SMPTE ST 299-1:2009
Revision of SMPTE 299M-2004

SMPTE STANDARD

NOTE: This document previously numbered SMPTE 299-2009; Content Unchanged

24-Bit Digital Audio Format for SMPTE 292 Bit-Serial Interface

Table of Contents

Approved May 29, 2009
Document renumbered July 21, 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 Standard 299-1 was prepared by Technology Committee 32NF.

Intellectual Property

SMPTE draws attention to the fact that it is claimed that compliance with this Standard may involve the use of one or more patents or other intellectual property rights (collectively, "IPR"). The Society takes no position concerning the evidence, validity, or scope of this IPR.

Each holder of claimed IPR has assured the Society that it is willing to License all IPR it owns, and any third party IPR it has the right to sublicense, that is essential to the implementation of this Standard to those (Members and non-Members alike) desiring to implement this Standard under reasonable terms and conditions, demonstrably free of discrimination. Each holder of claimed IPR has filed a statement to such effect with SMPTE. Information may be obtained from the Director, Standards & Engineering at SMPTE Headquarters.

Attention is also drawn to the possibility that elements of this Standard may be subject to IPR other than those identified above. The Society shall not be responsible for identifying any or all such IPR.

1 Scope

1.1 This standard defines the mapping of 24-bit AES digital audio data and associated control information into the ancillary data space of a serial digital video conforming to SMPTE 292. The audio data are derived from AES3, hereafter referred to as AES audio. The AES audio data may contain linear PCM audio or non-PCM data formatted according to SMPTE 337. Not all implementations compliant with this standard may support all sampling frequencies. Manufacturers are encouraged to indicate which sampling frequencies are supported (see Annex D).

1.2 Audio sampled at a clock frequency of either 48 kHz or 96 kHz locked (synchronous) to video, is the preferred implementation for intrastudio applications. As an option, this standard supports AES audio at synchronous or asynchronous sampling rates from 32 kHz to 48 kHz.

1.3 Audio channels are transmitted in groups of four, up to a maximum of 16 audio channels in the case of 32 kHz, 44.1 kHz or 48 kHz sampling, and up to a maximum of 8 audio channels in case of 96-kHz sampling. Each group is identified by a unique ancillary data ID.

1.4 Audio data packets are multiplexed (embedded) into the horizontal ancillary data space of the C_b/C_r data stream, and audio control packets are multiplexed into the horizontal ancillary data space of the Y data 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.

A conformant implementation according to this document is one that includes all mandatory provisions ("shall") and, if implemented, all recommended provisions ("should") as described. A conformant implementation need not implement optional provisions ("may") and need not implement them as described.

3 Normative References

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

AES3-2003, AES Standard for Digital Audio — Digital Input-Output Interfacing — Serial Transmission Format for Two-Channel Linearly Represented Digital Audio Data (AES3)

SMPTE 291M-2006, Television — Ancillary Data Packet and Space Formatting

SMPTE 292-2008, 1.5 Gb/s Signal/Data Serial Interface

SMPTE RP 168-2002, Definition of Vertical Interval Switching Point for Synchronous Video Switching

4 Definition of Terms

4.1 AES audio: All the VUCP (sample validity bit (V), channel status bit (C), user data bit (U), even parity bit (P)) data, audio data and auxiliary data, associated with one AES digital stream as defined in AES3.

4.2 AES frame: Two AES subframes: in the case of the 32 kHz to 48 kHz sampling subframes, one and two carry AES audio channel 1 and 2, respectively. In the case of 96 kHz sampling subframes, one and two carry successive samples of the same AES audio signal which is mandatory for 96 kHz application.

4.3 AES subframe: All data associated with one AES audio sample for one channel in a channel pair.

4.4 Ancillary data packet: A data packet as defined by SMPTE 291M.

4.5 audio clock phase data: Audio clock phase is indicated by the number of video clocks between the first word of EAV and the video sample appearing at the same time as the audio sample at the input to the formatter.

4.6 audio control packet: An ancillary data packet occurring once a field in an interlaced system and once a frame in a progressive system and containing data used in the process of decoding the audio data stream.

4.7 audio data: 29 bits: 24 bits of AES audio associated with one audio sample, including AES auxiliary data, plus VUCP bits and the Z flag which is derived from the preamble of AES3 stream. The Z bit is common to the two channels of an AES channel pair.

4.8 audio data packet: An ancillary data packet containing audio clock phase data, audio data for 2 channel pairs (4 channels) and error correction code. An audio data packet shall contain audio data of one sample associated with each audio channel.

4.9 audio frame number: A number, starting at 1, for each frame within the audio frame sequence. For the example in § 4.9, 48 kHz sampling at 30.00/1.001 frames/s system, the frame numbers would be 1, 2, 3, 4 and 5.

4.10 audio frame sequence: The number of video frames required for an integer number of audio samples in synchronous operation. As an example, the audio frame sequence for synchronous 48 kHz sampling at 30.00/1.001 frames/s system is 5 frames.

4.11 audio group: Consists of two channel pairs that are contained in one ancillary data packet. Each audio group has a unique ID as defined in § 6.1 and § 7.1. Audio groups are numbered 1 through 4.

4.12 auxiliary data: Four bits of data associated with one AES audio sample defined as auxiliary data by AES3. The four bits can be used to extend the resolution of the audio sample.

4.13 channel pair: Two digital audio channels, derived from the same AES audio source.

4.14 data ID: A word in the ancillary data packet that identifies the use of the data therein.

4.15 error correction code: BCH (31, 25) code (an error-correction method) in each bit sequence of b0-b7. Errors between the first word of ancillary data flag (ADF) through the last word of audio data of channel 4 (CH4) in user data words (UDW) can be corrected or detected within the capability of this code.

4.16 horizontal ancillary data block: An ancillary data space located in the digital line blanking interval of one television line.

4.17 synchronous audio: Audio is defined as being clock synchronous with video if the sampling rate of audio is such that the number of audio samples occurring within an integer number of video frames is itself a constant integer number. Examples are shown in Table 1.

Table 1 – Examples of audio samples per frame for synchronous audio

	Samples/frame				
Audio sampling rate	30.00 frame/s	30.00/1.001 frame/s	25.00 frames/s	24.00 frames/s	24.00/1.001 frames/s
96.0 kHz	3200/1	16016/5	3840/1	4000/1	4004/1
48.0 kHz	1600/1	8008/5	1920/1	2000/1	2002/1
44.1 kHz	1470/1	147147/100	1764/1	3675/2	147147/80
32.0 kHz	3200/3	16016/15	1280/1	4000/3	4004/3

AES11 provides specific recommendations for audio and video synchronization.

Note: Implementations of this standard may achieve synchronous or asynchronous operation through the use of sample rate converters. In the context of this standard, synchronous audio applies to the AES audio stream that is directly mapped into the ancillary data space, which may or may not be the AES audio stream present on device interfaces. It is recommended that product manufacturers clearly state when sample rate conversion is used to support multiple sample rates and/or asynchronous operation. It is also recommended that the use of sample rate conversion be user selectable. For example, when the AES audio data contains SMPTE 337 formatted data the use of sample rate conversion will corrupt the SMPTE 337 data (see Annex A). This recommendation applies to both multiplexing (embedding) and demultiplexing (receiving) devices.

5 Overview

5.1 The modes of transmission carried in an audio data packet defined in § 6 shall be the TWO-CHANNEL MODE at all sampling frequencies from 32 kHz to 48 kHz, and the SINGLE CHANNEL DOUBLE SAMPLING FREQUENCY MODE at the sampling frequency of 96 kHz as defined in AES3. Audio data channels 1~4 (CH1~ CH4) defined in § 6.2.2 carry two AES audio channel pairs (AES1 channel 1 and 2 and AES2 channel 1 and 2) in the case of 32 kHz to 48 kHz sampling. For 96 kHz sampling, two successive samples of two AES audio channels (AES1 channel 1 1st and 2nd sample and AES2 channel 1 1st and 2nd sample) shall be carried.

Note: AES3 document allows for different audio samples configurations of the digital audio. Single sample mode or successive sample modes are permitted in AES3 without regard to used sampling frequency.

5.2 The 32 kHz, 44.1 kHz or 48 kHz sampling audio data derived from two channel pairs shall be configured in an audio data packet as shown in Figure 1. Both channels of a channel pair are derived from the same AES audio source. The number of samples per channel used for one audio data packet shall be constant and is equal to one. The number of audio data packets in a given group shall be less than or equal to Na in a horizontal ancillary data block. The definition and examples of Na are described in § 6.3.3.

5.3 Figure 2 shows the audio data packet at the sampling rate of 96 kHz. AES subframes 1 and 2 carry successive samples of the same AES audio signal. Both channels shall be derived from the same AES audio source. The number of samples per channel used for one audio data packet shall be constant and equal to two. The number of audio data packets in a given group is less than or equal to Na/2 in a horizontal ancillary data block. The definition and examples of Na are described in § 6.3.3.

5.4 Two types of ancillary data packets carrying AES audio information are defined and formatted per SMPTE 292. Each audio data packet shall carry all of the information in the AES bit stream as defined by AES3. The audio data packet shall be located in the horizontal ancillary data space of the C_b/C_r data stream. An audio control packet shall be transmitted once per field in an interlaced system and once per frame in a progressive system in the horizontal ancillary data space of the second line after the switching point of the Y data stream.

5.5 Data ID shall be defined for four separate packets of each packet type. This allows for up to eight channel pairs. In this standard, the audio groups are numbered 1 through 4 and the channels are numbered 1 through 16. Channels 1 through 4 are in group 1, channels 5 through 8 are in group 2, and so on. Table 2 defines the relationship between CH1~CH4 (UDW2~UDW17) in the audio data packet and the channel/sample number for 32 kHz to 48 kHz sampling and 96-kHz sampling respectively.

5.6 The audio data packet and audio control packet shall be located in SMPTE 292 transport HANC space that is equal to 268 clock pulses at 30 Hz video frame rate.

