INTERNATIONAL STANDARD

ISO/IEC 13818-1

Second edition
2000-12-01

AMENDMENT 3
2004-11-01

Information technology — Generic coding of moving pictures and associated audio information: Systems

AMENDMENT 3

Technologies de l'information — Codage générique des images animées et du son associé: Systèmes

AMENDEMENT 3

Reference number
ISO/IEC 13818-1:2000/Amd.3:2004(E)

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

CONTENTS

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Amendment 3 to ISO/IEC 13818-1:2000 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*, in collaboration with ITU-T. The identical text is published as ITU-T Rec. H.222.0/Amd. 3.

INTERNATIONAL STANDARD
ITU-T RECOMMENDATION

Information technology – Generic coding of moving pictures and associated audio information: Systems

Amendment 3

Transport of AVC video data over ITU-T Rec. H.222.0 | ISO/IEC 13818-1 streams

1) Subclause 1.2.2

Add the following "paired" reference to subclause 1.2.2:

– ITU-T Recommendation H.264 (2003), *Advanced video coding for generic audiovisual services*.

ISO/IEC 14496-10:2003, *Information technology – Coding of audio-visual objects – Part 10: Advanced video coding.*

2) Subclause 2.1.1

Add to the definition for access unit in subclause 2.1.1:

For the definition of an access unit for ITU-T Rec. H.264 | ISO/IEC 14496-10 video, see the AVC access unit definition in 2.1.3.

3) New subclauses 2.1.2 to 2.1.7

Insert the following definitions as subclauses 2.1.2 to 2.1.7 and renumber existing ones accordingly:

2.1.2 AVC 24-hour picture (system): An AVC access unit with a presentation time that is more than 24 hours in the future. For the purpose of this definition, AVC access unit n has a presentation time that is more than 24 hours in the future if the difference between the initial arrival time $t_{ai}(n)$ and the DPB output time $t_{o,dpb}(n)$ is more than 24 hours.

2.1.3 AVC access unit (system): An access unit as defined for byte streams in ITU-T Rec. H.264 | ISO/IEC 14496-10 with the constraints specified in 2.14.1.

2.1.4 AVC Slice (system): A byte_stream_nal_unit as defined in ITU-T Rec. H.264 | ISO/IEC 14496-10 with nal_unit_type values of 1 or 5, or a byte_stream_nal_unit data structure with nal_unit_type value of 2 and any associated byte_stream_nal_unit data structures with nal_unit_type equal to 3 and/or 4.

2.1.5 AVC still picture (system): An AVC still picture consists of an AVC access unit containing an IDR picture, preceded by SPS and PPS NAL units that carry sufficient information to correctly decode the IDR picture. Preceding an AVC still picture, there shall be another AVC still picture or an End of Sequence NAL unit terminating a preceding coded video sequence.

2.1.6 AVC video sequence (system): Coded video sequence as defined in ITU-T Rec. H.264 | ISO/IEC 14496-10, clause 3.27.

2.1.7 AVC video stream (system): An ITU-T Rec. H.264 | ISO/IEC 14496-10 stream. An AVC video stream consists of one or more AVC video sequences.

4) Subclause 2.1.52

Replace the still picture definition in subclause 2.1.52:

2.1.52 still picture: A coded still picture consists of a video sequence containing exactly one coded picture which is intra-coded. This picture has an associated PTS and the presentation time of succeeding pictures, if any, is later than that of the still picture by at least two picture periods.

by:

2.1.52 still picture: A still picture consists of a video sequence, coded as defined in ITU-T Rec. H.262 | ISO/IEC 13818-2, ISO/IEC 11172-2 or ISO/IEC 14496-2, that contains exactly one coded picture which is intra-coded. This picture has an associated PTS and in case of coding according to ISO/IEC 11172-2, ITU-T Rec. H.262 | ISO/IEC 13818-2 or ISO/IEC 14496-2, the presentation time of succeeding pictures, if any, is later than that of the still picture by at least two picture periods.

5) New subclause 2.4.2.8

Add after subclause 2.4.2.7:

2.4.2.8 T-STD extensions for carriage of ITU-T Rec. H.264 | ISO/IEC 14496-10 Video

To define the decoding in the T-STD of ITU-T Rec. H.264 | ISO/IEC 14496-10 video streams carried in a Transport Stream, the T-STD model needs to be extended. The T-STD extension and T-STD parameters for decoding of ITU-T Rec. H.264 | ISO/IEC 14496-10 video streams are defined in 2.14.3.1.

6) Subclause 2.4.3.5

a) Replace in the semantics of discontinuity_indicator under subclause 2.4.3.5 starting from the 5th paragraph:

For the purpose of this clause, an elementary stream access point is defined as follows:

- Video – The first byte of a video sequence header.
- Audio – The first byte of an audio frame.

After a continuity counter discontinuity in a Transport packet which is designated as containing elementary stream data, the first byte of elementary stream data in a Transport Stream packet of the same PID shall be the first byte of an elementary stream access point or in the case of video, the first byte of an elementary stream access point or a sequence_end_code followed by an access point.

by:

For the purpose of this clause, an elementary stream access point is defined as follows:

- ISO/IEC 11172-2 video and ITU-T Rec. H.262 | ISO/IEC 13818-2 video – The first byte of a video sequence header.
- ISO/IEC 14496-2 visual – The first byte of the visual object sequence header.
- ITU-T Rec. H.264 | ISO/IEC 14496-10 video – The first byte of an AVC access unit. The SPS and PPS parameter sets referenced in this and all subsequent AVC access units in the coded video stream shall be provided after this access point in the byte stream and prior to their activation.
- Audio – The first byte of an audio frame.

After a continuity counter discontinuity in a Transport packet which is designated as containing elementary stream data, the first byte of elementary stream data in a Transport Stream packet of the same PID shall be the first byte of an elementary stream access point. In the case of ISO/IEC 11172-2, or ITU-T Rec. H.262 | ISO/IEC 13818-2 or ISO/IEC 14496-2 video, the first byte of an elementary stream access point may also be the first byte of a sequence_end_code followed by an elementary stream access point.

b) Replace in the semantics of random_access_indicator under subclause 2.4.3.5:

Specifically, when the bit is set to '1', the next PES packet to start in the payload of Transport Stream packets with the current PID shall contain the first byte of a video sequence header if the PES stream type (refer to Table 2-29) is 1 or 2, or shall contain the first byte of an audio frame if the PES stream type is 3 or 4. In addition, in the case of video, a presentation timestamp shall be present in the PES packet containing the first picture following the sequence header.

by:

Specifically, when the bit is set to '1', the next PES packet to start in the payload of Transport Stream packets with the current PID shall contain an elementary stream access point as defined in the semantics for the discontinuity_indicator field. In addition, in the case of video, a presentation timestamp shall be present for the first picture following the elementary stream access point.

c) Replace in the semantics of elementary_stream_priority_indicator under subclause 2.4.3.5:

In the case of video, this field may be set to '1' only if the payload contains one or more bytes from an intra-coded slice.

by:

In the case of ISO/IEC 11172-2 or ITU-T Rec. H.262 | ISO/IEC 13818-2 or ISO/IEC 14496-2 video, this field may be set to '1' only if the payload contains one or more bytes from an intra-coded slice.

