

SMPTE RECOMMENDED PRACTICE

VC-4 Bitstream Transport Encodings



Table of Contents	Page
Foreword	3
Intellectual Property	3
Introduction.....	3
1 Scope	4
2 Conformance Notation	4
3 Normative References	4
4 Overview (Informative).....	5
5 Notation.....	5
5.1 Abbreviations	5
5.2 Definition of Terminology	6
5.3 Mnemonics.....	7
6 MPEG-2 Transport Stream Encoding	7
6.1 Signaling of VC-4 Elementary Streams	7
6.2 Encapsulation of VC-4 Elementary Streams.....	13
6.3 T-STD Buffer Model for VC-4 Streams	18
7 MPEG-2 Program Stream Encoding.....	23
7.1 Signaling of VC-4 Elementary Streams	23
7.2 Encapsulation of VC-4 Elementary Streams.....	24
7.3 P-STD Buffer Model for VC-4 Streams.....	24
Annex A Graphics of Syntax for this SMPTE Recommended Practice (Informative).....	29
A.1 Introduction	29
A.2 PES Packet.....	29
A.3 TS Program Map Section.....	30
A.4 Program Stream Map.....	31
Annex B Bibliography (Informative)	32

Table of Figures

Figure 1 – Transport Stream System Target Decoder for VC-4 streams.....	19
Figure 2 – Program Stream System Target Decoder for VC-4 streams	25
Figure A.1 – PES packet syntax diagram for SMPTE elementary stream	29
Figure A.2 – TS Program Map Section diagram for VC-4 elementary stream	30
Figure A.3 – Program Stream Map diagram for VC-4 elementary stream	31

Table of Tables

Table 1 – Syntax for registration descriptor.....	8
Table 2 – Lists of values applicable to the subdescriptor_tag.....	8
Table 3 – Syntax for the profile and level sub-descriptor	9
Table 4 – Syntax for the alignment sub-descriptor	9
Table 5 – Alignment type for VC-4 elementary stream	10
Table 6 – Syntax for the buffer size sub-descriptor	10
Table 7 – Syntax for the scalable extension sub-descriptor.....	11
Table 8 – Lists of values applicable to the frame_field_coding_mode/base_layer_frame_field_coding_mode	12
Table 9 – Elements and notations for T-STD model	19
Table 10 – Elements and notations for P-STD model.....	26

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 RP 2058-3 was prepared by Technology Committee 10E on Essence.

Intellectual Property

SMPTE draws attention to the fact that it is claimed that compliance with this Recommended Practice 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 Recommended Practice to those (Members and non-Members alike) desiring to implement this Recommended Practice 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 Recommended Practice may be subject to IPR other than those identified above. The Society shall not be responsible for identifying any or all such IPR.

Introduction

This section is entirely informative and does not form an integral part of this Engineering Document.

The purpose of this document is to define a minimum set of rules for the carriage of a layered video extension elementary stream in MPEG-2 protocols where an elementary stream from the base layer is specified by other video coding standards, such as ITU-T H.262 (MPEG-2) or ITU-T H.264 (MPEG-4 AVC), etc. In particular, this document is aimed to provide a generic method for carrying enhancement elementary streams from the SMPTE Layered Video Extension (VC-4) coding and an elementary stream from the base layer over MPEG-2 Transport Streams. It is also aimed to provide a generic method to carry enhancement streams from the SMPTE Layered Video Extension (VC-4) coding and the base layer stream in an MPEG-2 Program Stream.

1 Scope

This document describes the protocols for carrying video elementary streams conforming to the SMPTE Layered Video Extension (VC-4) over MPEG-2 Transport Streams and MPEG-2 Program Streams. This document does not describe the video bitstream syntax. It describes how to encapsulate and signal the SMPTE Layered Video Extension (VC-4) bitstream using ISO/IEC 13818-1:2007 and ISO/IEC 13818-1:2007/Amd3:2009.

