

SCTE • ISBE[®]

S T A N D A R D S

Digital Video Subcommittee

AMERICAN NATIONAL STANDARD

ANSI/SCTE 194-1 2018

**DTS-HD AUDIO SYSTEM – Part 1:
Coding Constraints for Cable Television**

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>.

All Rights Reserved

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

Table of Contents

Title	Page Number
NOTICE.....	2
Table of Contents.....	3
1. Scope.....	4
2. Normative References.....	4
2.1. SCTE References.....	4
2.2. Standards from Other Organizations.....	4
2.3. Published Materials.....	4
3. Informative References.....	4
3.1. SCTE References.....	4
3.2. Standards from Other Organizations.....	4
3.3. Published Materials.....	4
4. Compliance Notation.....	5
5. Abbreviations and Definitions.....	5
5.1. Abbreviations.....	5
5.2. Definitions.....	5
6. DTS-HD Coding Constraints.....	6
6.1. Stream Architecture.....	6
6.1.1. Overview.....	6
6.1.2. Core Substream.....	7
6.1.3. Extension substream.....	7
6.2. Bitstream Characteristics.....	8
6.3. Metadata Support.....	9
6.3.1. Dialog Normalization.....	9
6.3.2. Dynamic range compression.....	9
6.3.3. Down-mixing parameters.....	10

List of Tables

Title	Page Number
Table 1 - Composition of Core Substream.....	7
Table 2 - Construction of Core and Dependent Extension Substreams.....	7
Table 3 - Independent Assets in Extension Substreams.....	8
Table 4 - Summary of Channel, Bitrate and Sampling Frequency.....	8
Table 5 - Dynamic Range Metadata in the Extension Substream.....	10

1. Scope

This document describes the coding constraints of the DTS-HD audio system and identifies the normative references that apply. The carriage of the streams described in this specification is defined in SCTE 194-2 [3].

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

- No normative references are applicable.

2.2. Standards from Other Organizations

- [1] ETSI TS 102 114 v1.5.1, “DTS Coherent Acoustics; Core and Extensions with Additional Profiles”
- [2] ITU-R Recommendation BS.1770-4, “Algorithms to measure audio programme loudness and true-peak audio level”

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

- [3] SCTE 194-2 2018, “DTS-HD Audio System – Part 2 : Constraints for Carriage over MPEG-2 Transport”

3.2. Standards from Other Organizations

- [4] ATSC A/85:2018, “ATSC Recommended Practice: Techniques for Establishing and Maintaining Audio Loudness for Digital Television”

3.3. Published Materials

- No informative references are applicable.

4. Compliance Notation

<i>shall</i>	This word or the adjective “ required ” means that the item is an absolute requirement of this document.
<i>shall not</i>	This phrase means that the item is an absolute prohibition of this document.
<i>forbidden</i>	This word means the value specified shall never be used.
<i>should</i>	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.
<i>should not</i>	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.
<i>may</i>	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.
<i>deprecated</i>	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
LBR	DTS-HD extension used to implement the low bit rate coding profile
LFE	Low Frequency Effects
SCTE	Society of Cable Telecommunications Engineers
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

5.2. Definitions

audio frame	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
asset	A segregated block of coded audio containing one or more audio channels
audio stream	A sequence of synchronized audio frames

chunk	A range of bytes that can be reliably parsed from the bitstream, often formatted having a start code, a byte size and a checksum verification
core	The primary component of a core substream, excluding extension coding blocks
coding block	A part of an asset that <i>may</i> or <i>may not</i> be decoded independently
core substream	An audio stream component that adheres to the original DTS Coherent Acoustics definition (see Chapter 5 of TS 102 114 [1])
DIALNORM	Parameter in the core header that <i>may</i> be used to calculate dialog normalization correction.
duration	Time represented by one decoded audio frame, <i>may</i> be represented in audio samples per channel at a specific audio sampling frequency or in seconds
DYNF	Parameter in the core header indicating the presence of dynamic range compression coefficients at the beginning of each subframe.
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 [1])
LKFS	Loudness, K-weighted, relative to full scale, measured with equipment that implements the algorithm specified by ITU-R Rec. BS.1770 [2]; a unit of LKFS is equivalent to a decibel
PES payload	Portion of the PES packet following the PES header
presentation	The selection within the bitstream being rendered to output channels.
substream	Sequence of synchronized frames comprising one of the logical components of the audio stream
subframe	One synchronized frame of a given substream
subsubframe	Atomic unit of the DTS core containing 256 samples.
X96	Extension that doubles the sampling frequency of the core component.