In the case of ITU-T Rec. H.264 | ISO/IEC 14496-10 video, this field may be set to '1' only if the payload contains one or more bytes from a slice with slice_type set to 2, 4, 7, or 9.

d) Replace in the semantics of splice_countdown under subclause 2.4.3.5:

For the purpose of this subclause, an access point is defined as follows:

- Video – The first byte of a video_sequence_header.
- Audio – The first byte of an audio frame.

by:

For the definition of an elementary stream access point, see the semantics of discontinuity_indicator in 2.4.3.5.

e) Replace in the semantics of seamless_splice_flag under subclause 2.4.3.5 the sentences:

When this flag is set, if the elementary stream carried in this PID is an audio stream, the splice_type field shall be set to '0000'. If the elementary stream carried in this PID is a video stream, it shall fulfil the constraints indicated by the splice_type value.

by:

When this flag is set, and if the elementary stream carried in this PID is not an ITU-T Rec. H.262 | ISO/IEC 13818-2 video stream, then the splice_type field shall be set to '0000'. If the elementary stream carried in this PID is an ITU-T Rec. H.262 | ISO/IEC 13818-2 video stream, it shall fulfil the constraints indicated by the splice_type value.

f) Replace in the semantics of splice_type under subclause 2.4.3.5 the sentences:

If the elementary stream carried in that PID is an audio stream, this field shall have the value '0000'. If the elementary stream carried in that PID is a video stream, this field indicates the conditions that shall be respected by this elementary stream for splicing purposes.

by:

If the elementary stream carried in that PID is not an ITU-T Rec. H.262 | ISO/IEC 13818-2 video stream, then this field shall have the value '0000'. If the elementary stream carried in that PID is an ITU-T Rec. H.262 | ISO/IEC 13818-2 video stream, then this field indicates the conditions that shall be respected by this elementary stream for splicing purposes.

7) Subclause 2.4.3.7

a) Replace Table 2-18 in subclause 2.4.3.7 by:

Table 2-18 – Stream_id assignments

Stream_id	Note	stream coding
1011 1100	1	program_stream_map
1011 1101	2	private_stream_1
1011 1110		padding_stream
1011 1111	3	private_stream_2
110x xxxx		ISO/IEC 13818-3 or ISO/IEC 11172-3 or ISO/IEC 13818-7 or ISO/IEC 14496-3 audio stream number x xxxx
1110 xxxx		ITU-T Rec. H.262 | ISO/IEC 13818-2, ISO/IEC 11172-2, ISO/IEC 14496-2 or ITU-T Rec. H.264 | ISO/IEC 14496-10 video stream number xxxx
1111 0000	3	ECM_stream
1111 0001	3	EMM_stream
1111 0010	5	ITU-T Rec. H.222.0 | ISO/IEC 13818-1 Annex A or ISO/IEC 13818-6_DSMCC_stream
1111 0011	2	ISO/IEC_13522_stream
1111 0100	6	ITU-T Rec. H.222.1 type A
1111 0101	6	ITU-T Rec. H.222.1 type B
1111 0110	6	ITU-T Rec. H.222.1 type C
1111 0111	6	ITU-T Rec. H.222.1 type D
1111 1000	6	ITU-T Rec. H.222.1 type E
1111 1001	7	ancillary_stream
1111 1010		ISO/IEC 14496-1_SL-packetized_stream
1111 1011		ISO/IEC 14496-1_FlexMux_stream
1111 1100		metadata stream
1111 1101		extended_stream_id
1111 1110		reserved data stream
1111 1111	4	program_stream_directory

The notation x means that the values '0' or '1' are both permitted and results in the same stream type. The stream number is given by the values taken by the x's.

NOTE 1 – PES packets of type program_stream_map have unique syntax specified in 2.5.4.1.

NOTE 2 – PES packets of type private_stream_1 and ISO/IEC_13552_stream follow the same PES packet syntax as those for ITU-T Rec. H.262 | ISO/IEC 13818-2 video and ISO/IEC 13818-3 audio streams.

NOTE 3 – PES packets of type private_stream_2, ECM_stream and EMM_stream are similar to private_stream_1 except no syntax is specified after PES_packet_length field.

NOTE 4 – PES packets of type program_stream_directory have a unique syntax specified in 2.5.5.

NOTE 5 – PES packets of type DSM-CC_stream have a unique syntax specified in ISO/IEC 13818-6.

NOTE 6 – This stream_id is associated with stream_type 0x09 in Table 2-29.

NOTE 7 – This stream_id is only used in PES packets, which carry data from a Program Stream or an ISO/IEC 11172-1 System Stream, in a Transport Stream (refer to 2.4.3.7).

b) Replace the semantics of data_alignment_indicator in subclause 2.4.3.7 by:

data_alignment_indicator – This is a 1-bit flag. When set to a value of '1', it indicates that the PES packet header is immediately followed by the video syntax element or audio sync word indicated in the data_stream_alignment_descriptor in 2.6.10 if this descriptor is present. If set to a value of '1' and the descriptor is not present, alignment as indicated in alignment_type '01' in Table 2-47, Table 2-48 or Table AMD3-1 is required. When set to a value of '0', it is not defined whether any such alignment occurs or not.

c) Replace in the semantics of PTS in subclause 2.4.3.7:

In the case of video, if a PTS is present in a PES packet header it shall refer to the access unit containing the first picture start code that commences in this PES packet. A picture start code commences in PES packet if the first byte of the picture start code is present in the PES packet.

For audio presentation units (PUs), video PUs in low_delay sequences, and B-pictures, the presentation time $tp_n(k)$ shall be equal to the decoding time $td_n(k)$.

For I- and P-pictures in non-low_delay sequences and in the case when there is no decoding discontinuity between access units (AUs) k and k', the presentation time $tp_n(k)$ shall be equal to the decoding time $td_n(k')$ of the next transmitted I- or P-picture (refer to 2.7.5). If there is a decoding discontinuity, or the stream ends, the difference between $tp_n(k)$ and $td_n(k)$ shall be the same as if the original stream had continued without a discontinuity and without ending.