This document is not concerned with the carriage of audio elementary streams or data elementary streams that may be multiplexed with the Layered Video Extension elementary stream in the same MPEG-2 Transport Stream or the same MPEG-2 Program Stream. The synchronization of audio access units in an audio stream with Layered Video Extension video access units in a video elementary stream should be achieved by means of the synchronization layer specified by ISO/IEC 13818-1.

Any field of the MPEG-2 Transport Stream packet header or MPEG-2 PES packet header that is not mentioned in this document shall keep its default meaning as defined in ISO/IEC 13818-1:2007 and ISO/IEC 13818-1:2007/Amd3:2009

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.

Unless otherwise specified, the order of precedence of the types of normative information in this document shall be as follows: Normative prose shall be the authoritative definition; Tables shall be next; followed by formal languages; then figures; and then any other language forms.

3 Normative References

Note: All references in this document to other SMPTE documents use the current numbering style (e.g. SMPTE ST 421:2006) although, during a transitional phase, the document as published (printed or PDF) may bear an older designation (such as SMPTE 421M-2006). Documents with the same root number (e.g. 421) and publication year (e.g. 2006) are functionally identical.

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

SMPTE ST 421:2006, Television — VC-1 Compressed Video Bitstream Format and Decoding Process

Amendment 1:2007 to SMPTE ST 421:2006

Amendment 2:2011 to SMPTE ST 421:2006

SMPTE ST 2058-1:2011, VC- 4 Layered Video Extension Bitstream Format and Decoding Process

SMPTE RP 227-2010, VC-1 Bitstream Transport Encodings

ISO/IEC 13818-1:2007, Information Technology — Generic Coding of Moving Pictures and Associated Audio Information: Systems (3rd Edition)

ISO/IEC 13818-1:2007/Amd3:2009, Transport of Scalable Video over Recommendation ITU-T H.222.0|ISO/IEC 13818-1

4 Overview (Informative)

This document depends on ISO/IEC 13818-1:2007 and ISO/IEC 13818-1:2007/Amd3:2009 and defines extensions that enable the encapsulation and signaling of SMPTE Layered Video Extension (VC-4) elementary streams. This results in implementations compatible with ISO/IEC 13818-1:2007 and ISO/IEC 13818-1:2007/Amd3:2009.

The bitstream specified by the SMPTE Layered Video Extension (VC-4) carries residual signals that when integrated with the video content from the base layer can provide enhanced quality as well as enhanced fidelity in bit depth, resolution and color difference. Syntax elements specifying the control parameters for the picture reconstruction or integration process are also carried in the same bitstreams. The 8-bit 4:2:0 pictures can be pictures decoded, for example, from SMPTE ST 421 compliant bitstreams as well as content decoded from bitstreams specified by other video coding standards, such as ITU-T H.262 (MPEG-2) or ITU-T H.264 (MPEG-4 AVC), etc. The 8-bit 4:2:0 pictures are referred to as base layer pictures in this document. The residual bitstreams can be decoded independently from the base layer picture decoding process.

This document specifies how to map SMPTE Layered Video Extension (VC-4) elementary streams to MPEG-2 Transport Streams and how to map SMPTE Layered Video Extension (VC-4) elementary streams to MPEG-2 Program Streams. In addition, this document describes how to encapsulate SMPTE Layered Video Extension (VC-4) elementary streams to packetized elementary streams (PESs). The definition and operation of the hypothetical buffer model for SMPTE Layered Video Extension (VC-4) elementary streams are specified as ISO/IEC 13818-1:2007 and ISO/IEC 13818-1:2007/Amd3:2009 for other video coding standards.

The document employs the similar method to SMPTE RP 227 (VC-1 Bitstream Transport Encodings) to map the SMPTE Layered Video Extension (VC-4) bitstream to MPEG-2 System protocols. The document also pursues the consistency with the previous SMPTE document.

5 Notation

5.1 Abbreviations

This section provides a list of acronyms used in this document. Those are defined in ISO/IEC 13818-1.

