Dolby® Digital Plus Streams Within the ISO Base Media File Format

Issue 0.9 (Preliminary)

Confidential Information

Dolby Laboratories Licensing Corporation

Corporate Headquarters

Dolby Laboratories, Inc.
Dolby Laboratories Licensing Corporation
100 Potrero Avenue
San Francisco, CA 94103-4813 USA
Telephone 415-558-0200
Fax 415-863-1373
www.dolby.com

European Licensing Liaison Office

Dolby Laboratories, Inc.
Wootton Bassett
Wiltshire SN4 8QJ England
Telephone 44-1793-842100
Fax 44-1793-842101

Asia

Dolby Japan K.K.
NBF Higashi-Ginza Square 3F
13–14 Tsukiji 1-Chome, Chuo-ku
Tokyo 104-0045 Japan
Telephone 81-3-3524-7300
Fax 81-3-3524-7389
www.dolby.co.jp

Dolby Laboratories Hong Kong Limited
Unit 5407, Central Plaza
18 Harbour Road
Wanchai, Hong Kong
Telephone 852-2519-0888
Fax 852-2519-8988

Dolby Laboratories International Services (Shanghai) Co., Ltd.
03–07a, Floor 18
The Center
989 Changle Road
Shanghai 200031 China
Telephone 86-21-6113-3456
Fax 86-21-6113-3400
www.dolby.com.cn

S09/21199
Issue 0.9 (Preliminary)

Table of Contents

1 Scope

This document describes the requirements to store Dolby® Digital Plus (Enhanced AC-3) streams in files conforming to the ISO base media file format (defined in ISO/IEC 14496-12). Variants of the base file format (such as .mov or .mp4) are commonly used for the delivery of content in online video and audio services. Within this document the ISO media file is referred to generically as an MP4 file. This document covers:

- The steps necessary to correctly packetize a Dolby Digital Plus stream for multiplexing and storage in an MP4 file
- The data needed to identify a Dolby Digital Plus stream within an MP4 file
- The steps needed to demultiplex a Dolby Digital Plus stream from an MP4 file.

2 References

[1] ETSI TS 102 366 v. 1.2.1 – Digital Audio Compression (AC-3, Enhanced AC-3) Standard

[2] ISO/IEC 14496-12:2005 – ISO base media file format

3 Packetization of Dolby Digital Plus streams for MP4 file storage

3.1 Dolby Digital Plus Bitstream Structure

A Dolby Digital Plus bitstream is constructed from one or more substreams. Each substream can carry up to 5.1 channels of audio, and is constructed from a sequence of syncframes. The use of multiple substreams within a single Dolby Digital Plus bitstream allows delivery of a single program with more than 5.1 channels, multiple independent programs, or a combination of both. The substreams are time-multiplexed into the Dolby Digital Plus bitstream. Section E.2.8 of [1] provides detailed information on the use and structure of multiple substreams within a Dolby Digital Plus bitstream.

The syncframes that make up each substream are themselves constructed from a number of smaller units referred to as blocks, with each block representing 256 samples of audio from each channel carried in the substream. A syncframe may contain one, two, three, or six blocks of audio data (representing 256, 512, 768, or 1536 samples of audio respectively).

3.2 Constraints on Dolby Digital Plus streams within MP4 Files

Annex F of [1] details the constraints that must be applied to the Dolby Digital Plus bitstream for storage within the ISO base media file format. If the requirements stated in this specification are not met by a Dolby Digital Plus bitstream, then the MP4 multiplexer should halt processing and inform the user that the stream is not compliant.

Additionally, while the ISO base media file format can support Dolby Digital Plus bitstreams up to a practically unlimited data rate, it is strongly recommended that the maximum bit rate of a Dolby Digital Plus bitstream within an MP4 file is limited to 3024 kbps at a sampling rate of 48 kHz. It is recommended that if a Dolby Digital Plus stream is encountered that exceeds this bit rate, the MP4 multiplexer should provide a warning to the user, but if the stream is otherwise compliant the multiplexer can still multiplex the Dolby Digital Plus stream.

3.3 Contents of the MP4 Audio Sample

Each Audio Sample must contain all the data needed by the Dolby Digital Plus decoder to produce 1,536 samples of decoded audio for each audio channel present in the bitstream. This means that six blocks of audio data from each of the substreams present in the Dolby Digital Plus bitstream must be included in the Audio Sample. Using multiple data packets to deliver six blocks of audio data from each substream, or placing more than six blocks of audio data from each substream within a single Audio Sample is not permitted.