NOTE 1 – A low_delay sequence is a video sequence in which the low_delay flag is set (refer to 6.2.2.3 of ITU-T Rec. H.262 | ISO/IEC 13818-2).

by:

In the case of ISO/IEC 11172-2 video, ITU-T Rec. H.262 | ISO/IEC 13818-2 video, or ISO/IEC 14496-2 video, if a PTS is present in a PES packet header, it shall refer to the access unit containing the first picture start code that commences in this PES packet. A picture start code commences in a PES packet if the first byte of the picture start code is present in the PES packet. For I- and P-pictures in non-low_delay sequences and in the case when there is no decoding discontinuity between access units (AUs) k and k', the presentation time $tp_n(k)$ shall be equal to the decoding time $td_n(k')$ of the next transmitted I- or P-picture (refer to 2.7.5). If there is a decoding discontinuity, or the stream ends, the difference between $tp_n(k)$ and $td_n(k)$ shall be the same as if the original stream had continued without a discontinuity and without ending.

NOTE 1 – A low_delay sequence is an ITU-T Rec. H.262 | ISO/IEC 13818-2 or ISO/IEC 14496-2 video sequence in which the low_delay flag is set to '1' (refer to 6.2.2.3 of ITU-T Rec. H.262 | ISO/IEC 13818-2 and to 6.2.3 of ISO/IEC 14496-2).

For ITU-T Rec. H.264 | ISO/IEC 14496-10 video, if a PTS is present in the PES packet header, it shall refer to the first AVC access unit that commences in this PES packet. An AVC access unit commences in a PES packet if the first byte of the AVC access unit is present in the PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of ITU-T Rec. H.264 | ISO/IEC 14496-10, for each decoded AVC access unit, the PTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal DPB output time in the HRD, defined herein as $t_{o,n,dpb}(n) = t_{r,n}(n) + t_c * dpb\_output\_delay(n)$, where $t_{r,n}(n)$, t_c, and dpb_output_delay(n) are defined as in Annex C of ITU-T Rec. H.264 | ISO/IEC 14496-10.

NOTE 2 – Different clocks may be used for derivation of PTS and $t_{o,n,dpb}(n)$.

The presentation time $tp_n(k)$ shall be equal to the decoding time $td_n(k)$ for:

- audio access units;
- access units in ITU-T Rec. H.262 | ISO/IEC 13818-2 or ISO/IEC 14496-2 low delay video sequences;
- B-pictures in ISO/IEC 11172-2, ITU-T Rec. H.262 | ISO/IEC 13818-2 or ISO/IEC 14496-2 video streams.

d) Replace in the semantics of DTS in subclause 2.4.3.7:

In the case of video, if a DTS is present in a PES packet header it shall refer to the access unit containing the first picture start code that commences in this PES packet. A picture start code commences in PES packet if the first byte of the picture start code is present in the PES packet.

by:

In the case of ISO/IEC 11172-2 video, ITU-T Rec. H.262 | ISO/IEC 13818-2 video, or ISO/IEC 14496-2 video, if a DTS is present in a PES packet header, it shall refer to the access unit containing the first picture start code that commences in this PES packet. A picture start code commences in a PES packet if the first byte of the picture start code is present in the PES packet.

For ITU-T Rec. H.264 | ISO/IEC 14496-10 video, if a DTS is present in the PES packet header, it shall refer to the first AVC access unit that commences in this PES packet. An AVC access unit commences in a PES packet if the first byte of the AVC access unit is present in the PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of ITU-T Rec. H.264 | ISO/IEC 14496-10, for each AVC access unit the DTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal CPB removal time $t_{r,n}(n)$ in the HRD, as defined in Annex C of ITU-T Rec. H.264 | ISO/IEC 14496-10.

NOTE 3 – Different clocks may be used for derivation of DTS and $t_{r,n}(n)$.

e) *Add to the semantics of P-STD_buffer_size under subclause 2.4.3.7:*

The size BS_n shall be larger than or equal to the size of the CPB signalled by the CpbSize[cpb_cnt_minus1] specified by the NAL hrd_parameters() in the AVC video stream. If the NAL hrd_parameters() are not present in the AVC video stream, then BS_n shall be larger than or equal to the size of the NAL CPB for the byte stream format defined in Annex A of ITU-T Rec. H.264 | ISO/IEC 14496-10 as 1200 × MaxCPB for the applied level.

8) Subclause 2.4.4.10

Replace Table 2-29 in subclause 2.4.4.10 by:

Table 2-29 – Stream type assignments

Value	Description
0x00	ITU-T | ISO/IEC Reserved
0x01	ISO/IEC 11172-2 Video
0x02	ITU-T Rec. H.262 | ISO/IEC 13818-2 Video or ISO/IEC 11172-2 constrained parameter video stream
0x03	ISO/IEC 11172-3 Audio
0x04	ISO/IEC 13818-3 Audio
0x05	ITU-T Rec. H.222.0 | ISO/IEC 13818-1 private_sections
0x06	ITU-T Rec. H.222.0 | ISO/IEC 13818-1 PES packets containing private data
0x07	ISO/IEC 13522 MHEG
0x08	ITU-T Rec. H.222.0 | ISO/IEC 13818-1 Annex A DSM-CC
0x09	ITU-T Rec. H.222.1
0x0A	ISO/IEC 13818-6 type A
0x0B	ISO/IEC 13818-6 type B
0x0C	ISO/IEC 13818-6 type C
0x0D	ISO/IEC 13818-6 type D
0x0E	ITU-T Rec. H.222.0 | ISO/IEC 13818-1 auxiliary
0x0F	ISO/IEC 13818-7 Audio with ADTS transport syntax
0x10	ISO/IEC 14496-2 Visual
0x11	ISO/IEC 14496-3 Audio with the LATM transport syntax as defined in ISO/IEC 14496-3/AMD-1
0x12	ISO/IEC 14496-1 SL-packetized stream or FlexMux stream carried in PES packets
0x13	ISO/IEC 14496-1 SL-packetized stream or FlexMux stream carried in ISO/IEC14496_sections
0x14	ISO/IEC 13818-6 Synchronized Download Protocol
0x15	Metadata carried in PES packets
0x16	Metadata carried in metadata_sections
0x17	Metadata carried in ISO/IEC 13818-6 Data Carousel
0x18	Metadata carried in ISO/IEC 13818-6 Object Carousel
0x19	Metadata carried in ISO/IEC 13818-6 Synchronized Download Protocol
0x1A	IPMP stream (defined in ISO/IEC 13818-11, MPEG-2 IPMP)
0x1B	AVC video stream as defined in ITU-T Rec. H.264 | ISO/IEC 14496-10 Video
0x1C-0x7E	ITU-T Rec. H.222.0 | ISO/IEC 13818-1 Reserved
0x7F	IPMP stream
0x80-0xFF	User Private

9) Subclause 2.5.2.4

Add in subclause 2.5.2.4 "PES streams" the sentence:

– *For ITU-T Rec. H.264 | ISO/IEC 14496-10 video:*

$$BS_n = 1200 \times MaxCPB[level] + BS_{oh}$$

Where MaxCPB[level] is defined in Table A.1 (Level Limits) in ITU-T Rec. H.264 | ISO/IEC 14496-10 for each level.