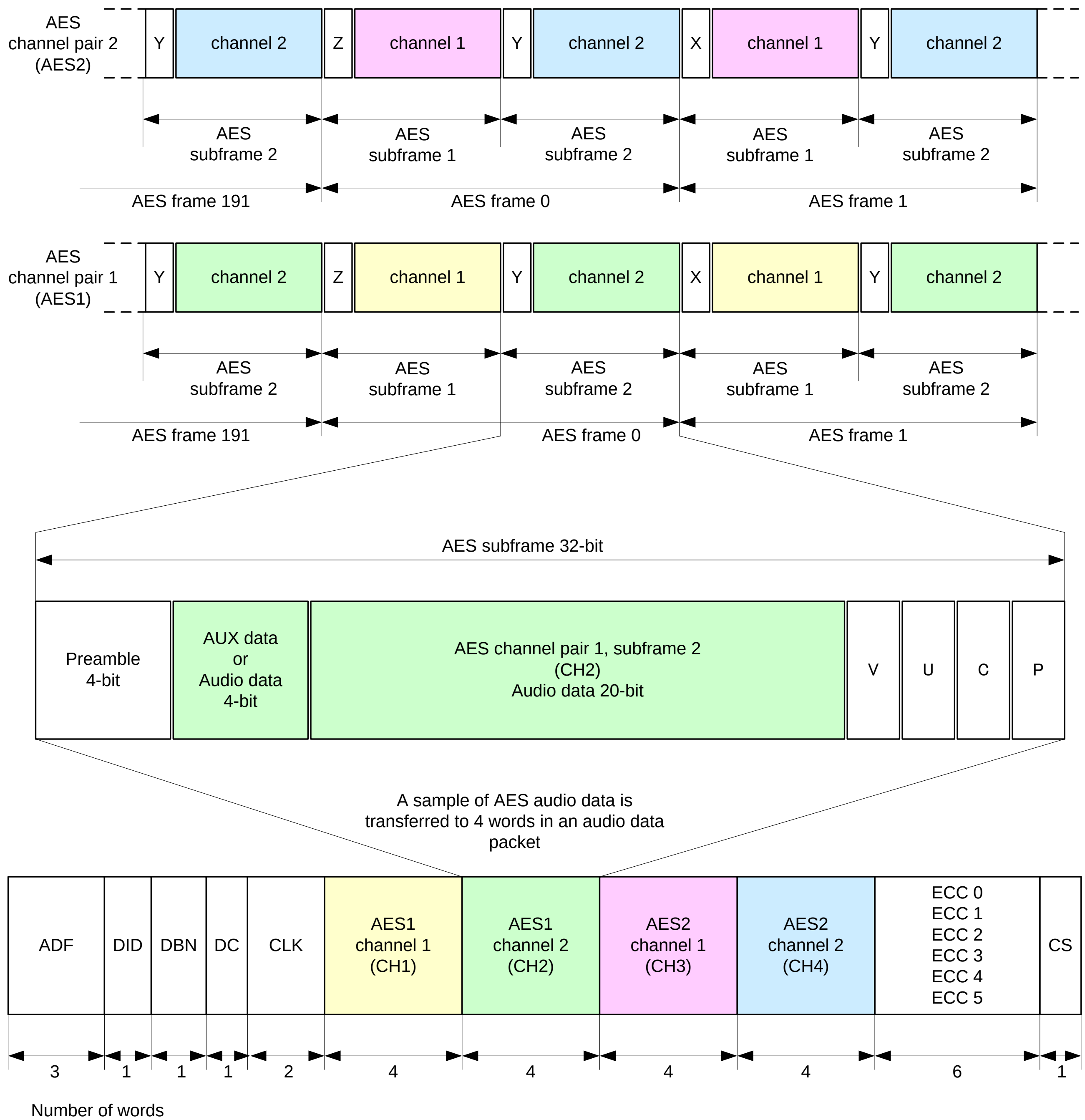


Figure 1 – Relationship between AES audio and audio data packets at sampling rates of 32 kHz, 44.1 kHz or 48 kHz

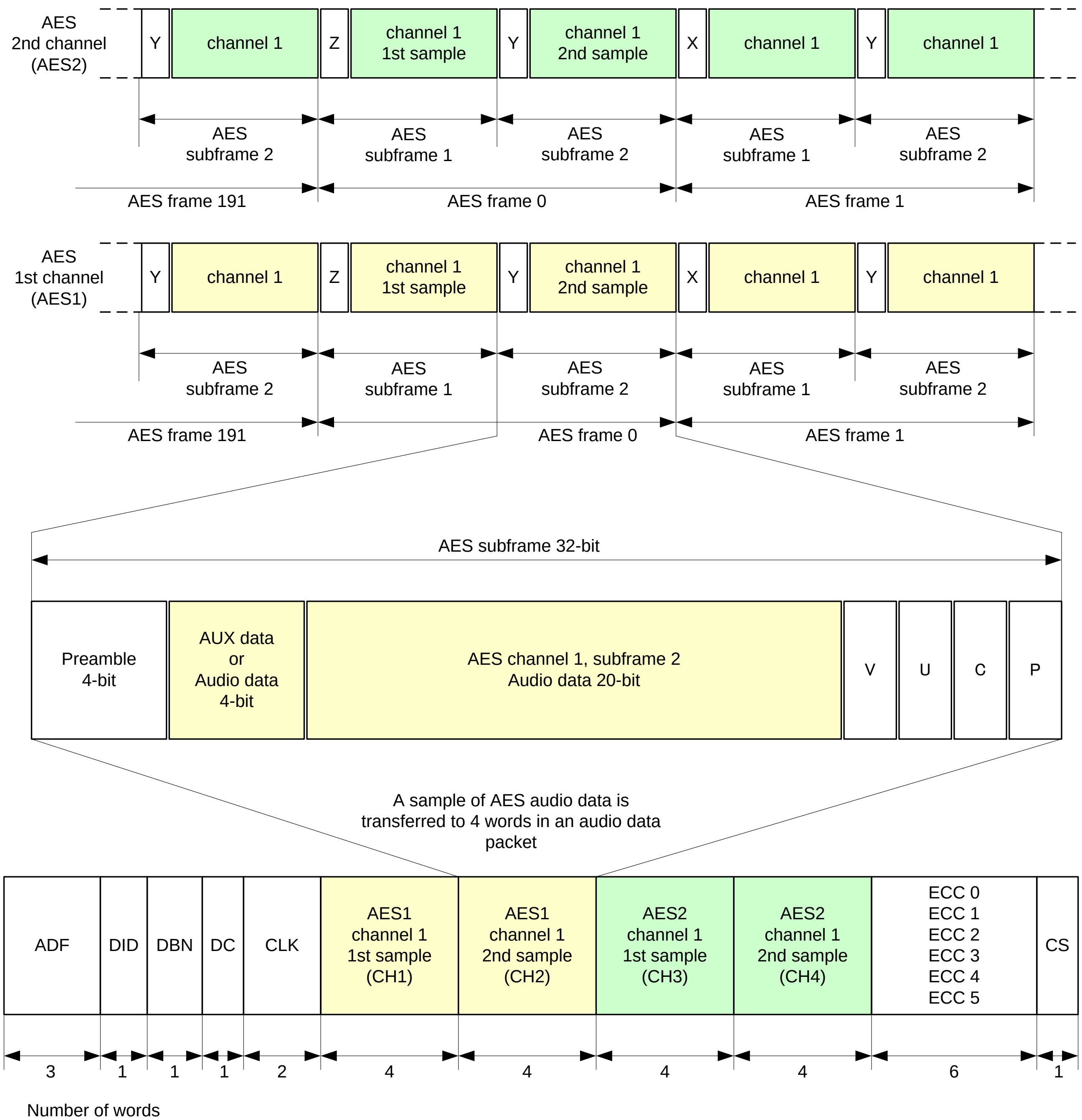


Figure 2 – Relationship between AES audio and audio data packets at a sampling rate of 96 kHz

Table 2 – Relationship between audio data packets and the channel/sample number of 32 kHz to 48 kHz and 96 kHz sampling

	Audio Group 1			
Audio sampling rate	UDW2~UDW5 CH1	UDW6~UDW9 CH2	UDW10~UDW13 CH3	UDW14~UDW17 CH4
32.0 kHz, 44.1 kHz or 48.0 kHz	AES1 channel 1	AES1 channel 2	AES2 channel 1	AES2 channel 2
96.0 kHz	AES1 channel 1 1st sample	AES1 channel 1 2nd sample	AES2 channel 1 1st sample	AES2 channel 1 2nd sample

6 Audio Data Packet

6.1 Structure of Audio Data Packet

6.1.1 The structure of the audio data packet shall be as shown in Figure 3. Audio data packets shall be formatted according to the requirements of SMPTE 291M and shall include ancillary data flag (ADF), data identification (DID), data block number (DBN), data count (DC), user data words (UDW) and checksum (CS) fields as specified in SMPTE 291M. DC is always 218_h.

