SMPTE 302M-2002
Revision of
SMPTE 302M-1998

SMPTE STANDARD

for Television — Mapping of AES3 Data into an MPEG-2 Transport Stream

1 Scope

1.1 This standard specifies the method of transporting AES3 data in an MPEG-2 transport stream for television applications. For these television applications, the AES3 frame rate is 48 kHz.

1.2 Some applications may require linear PCM (pulse code modulated) digital audio in conjunction with compressed video specified in the MPEG-2 4:2:2 profile. The MPEG audio standard defines compressed audio, but does not define uncompressed audio for carriage in an MPEG-2 transport system. This standard augments the MPEG standards to address the requirement to carry AES3 streams, which may consist of linear PCM audio (at 48 ksample/s), or data carried within the AES3 format.

2 Normative references

The following documents 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.

AES3-1992, Digital Audio Engineering — Serial Transmission Format for Two-Channel Linearly Represented Digital Audio Data

SMPTE 337M-2000, Television — Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface

SMPTE EG 32-1996, Emphasis of AES/EBU Audio in Television Systems and Preferred Audio Sampling Rate

ISO/IEC 13818-1:1996, Information Technology — Generic Coding of Moving Pictures and Associated Audio Information: Systems

3 Introduction

MPEG-2 transport streams convey one or more programs of coded data, and may be constructed from one or more elementary coded data streams, program streams, or other transport streams. This standard specifies the method to transport AES3 streams, which may consist of linear PCM digital audio, or data, within an MPEG-2 transport stream. The specifications are described in terms of a model which starts with AES3 data, constructs elementary streams (ES) from the AES3 data, then constructs packetized elementary streams (PES) from the elementary streams, and finally constructs MPEG-2 transport streams (MTS) from the

Approved
August 5, 2002

packetized elementary streams. Although this model is used to describe the transport of AES3 streams in MPEG-2 transport streams, the model is not mandatory. MPEG-2 transport streams may be constructed by any method which results in a valid stream.

4 SMPTE AES3 elementary streams

4.1 SMPTE AES3 elementary streams shall consist of AES3 data words, which may be derived from AES3 subframes, together with validity, user, and channel status (V,U,C) bits and a framing (F) bit. The F bit replaces the parity (P) bit normally carried in AES3 subframes, and is used to indicate the start of a block of AES3 channel status (C) and user (U) bit information (see paragraphs 4.5 and 4.6 and figure 3).

4.2 There may be 1, 2, 3, or 4 AES3 data streams conveyed in a single audio elementary stream and corresponding packetized elementary stream. Multiple packetized elementary streams may be used in applications requiring carriage of a larger number of AES3 streams. Where multiple packetized elementary streams are used to convey multiple AES3 streams, the ordering of AES3 subframes within a packetized elementary stream is maintained within the packetized elementary stream, while the ordering across different packetized elementary streams can be maintained through the value of channel_identification (see 5.6) in the elementary stream header.

4.3 The data word size shall be 16, 20, or 24 bits per word. All AES3 streams in one SMPTE AES3 elementary stream shall have the same word size.

4.4 The word rate shall be 48 kHz. This rate shall be locked to the 27-MHz transport clock. All channels in a SMPTE AES3 elementary stream shall have the same word frequency.

4.5 AES3 streams consist of a sequence of AES3 frames. A set of 192 sequential frames forms a block. Individual AES3 frames consist of a pair of AES3 subframes, designated the A subframe and the B subframe. AES3 subframes consist of data for a single data word, as well as additional data including a validity bit (V), a user data bit (U), a channel status bit (C), a parity bit (P), and four auxiliary sample bits (aux). AES3 subframes may include 24-bit data words as shown in figure 1, 20-bit data words as shown in figure 2, or 16-bit data words as shown in figure 3. AES3 subframes which carry 16-bit data words use a 20-bit word with the four least significant bits (LSBs) of the 20-bit word set to 0.