10) New subclause 2.5.2.7

Add after subclause 2.5.2.6:

2.5.2.7 P-STD extensions for carriage of ITU-T Rec. H.264 | ISO/IEC 14496-10 Video

For decoding of ITU-T Rec. H.264 | ISO/IEC 14496-10 video streams carried in a Program Stream in the P-STD model, see 2.14.3.2.

11) Subclause 2.5.3.6

a) Replace in the semantics of the system_video_lock_flag in subclause 2.5.3.6:

The system_video_lock_flag is a 1-bit field indicating that there is a specified, constant rational relationship between the video frame rate and the system clock frequency in the system target decoder. Subclause 2.5.2.1 defines system_clock_frequency and the video frame rate is specified in ITU-T Rec. H.262 | ISO/IEC 13818-2. The system_video_lock_flag may only be set to '1' if, for all presentation units in all video elementary streams in the ITU-T Rec. H.222.0 | ISO/IEC 13818-1 program, the ratio of system_clock_frequency to the actual video frame rate, SCFR, is constant and equal to the value indicated in the following table at the nominal frame rate indicated in the video stream.

by:

The system_video_lock_flag is a 1-bit field indicating that there is a specified, constant rational relationship between the video time base and the system clock frequency in the system target decoder. The system_video_lock_flag may only be set to '1' if, for all presentation units in all video elementary streams in the ITU-T Rec. H.222.0 | ISO/IEC 13818-1 program, the ratio of system_clock_frequency to the frequency of the actual video time base is constant.

For ISO/IEC 11172-2 and ITU-T Rec. H.262 | ISO/IEC 13818-2 video streams, if the system_video_lock_flag is set to '1', then the ratio of system_clock_frequency to the actual video frame rate, SCFR, shall be constant and equal to the value indicated in the following table at the nominal frame rate indicated in the video stream.

For ISO/IEC 14496-2 video streams, if the system_video_lock_flag is set to '1', then the time base of the ISO/IEC 14496-2 video stream, as defined by vop_time_increment_resolution, shall be locked to the STC and shall be exactly equal to N times system_clock_frequency divided by K, with N and K integers that have a fixed value within each visual object sequence, with K greater than or equal to N.

For ITU-T Rec. H.264 | ISO/IEC 14496-10 video streams, the frequency of the AVC time base is defined by the AVC parameter time_scale. If the system_video_lock_flag is set to '1' for an AVC video stream, then the frequency of the AVC time base shall be locked to the STC and shall be exactly equal to N times system_clock_frequency divided by K, with N and K integers that have a fixed value within each AVC video sequence, with K greater than or equal to N.

b) Replace the semantics of video_bound in subclause 2.5.3.6 by:

The video_bound is a 5-bit integer in the inclusive range from 0 to 16 and is set to a value greater than or equal to the maximum number of video streams in the Program Stream of which the decoding processes are simultaneously active. For the purpose of this subclause, the decoding process of a video stream is active if one of the buffers in the P-STD model is not empty, or if a Presentation Unit is being presented in the P-STD model.

12) Subclause 2.5.5

Add the following semantics in subclause 2.5.5 "Program Stream directory" immediately after NOTE 2:

Directory entries may be required to reference IDR picture or pictures associated with a recovery point SEI message in an AVC video stream. Each such directory entry shall refer to the first byte of an AVC access unit.

13) Subclause 2.6.1

Replace Table 2-39 in subclause 2.6.1 by:

Table 2-39 – Program and program element descriptors

descriptor_tag	TS	PS	Identification
0	n/a	n/a	Reserved
1	n/a	n/a	Reserved
2	X	X	video_stream_descriptor
3	X	X	audio_stream_descriptor
4	X	X	hierarchy_descriptor
5	X	X	registration_descriptor
6	X	X	data_stream_alignment_descriptor
7	X	X	target_background_grid_descriptor
8	X	X	Video_window_descriptor
9	X	X	CA_descriptor
10	X	X	ISO_639_language_descriptor
11	X	X	System_clock_descriptor
12	X	X	Multiplex_buffer_utilization_descriptor
13	X	X	Copyright_descriptor
14	X		Maximum_bitrate_descriptor
15	X	X	Private_data_indicator_descriptor
16	X	X	Smoothing_buffer_descriptor
17	X		STD_descriptor
18	X	X	IBP_descriptor
19-26	X		Defined in ISO/IEC 13818-6
27	X	X	MPEG-4_video_descriptor
28	X	X	MPEG-4_audio_descriptor
29	X	X	IOD_descriptor
30	X		SL_descriptor
31	X	X	FMC_descriptor
32	X	X	External_ES_ID_descriptor
33	X	X	MuxCode_descriptor
34	X	X	FmxBufferSize_descriptor
35	X		MultiplexBuffer_descriptor
36	X	X	Content_labeling_descriptor
37	X	X	Metadata_pointer_descriptor
38	X	X	Metadata_descriptor
39	X	X	Metadata_STD_descriptor
40	X	X	AVC video descriptor
41	X	X	IPMP_descriptor (defined in ISO/IEC 13818-11, MPEG-2 IPMP)
42	X	X	AVC timing and HRD descriptor
43-63	n/a	n/a	ITU-T Rec. H.222.0 | ISO/IEC 13818-1 Reserved
64-255	n/a	n/a	User Private

14) Subclause 2.6.6

Replace in subclause 2.6.6 'Hierarchy descriptor':

The hierarchy descriptor provides information to identify the program elements containing components of hierarchically-coded video and audio, and private streams which are multiplexed in multiple streams as described in this Recommendation | International Standard, in ITU-T Rec. H.262 | ISO/IEC 13818-2 and in ISO/IEC 13818-3. (See Table 2-43.)

by:

The hierarchy descriptor provides information to identify the program elements containing components of hierarchically-coded video, audio, and private streams. (See Table 2-43.)