CRC	Cyclic Redundancy Check
DTS	Decoding Time Stamp

ES	Elementary Stream
ESCR	Elementary Stream Clock Reference
PAT	Program Association Table
PCR	Program Clock Reference
PES	Packetized Elementary Stream
PID	Packet Identifier
PMT	Program Map Table
PS	Program Stream
PSI	Program Specific Information
PSM	Program Stream Map
P-STD	Program Stream System Target Decoder
PTS	Presentation Time Stamp
PU	Presentation Unit
SCR	System Clock Reference
VC-4	Shortened Name of SMPTE Layered Video Extension
TS	Transport Stream
T-STD	Transport Stream System Target Decoder

5.2 Definition of Terminology

5.2.1 Data Access Unit Definition

The following definition for an access unit shall be applicable for both MPEG-2 Transport Stream and MPEG-2 Program Stream.

An access unit is defined to be a coded representation of a single picture in a VC-4 elementary stream. An access unit shall include all coded data for a picture and any flushing bits that follow it, up to not including the start code of the next access unit. The start code of the next access unit shall be the byte of a sequence start code, frame start code or field start code. If a picture is not preceded by a sequence header, the access unit shall start with the picture start code. Otherwise the access unit shall start with the first byte of sequence start code.

5.2.2 Base Layer

A picture that is decoded from the base layer stream, specified by other video coding standards, such as ITU-T H.262 (MPEG-2) or ITU-T H.264 (MPEG-4 AVC), etc, and is used in reconstructing a new picture when integrating with the residual picture decoded from the residual layer stream. The synchronization of the base layer picture and other residual layer streams is expected to be signaled in the system layer.

5.2.3 Base Layer Stream

A sequence of bits that forms the representation of coded picture at the base layer consists of one or more video sequences which are specified by other video coding standards, such as ITU-T H.262 (MPEG-2) or ITU-T H.264 (MPEG-4 AVC), etc.

5.2.4 Enhancement Layer

A residual layer has at least one lower layer. The lowest layer is a base layer. The number of enhancement layer should be greater than or equal to one and each enhancement layer is a residual layer.

5.2.5 Enhancement Layer Stream

A VC-4 residual enhancement layer stream consists of one or more VC-4 residual sequences. A VC-4 residual enhancement layer stream may also result from re-assembling enhancement residual sub-bitstreams.

5.2.6 VC-4 Elementary Stream

The VC-4 coding may generate multiple enhancement layer streams. Each enhancement layer stream shall be an elementary stream and conform to the profile and the level defined in Annex D of SMPTE ST 2058-1. A scalable video stream set or a scalable video service consists of the video sub-stream from base layer and these multiple enhancement layer streams. The VC-4 elementary stream is defined as one enhancement layer stream among multiple enhancement layer streams coded by SMPTE ST 2058-1. A VC-4 elementary stream is also called a VC-4 video sub-stream. This elementary stream consists of the data which are associated with the same LAYER_ID defined in SMPTE ST 2058-1.

5.2.7 VC-4 Video Sub-Stream

A VC-4 video sub-stream is one stream among multiple enhancement layer streams coded by SMPTE ST 2058-1. This sub-stream is treated as an elementary stream, and has a unique PID or **elementary_PID**.

5.2.8 Video Sub-Stream

A video sub-stream is defined as one elementary stream among multiple elementary streams which make a scalable video stream set. The video sub-stream from base layer is specified by other video coding standards, such as ITU-T H.262 (MPEG-2) or ITU-T H.264 (MPEG-4 AVC), etc. Rest enhancement video sub-streams are from SMPTE-LVE coding. Each video sub-stream is assigned by **hierarchy_layer_index** of **hierarchy_descriptor()**.

5.3 Mnemonics

The following mnemonics are defined to describe the different data types used in the coded bitstream:

uimsbf Unsigned integer, most-significant bit first.