4.6 Validity, user, and channel status (V,U,C) bits shall be carried in the MPEG transport stream specified by this standard.

4.7 Framing information signaling the first AES3 subframe of a block is carried by a framing (F) bit that replaces the parity bit normally carried in the AES3 stream. The framing bit shall be set to 1 for the first A subframe of an AES3 block (indicated by the Z preamble in the AES3 stream) and to 0 for all other subframes (including all B subframes). Note that when multiple AES3 data streams are conveyed within one SMPTE AES3 elementary stream, each individual AES3 stream that is carried will have the F bit set to 1 for one A subframe out of every 192 A subframes; the F bit in all B subframes will always be set to 0. It is not necessary for all of the individual AES3 streams conveyed by the SMPTE AES3 data elementary stream have the same phasing of the 192 frame Blocks.

4.8 Elementary streams derived from AES3 data words and V,U,C,F bits shall be constructed as shown in figure 4. Elementary streams shall be composed of 302M word groups. Each 302M word group shall contain one AES3 data word together with the associated V,U,C and F bits for each subframe data channel carried in the elementary stream (i.e., one A subframe or one B subframe from each AES3 stream). The bits of the AES3 data words shall be the LSBs of the 302M data word; V shall be more significant, then U, then C, and the F bit shall be the MSB. The bits of the 302M data words shall be sent in the same order as in the AES3 sub-frame as shown in figures 1, 2, and 3 (i.e., LSB first). The 302M data words within each 302M data word group shall be in the order of their channel number (AES3 stream 1 subframe A, followed by AES3 stream 1 subframe B, followed by AES3 stream 2 subframe A, etc.). Successive 302M data word groups shall contain successive 302M data words from each of the channels carried in the elementary stream.