6. DTS-HD Coding Constraints

The DTS-HD bitstream *shall* be compliant with that defined in TS 102 114 [1]. Additional information and constraints for DTS HD bitstreams are defined in the following sub-clauses.

6.1. Stream Architecture

6.1.1. Overview

The DTS-HD bitstream *may* be composed of a core substream, or an extension substream, or both a core and extension substream. The core substream consist of a sequence of synchronized core frames, where

each core frame consists of a header, coded data for up to 5.1 channels of audio and up to one extension. A core substream contains a single asset.

The extension substream also consists of a series of synchronized extension frames. If both a core and extension substream exist, the core and extension frames are interleaved in presentation order. An extension frame consists of an extension header, and coded data for up to eight assets. Only the first asset *may* be dependent on the core substream.

6.1.2. Core Substream

The core substream is specified in Chapter 5 of TS 102 114 [1]. Each core substream contains one audio asset consisting of one core coding block and up to one additional coding extension block, as shown in Table 1. For more information on the construction of the core substream, see Chapter 5.3 in [1].

Table 1 - Composition of Core Substream

Core substream			
Core	XCH	X96	XXCH
✓			
✓	✓		
✓			✓
✓		✓	

6.1.3. Extension substream

The extension substream header is described in Chapter 7 of TS 102 114 [1]. The extension components XCH, XXCH, X96, and XBR are described in Chapter 6 of [1]. The XLL extension is described in chapter 8 of [1] and LBR is described in Chapter 9 of [1].

An extension substream *may* contain up to 8 audio assets. A description of audio presentations and audio assets is described in Chapter 7.2 of TS 102 114 [1]. Only the first asset of the first extension substream, (asset 0 of substream 0), *may* be dependent on the core substream. In this case, only the first asset in the extension substream is associated with the core. The permitted combinations of audio extensions are shown in Table 2.

Table 2 - Construction of Core and Dependent Extension Substreams

Core substream				Asset[0] of Substream[0]				
Core	XCH	X96	XXCH	XXCH	X96	XBR	XLL	LBR
✓				✓				
✓						✓		
✓	✓					✓		
✓			✓			✓		
✓				✓		✓		
✓					✓			
✓	✓				✓			
✓			✓		✓			

Core substream				Asset[0] of Substream[0]				
Core	XCH	X96	XXCH	XXCH	X96	XBR	XLL	LBR
✓				✓	✓			
✓							✓	
✓	✓						✓	
✓		✓					✓	

The remaining available assets in substream 0, and all assets in substreams 1 to 3, *shall* be constructed according to Table 3.

Table 3 - Independent Assets in Extension Substreams

Core	XXCH	X96	XLL	LBR
			✓	
				✓
✓				
✓	✓			
✓			✓	

6.2. Bitstream Characteristics

For each asset, the allowable range of bit rates per channel configuration *shall* be according to Table 4. For all cases a complete DTS-HD bitstream, (including all contained assets), *shall not* exceed 24,576 kbps.

Table 4 - Summary of Channel, Bitrate and Sampling Frequency

Number of channels	Range of bitrates (kbps)	Sampling frequencies (kHz)
Mono	32 - 1524	32, 44.1, 48, 96
Stereo	48 - 2046	32, 44.1, 48, 96
5.x	128 - 4224	44.1, 48, 96
7.x or greater	447 - 6075	44.1, 48, 96
Lossless (all channel configurations)	VBR up to 24,576	44.1, 48, 96, 192
x indicates presence of LFE channel, and <i>may</i> be 1 or 0. x = 0 indicates LFE is not present		

Audio sampling frequencies *shall* be constrained according to Table 4.

Note that 192 kHz sampling frequency is only supported for lossless coding, where it is coded in two equal frequency bands. A decoder *may* be implemented such that this signal can be properly decoded at 96 kHz by decoding and rendering only the lower band. Similarly, for audio coded at 96 kHz using the X96 extension, a core component coded at 48 kHz *shall* always be present such that a presentation can be properly rendered at 48 kHz.