6 MPEG-2 Transport Stream Encoding

The bitstream syntax is defined in separate tables that use a C-like pseudo-code notation to expose the sequential order of data fields in the stream.

6.1 Signaling of VC-4 Elementary Streams

MPEG-2 elementary streams carrying VC-4 residual elementary stream data shall be signaled in a PMT as defined in ISO/IEC 13818-1. The syntax and semantics of these signaling parameters are specified in this section.

6.1.1 Stream Type

The **stream_type** value in the PMT describing a VC-4 elementary stream should be set to **0xEB**. This value indicates that this elementary stream is considered to be a privately-handled stream. The scope of this private value shall be captured by means of the **registration_descriptor()** of ISO/IEC 13818-1.

6.1.2 Registration Descriptor for VC-4

At least one MPEG-2 **registration_descriptor()** shall be present in the inner descriptor loop of the MPEG-2 program element listed in the **TS_program_map_section()** corresponding to the VC-4 elementary stream. The syntax and semantics for this descriptor appears in Table 1 and in the subsequent text.

Table 1 – Syntax for registration descriptor

Syntax	Bits	Format
<pre> registration_descriptor() { descriptor_tag descriptor_length format_identifier for(i = 0; i < K; i++) { subdescriptor() } }</pre>	<p>8</p> <p>8</p> <p>32</p> <p>N*8</p>	<p>0x05</p> <p>uimsbf</p> <p>0x56432D34</p> <p>uimsbf</p>

descriptor_tag – This 8-bit field shall be set to 0x05 to identify this data structure as an MPEG-2 registration descriptor.

descriptor_length – This 8-bit field shall specify the number of bytes the descriptor immediately following the **descriptor_length** field.

format_identifier – The value for this 32-bit field shall be set to 0x56432D34 (“VC-4”).

subdescriptor() – A data structure of up to 251 bytes of additional information whose syntax is determined in a case-by-case basis. The first field of a **subdescriptor()** shall be an 8-bit value known as the **subdescriptor_tag** which identifies syntax and semantics for a particular sub-descriptor. Table 2 defines values for the **subdescriptor_tag**.

Table 2 – Lists of values applicable to the subdescriptor_tag

Assigned Value	Description
0x00	Indicates a null sub-descriptor as SMPTE RP 227. A null sub-descriptor consists only of an 8-bit subdescriptor_tag field.
0x01	Profile and level sub-descriptor defined as part of this document.
0x02	Alignment sub-descriptor defined as part of this document
0x03	Buffer size sub- descriptor defined as part of this document.
0x04	SMPTE Reserved for future applications.
0x05	Scalability extension sub-descriptor defined as part of this document.
0x06 ~ 0xFE	SMPTE Reserved for future applications.
0xFF	Indicates a null sub-descriptor as SMPTE RP 227. A null sub-descriptor consists only of an 8-bit subdescriptor_tag field.

In Table 1, the sub-descriptors listed in Table 2 shall always appear in increasing order of their **subdescriptor_tag** value like SMPTE RP 227.

If an unknown sub-descriptor is found, the sub-descriptor and subsequent sub-descriptors shall be ignored.

6.1.3 Profile and Level Sub-descriptor

profile_level_subdescriptor() specifies the profile and level used to decode current layer stream which is one elementary layer stream among the several VC-4 enhancement elementary streams. Syntax for this sub-descriptor is defined in Table 3. The **profile_level_subdescriptor()** signals the profile and level for the associated VC-4 elementary stream.

Table 3 – Syntax for the profile and level sub-descriptor

Syntax	Bits	Format
profile_level_subdescriptor() {		
subdescriptor_tag	8	0x01
profile	4	uimsbf
level	4	uimsbf
}		

subdescriptor_tag – This 8-bit field shall be set to 0x01 to identify this data structure as the profile and level sub-descriptor.

profile – This 4-bit field shall specify the profile used for encoding a VC-4 elementary video stream or sub-stream. This field has the same value as the syntax **PROFILE** of the sequence-level header of the corresponding VC-4 elementary stream (see Section 7.1.2 in SMPTE ST 2058-1).

level – This 4-bit field shall specify the level used for encoding a VC-4 elementary video stream or sub-stream. This field has the same value as the syntax **LEVEL** of the sequence-level header of the corresponding VC-4 elementary stream (see Section 7.1.3 in SMPTE ST 2058-1).