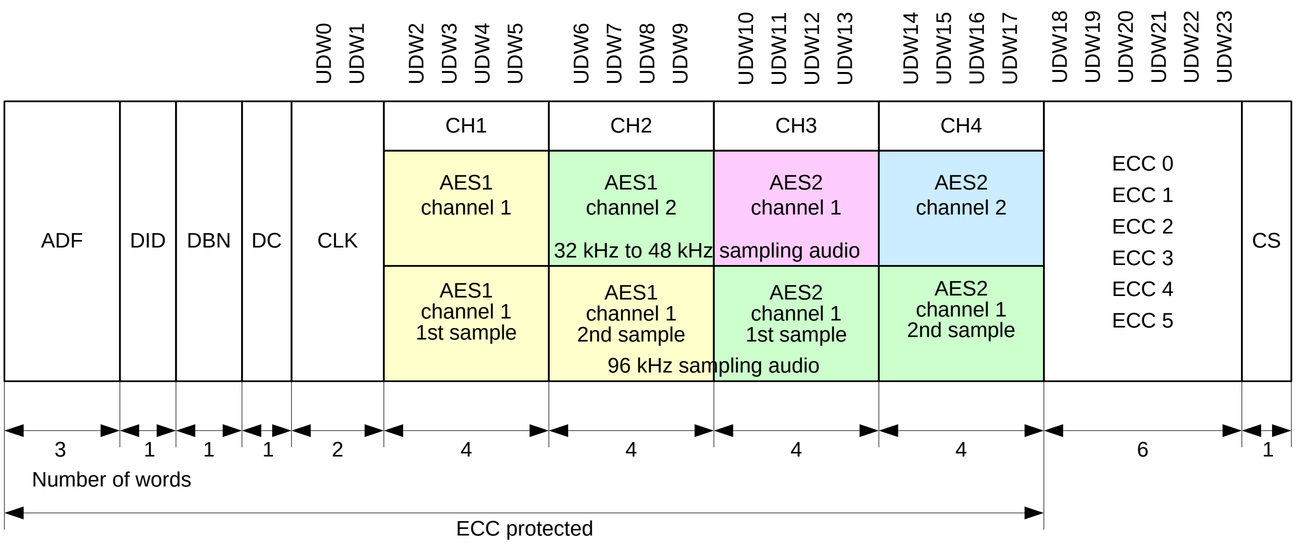


Figure 3 – Structure of audio data packets

6.1.2 The DID value shall be defined as $2E7_h$ for audio group 1 (channel 1~4), $1E6_h$ for audio group 2 (channel 5~8), $1E5_h$ for audio group 3 (channel 9-12) and $2E4_h$ for audio group 4 (channel 13~16), respectively.

6.1.3 The UDW is defined in § 6.2. In this standard, UDWx means the xth user data word. There shall always be 24 words in the UDW of an audio data packet; i.e., UDW0, UDW1 … UDW22, UDW23.

6.1.4 All audio channels in a given audio group shall have identical sampling rate, identical sampling phase and identical synchronous/asynchronous status.

6.1.5 For a given audio data packet, one sample of the audio data of each channel (CH1~CH4) shall be transmitted. Even when only one of the four channels (CH1~CH4) is active, all audio data of the 4 channels shall be transmitted. In such case, the value of audio data, V, U, C and P bits of all inactive channels shall be set to zero.

6.2 Structure of User Data Words (UDW)

The UDW shall consist of the three kinds of data defined in § 6.2.1, § 6.2.2 and § 6.2.3. The description in this clause covers only audio group 1. The description for audio groups 2, 3 and 4 is similar to that for audio group 1 where channels 5, 9 and 13 correspond to channel 1; channels 6, 10 and 14 correspond to channel 2; channels 7, 11 and 15 correspond to channel 3; and channels 8, 12 and 16 correspond to channel 4, respectively.

6.2.1 CLK (audio clock phase data)

6.2.1.1 Bit assignment of CLK shall be as shown in Table 3. Valid CLK data is required.

Table 3 – Bit assignment of CLK

Bit number	UDW0	UDW1
b9 (MSB)	Not b8	Not b8
b8	Even parity [1)]	Even parity [1)]
b7	ck7 audio clock phase data	Reserved (set to 0)
b6	ck6 audio clock phase data	Reserved (set to 0)
b5	ck5 audio clock phase data	ck12 audio clock phase data (MSB)
b4	ck4 audio clock phase data	mpf multiplex position flag
b3	ck3 audio clock phase data	ck11 audio clock phase data
b2	ck2 audio clock phase data	ck10 audio clock phase data
b1	ck1 audio clock phase data	ck9 audio clock phase data
b0 (LSB)	ck0 audio clock phase data (LSB)	ck8 audio clock phase data

[1)] Even parity for b0 through b7.

6.2.1.2 The bits of ck0 to ck12 shall indicate the number of video clocks between the first word of EAV and the video sample at the same time that audio sample appears at the input of the formatter. For example, the value of ck0~12 is in the range of 0 to 8191 for systems that use 74.25 MHz or 74.25/1.001 MHz clocks covered by SMPTE 292. Examples of the relationship among video, sampling instants of digital audio and audio clock phase data for a 1080/60i system are shown in Figure 4 (30 Hz frame rate), Figure 5 (30/1.001 Hz frame rate), and Figure 6 (96 kHz sampling and 30 Hz frame rate).

In the case of 96 kHz sampling, CLK (the audio clock phase data) indicates the number of video clocks between the first word of EAV and the video sample at the same time that the second audio sample of the successive two samples of the same AES audio signal appears at the input of the formatter.

Note: Designers should recognize that some existing equipment may not recognize or support bit ck12 (see Annex B).

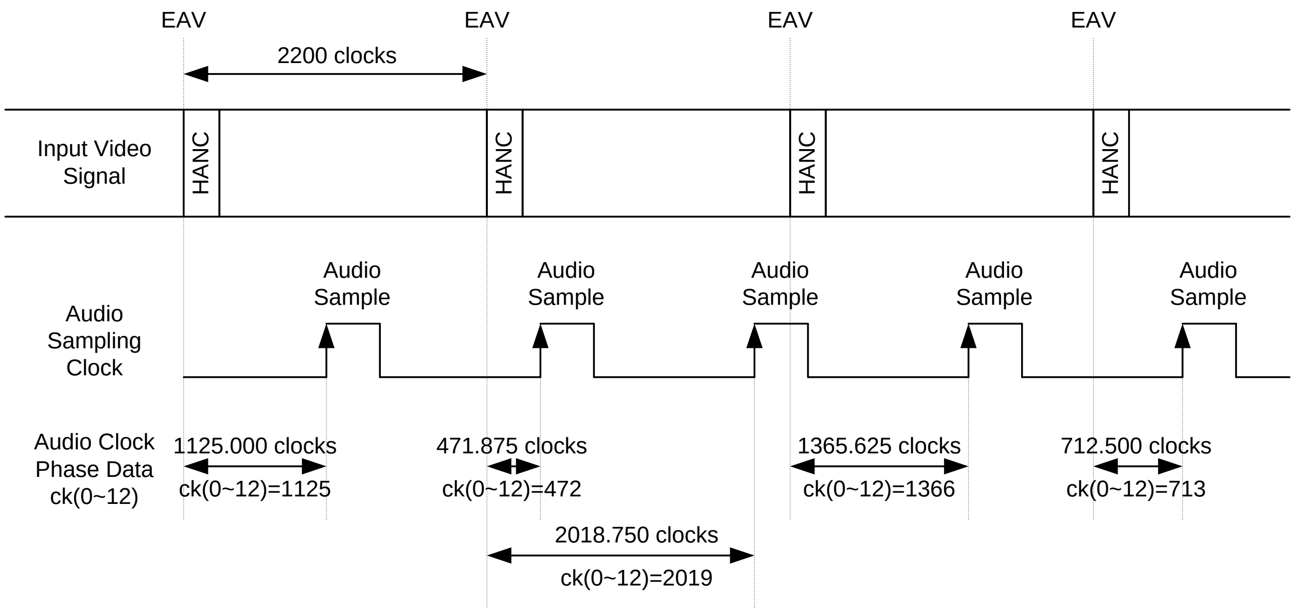


Figure 4 – Relationship between video lines, sampling instants of digital audio and audio clock phase data (informative example – 1080/60i system with 48 kHz audio sampling rate and 30.00 Hz video frame rate) [1]

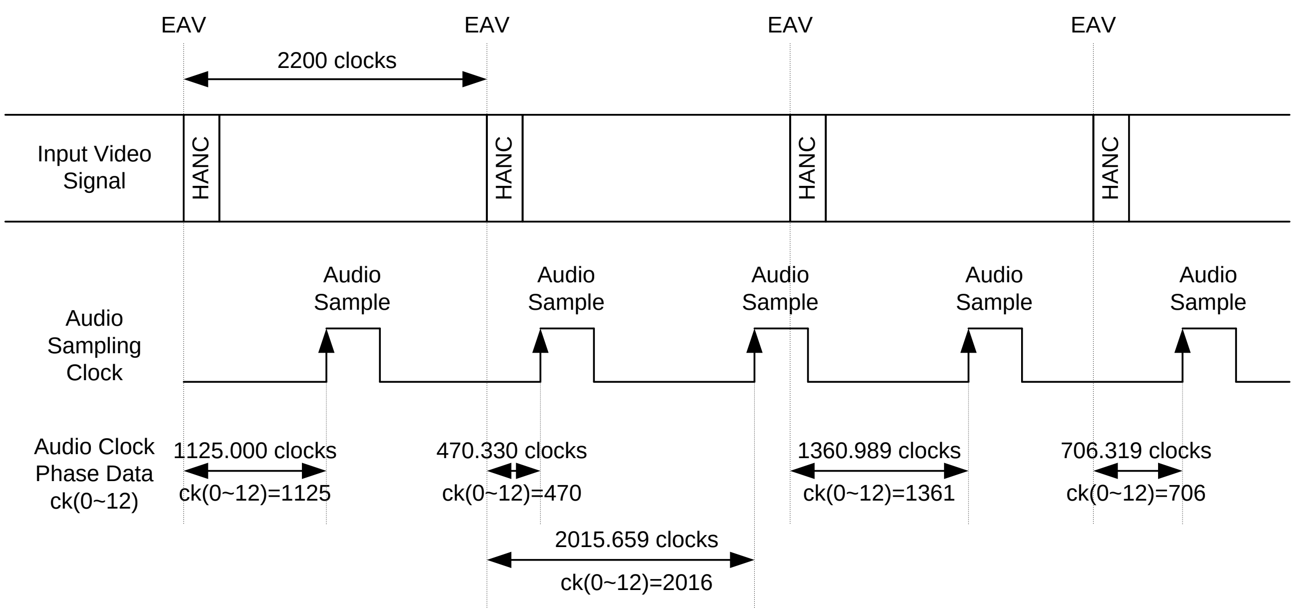


Figure 5 – Relationship between video lines, sampling instants of digital audio and audio clock phase data (informative example – 1080/60i system with 48 kHz audio sampling rate and 30.00/1.001 Hz video frame rate) [1]

[1] Note: In Figure 4 and Figure 5, "clocks" refers to the video sampling clock as defined in § 6.2.1.2.