15) Subclause 2.6.7

Replace the following entries in Table 2-44 'Hierarchy_type field values' in subclause 2.6.7:

ITU-T Rec. H.262 | ISO/IEC 13818-2 Spatial Scalability
ITU-T Rec. H.262 | ISO/IEC 13818-2 SNR Scalability
ITU-T Rec. H.262 | ISO/IEC 13818-2 Temporal Scalability
ITU-T Rec. H.262 | ISO/IEC 13818-2 Data partitioning
ISO/IEC 13818-3 Extension bitstream
ITU-T Rec.H222.0 | ISO/IEC 13818-1 Private Stream
ITU-T Rec. H.262 | ISO/IEC 13818-2 Multi-view Profile

by, respectively:

Spatial Scalability
SNR Scalability
Temporal Scalability
Data partitioning
Extension bitstream
Private Stream
Multi-view Profile

16) Subclause 2.6.11

a) Replace in the semantics of alignment_type in subclause 2.6.11:

Table 2-47 describes the video alignment type when the data_alignment_indicator in the PES packet header has a value of '1'. In each case of alignment_type value the first PES_packet_data_byte following the PES header shall be the first byte of a start code of the type indicated in Table 2-47. At the beginning of a video sequence, the alignment shall occur at the start code of the first sequence header.

NOTE – Specifying alignment type '01' from Table 2-47 does not preclude the alignment from beginning at a GOP or SEQ header.

The definition of access unit for video data is given in 2.1.1.

by:

Table 2-47 describes the alignment type for ISO/IEC 11172-2 video, ITU-T Rec. H.262 | ISO/IEC 13818-2 video, or ISO/IEC 14496-2 visual streams when the data_alignment_indicator in the PES packet header has a value of '1'. For these video streams, the first PES_packet_data_byte following the PES header shall be the first byte of a start code of the type indicated in Table 2-47. At the beginning of a video sequence, the alignment shall occur at the start code of the first sequence header.

NOTE – Specifying alignment type '01' from Table 2-47 does not preclude the alignment from beginning at a GOP or SEQ header.

The definition of an access unit is given in 2.1.1.

b) Insert the following text and table AMD3-1 immediately after Table 2-47 in subclause 2.6.11:

Table AMD3-1 describes the alignment type for ITU-T Rec. H.264 | ISO/IEC 14496-10 video when the data_alignment_indicator in the PES packet header has a value of '1'. In this case the first PES_packet_data_byte following the PES header shall be the first byte of an AVC access unit or the first byte of an AVC slice, as signalled by the alignment_type value.

Table AMD3-1 – AVC video stream alignment values

Alignment type	Description
00	Reserved
01	AVC slice or AVC access unit
02	AVC access unit
03-FF	Reserved

17) Subclause 2.6.32

Replace in subclause 2.6.32 'STD descriptor':

This descriptor is optional and applies only to the T-STD model and to video elementary streams, and is used as specified 2.4.2. This descriptor does not apply to Program Streams (see Table 2-60).

by:

This descriptor is optional and applies only to the T-STD model and to ITU-T Rec. H.262 | ISO/IEC 13818-2 video elementary streams, and is used as specified in 2.4.2. This descriptor does not apply to Program Streams (see Table 2-60).

18) Subclause 2.6.34

Replace in subclause 2.6.34 'IBP descriptor':

This optional descriptor provides information about some characteristics of the sequence of frame types in the video sequence (see Table 2-61).

by:

This optional descriptor provides information about some characteristics of the sequence of frame types in an ISO/IEC 11172-2, ITU-T Rec. H.262 | ISO/IEC 13818-2, or ISO/IEC 14496-2 video stream (see Table 2-61).

19) New subclauses 2.6.64-2.6.67

Add after subclause 2.6.63:

2.6.64 AVC video descriptor

For ITU-T Rec. H.264 | ISO/IEC 14496-10 video streams, the AVC video descriptor provides basic information for identifying coding parameters of the associated AVC video stream, such as on profile and level parameters included in the SPS of an AVC video stream.

The AVC video descriptor also signals the presence of AVC still pictures and the presence of AVC 24-hour pictures in the AVC video stream. If this descriptor is not included in the PMT for an AVC video stream in a transport stream or in the PSM, if present, for an AVC video stream in a program stream, then such AVC video stream shall not contain AVC still pictures and shall not contain AVC 24-hour pictures. (See Table AMD3-2.)

Table AMD3-2 – AVC video descriptor

Syntax	No. of bits	Mnemonic
AVC_video_descriptor () {		
descriptor_tag	**8**	**uimsbf**
descriptor_length	**8**	**uimsbf**
profile_idc	**8**	**uimsbf**
constraint_set0_flag	**1**	**bslbf**
constraint_set1_flag	**1**	**bslbf**
constraint_set2_flag	**1**	**bslbf**
AVC_compatible_flags	**5**	**bslbf**
level_idc	**8**	**uimsbf**
AVC_still_present	**1**	**bslbf**
AVC_24_hour_picture_flag	**1**	**bslbf**
reserved	**6**	**bslbf**
}		

2.6.65 Semantic definition of fields in AVC video descriptor

profile_idc, constraint_set0_flag, constraint_set1_flag, constraint_set2_flag, AVC_compatible_flags and level_idc – These fields, with the exception of AVC_compatible_flags shall be coded according to the semantics for these fields defined in ITU-T Rec. H.264 | ISO/IEC 14496-10. The semantics of AVC_compatible_flags are exactly equal to the semantics of the field(s) defined for the 5 bits between the constraint_set2 flag and the level_idc field in the Sequence Parameter Set, as defined in ITU-T Rec. H.264 | ISO/IEC 14496-10. The entire AVC video stream to which the AVC descriptor is associated shall conform to the profile, level and constraints signalled by these fields.

NOTE – In one or more sequences in the AVC video stream the level may be lower than the level signalled in the AVC video descriptor, while also a profile may occur that is a subset of the profile signalled in the AVC video descriptor. However, in the entire AVC video stream, only tools shall be used that are included in the profile signalled in the AVC video descriptor, if present. For example, if the main profile is signalled, then the baseline profile may be used in some sequences, but only using those tools that are in the main profile. If the sequence parameter sets in an AVC video stream signal different profiles, and no additional constraints are signalled, then the stream may need examination to determine which profile, if any, the entire stream conforms to. If an AVC video descriptor is to be associated with an AVC video stream that does not conform to a single profile, then the AVC video stream must be partitioned into two or more sub-streams, so that AVC video descriptors can signal a single profile for each such sub-stream.

AVC_still_present – This 1-bit field when set to '1' indicates that the AVC video stream may include AVC still pictures. When set to '0', then the associated AVC video stream shall not contain AVC still pictures.

AVC_24_hour_picture_flag – This 1-bit flag when set to '1' indicates that the associated AVC video stream may contain AVC 24-hour pictures. For the definition of an AVC 24-hour picture, see 2.1.2. If this flag is set to '0', the associated AVC video stream shall not contain any AVC 24-hour picture.