6.1.4 Alignment Sub-descriptor

This sub-descriptor defines which type of alignment is inserted between the coded byte sequence and a PES packet by an 8-bit field. Syntax for this sub-descriptor is defined in Table 4.

Table 4 – Syntax for the alignment sub-descriptor

Syntax	Bits	Format
alignment_subdescriptor() {		
subdescriptor_tag	8	0x02
alignment_type	8	uimsbf
}		

subdescriptor_tag – This 8-bit field shall be set to 0x02 to identify this data structure as the alignment sub-descriptor.

alignment_type – This 8-bit field shall specify the alignment type used for encoding a VC-4 elementary video stream or sub-stream. Table 5 lists available alignment types.

Table 5 – Alignment type for VC-4 elementary stream

Alignment Type	Description
0x00	SMPTE Reserved
0x01	Slice or Video Access Unit
0x02	Video Access Unit
0x03	Sequence
0x04	Frame
0x05	Field
0x06 ~ 0xFF	SMPTE Reserved

6.1.5 Buffer Size Sub-descriptor

buffer_size_subdescriptor() specifies the minimum size of the elementary stream buffer EB_n (in T-STD) or B_n (in P-STD) needed in the decoder to decode the ES conveyed in the MPEG-2 program element associated with this sub-descriptor. The buffer size associated with the VC-4 profile and level shall be assumed if this sub-descriptor is not present and if there are no HRD parameters specified in the sequence header. Table 6 shows how to specify the buffer size of the elementary stream associated with this sub-descriptor.

Table 6 – Syntax for the buffer size sub-descriptor

Syntax	Bits	Format
<i>buffer_size_subdescriptor()</i> {		
subdescriptor_tag	8	0x03
reserved	4	0xF
hrd_buffer_size_exponent	4	uimsbf
hrd_buffer_size_mantissa	16	uimsbf
}		

subdescriptor_tag – This 8-bit field shall be set to 0x03 to identify this data structure as the buffer size sub-descriptor.

hrd_buffer_size_exponent – This 4-bit field shall specify the value of the exponent to use in calculating EBS_n the size of the video elementary stream buffer EB_n in T-STD or BS_n the size of the video elementary stream buffer B_n in p-STD. The range of this value is from 4 to 19.

hrd_buffer_size_mantissa – This 16-bit field shall specify the value of the mantissa to use in calculating EBS_n the size of the video elementary stream buffer EB_n in T-STD or BS_n the size of the video elementary stream buffer B_n in p-STD. The range of this value is from 1 to 2^{16} .

Thus, EBS_n or $BS_n = (\text{hrd_buffer_size_mantissa} + 1) \ll (\text{hrd_buffer_size_exponent} + 4)$. The resulting video elementary stream buffer size shall not exceed the maximum video elementary stream buffer size for the profile and level associated with the VC-4 elementary stream.

6.1.6 Scalability Extension Sub-descriptor

scalability_extension_subdescriptor() shall specify the information about the associated video sub-bitstreams and the format of the base video stream.