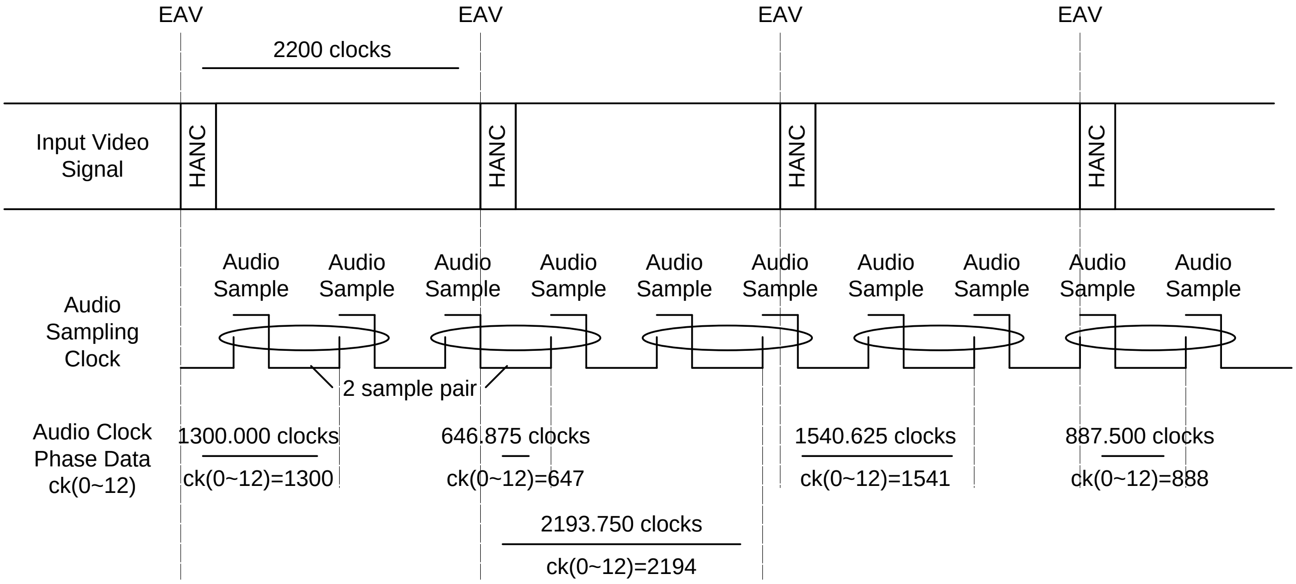


Figure 6 – Relationship between video lines, sampling instants of digital audio and audio clock phase data (informative example – 1080/60i system with 96 kHz audio sampling rate and 30.00 Hz video frame rate)[1]

[1] Note: In Figure 4, Figure 5 and Figure 6, "clocks" refers to the video sampling clock as defined in § 6.2.1.2.

6.2.1.3 The formatter shall place the audio data packet in the horizontal ancillary space following the video line during which the audio sample occurred. Following a switching point, the audio data packet shall be delayed one additional line to prevent data corruption.

Flag bit *mpf* defines the audio data packet position in the multiplexed output stream relative to the associated video data.

When bit *mpf* = 0, it shall indicate that the audio data packet is located immediately after the video line during which the audio sample occurred.

When bit *mpf* = 1, it shall indicate that the audio data packet is located in the second line following the video line during which the audio sample occurred.

The relationship between the multiplex position flag (mpf) and the multiplex position of the audio data packet is shown in Figure 7 and Figure 8.

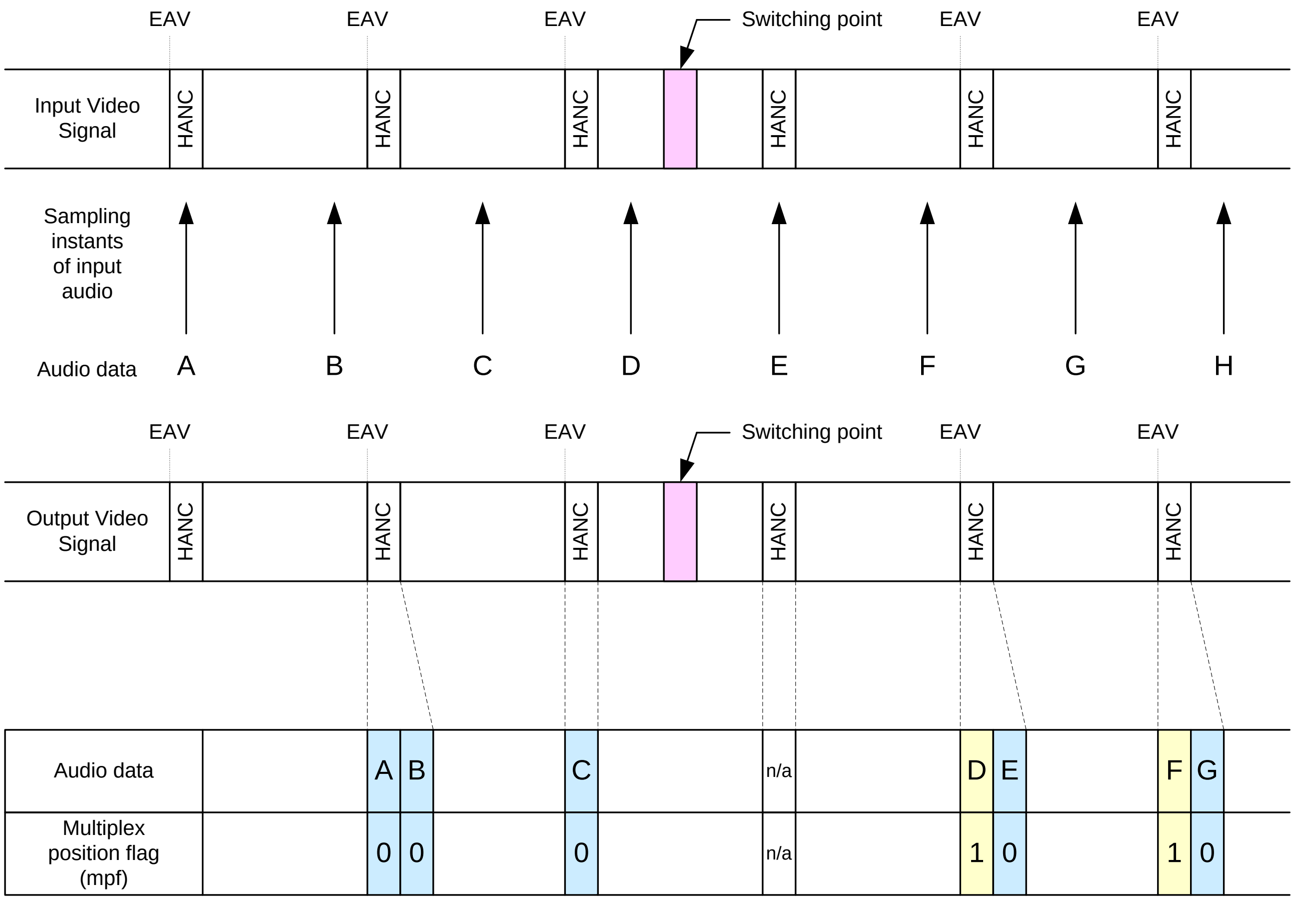
In the case of 96 kHz sampling, *mpf* shall be defined according to the position of the second sample of the successive two samples of the same AES audio signal.

6.2.2 CHn (audio data)

6.2.2.1 The bit assignment of CHn (n = 1~4) shall be as shown in Table 4. All bits of an AES subframe shall be transparently transferred to four consecutive UDW words (UDW4n-2, UDW4n-1, UDW4n, UDW4n+1). UDW2 through UDW17 are always used for CHn in audio data packets.

6.2.2.2 Bit 3 of UDW2 and UDW10 indicates the status of the Z flag which corresponds to the AES block sync. The Z flag bit in UDW2 shall be associated with CH1 and CH2, and the Z flag bit in UDW10 shall be associated with CH3 and CH4,

6.2.2.3 Bits b0 through b2 in UDW2, UDW6, UDW10 and UDW14, and bit b3 in UDW6 and UDW14 shall be set to zero.



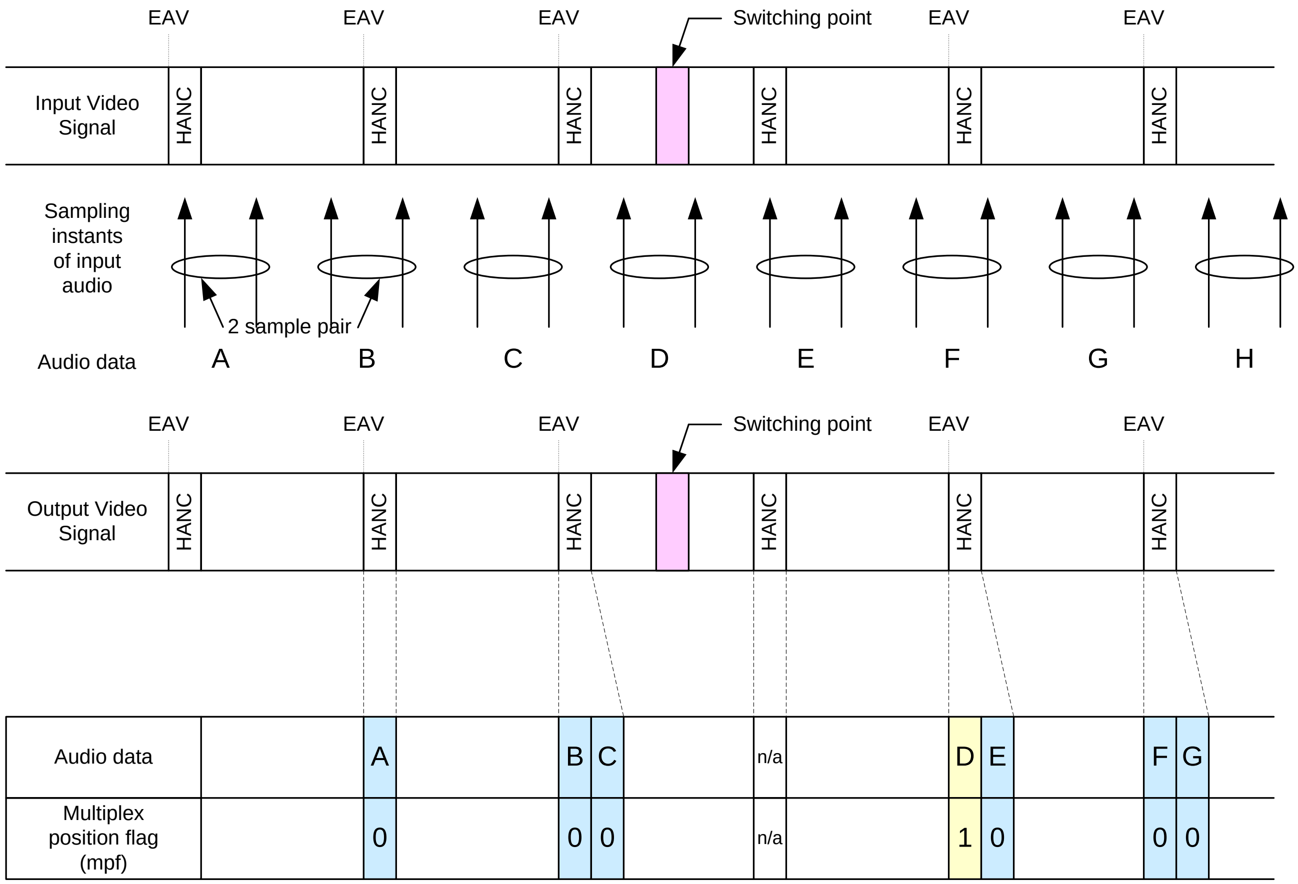
Notes:

1. For example, for samples A, B, C, E and G, mpf = 0 because the ancillary data packet is multiplexed in the horizontal ancillary data space of the next line relative to the input timing of the audio sample.

2. N/A shows that the line subsequent to the switching point precludes the insertion of ancillary data packets.

3. For example, for samples D and F, mpf = 1 because the ancillary data packet is multiplexed in the horizontal ancillary data space of the second line relative to the input timing of the audio sample.

Figure 7 – Relationship between the multiplex position flag (mpf) and the multiplex position of 32 kHz to 48 kHz sampling audio data packets



Notes:

1. For example, for samples A, B, C, E, F and G, mpf = 0 because the ancillary data packet is multiplexed in the horizontal ancillary data space of the next line relative to the input timing of the audio sample.

2. N/A shows that the line subsequent to the switching point precludes the insertion of ancillary data packets.

3. For example, for samples D, mpf = 1 because the ancillary data packet is multiplexed in the horizontal ancillary data space of the second line relative to the input timing of the audio sample.

Figure 8 – Relationship between the multiplex position flag (mpf) and the multiplex position of 96 kHz sampling audio data packets

Table 4 – Bit-assignment of audio data (CHn)

	Bit number	UDW2	UDW3	UDW4	UDW5
CH1	b9 (MSB)	Not b8	Not b8	Not b8	Not b8
	b8	Even parity [1]	Even parity [1]	Even parity [1]	Even parity [1]
	b7	aud_1 3	aud_1 11	aud_1 19	P_1
	b6	aud_1 2	aud_1 10	aud_1 18	C_1
	b5	aud_1 1	aud_1 9	aud_1 17	U_1
	b4	aud_1 0 (LSB)	aud_1 8	aud_1 16	V_1
	b3	Z	aud_1 7	aud_1 15	aud_1 23(MSB)
	b2	0	aud_1 6	aud_1 14	aud_1 22
	b1	0	aud_1 5	aud_1 13	aud_1 21
	b0 (LSB)	0	aud_1 4	aud_1 12	aud_1 20
	Bit number	UDW6	UDW7	UDW8	UDW9
CH2	b9 (MSB)	Not b8	Not b8	Not b8	Not b8
	b8	Even parity [1]	Even parity [1]	Even parity [1]	Even parity [1]
	b7	aud_2 3	aud_2 11	aud_2 19	P_2
	b6	aud_2 2	aud_2 10	aud_2 18	C_2
	b5	aud_2 1	aud_2 9	aud_2 17	U_2
	b4	aud_2 0 (LSB)	aud_2 8	aud_2 16	V_2
	b3	0	aud_2 7	aud_2 15	aud_2 23(MSB)
	b2	0	aud_2 6	aud_2 14	aud_2 22
	b1	0	aud_2 5	aud_2 13	aud_2 21
	b0 (LSB)	0	aud_2 4	aud_2 12	aud_2 20
	Bit number	UDW10	UDW11	UDW12	UDW13
CH3	b9 (MSB)	Not b8	Not b8	Not b8	Not b8
	b8	Even parity [1]	Even parity [1]	Even parity [1]	Even parity [1]
	b7	aud_3 3	aud_3 11	aud_3 19	P_3
	b6	aud_3 2	aud_3 10	aud_3 18	C_3
	b5	aud_3 1	aud_3 9	aud_3 17	U_3
	b4	aud_3 0 (LSB)	aud_3 8	aud_3 16	V_3
	b3	Z	aud_3 7	aud_3 15	aud_3 23(MSB)
	b2	0	aud_3 6	aud_3 14	aud_3 22
	b1	0	aud_3 5	aud_3 13	aud_3 21
	b0 (LSB)	0	aud_3 4	aud_3 12	aud_3 20
	Bit number	UDW14	UDW15	UDW16	UDW17
CH4	b9 (MSB)	Not b8	Not b8	Not b8	Not b8
	b8	Even parity [1]	Even parity [1]	Even parity [1]	Even parity [1]
	b7	aud_4 3	aud_4 11	aud_4 19	P_4
	b6	aud_4 2	aud_4 10	aud_4 18	C_4
	b5	aud_4 1	aud_4 9	aud_4 17	U_4
	b4	aud_4 0 (LSB)	aud_4 8	aud_4 16	V_4
	b3	0	aud_4 7	aud_4 15	aud_4 23(MSB)
	b2	0	aud_4 6	aud_4 14	aud_4 22
	b1	0	aud_4 5	aud_4 13	aud_4 21
	b0 (LSB)	0	aud_4 4	aud_4 12	aud_4 20

Notes:

1 Even parity for b0 through b7.
2 Z = AES block sync.
3 Un = AES user bit of CHn.
4 Pn = AES parity bit of CHn.
5 aud (0-23) =24-bit AES audio data of CHn.
6 Vn = AES sample validity bit of CHn.
7 Cn = AES channel status bit of CHn.
8 Value of Vn, Un, Cn and Pn is equal to that of AES subframe, respectively.

6.2.3 ECC (Error correction codes)

6.2.3.1 ECC shall be used to correct or detect errors in 24 words from the first word of ADF through UDW17. The error correction code is BCH (31, 25) code. The BCH code shall be formed for each bit sequence of b0 – b7,. The ECC shall consist of six words determined by the polynomial generator equation:

$$ECC(X) = (X+1)(X^5+X^2+1) = X^6+X^5+X^3+X^2+X+1$$

The initial value of all FFn shall be set to zero. The calculation shall start at the first word of ADF and shall end at the final word of CH4 (UDW17) for each bit of b0 to b7, respectively. The remaining data in the FFn shall be ECCn (n = 0~5).

Note: FFn is the flip flop number. For example, the data of FF0 is ECC0; and the data of FF5 is ECC5.

6.2.3.2 Bit assignment of ECC shall be as shown in Table 5. An example of the block diagram of the BCH-code information circuit is shown in Figure 9.

Table 5 – Bit assignment of ECC

	UDW18	UDW19	UDW20	UDW21	UDW22	UDW23
Bit number	ECC0	ECC1	ECC2	ECC3	ECC4	ECC5
b9 (MSB)	Not b8	Not b8	Not b8	Not b8	Not b8	Not b8
b8	Even parity [1)]	Even parity [1)]	Even parity [1)]	Even parity [1)]	Even parity [1)]	Even parity [1)]
b7	ecc0 7	ecc1 7	ecc2 7	ecc3 7	ecc4 7	ecc5 7
b6	ecc0 6	ecc1 6	ecc2 6	ecc3 6	ecc4 6	ecc5 6
b5	ecc0 5	ecc1 5	ecc2 5	ecc3 5	ecc4 5	ecc5 5
b4	ecc0 4	ecc1 4	ecc2 4	ecc3 4	ecc4 4	ecc5 4
b3	ecc0 3	ecc1 3	ecc2 3	ecc3 3	ecc4 3	ecc5 3
b2	ecc0 2	ecc1 2	ecc2 2	ecc3 2	ecc4 2	ecc5 2
b1	ecc0 1	ecc1 1	ecc2 1	ecc3 1	ecc4 1	ecc5 1
b0 (LSB)	ecc0 0	ecc1 0	ecc2 0	ecc3 0	ecc4 0	ecc5 0

[1)] Even parity for b0 through b7.