2.6.66 AVC timing and HRD descriptor

The AVC timing and HRD descriptor provides timing and HRD parameters of the associated AVC video stream. For each AVC video stream carried in an ITU-T Rec. H.222.0 | ISO/IEC 13818-1 stream, the AVC timing and HRD descriptor shall be included in the PMT or in the PSM, if PSM is present in the program stream, unless the AVC video stream carries VUI parameters with the timing_info_present_flag set to '1':

- for each IDR picture; and
- for each picture that is associated with a recovery point SEI message.

Absence of the AVC timing and HRD descriptor in the PMT for an AVC video stream signals usage of the leak method in the T-STD is defined in 2.14.3.1 for the transfer from MB_n to EB_n, but such usage can also be signalled by the hrd_management_valid_flag set to '0' in the AVC timing and HRD descriptor. If the transfer rate into buffer EB_n can be determined from HRD parameters contained in an AVC video stream, and if this transfer rate is used in the T-STD for the transfer between MB_n to EB_n, then the AVC timing and HRD descriptor with the hrd_management_valid_flag set to '1' shall be included in the PMT for that AVC video stream. (See Table AMD3-3.)

Table AMD3-3 – AVC timing and HRD descriptor

Syntax	No. of bits	Mnemonic
AVC timing and HRD descriptor () {		
descriptor_tag	**8**	**uimsbf**
descriptor_length	**8**	**uimsbf**
hrd_management_valid_flag	**1**	**bslbf**
reserved	**6**	**bslbf**
picture_and_timing_info_present	**1**	**bslbf**
if (picture_and_timing_info_present) {		
90kHz_flag	**1**	**bslbf**
reserved	**7**	**bslbf**
if (90kHz_flag == '0') {		
N	**32**	**uimsbf**
K	**32**	**uimsbf**
}		
num_units_in_tick	**32**	**uimsbf**
}		
fixed_frame_rate_flag	**1**	**bslbf**
temporal_poc_flag	**1**	**bslbf**
picture_to_display_conversion_flag	**1**	**bslbf**
reserved	**5**	**bslbf**
}		

2.6.67 Semantic definition of fields in AVC timing and HRD descriptor

hrd_management_valid_flag – This 1-bit field is only defined for use in transport streams.

When the AVC timing and HRD descriptor is associated to an AVC video stream carried in a transport stream, then the following applies. If the hrd_management_valid_flag is set to '1', then Buffering Period SEI and Picture Timing SEI messages, as defined in Annex C of ITU-T Rec. H.264 | ISO/IEC 14496-10, shall be present in the associated AVC video stream. These Buffering Period SEI messages shall carry coded initial_cpb_removal_delay and initial_cpb_removal_delay_offset values for the NAL HRD. If the hrd_management_valid_flag is set to '1', then the transfer of each byte from MB_n to EB_n in the T-STD shall be according to the delivery schedule for that byte into the CPB in the NAL HRD, as determined from the coded initial_cpb_removal_delay and initial_cpb_removal_delay_offset values for SchedSelIdx = cpb_cnt_minus1. When the hrd_management_valid_flag is set to '0', the leak method as defined in 2.14.3.1 shall be used for the transfer from MB_n to EB_n in the T-STD.

When the AVC timing and HRD descriptor is associated to an AVC video stream carried in a program stream, then the meaning of the hrd_management_valid_flag is not defined.

picture_and_timing_info_present – This 1-bit field when set to '1' indicates that the 90kHz_flag and parameters for accurate mapping to 90-kHz system clock are included in this descriptor.

90kHz_flag, N, K – The 90kHz_flag when set to '1' indicates that the frequency of the AVC time base is 90 kHz. For an AVC video stream the frequency of the AVC time base is defined by the AVC parameter time_scale in VUI parameters, as defined in Annex E of ITU-T Rec. H.264 | ISO/IEC 14496-10. The relationship between the AVC time_scale and the STC shall be defined by the parameters N and K in this descriptor as follows.

$$time\_scale = \frac{(N \times system\_clock\_frequency)}{K}$$

where time_scale denotes the exact frequency of the AVC time base, with K larger than or equal to N.

If the 90kHz_flag is set to '1', then N equals 1 and K equals 300. If the 90kHz_flag is set to '0', then the values of N and K are provided by the coded values of the N and K fields.

NOTE 1 – This allows mapping of time expressed in units of time_scale to 90 kHz units, as needed for the calculation of PTS and DTS timestamps, for example in decoders for AVC access units for which no PTS or DTS is encoded in the PES header.

num_units_in_tick – Coded exactly in the same way as the num_units_in_tick field in VUI parameters in Annex E of ITU-T Rec. H.264 | ISO/IEC 14496-10. The information provided by this field shall apply to the entire AVC video stream to which the AVC timing and HRD descriptor is associated.

fixed_frame_rate_flag – Coded exactly in the same way as the fixed_frame_rate_flag in VUI parameters in Annex E of ITU-T Rec. H.264 | ISO/IEC 14496-10. When this flag is set to '1', it indicates that the coded frame rate is constant within the associated AVC video stream. When this flag is set to '0', no information about the frame rate of the associated AVC video stream is provided in this descriptor.

temporal_poc_flag – When the temporal_poc_flag is set to '1' and the fixed_frame_rate_flag is set to '1', then the associated AVC video stream shall carry Picture Order Count (POC) information (PicOrderCnt) whereby pictures are counted in units of $\Delta t_{fi,dpb}(n)$, where $\Delta t_{fi,dpb}(n)$ is specified in equation E-10 of ITU-T Rec. H.264 | ISO/IEC 14496-10. When the temporal_poc_flag is set to '0', no information is conveyed regarding any potential relationship between the POC information in the AVC video stream and time.

NOTE 2 – This reduces the overhead necessary to signal timing for each access unit. An effective PTS and DTS can be calculated for access units for which no explicit PTS/DTS is carried. Repetition of most recently presented field of the appropriate parity (or frame) is implied when the difference between the PTSs of the current and the next picture is greater than $2 \times \Delta t_{fi,dpb}$ (or greater than $\Delta t_{fi,dpb}$ when *frame_mbs_only_flag* is equal to 1).

picture_to_display_conversion_flag – This 1-bit field when set to '1' indicates that the associated AVC video stream may carry display information on coded pictures by providing the pic_struct field in picture_timing SEI messages (see Annex D of ITU-T Rec. H.264 | ISO/IEC 14496-10) and/or by providing the Picture Order Count (POC) information (PicOrderCnt), whereby pictures are counted in units of $\Delta t_{fi,dpb}(n)$ (see also the semantics of temporal_poc_flag), so that timing information for a successive AVC access unit can be derived from the previous picture in decoding or presentation order.