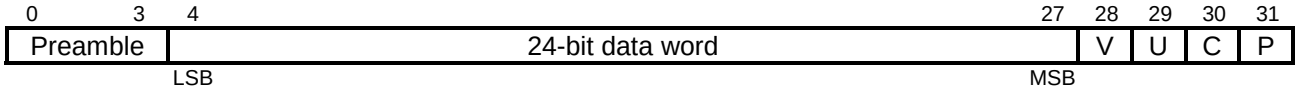

Figure 1 – 24-bit AES3 subframe

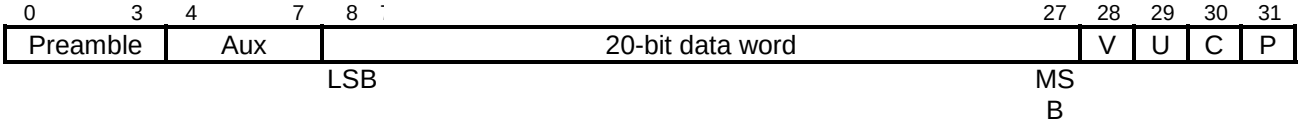

Figure 2 – 20-bit AES3 subframe

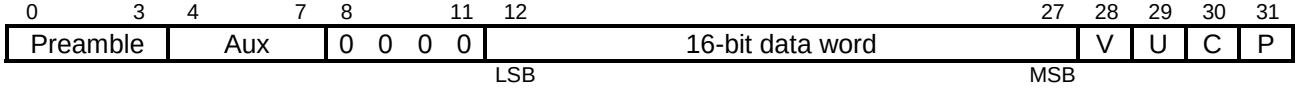

Figure 3 – 16-bit AES3 subframe

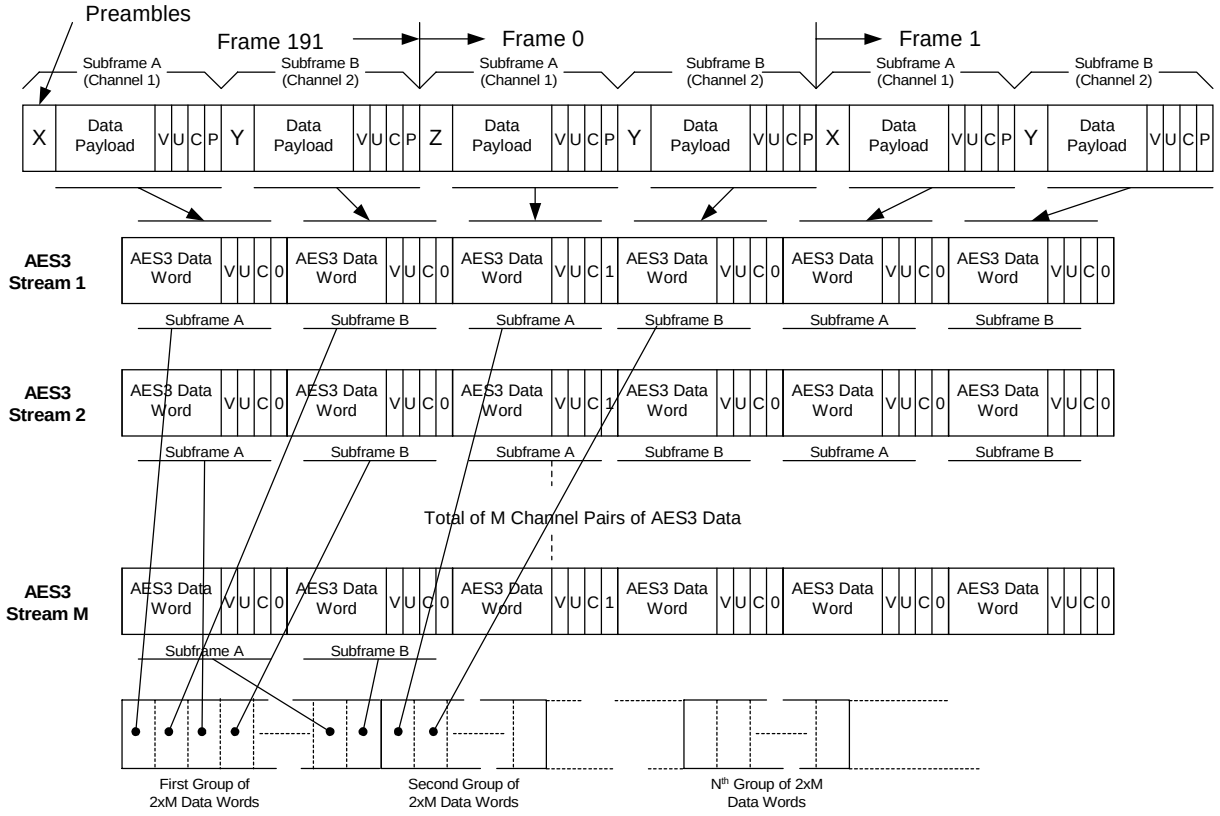

Figure 4 – SMPTE AES3 elementary stream

4.9 The 302M data words shall be packed, leaving no unused data space (see figure 5). In the case of 20-bit AES3 data word resolution, the AES3 data word plus VUCF bits for each channel are inherently packed since the sum of payload sample word length plus VUCF bits are multiples of 8 bits. In the case of 16- and 24-bit AES3 data word resolution, the AES3 data word plus VUCF bits shall be packed without sample byte alignment. Since AES3 frames carry a pair of AES3 data words, the resulting length of a 16- or 24-bit AES3 data payload plus VUCF for each AES3 frame will be an integer number of bytes (see figure 5).