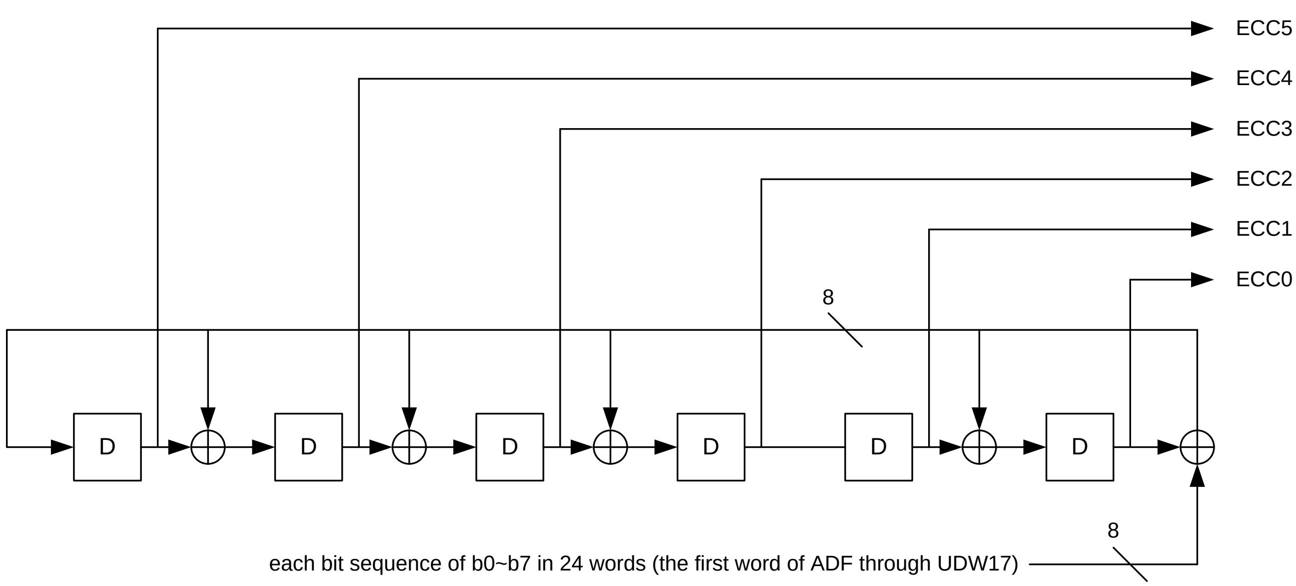


Figure 9 – Block diagram of the BCH-code formation circuitry (informative example)

6.3 Multiplexing of Audio Data Packet

6.3.1 Only the horizontal ancillary data space of the color-difference data stream (C_b/C_r) shall be used for transmission of the audio data packet.

6.3.2 The audio data packet shall not be multiplexed into the horizontal ancillary data space of the line subsequent to the switching point defined by the source format. As an example, the ancillary data space available for audio data packet in the 1080/60i system is shown in Figure 10.

6.3.3 The number of samples per audio channel which can be multiplexed in one horizontal ancillary data space shall be less than or equal to Na (Number of audio samples), where Na is defined in the following pseudocode:

```
No = Int (audio sample rate/line frequency) + 1
if (( No × (the number of total lines per video frame – the number of switching line per video frame)
   < (the number of audio samples per video frame )
then  Na = No + 1
else  Na = No
if (audio sampling rate == 96 kHz)  Na = Even(Na)
```

The function Even(n) returns the smallest even number that is greater than or equal to n. For example, Even(123)=124, Even(98)=98.

When two or more samples of the audio data are transmitted in one horizontal ancillary data block, the packet of the audio sample which appears earlier at the input of the formatter shall be transmitted first.

Note: Some video formats may require up to 8 samples per data block (i.e. Na=8).

6.3.4 An audio data packet shall be multiplexed in the horizontal ancillary data space of the first or second line following the line during which the audio sample occurred at the input of the formatter.

Note: Audio phase needs to be maintained across the audio groups carrying multiple-channel audio.

6.3.5 The audio data packet shall be multiplexed following the CRC, which is defined in SMPTE 292.

6.3.6 When more than two audio data packets are transmitted in one horizontal ancillary data block, the audio data packets shall be contiguous with each other.

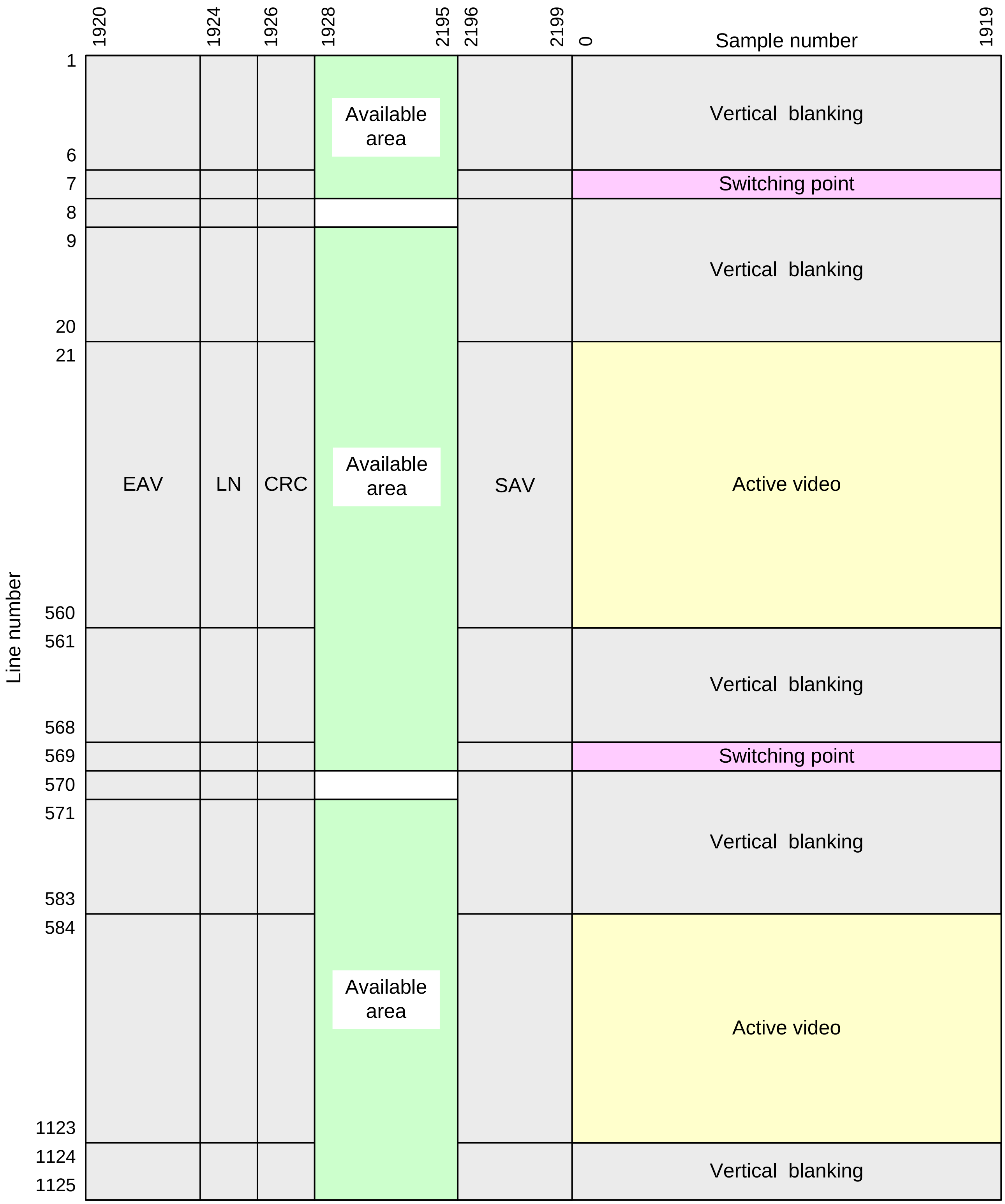


Figure 10 – Ancillary data space of C_b/C_r data stream available for transmission of audio data packets (1080/60i system)

7 Audio Control Packet

7.1 Structure of Audio Control Packet

7.1.1 The structure of the audio control packet shall be as shown in figure 11. Audio control packets shall be formatted according to the requirements of SMPTE 291M and shall include ancillary data flag (ADF), data identification (DID), data block number (DBN), data count (DC), user data words (UDW) and checksum (CS) fields as specified in SMPTE 291M. DC is always $10B_h$ and DBN is always 200_h.

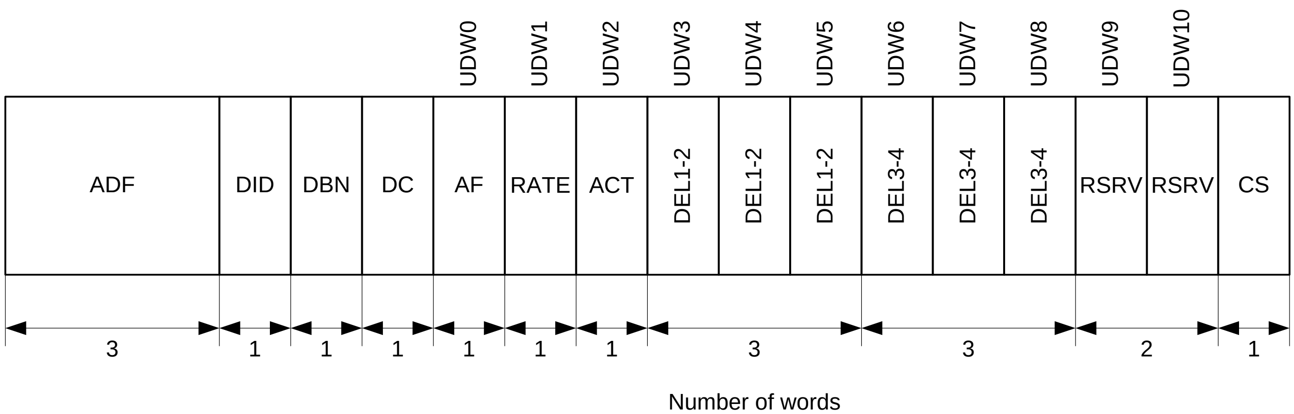


Figure 11 – Structure of audio control packet

7.1.2 The DID shall be defined as $1E3_h$ for audio group 1 (channels 1~4), $2E2_h$ for audio group 2 (channels 5~8), $2E1_h$ for audio group 3 (channel 9~12) and $1E0_h$ for audio group 4 (channel 13~16), respectively.

7.1.3 The UDW is defined in § 7.2. In this standard, UDWx means the xth user data word. There shall be always 11 words in the UDW of an audio control packet; i.e., UDW0, UDW1 … UDW9, UDW10.

7.2 Structure of User Data Words (UDW)

The UDW consists of five types of data defined in § 7.2.1 through § 7.2.5. The description in this section covers only audio group 1. The description for audio groups 2, 3 and 4 is similar to audio group 1 where channels 5, 9 and 13 correspond to channel 1; channels 6, 10 and 14 correspond to channel 2; channels 7, 11 and 15 correspond to channel 3; channels 8, 12 and 16 correspond to channel 4, respectively.