When the picture_to_display_conversion_mode_flag is set to '0', then picture timing SEI messages in the AVC video stream, if present, shall not contain the pic_struct field, and hence the pic_struct_present_flag shall be set to '0' in the VUI parameters in the AVC video stream.

20) Subclause 2.7.4

Replace in subclause 2.7.4 "Frequency of presentation timestamp coding" the sentence:

In the case of still pictures the 0,7 s constraint does not apply

by:

The 0.7 s constraint does not apply in the case of:

- still pictures as defined in 2.1;
- AVC still pictures;
- AVC access units with a very low frame rate, where the presentation time of subsequent access units differs by more than 0.7 s. In this particular case, the VUI parameters num_units_in_tick and time_scale shall be present either in the AVC video stream or in an AVC-timing and HRD descriptor associated to the AVC video stream.

NOTE – The presentation time of an AVC access unit is equivalent to the DPB output time $t_{o,dpb}(n)$ defined in Annex C of ITU-T Rec. H.264 | ISO/IEC 14496-10.

21) Subclause 2.7.5

Insert the following at the end of subclause 2.7.5:

For each AVC 24-hour picture, no explicit PTS and DTS value shall be encoded in the PES header. For such AVC access unit, decoders shall infer the presentation time from the parameters within the AVC video stream. Therefore, each AVC video stream that contains one or more AVC 24-hour picture(s):

- shall either carry picture timing SEI messages with coded values of cpb_removal_delay and dpb_output_delay; or
- shall carry VUI parameters with the fixed_frame_rate_flag set to '1' and shall carry Picture Order Count (POC) information (PicOrderCnt) whereby pictures are counted in units of $\Delta t_{fi,dpb}(n)$, where $\Delta t_{fi,dpb}(n)$ is specified in equation E-10 of ITU-T Rec. H.264 | ISO/IEC 14496-10.

NOTE 1 – The requirements in the second bullet are met if an AVC timing and HRD descriptor is associated with the AVC video stream with the fixed_frame_rate_flag set to '1' and the temporal_poc_flag set to '1'.

The following applies to AVC access units in an AVC video stream carried in an ITU-T Rec. H.222.0 | ISO/IEC 13818-1 stream. For each AVC access unit that does not represent an AVC 24-hour picture, a PES header with a coded PTS and, if applicable, DTS value shall be provided, unless all conditions expressed under one of the following four bullets are true:

- In the AVC video sequence the following SEI messages are present, as signalled by VUI parameters:
  a) picture timing SEI messages providing the cpb_removal_delay and the dpb_output_delay parameters; and

b) buffering period SEI messages providing the initial_cpb_removal_delay and the initial_cpb_removal_delay_offset parameters.

NOTE 2 – When picture timing SEI messages are present in the AVC video sequence, then these messages are present for each AVC access unit, as required by ITU-T Rec. H.264 | ISO/IEC 14496-10. When buffering period SEI messages are present in the AVC video sequence, then these messages shall be present for each IDR access unit and for each access unit that is associated with a recovery point SEI message, as required by ITU-T Rec. H.264 | ISO/IEC 14496-10.

- An AVC timing and HRD descriptor is associated with the AVC video stream and in this descriptor the fixed_frame_rate_flag is set to '1' and the temporal_poc_flag is set to '1'.
- An AVC timing and HRD descriptor is associated with the AVC video stream and in this descriptor the fixed_frame_rate_flag is set to '1', the picture_to_display_conversion_flag is set to '1', the temporal_poc_flag is set to '0' and in the AVC video sequence picture timing SEI messages with the pic_struct field are present.

  NOTE 3 – In this specific case the pic_struct field is used to determine subsequent PTS values.
- An AVC timing and HRD descriptor is associated with the AVC video stream and in this descriptor the fixed_frame_rate_flag is set to '1' and the temporal_poc_flag is set to '0' and the picture_to_display_conversion_flag is set to '0'.

  NOTE 4 – In this case the POC information in the AVC video stream is used to determine the subsequent PTS values.

22) Subclause 2.7.6

Replace in subclause 2.7.6 "Timing constraints for scalable coding" the text:

If an audio sequence is coded using an ISO/IEC 13818-3 extension bitstream, corresponding decoding/presentation units in the two layers shall have identical PTS values.

If a video sequence is coded as a SNR enhancement of another sequence, as specified in 7.8 of ITU-T Rec. H.262 | ISO/IEC 13818-2, the set of presentation times for both sequences shall be the same.

If a video sequence is coded as two partitions, as specified in 7.10 of ITU-T Rec. H.262 | ISO/IEC 13818-2, the set of presentation times for both partitions shall be the same.

If a video sequence is coded as a spatial scalable enhancement of another sequence, as specified in 7.7 of ITU-T Rec. H.262 | ISO/IEC 13818-2, the following shall apply:

- If both sequences have the same frame rate, the set of presentation times for both sequences shall be the same.

  NOTE – This does not imply that the picture coding type is the same in both layers.
- If the sequences have different frame rates, the set of presentation times shall be such that as many presentation times as possible shall be common to both sequences.
- The picture from which the spatial prediction is made shall be one of the following:
  - the coincident or most recently decoded lower layer picture;
  - the coincident or most recently decoded lower layer picture that is an I- or P-picture;
  - the second most recently decoded lower layer picture that is an I- or P-picture, and provided that the lower layer does not have low_delay set to '1'.

If a video sequence is coded as a temporally scalable enhancement of another sequence, as specified in 7.9 of ITU-T Rec. H.262 | ISO/IEC 13818-2, the following lower layer pictures may be used as the reference. Times are relative to presentation times:

- the coincident or most recently presented lower layer picture;
- the next lower layer picture to be presented.

by:

If an audio sequence is coded using an extension bitstream, such as specified in ISO/IEC 13818-3, then corresponding decoding/presentation units in the two layers shall have identical PTS values.

If a video sequence is coded as an SNR enhancement of another sequence, such as specified in 7.8 of ITU-T Rec. H.262 | ISO/IEC 13818-2, then the set of presentation times for both sequences shall be the same.

If a video sequence is coded as two partitions, such as specified in 7.10 of ITU-T Rec. H.262 | ISO/IEC 13818-2, then the set of presentation times for both partitions shall be the same.

If a video sequence is coded as a spatial scalable enhancement of another sequence, such as specified in 7.7 of ITU-T Rec. H.262 | ISO/IEC 13818-2, then the following shall apply:

- If both sequences have the same frame rate, the set of presentation times for both sequences shall be the same.