Table 7 – Syntax for the scalable extension sub-descriptor

Syntax	Bits	Format
<i>scalability_extension_subdescriptor()</i> {		
subdescriptor_tag	8	0x05
layer_id	4	uimsbf
frame_field_coding_mode	4	uimsbf
sub_layer_id_min	4	uimsbf
sub_layer_id_max	4	uimsbf
reserved	8	0xFF
max_gop_length	14	uimsbf
hierarchy_embedded_layer_id	4	uimsbf
hierarchy_embedded_layer_max_gop_length	14	uimsbf
if(layer_id == 0x1) {		
base_layer_stream_type	8	uimsbf
if(base_layer_stream_type > 0x80){		
base_layer_format_identifier	8	uimsbf
}		
base_layer_frame_field_coding_mode	4	uimsbf
reserved	4	0xF
}		
reserved	16	0xFFFF
}		

subdescriptor_tag – This 8-bit field shall be set to 0x05 to identify this data structure as the scalability extension sub-descriptor.

layer_id – This 4-bit field shall identify the layer of the VC-4 stream associated with this descriptor among all enhancement layer streams. This value shall be **LAYER_ID_SEQ** in the sequence header, **LAYER_ID** in the picture header or **LAYER_ID_SLICE** in the slice header of this VC-4 elementary stream.

frame_field_coding_mode – This 4-bit field shall specify this video elementary stream is coded with which coding mode among the interlaced field mode, the interlaced frame mode, the interlaced frame/field mixture mode or the progressive frame mode. Table 8 defines values for the **frame_field_coding_mode**.

sub_layer_id_min – This 4-bit field shall indicate the minimum value of the **QUALITY_LAYER_ID** of this VC-4 elementary stream.

sub_layer_id_max – This 4-bit field shall indicate the maximum value of the **QUALITY_LAYER_ID** of this VC-4 elementary stream.

Table 8 – Lists of values applicable to the frame_field_coding_mode/base_layer_frame_field_coding_mode

Assigned Value	Description
0x00	Interlaced separated-field mode which means all pictures of interlaced sequence are coded with the field picture mode. The pair of two field pictures is not to be coded sequentially.
0x01	Interlaced paired-field mode which means all pictures of interlaced sequence are coded with the field picture mode. The pair of two field pictures is to be coded sequentially.
0x02	Interlaced frame mode which means all interlaced pictures of interlaced sequence are coded with the frame picture mode. Here, Two successive fields are considered as one frame.
0x03	Interlaced frame/field mixture mode which means some pictures of interlaced sequences are coded with field picture mode and rest pictures are coded with the frame picture mode. The pair of two field pictures is coded sequentially.
0x04	Progressive frame mode which means all pictures of progressive sequences are coded with the frame picture mode.
0x05 ~ 0x07	SMPTE Reserved

max_gop_length – This 14-bit unsigned integer field shall specify the maximum number of the coded pictures between any two consecutive intra-pictures in the current layer elementary stream. The value of 0 is forbidden.

hierarchy_embedded_layer_id – This 4-bit field shall identify the embedded layer referred by the layer of the VC-4 stream associated with this descriptor among all enhancement layer streams. This value shall be REF_LAYER_ID in the picture header of this VC-4 elementary stream.

hierarchy_embedded_layer_max_gop_length – This 14-bit unsigned integer field shall specify the maximum number of the coded pictures between any two consecutive intra-pictures in the embedded layer elementary stream referred by the current layer. The value of 0 is forbidden. This value shall be equal to a multiple of **max_gop_length** (**hierarchy_embedded_layer_max_gop_length** == $N * \text{max_gop_length}$ where N is any integer number).

base_layer_stream_type – This 8-bit field shall identify the stream type of the base layer video elementary stream.

base_layer_format_identifier – This 8-bit field shall identify the format type of the base layer video elementary stream when the stream type of the base layer video elementary stream is ‘User Private’ (See subsection 2.4.4.9 in ISO/IEC 13818-1:2007 and ISO/IEC 13818-1/Amd3:2009). This value shall be same as **format_identifier** of the **registration_descriptor()** of the base elementary stream.

base_layer_frame_field_coding_mode – This 4-bit field shall specify the base layer video elementary stream is coded with which coding mode among the interlaced field mode, the interlaced frame mode, the interlaced frame/field mixture mode or the progressive frame mode. Table 8 defines values for the **base_layer_frame_field_coding_mode**.