7.2.1 AF (audio frame number data)

7.2.1.1 The Audio frame number data (AF) provides a sequential numbering of audio frames to indicate where they fall in the progression of non-integer number of samples per video frame (audio frame sequence). The first number of the sequence shall be always 1 and the final number shall be equal to the length of the audio frame sequence. A value of AF equal to all zeros shall indicate that frame numbering is not available.

7.2.1.2 The bit-assignment of the AF shall be as shown in Table 6. The AF shall be a common value for all channels in a given audio group.

Table 6 – Bit-assignment of AF

Bit number	UDW0
	AF
b9 (MSB)	not b8
b8	f8 audio frame number (MSB)
b7	f7 audio frame number
b6	f6 audio frame number
b5	f5 audio frame number
b4	f4 audio frame number
b3	f3 audio frame number
b2	f2 audio frame number
b1	f1 audio frame number
b0 (LSB)	f0 audio frame number (LSB)

7.2.1.3 For correct use of the audio frame number, the audio frame sequence shall be defined. Four synchronous sampling rates are defined in this standard (see § 4.16).

All audio frame sequences shall be based on two integer numbers of samples per frame (m and m+1) with audio frame numbers starting at 1 and proceeding to the end of the sequence. Odd-numbered audio frame (1, 3, 5, etc.) have the larger integer number of samples and even-numbered audio frames (2, 4, 6, etc.) have the smaller integer number of samples with the exception tabulated in Table 7.

Table 7 – Exceptions to audio frame sequences

Television system	Sampling rate (kHz)	Frame sequence	Basic numbering		Exceptions	
			Samples per odd audio frame (m)	Samples per even audio frame (m+1)	Frame number	Number of samples
30 frame/s	96.0	1	3200		none	
	48.0	1	1600		none	
	44.1	1	1470		none	
	32.0	3	1067	1066	none	
30.00/1.001 frame/s	96.0	5	3204	3202 [1)]	none	
	48.0	5	1602	1601	none	
	44.1	100	1472	1471	23, 47, 71	1471
	32.0	15	1068	1067	4, 8, 12	1068

[1)] Successive samples are carried in audio data packet.

Note: Receiver designers should be aware that some existing equipment may not conform to the sequence restriction specifications of § 7.2.1.3. Receivers should have the ability to receive audio data sequences correctly even when § 7.2.1.3 is not implemented.

7.2.1.4 When channel pairs in a given audio group are operating in asynchronous mode, the AF word in the audio control packet is not used and b0 – b8 should be set to zero.

7.2.2 RATE (Sampling rate)

7.2.2.1 The sampling rate for all channel pairs is defined by the word RATE. The bit assignment of RATE shall be as shown in Table 8.

Table 8 – Bit assignment of RATE

Bit number	UDW1
	RATE
b9 (MSB)	not b8
b8	Reserved (set to 0)
b7	Reserved (set to 0)
b6	Reserved (set to 0)
b5	Reserved (set to 0)
b4	Reserved (set to 0)
b3	X2 (MSB)
b2	X1 Rate code
b1	X0 (LSB)
b0 (LSB)	asx 0 = synchronous audio 1 = asynchronous audio

7.2.2.2 The sync mode bit asx, when set to one, shall indicate that the channel pairs in a given audio group are operating asynchronously.

7.2.2.3 The rate code shall be as defined in Table 9.

Table 9 – Assignment of rate code

X2	X1	X0	Sample rate
0	0	0	48.0 kHz
0	0	1	44.1 kHz
0	1	0	32.0 kHz
1	0	0	96.0 kHz
0	1	1	Reserved
1	0	1	Reserved
1	1	0	Reserved
1	1	1	Free running

7.2.3 ACT

The word ACT shall indicate the active channels. Bits a1 to a4 shall be set to one for each active channel in a given audio group; otherwise, they shall be set to zero. The bit assignment of ACT shall be as defined in Table 10.

Table 10 – Bit assignment of ACT

Bit number	UDW2
	ACT
b9 (MSB)	Not b8
b8	Even parity [1)]
b7	Reserved (set to 0)
b6	Reserved (set to 0)
b5	Reserved (set to 0)
b4	Reserved (set to 0)
b3	a4 active: 1, inactive: 0 (CH4)
b2	a3 active: 1, inactive: 0 (CH3)
b1	a2 active: 1, inactive: 0 (CH2)
b0 (LSB)	a1 active: 1, inactive: 0 (CH1)

[1)] Even parity for b0 through b7.

7.2.4 DELm-n

7.2.4.1 The words DELm-n shall indicate the amount of accumulated audio processing delay relative to video, measured in audio sample intervals, for each channel pair of CHm and CHn.

In the case of 96 kHz sampling, DELm-n shall indicate the amount of accumulated audio processing delay relative to video measured in audio sample intervals for the successive two samples of the same AES audio signal carried in CH1, CH2 and CH3, CH4.

7.2.4.2 The bit assignment of DELm-n shall be as shown in Table 11. The *e* bit shall be set to one to indicate valid audio delay data otherwise it shall be zero. The delay words are referenced to the point where the AES/EBU data are input to the formatter. The delay words shall represent the average delay value, inherent in the formatting process, over a period no less than the length of the audio frame sequence plus any preexisting audio delay.

Table 11 – Bit assignment of DELm-n

Bit number	UDW3	UDW4	UDW5	UDW6	UDW7	UDW8
	DEL1-2			DEL3-4		
b9 (MSB)	Not b8	Not b8	Not b8	Not b8	Not b8	Not b8
b8	del 7	del 16	del 25 (±)	del 7	del 16	del 25 (±)
b7	del 6	del 15	del 24 (MSB)	del 6	del 15	del 24 (MSB)
b6	del 5	del 14	del 23	del 5	del 14	del 23
b5	del 4	del 13	del 22	del 4	del 13	del 22
b4	del 3	del 12	del 21	del 3	del 12	del 21
b3	del 2	del 11	del 20	del 2	del 11	del 20
b2	del 1	del 10	del 19	del 1	del 10	del 19
b1	del 0 (LSB)	del 9	del 18	del 0 (LSB)	del 9	del 18
b0 (LSB)	*e*	del 8	del 17	*e*	del 8	del 17

7.2.4.3 The audio delay data (del 0 – del 25) shall be represented in the format of 26-bit two's complement. Positive values shall indicate that the video leads the audio.

7.2.5 RSRV

7.2.5.1 The words marked RSRV are reserved for future use.

7.2.5.2 The bit assignment of RSRV word shall be as shown in Table 12.

Table 12 – Bit assignment of RSRV

	UDW9	UDW10
Bit number	RSRV	RSRV
b9 (MSB)	Not b8	Not b8
b8	Reserved (set to 0)	Reserved (set to 0)
b7	Reserved (set to 0)	Reserved (set to 0)
b6	Reserved (set to 0)	Reserved (set to 0)
b5	Reserved (set to 0)	Reserved (set to 0)
b4	Reserved (set to 0)	Reserved (set to 0)
b3	Reserved (set to 0)	Reserved (set to 0)
b2	Reserved (set to 0)	Reserved (set to 0)
b1	Reserved (set to 0)	Reserved (set to 0)
b0 (LSB)	Reserved (set to 0)	Reserved (set to 0)

7.3 Multiplexing of the Audio Control Packets

7.3.1 The audio control packets shall be transmitted once every field in an interlaced system and once per frame in a progressive system.

7.3.2 The audio control packets shall be transmitted in the horizontal ancillary data space of the second line after the switching point of Y data stream. For example, since the switching point for 1080/60i system exists in line 7 and 569, the audio control packets are transmitted in the horizontal ancillary data space of line 9 and line 571 of the Y data stream. Ancillary data space available for the transmission of audio control packets for this example is shown in Figure 12.

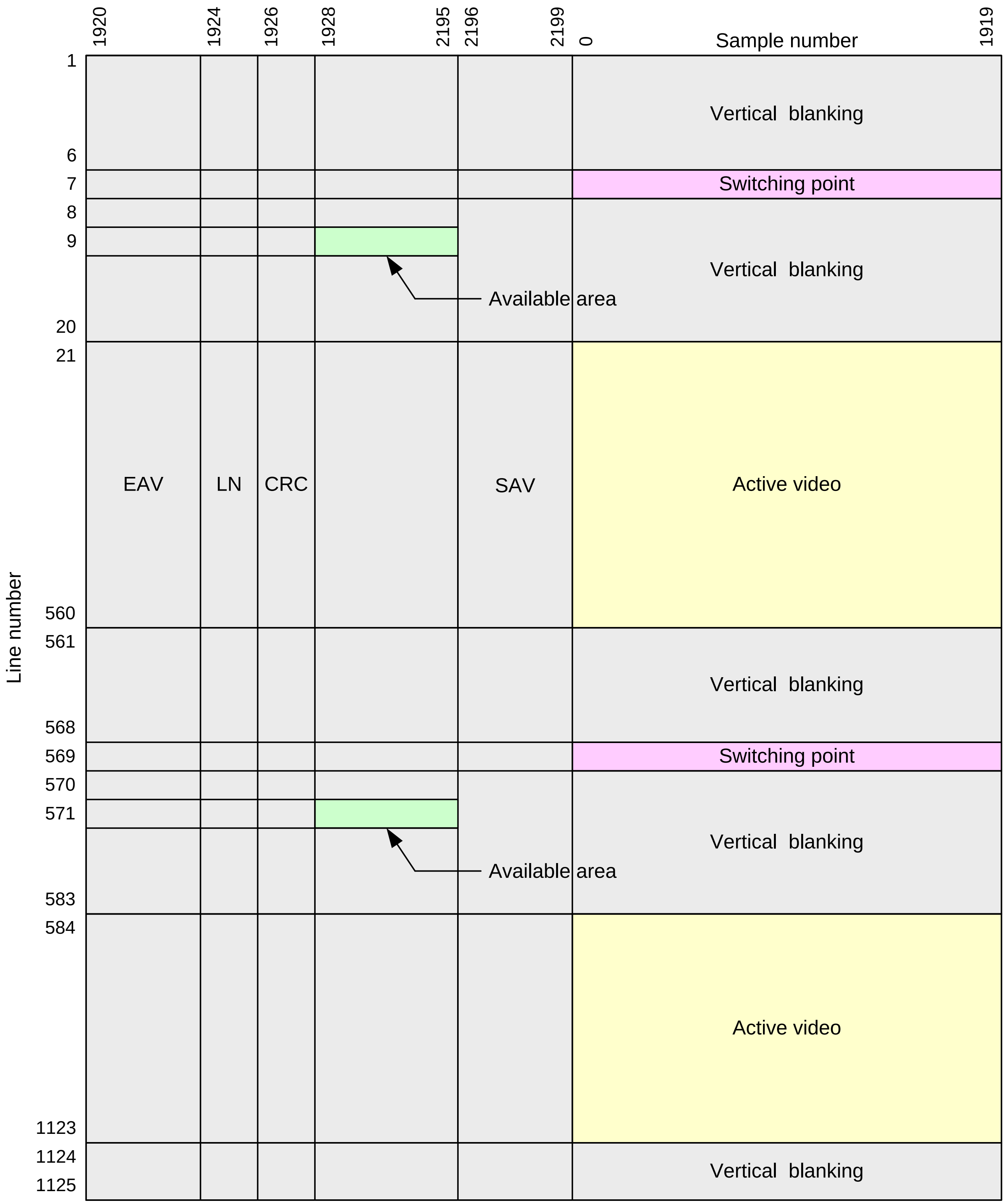


Figure 12 – Ancillary data space of Y data stream available for transmission of audio control packets (1080/60i system)

