

language_code_flag: this flag **shall** be set to 1 when ISO_639_language_code field is present.

ISO_639_language_code: this 24-bit language code conforms to the ASCII language codes described in ISO/IEC 639-2 [5].

6.2. DTS-HD PES Packet Encapsulation

6.2.1. Stream ID

All DTS and DTS-HD elementary streams *shall* use a stream_id of 0xBD, indicating private stream 1, in accordance with ISO/IEC 13818-1 [4]. Multiple DTS / DTS-HD streams *may* share the same value of stream_id since each stream is carried with a unique PID value. The mapping of values of PID to stream_type is indicated in the transport stream PMT.

6.2.2. Audio Access Unit Alignment in the PES packet

A valid sync word *shall* be aligned with the start of the PES packet data area. Valid DTS sync words are listed in Table 10. Since the sync word is aligned to the start of the PES packet, Data_Alignment_Indicator *shall* be set to 1 as defined in 13818-1 [4].

Table 10 - DTS-HD Sync Words

name	sync word	description
DTS_SYNCWORD_CORE	0x7ffe8001	core substream
DTS_SYNCWORD_SUBSTREAM	0x64582025	extension substream

When a core substream is present, DTS_SYNCWORD_CORE *shall* be aligned to the beginning of the PES payload. When only an extension substream is present, DTS_SYNCWORD_SUBSTREAM *shall* be aligned to the beginning of the PES payload.

A PES packet of DTS audio *shall* contain at least one complete audio access unit. If a DTS-HD stream contains both core and extension substreams then both access units *shall* be included in the same PES packet as illustrated in Figure 1. Multiple complete access units are permitted in a PES packet only when the ES consists of a single substream.

If multiple substreams are present, the access units *shall* maintain an interleaved order of presentation, as illustrated below in Figure 1. The header parameters FSIZE and substream_length indicate the number of bytes in the core substream and extension substream frames respectively, as illustrated in Figure 1.
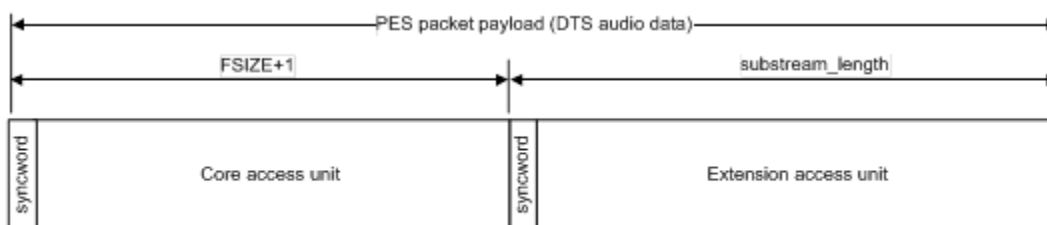

Figure 1 - PES Packet Payload