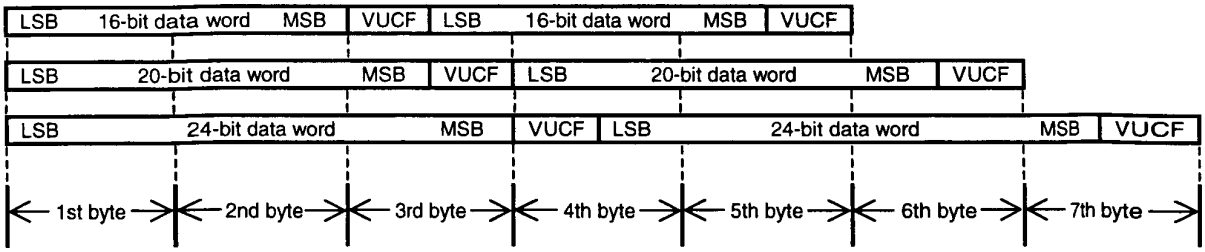

Figure 5 – Packing of AES3 data into ES bytes

5 Packetized elementary stream

5.1 PES packets for SMPTE AES3 data shall follow the specifications in ISO/IEC 13818-1 and as noted in this standard.

5.2 SMPTE AES3 data PES packets shall have an MPEG-2 PES header as described in ISO/IEC 13818-1, clauses 2.4.3.6 (PES packet) and 2.4.3.7 (semantic definition of fields in PES packet). Additional SMPTE AES3 data header information shall appear at the beginning of the MPEG-2 PES packet payload as shown in figure 6.

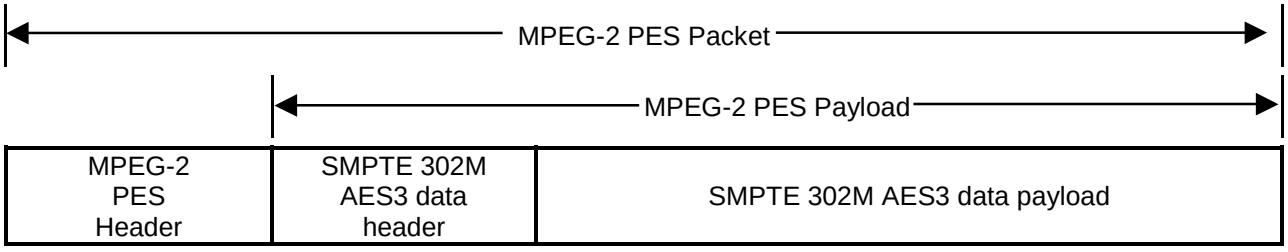

Figure 6 – Audio data PES packets

5.3 The SMPTE AES3 data PES packets shall conform to the semantic definition for private_stream_1 in ISO/IEC 13818-1, clause 2 .4 .3 .7 . The semantic definition for private_stream_1 includes presentation time stamps (PTS) and other information. (Note that the semantic definition for private_stream_2, which is an abbreviated header, is not used in this standard.)

5.4 SMPTE AES3 data PES packets shall have PTS_DTS_flags set to XO.

5.5 In SMPTE AES3 data PES packets, the ESCR_flag, ES_rate_flag, DSM_trick_mode_flag, additional_ copy_info_flag, and PES_extension_flag shall be set to 0.

5.6 Each SMPTE AES3 data PES packet payload shall include a SMPTE AES3 data header which shall be carried at the beginning of the associated MPEG-2 PES payload. This header shall include audio_packet_size, number_channels, channel_identification, and bits_per_sample fields as described below. This header shall apply to all of the AES3 subframe channels in the PES packet.

NOTE – The channel_identification may prove useful when multiple packetized elementary streams are used to convey multiple AES3 data streams in separate SMPTE AES data elementary streams. (If more than four AES3 data streams must be conveyed, it is necessary to employ multiple SMPTE AES3 data elementary streams.) For example, suppose a 12 channel audio program contained in six AES3 streams must be carried in an MPEG-2 transport stream. The first four AES3 streams containing audio channels 0-7 could be carried by one 302M stream with channel_identification set to a value of 0, and the final two AES3 streams containing audio channels 8-11 could be carried by a second 302M stream with channel_identification set to a value of 8.