  NOTE – This does not imply that the picture coding type is the same in both layers.
- If the sequences have different frame rates, the set of presentation times shall be such that as many presentation times as possible shall be common to both sequences.
- The picture from which the spatial prediction is made shall be one of the following:
  - the coincident or most recently decoded lower layer picture;
  - the coincident or most recently decoded lower layer picture that is an I- or P-picture;
  - the second most recently decoded lower layer picture that is an I- or P-picture, and provided that the lower layer does not have the low_delay flag set to '1'.

If a video sequence is coded as a temporally scalable enhancement of another sequence, such as specified in 7.9 of ITU-T Rec. H.262 | ISO/IEC 13818-2, then the following lower layer pictures may be used as the reference. Times are relative to presentation times of:

- the coincident or most recently presented lower layer picture;
- the next lower layer picture to be presented.

23) Subclause 2.7.9

Replace the following text under "Decoder Buffer Size" in subclause 2.7.9:

In the case of a video elementary stream in a CSPS, the following applies:

BS_n has a size which is equal to the sum of the size of the video buffer verifier (vbv) as specified in ITU-T Rec. H.262 | ISO/IEC 13818-2 and an additional amount of buffering BS_{add}. BS_{add} is specified as:

$$BS_{add} \leq MAX\,[6 \times 1024,\ R_{vmax} \times 0{,}001] \text{ bytes}$$

where R_{vmax} is the maximum video bit rate of the video elementary stream.

by:

In the case of an ITU-T Rec. H.262 | ISO/IEC 13818-2 or ISO/IEC 11172-2 video elementary stream in a CSPS, the following applies:

BS_n has a size which is equal to the sum of the size of the Video Buffer Verifier (VBV) as specified in the ITU-T Rec. H.262 | ISO/IEC 13818-2 or ISO/IEC 11172-2 stream, respectively, and an additional amount of buffering BS_{add}. BS_{add} is specified as:

$$BS_{add} \leq MAX\,[6 \times 1024,\ R_{vmax} \times 0.001] \text{ bytes}$$

where R_{vmax} is the maximum bit rate of the ITU-T Rec. H.262 | ISO/IEC 13818-2 or ISO/IEC 11172-2 video elementary stream.

In the case of an ITU-T Rec. H.264 | ISO/IEC 14496-10 video elementary stream in a CSPS, the following applies:

BS_n has a size which is equal to the sum of cpb_size and an additional amount of buffering BS_{add}. BS_{add} is specified as:

$$BS_{add} \leq MAX\,[6 \times 1024,\ R_{vmax} \times 0.001] \text{ bytes}$$

where R_{vmax} is the maximum video bit rate of the AVC video stream, and

where cpb_size is the CpbSize[cpt_cnt_minus1] size of the CPB for the byte stream format signalled in the NAL hrd_parameters() in the AVC video stream. If the NAL hrd_parameters() are not present in the AVC video stream, then the cpb_size shall be the size defined as 1200 × MaxCPB in Annex A of ITU-T Rec. H.264 | ISO/IEC 14496-10 for the applied level.

24) Subclause 2.7.10

a) Replace in subclause 2.7.10 "Transport Stream" the text:

For all presentation units in all video elementary streams in the Transport Stream, the ratio of system_clock_frequency to the actual video frame rate, SCFR, is constant and equal to the value indicated in the following table at the nominal frame rate indicated in the video stream.

by:

For all presentation units in each ISO/IEC 11172-2 video and ITU-T Rec. H.262 | ISO/IEC 13818-2 video stream in the Transport Stream, the ratio of system_clock_frequency to the actual video frame rate, SCFR, is constant and equal to the value indicated in the following table at the nominal frame rate indicated in the video stream.

b) Add trailing paragraph to subclause 2.7.10:

For ISO/IEC 14496-2 video streams carried in a Transport Stream, the time base of the ISO/IEC 14496-2 video stream, as defined by vop_time_increment_resolution, shall be locked to the STC and shall be exactly equal to N times system_clock_frequency divided by K, with N and K integers that have a fixed value within each visual object sequence, with K greater than or equal to N.

For ITU-T Rec. H.264 | ISO/IEC 14496-10 video streams, the time base of the ITU-T Rec. H.264 | ISO/IEC 14496-10 video stream shall be locked to the system clock frequency. The frequency of the AVC time base is defined by the AVC parameter time_scale, and this frequency shall be exactly equal to N times system_clock_frequency divided by K, with N and K integers that have a fixed value within each AVC video sequence and K greater than or equal to N. For example, if the time_scale is set to 90 000, then the frequency of the AVC time base is exactly equal to system_clock_frequency divided by 300.

25) Subclause 2.11.1

Replace subclause 2.11.1 by:

2.11.1 Introduction

An ITU-T Rec. H.222.0 | ISO/IEC 13818-1 stream may carry individual ISO/IEC 14496-2 and 14496-3 elementary streams as well as ISO/IEC 14496-1 audiovisual scenes with its associated streams. Typically, the ISO/IEC 14496 streams will be elements of an ITU-T Rec. H.222.0 | ISO/IEC 13818-1 program, as defined by the PMT in a Transport Stream and the PSM in a Program Stream.

For the carriage of ISO/IEC 14496 data in Transport Streams and Program Streams, distinction is made between individual elementary streams and an ISO/IEC 14496-1 audiovisual scene with its associated streams. For carriage of individual ISO/IEC 14496-2 and 14496-3 elementary streams, only system tools from ITU-T Rec. H.222.0 | ISO/IEC 13818-1 are used, as defined in 2.11.2. For carriage of an audiovisual ISO/IEC 14496-1 scene and associated ISO/IEC 14496 elementary streams, contained in ISO/IEC 14496-1 SL_packetized streams or FlexMux streams, tools from both ITU-T Rec. H.222.0 | ISO/IEC 13818-1 and from ISO/IEC 14496-1 are used, as defined in 2.11.3.

Carriage of ITU-T Rec. H.264 | ISO/IEC 14496-10 video over ITU-T Rec. H.222.0 | ISO/IEC 13818-1 streams is specified in 2.14.

26) Subclause 2.11.2.1

Replace the last paragraph of subclause 2.11.2.1 by:

Carriage of individual ISO/IEC 14496-2 and ISO/IEC 14496-3 elementary streams in PES packets shall be identified by appropriate stream_id and stream_type values, indicating the use of ISO/IEC 14496-2 Visual or 14496-3 Audio. In addition, such carriage shall be signalled by the MPEG-4_video descriptor or MPEG-4_audio descriptor, respectively. These descriptors shall be conveyed in the descriptor loop for the respective elementary stream entry in the Program Map Table in case of a Transport Stream or in the Program Stream Map, when present, in case of a Program Stream. ITU-T Rec. H.222.0 | ISO/IEC 13818-1 does not specify presentation of ISO/IEC 14496-2 and ISO/IEC 14496-3 elementary streams in the context of a program.