6.1.7 Hierarchy Descriptor

The **hierarchy_descriptor()** of ISO/IEC 13818-1:2007 and ISO/IEC 13818-1:2007/Amd3:2009 is designed to provide information which identifies the program elements containing components of hierarchically coded

video, audio, and private streams. This document employs this descriptor to specify the dependency information and scalability information between the current elementary stream and the embedded lower layer in the inner descriptor loop of the MPEG-2 element listed in the ***TS_program_map_section()*** corresponding to the VC-4 elementary stream..

6.1.8 Other ISO/IEC 13818-1 Elementary Stream Descriptors

ISO/IEC 13818-1:2007 and ISO/IEC 13818-1:2007/Amd3:2009 define various descriptors that under certain conditions could be used to provide additional information elementary streams in a PMT. Some of these descriptors are used for conveying MPEG-2 video, MPEG-2 audio, MPEG-4 video or H.264 video streams, so these descriptors cannot be used for VC-4 elementary streams. Other descriptors provide complementary functionality (e.g. conditional access) and may be used for VC-4 elementary streams. Descriptors defined in ISO/IEC 13818-1:2007 and ISO/IEC 13818-1:2007/Amd3:2009 may be associated with VC-4 elementary streams include:

target_background_grid_descriptor()
video_window_descriptor()
CA_descriptor()
ISO_639_language_descriptor()
system_clock_descriptor()
multiplex_buffer_utilization_descriptor()
copyright_descriptor()
maximum_bitrate_descriptor()
private_data_indicator_descriptor()
smoothing_buffer_descriptor()
STD_descriptor()
ibp_descriptor()

If the ***data_alignment_indicator*** field in the PES packet header is set to 1 and the ***data_stream_alignment_descriptor()*** is present, this ***data_stream_alignment_descriptor()*** specifies type of alignment. However, for conveying VC-4 elementary streams, the ***alignment_subdescriptor()*** defined in Section 6.1.4 shall be used instead of the ***data_stream_alignment_descriptor()***.

6.1.9 PMT Descriptor for Program or Services

ISO/IEC 13818-1:2007 and ISO/IEC 13818-1:2007/Amd3:2009 define descriptors that apply at the MPEG-2 program level. MPEG-2 program-level descriptors appear in the program-level loop of a PMT (also known as the outer loop). Systems that carry VC-4 elementary streams in a program (or service) shall follow the rules and recommendations for program-level descriptors defined in the MPEG-2 suite of documents.

Compliance with other standards that build on top of MPEG-2 Transport Streams, like digital television standards, may require support for additional program-level descriptors which may appear in a PMT for a program using VC-4 elementary streams.

6.2 Encapsulation of VC-4 Elementary Streams

VC-4 coded data encapsulated in PES packets shall comply with PES encapsulation procedures defined in ISO/IEC 13818-1:2007 and ISO/IEC 13818-1:2007/Amd3:2009. This section describes how to use PES packet header fields for conveying VC-4 elementary streams. The PES packet header fields in this section are the fields for which the semantics are modified or/and enhanced when the PES packet payload includes VC-4 elementary data bytes. Any other fields in the PES packet header shall assume their original semantics as defined in ISO/IEC 13818-1:2007 and ISO/IEC 13818-1:2007/Amd3:2009.

6.2.1 Stream ID

For VC-4 elementary streams encapsulated in PES packets, the value for **stream_id** shall be set to be 0xFD to identify that these PES packets employ an extended syntax to permit additional stream types to be identified. Each elementary sub-video stream in enhancement stream has a **stream_id_extension** explained in Section 6.2.6

6.2.2 PES Packet Length

ISO/IEC 13818-1 defines **PES_packet_length** which is a 16 bit field specifying the number of bytes in the PES packet following the last byte of the field. A value of 0 indicates that the PES packet length is neither specified nor bounded and is allowed only in PES packets whose payload consists of bytes from a video elementary stream in Transport Stream packets. Here, the semantics associated with **PES_packet_length** field value 0 shall be VC-4 video elementary streams.