5.7 The SMPTE AES3 data header data shall be as defined in table 1. (Note that uimsbf means unsigned integer with most significant bit first, as used in ISO/IEC 13818-1.)

Table 1 – SMPTE AES3 data elementary stream header

audio_packet_size	16 uimsbf [1)]	SMPTE AES3 payload packet size in bytes exclusive of SMPTE header
number_channels	2 uimsbf	00 = 2 data channels 01 = 4 data channels 10 = 6 data channels 11 = 8 data channels
channel_identification	8 uimsbf	channel number of first data channel [2)]
bits_per_sample	2 uimsbf	00 = 16 bits/data word 01 = 20 bits/data word 10 = 24 bits/data word 11 = reserved
alignment bits	4 uimsbf	SMPTE header byte alignment (reserved, set to 0)
NOTES [1)] Unsigned integer with most significant bit first [2)] See 5.6		

5.8 SMPTE AES3 data PES packets shall be integer numbers of bytes in length.

5.9 Groups of AES3 data words associated with individual video frames shall be collected into AES3 access units corresponding to the video frames. The payload of each AES3 data PES packet shall contain one AES3 access unit. The PES packet payload length shall be set according to the size of the associated AES3 access unit. In the case of video frame rates which do not have a simple integer relationship to the AES3 clock, not all AES3 access units will contain the same number of AES3 data words.

For example, ANSI/SMPTE 272M specifies that, at 29.97 video frames per second and with 48-ksamples/s audio, the audio samples per frame of 8008/5 will be in a sequence of 1602, 1601, 1602, 1601, 1602 samples per frame. In the case of 25 video frames per second, an integer number of audio samples per frame results (1920).

5.10 Each PES packet shall carry a presentation time stamp (PTS). The value of the PTS shall be identical to the value of the PTS that applies to a corresponding video frame. The accuracy of the PTS with respect to the actual payload shall be within ± 1 ms. Decoders shall reproduce the 302M AES3 stream with a time accuracy, relative to the associated video, of within ± 1 ms.

6 MPEG-2 transport system

6.1 MPEG-2 transport streams (MTS) incorporating SMPTE AES3 data PES packets shall conform to the specifications of ISO/IEC 13818-1 (systems).

6.1.1 The semantics of the stream_type field are described in ISO/IEC 13818-1[1] table 2-29. The value of stream_type for a 302M elementary stream in PES packets shall be 0x06 (indicating PES packets containing private data).

6.2 The registration_descriptor shall be used to identify formats of private data, as described in ISO/IEC 13818-1, clause 2.6.8. AES3 data as specified in this standard shall use the format_ identifier of 0x42535344. The registration_descriptor can carry additional_identification_info bytes; these bytes are not defined in this standard and should not be present in 302M streams that are compliant with this version of this standard. Decoders need not interpret additional_identification_info bytes if they are present. Decoders shall be able to accept bit streams that contain additional_identification_info bytes (as these bytes may be defined in a future revision to this standard). The elementary stream descriptor loop in the PMT shall contain this registration descriptor.

6.3 MPEG-2 transport streams carrying 302M shall be compliant with the T-STD model specified in 13818-1, with a 302M elementary stream buffer (Bn) size of 65024 bytes ($2^{16} – 512$).

6.4 The transport buffer TBn is specified to be 512 bytes in ISO/IEC 13818-1 (systems). For 302M audio, the transport buffer shall be drained (Rxn) at 1.2 times [Rmax].

Annex A (informative)
Bibliography

ANSI/SMPTE 272M-1994, Television — Formatting AES/EBU Audio and Auxiliary Data into Digital Video Ancillary Data Space

EBU Tech. 3250 (1992), Specification of the Digital Audio Interface

ISO/IEC 13818-2:1996, Information Technology — Generic Coding of Moving Pictures and Associated Audio Information: Video

ITU-R BT.1359-1, Relative Timing of Sound and Vision for Broadcasting