As a Dolby Digital Plus syncframe may contain fewer than six blocks of audio data, it may be necessary to group multiple syncframes to accumulate the required six blocks. In addition, the Dolby Digital Plus syncframes must occur in the same sequence in the Audio Sample as they do in the Dolby Digital Plus bitstream. This is to ensure the correct operation of a downstream Dolby Digital Plus decoding device, particularly when this device is capable of converting the Dolby Digital Plus bitstream to Dolby Digital. Conversion requires the correct set of six blocks of audio data to produce a Dolby Digital syncframe.

Figure 1 shows the construction of the Audio Sample contents, including three examples of how Dolby Digital Plus data within the Audio Sample is structured for bitstreams with different configurations.

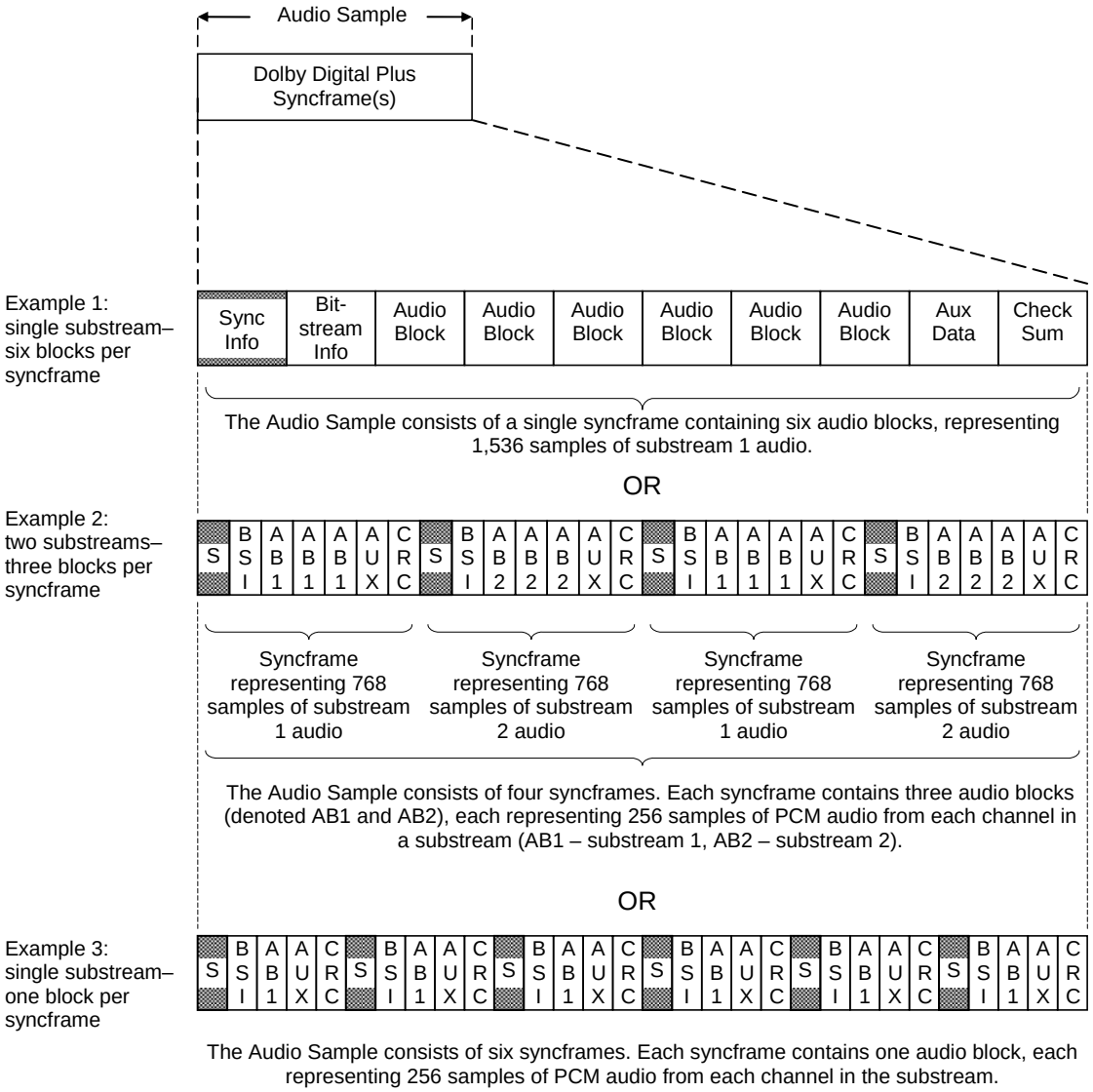

Figure 1 Dolby Digital Plus syncframes within the Audio Sample

3.4 Dolby Digital Plus Bitstream Parsing

The multiplexer used to construct the ISO media file must ensure that the correct syncframes are placed in each Audio Sample by monitoring the value of the `strmtyp`, `substreamid` and `numblkscod` parameters in each Dolby Digital Plus syncframe. When there are six blocks per syncframe (`numblkscod` is 3), the following procedure should be followed:

1. The multiplexer should scan the bitstream to find the first syncframe with a `strmtyp` value of zero or two, and a `substreamid` value of zero. This syncframe will begin the Audio Sample.