Bitstreams with channel configurations greater than 5.1 channels *shall* contain a 5.1 compatible configuration. It *should* be noted, for example, that a 7.1 channel bitstream can always be played back properly on a 5.1 only receiver. One method for performing this function is described in Chapter 8.5.2 of TS 102 114 [1].

6.3. Metadata Support

The DTS-HD bitstream supports a variety of metadata parameters, described in TS 102 114 [1]. The following sections provide a brief summary of the metadata controlling dialog normalization, dynamic range compression and down mixing.

Only one set of metadata exists for a given presentation. In the case of a core substream with a dependent extension substream, the core substream carries the metadata. When an independent presentation exists in the extension substream, the metadata is carried in the extension substream.

6.3.1. Dialog Normalization

Dialog normalization metadata **shall** be present.

For the case of the core substream, the dialog normalization metadata *may* be stored in the Rev2AUX data chunk, as described in clause 5.7.2 of TS 102 114 [1] as DIALNORM_rev2aux.

If DIALNORM_rev2aux exists in the core substream, then DIALNORM, as described in clause 5.3.1 of TS 102 114 [1] **shall** be ignored.

For the case of the extension substream, the dialog normalization metadata is stored in the parameter nuDialNormCode, as described in TS 102 114 [1] clause 7.4.2.2.

The value of the dialog normalization parameters in the DTS bit stream **shall** indicate the loudness of the corresponding encoded audio content (typically of the average spoken dialogue) using LKFS units. LKFS and its loudness measurement algorithm are specified in ITU-R Recommendation BS.1770 Annex 1 [2]. Receivers use the value of the dialog normalization parameter to adjust the reproduced audio level to normalize the loudness.

6.3.2. Dynamic range compression

Dynamic range compression metadata in the DTS-HD bitstream is optional.

For the case of the extension substream, the equivalent metadata is stored in the header of the audio asset descriptor block of the extension substream header as described in Table 5.

Table 5 - Dynamic Range Metadata in the Extension Substream

Syntax	No. of bits
<pre> bDRCMetadataRev2Present = (ExtractBits(1) == 1) ? TRUE : FALSE; if (bDRCMetadataRev2Present == TRUE) { DRCversion_Rev2 = ExtractBits(4); // one DRC value for each block of 256 samples nRev2_DRCs = nuExSSFrameDurationCode / 256; // assumes DRCversion_Rev2 == 1: for (subsubframe=0; subsubframe < nRev2_DRCs; subsubframe++) { DRCCoeff_Rev2[subsubframe] = dts_dynrng_to_db(ExtractBits(8)); } } </pre>	<p>1</p> <p>4</p> <p>8 * nRev2_DRCs</p>

where:

bDRCMetadataRev2Present is a Boolean parameter to signal the presence of DRC metadata.

DRCversion_Rev2 is a 4-bit version code. Currently only version 1 is supported, which is a single band compression curve.

nuExSSFrameDurationCode is defined in TS 102 114 [1] clause 7.4.1 and indicates the frame duration in samples.

DRCCoeff_Rev2 is an 8-bit code that represents a gain modifier to be applied against the subsubframe of samples. The same multiplier is applied across all channels. Coefficient smoothing is applied to prevent transient artifacts at subsubframe boundaries.

For the case of the core substream, dynamic range compression metadata is stored in the Rev2 auxiliary data chunk. Details of this metadata block are found in TS 102 114 [1], clause 5.7.2. A subsubFrameDRC_Rev2AUX[] parameter exists for each subsubframe of audio (every 5.33 ms). The DRC_Rev2AUX parameters are coded and treated identically to the DRCCoeff_Rev2 parameters in the extension substream.

DYNF *shall* always be set to 0.

6.3.3. Down-mixing parameters

Down-mixing parameters are optional in the DTS-HD bitstream.

For the case of a core substream, if down-mixing parameters are present, they *shall* be stored in the Auxiliary Data area. This is described in detail in TS 102 114 [1], clause 5.7.1.

For the case of an asset in an extension substream, if down-mixing parameters are present they *shall* be stored in the extension substream header as described in TS 102 114 [1] clause C.8.

In the event that no downmix parameters are included in the audio stream, default parameters defined in the decoder *shall* be used.