Annex A (Informative)
Recommendations for Handling of SMPTE 337 Non-PCM Data

While this standard is written in terms of the AES data containing linear PCM audio, the AES data may contain SMPTE 337 formatted data which may include compressed (bit-rate reduced) audio or other types of non-audio data. Implementers should take this into consideration and when possible include support for SMPTE 337 data compatibility. Users should be aware that not all devices compliant with the standard will properly handle SMPTE 337 data. This section contains recommendations for compatibility with SMPTE 337 data that apply to both implementers and users.

A.1 Levels of Operation

It is recommended that operation be restricted to 48-kHz synchronous modes when SMPTE 337 data is present in the AES data. Sample rate conversion should not be used to implement synchronous operation.

A.2 PCM Processing

Any PCM type processing performed on the AES data stream within multiplexing or receiving devices should be defeated or bypassed when SMPTE 337 data is present as such processing will corrupt the SMPTE 337 data. Examples of PCM processing include gain changes, sample rate conversion, truncation, dithering, cross-fades, etc.

A.3 AES Channel Status Data

AES channel status words contain useful information for detecting and identifying the presence of SMPTE 337 data with the AES data stream. It is recommended that all devices compliant with this standard maintain and transmit AES channel status information that is present on their input.

A.4 Additional Receiving Device Recommendations

Receiving devices can include specific processing for handling dynamic bitstream changes, such as when the received SMPTE 292 signal has been switched. This processing can include handling of receive buffer overflow and underflow conditions resulting from the switch, especially in the case of 30/1.001 frame/s systems. Nominally this processing is meant to minimize the audibility of the disruption for linear PCM signals. Typical processing can include periodic AES data word (audio sample) drops or repeats to maintain receive buffer fullness and PCM type processing of additional AES data words to minimize the audibility of the drops or repeats. This processing is sub-optimal when dealing with SMPTE 337 data.

If possible it is recommended that receiving devices include the ability to detect the presence of SMPTE 337 data and restrict drop or repeat locations to AES data words not containing SMPTE 337 data. Any PCM processing should also be disabled to minimize modification of AES data words. If detection of SMPTE 337 data is not possible it is recommended that *whenever possible* drop or repeat locations be restricted to the AES data words immediately adjacent to the vertical interval switching area.

Annex B (Informative)
Recommendations for Handling Legacy Implementations

The 720/24p video format, as defined in SMPTE 296M, defines a total line length of 4125 pixels, which requires a resolution of 13 bits for the CLK audio clock phase data words. Early versions of SMPTE 299M (ANSI/SMPTE 299M-1997 and earlier) offer only a maximum of 12 bits resolution (ck(0~11)), providing a maximum CLK value of 4095. SMPTE 299M-2004 added an additional bit in the audio data packet as the MSB (ck12) for the audio clock phase data, providing 13 bits resolution

Note: In ANSI/SMPTE 299M-1997 and earlier, ck12 was designated as the multiplex position flag. Since SMPTE 299M-2004, this bit has been renamed mpf and its bit position in UDW1 remains unchanged from ANSI/SMPTE 299M-1997.

Some legacy devices may not support this additional clock bit in the audio clock phase data. Some legacy audio formatting devices can hold the clock phase value at a maximum value when reached, until reset at the end of the line. This will produce a small amount of audio phase jitter for the period of one sample. Alternatively, the audio clock phase value can wrap around through zero in these legacy formatting devices.

To overcome these issues, it is recommended that audio receiver implementations should check for all of the above cases. On detection of the maximum value, a comparison can be made between previous clock phases and the correct position interpolated. If the clock phase data value starts to decrease within the same video line, it is recommended that the decoder should check to see if bit 5 (ck12) of UDW1 in the audio data packet is set. If ck12 is set, the correct 13-bit value of the audio clock phase data should be used. If ck12 is not set, the correct position should be interpolated.

Figure B.1 shows an example of the relationship between the video clock, audio sampling instants and the audio clock phase data for a 720/24p video signal with 48 kHz audio sampling.

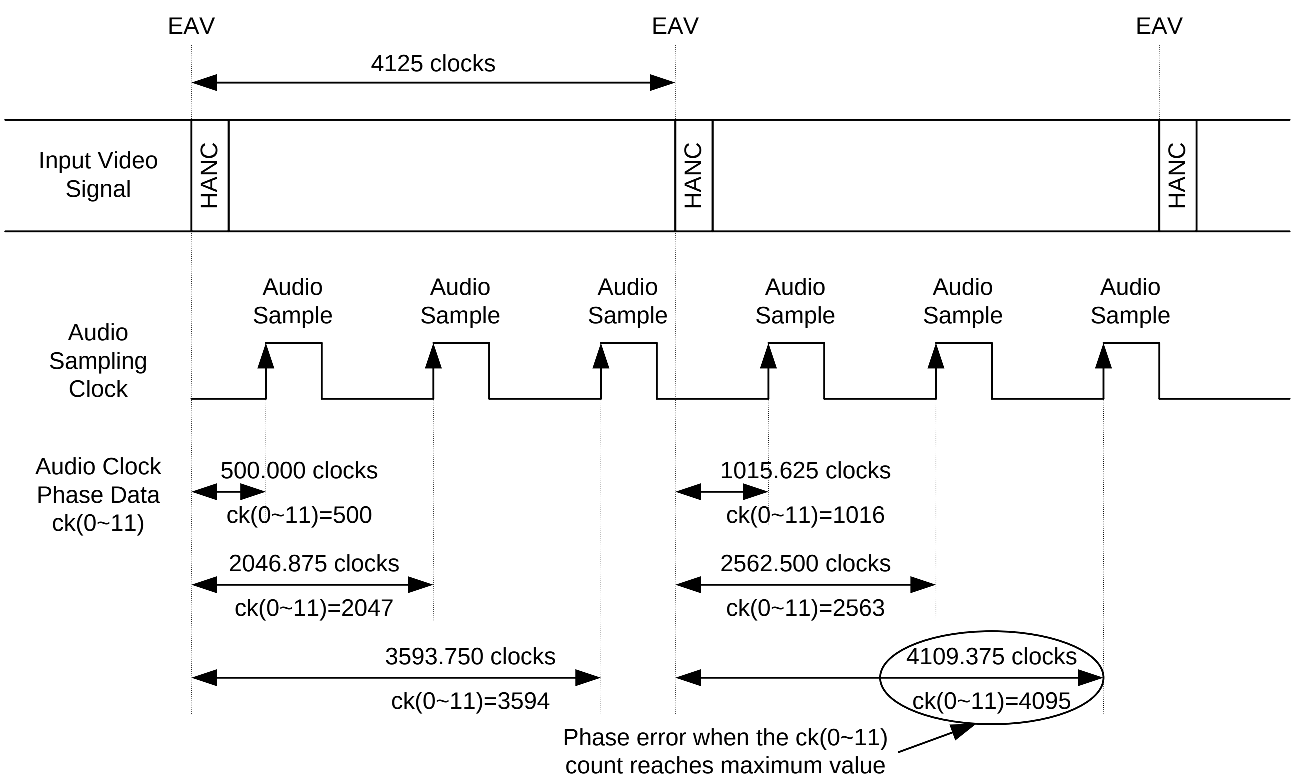


Figure B.1 – Relationship among video, audio sampling instants of digital audio and audio clock phase data (example – 720/24p system with 48 kHz audio sampling rate and 24.00 Hz video frame rate)

Annex C (Informative)
Bibliography

AES11-2003, AES Recommended Practice for Digital Audio Engineering — Synchronization of Digital Audio Equipment in Studio Operations

SMPTE 260M-1999 (Archived 2004), Television — 1125/60 High-Definition Production System — Digital Representation and Bit-Parallel Interface

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

SMPTE 274M-2008, Television — 1920 x 1080 Image Sample Structure, Digital Representation and Digital Timing Reference Sequences for Multiple Picture Rates

SMPTE 296M-2001, Television — 1280 x 720 Progressive Image Sample Structure — Analog and Digital Representation and Analog Interface

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

SMPTE 349M-2001 (Archived 2006), Television — Transport of Alternate Source Image Formats through SMPTE 292M

Annex D (Informative)
Additional Sampling Frequency

With the addition of 96 kHz sampling to this document, previous implementations may or may not support the additional sampling frequency. It is suggested that manufacturers indicate the level of support for this standard in the following way:

SMPTE 299 / 32,44,48 Would indicate support of the 32 kHz, 44.1 kHz and 48 kHz sampling.

SMPTE 299 / 48,96 Would indicate support of the 48 kHz and 96 kHz sampling.

In this way, end users may identify which product implementations best meet their needs.