2. Subsequent syncframes should be checked until the next syncframe with a `strmtyp` value of zero or two and `substreamid` value of zero is reached. The syncframe

immediately preceding this syncframe will complete the current Audio Sample; this syncframe will become the first syncframe of the next Audio Sample.

If the bitstream is constructed from syncframes containing fewer than six blocks of audio data, the Audio Sample will contain the number of syncframes needed to deliver six blocks of audio data from each of the substreams in the bitstream. Consequently, the process described above must be repeated multiple times to create the complete Audio Sample. For example, in the case of a bitstream that is constructed from syncframes, consisting of two blocks of audio data, the multiplexer repeats this process three times.

For bitstreams that are constructed from syncframes containing fewer than six blocks of audio, the multiplexer shall additionally ensure that when the `strmtyp` value of the first substream in the bitstream is zero, the first syncframe in the Audio Sample is always the syncframe where the `convsync` parameter is set to one. When the `strmtyp` value of the first substream in the bitstream is two, the first syncframe in the Audio Sample is always the syncframe where the `blkid` parameter is set to one.

This ensures that each Audio Sample contains all the parts of the bitstream required by the Dolby Digital Plus decoder to produce 1,536 samples of decoded audio for each audio channel present in the bitstream. In addition, this guarantees that each Audio Sample is independently decodable and convertible, achieving optimal error resilience.

4 Identification of Dolby Digital Plus bitstreams within an MP4 file

4.1 Overview

In the terminology of the ISO Base Media File Format specification [2], Dolby Digital Plus streams are treated as audio tracks. As a result, the basic structures defined within the specification to identify audio tracks are used for Dolby Digital Plus streams, with specific extensions, defined in Annex F of [1], that allow detailed information on the configuration of the bitstream to be presented in the moov box. The location of the boxes used to identify Dolby Digital Plus streams are shown in the hierarchical diagram below, together with the reference to the definition for each box – for clarity, only boxes directly related to Dolby Digital Plus audio streams are shown.

Table 1 MP4 File Box Hierarchy for Dolby Digital Plus audio streams

1	2	3	4	5	6	7	Reference
ftyp							[2]
moov							[2]
	mvhd						[2]
	trak						[2]
		tkhd					[2]
		mdia					[2]
			mdhd				[2]
			hdlr				[2]
			minf				[2]
				smhd			[2]
				dinf			[2]
					dref		[2]
				stbl			[2]
					EC3 Sample Entry (stsd)		Annex F of [1]
						EC3 Specific Box	Annex F of [1] and Section 4.2

This section provides additional details on the information specific to Dolby Digital Plus and how this is derived from the Dolby Digital Plus bitstream parameters themselves. Unless otherwise stated here, all requirements specified in [2] and Annex F of [1] shall be followed.

4.2 Deriving the contents of the EC3SpecificBox

Using the bitstream parsing process outlined in section 3.4, the necessary bitstream parameters can be read to enable the construction of the data inside the EC3SpecificBox.

Correction of error in ETSI TS 102 366 Annex F

The currently published specification for Annex F contains an error which could lead to the bitstream configuration of the first independent substream being left out of the EC3SpecificBox. The error is found in the Syntax definition in section F.6 in the following line:

```
for(i = 0; i < num_ind_sub; i++)
```

This line should instead be

```
for(i = 0; i < num_ind_sub + 1; i++)
```

When implementing an MP4 multiplexer, this correction must be implemented to ensure that the correct information is placed in EC3SpecificBox.

Data_rate

This parameter indicates the data rate of the entire Dolby Digital Plus bitstream in kbit/s. The value is the sum of the data rates of all the substreams in the Dolby Digital Plus bitstream. In the case where a bitstream uses variable data rate encoding, this parameter indicates the maximum data rate of the stream. The data rate of each substream is calculated using the following formula:

$$Data\_Rate\_Sub = \frac{(Frmsiz + 1) * fs}{numblks * 16}$$

Where:

- Frmsiz is the value of the `Frmsiz` parameter in the Dolby Digital Plus syncframe
- Fs is the sampling frequency of the Dolby Digital Plus bitstream in kHz (derived from the `fscod` parameter in the Dolby Digital Plus syncframe)
- Numblks is the number of audio blocks per syncframe (derived from the `numblkscod` parameter in the Dolby Digital Plus syncframe)

Num_ind_sub

This parameter indicates the number of independent substreams that are present in the Dolby Digital Plus bitstream. The value of num_ind_sub is equal to the value of the `substreamid` parameter found in the last independent substream of the bitstream. During bitstream parsing this will the syncframe with a `strmtyp` value of zero that precedes the syncframe with both a `strmtyp` value of zero and a `substreamid` value of zero (indicating that this syncframe belongs to the first independent substream of the bitstream).

Bsmod

The bsmod parameter is an optional parameter in a Dolby Digital Plus bitstream, and requires additional parsing to detect. Consequently it is optional to include the value of bsmod in the EC3SpecificBox if the Dolby Digital Plus bitstream contains only one independent substream. However, if more than one independent substream is found during parsing, the value of bsmod must be included in EC3SpecificBox for each substream to ensure that the system parsing the MP4 file can quickly identify the audio services that are present in the bitstream.

Num_dep_sub

This parameter indicates the number of dependent substreams that are associated with an independent substream. The value of this parameter is equal to the value of the substreamid parameter found in the syncframe with a strmtyp value of 1 (dependent substream) that immediately precedes a syncframe with a strmtyp value of 0 (independent substream).

Chan_loc

This bitfield is used to indicate channel locations beyond the standard 5.1 channels that are carried by the one or more dependent substreams that are associated with an independent substream. The contents of the bitfield are determined by parsing the chanmap bit field in each dependent substream, and indicating the corresponding channel locations from all of the dependent substreams associated with a particular independent substream. As this field is only used by the system to indicate the overall channel location of the bitstream, it is not necessary to reflect replacement channels in this field, so only new channels are indicated. Consequently when duplicate channel locations (indicating replacement channels) are found in the chanmap field, these can be ignored.

5 Demultiplexing a Dolby Digital Plus stream from an MP4 file

5.1 Overview

This section covers considerations for product designers when implementing an MP4 demultiplexer that supports Dolby Digital Plus bitstreams.

5.2 Buffering considerations

As the MP4 AudioSample for Dolby Digital Plus must contain all the audio data necessary to produce 1536 samples of decoded PCM from each substream, the size of the AudioSample may require special considerations for stream buffering. However if the upper data rate limit of 3024 kbps (at 48 kHz) is assumed then the maximum size of an AudioSample containing Dolby Digital Plus data will be no more than 100 kbits per AudioSample. Depending on the multiplexing strategies used during the creation of the MP4 file, multiple audio samples may be grouped together within chunks or movie fragments. The system should ensure that enough buffer memory is available for arriving audio data so that buffer overruns do not occur.

5.3 Conflicts between EC3SampleEntry and Dolby Digital Plus bitstream parameters

Only basic parameters describing the audio bitstream are present in the EC3SampleEntry box, and as such most are ignored, with data in the EC3SpecificBox being used to identify the Dolby Digital Plus stream configuration. The one exception is the SamplingRate field – as this field defines the timescale of the audio track, it is a requirement of both the ISO/IEC 14496-3 and ETSI TS 102 366 Annex F specifications that this parameter is set correctly based on the sample rate of the Dolby Digital Plus bitstream. If the decoding

device encounters a conflict between the SamplingRate field in the EC3SampleEntry box and the value of the `fscod` parameter in the Dolby Digital Plus bitstream, then it is likely that the entire audio track has been multiplexed incorrectly, and it is recommended that the decoding device halts demultiplexing and decoding of the MP4 file.

5.4 Conflicts between EC3SpecificBox and Dolby Digital Plus bitstream parameters

The EC3SpecificBox is intended to be used by the system for information only – for example to inform the system's OSD of the channel configuration of the audio stream, and (in the case where multiple independent substreams are present) to offer the user the ability to select between the different audio services that are being carried. It may also be useful for error recovery if there is an interruption in audio data delivery and the system cannot ascertain the configuration of the complete Dolby Digital Plus bitstream from the substream structure in the AudioSample due to CRC errors.

However, this information shall not be used to configure the audio decoder or the audio subsystem of the device, as it is possible that differences may occur between the EC3SpecificBox and the Dolby Digital Plus bitstream parameters. If a conflict is encountered, then the Dolby Digital Plus bitstream parameter value shall always take precedence over the value of the corresponding parameter in the EC3SpecificBox.