6.2.3 Data Alignment Indicator

This field is used independently or in conjunction with the **data_stream_alignment_descriptor()** defined in ISO/IEC 13818-1 to indicate the type of alignment that exists between a VC-4 elementary bitstream and the start of a PES packet.

For VC-4 elementary streams, the **data_alignment_indicator** shall follow the same semantics as in ISO/IEC 13818-1, except that **alignment_subdescriptor()** in Section 6.1.4 is applied. If there is a data alignment sub-descriptor associated with the VC-4 elementary stream in the PMT, the **data_alignment_indicator** in the PES header shall be set to 1. If **data_alignment_indicator** is set to 1 and there is no data alignment sub-descriptor associated with the VC-4 elementary stream in the PMT, the default alignment type value shall be equal to 0x01. When **data_alignment_indicator** is set to 0, it is not defined whether any such alignment occurs or not.

6.2.4 Time Stamp (PTS or DTS)

For VC-4 streams, PTS and DTS time stamps are used in exactly the same manner as they are defined for MPEG-2 video streams. Particularly, the values of the PTS and DTS fields shall be related to the first video access unit that starts in the payload of the PES packet.

6.2.5 Extension Flags

VC-4 elementary streams encapsulated in PES packets use the **stream_id_extension** defined in ISO/IEC 13818-1 to provide obvious identification of video packets. In order to use this field, the proper values for the three flags of the **PES_extension_flag**, the **PES_extension_flag_2** and the **stream_id_extension_flag** shall be assigned.

To enable the insertion of extensions in the PES packet header, the **PES_extension_flag** shall be set to 1.

After the insertion of extensions in the PES packet header, the **PES_extension_flag_2** shall be set to 1 to enable the insertion of the optional second group of extensions in the PES packet header.

Finally, in order to insert a valid **stream_id_extension** field, the **stream_id_extension_flag** in the optional second group of extensions in the PES packet header shall be set to 0.

6.2.6 Stream ID Extension

For VC-4 elementary streams, the **stream_id_extension** field defined in ISO/IEC 13818-1 shall have any value in the range from 0x60 to 0x6F. These values are defined within the allowed private range for the

private_stream in ISO/IEC 13818-1. All VC-4 enhancement streams of one scalable video service have the same **stream_id** and the same **stream_id_extenstion**. Each VC-4 elementary stream or VC-4 video sub-stream is identified by **layer_id**.

The **registration_descriptor()** for each VC-4 elementary stream and the combination of **stream_id** and **stream_id_extenstion** define the stream of PES packets carrying VC-4 data. The **elementary_PID** and the **hierarchy_descriptor()** enable to indentify each VC-4 elementary video sub-stream among all VC-4 enhancement streams of one scalable video service.

6.2.7 Discontinuity Indicator

For VC-4 elementary streams, the **discontinuity_indicator** field shall op the following semantics:

discontinuity_indicator – This is a 1-bit field which when set to 1 indicates the discontinuity state is true for the current Transport Stream packet. When the **discontinuity_inidicator** is set to 0 or is not present, the discontinuity state is false. The discontinuity indicator is used to indicate two types of discontinuities: system_time-base discontinuities and **continuity_counter** discontinuities.

A system time-base discontinuity is indicated by the use of the **discontinuity_indicator** in Transport Stream packets of a PID designated as a **PCR_PID** (refer to Section 2.4.4.9 in ISO/IEC 13818-1). When the discontinuity state is true for a Transport Stream packet of a PID designated as a **PCR_PID**, the next PCR in a Transport Stream packet with that same PID represents a sample of a new system time clock for the associated program. The system time-base discontinuity point is defined to be the instant in time when the first byte of a packet containing a PCR of a new system time-base arrives at the input of the Transport Target Decoder (T-STD). The **discontinuity_indicator** shall be set to 1 in the packet in which the system based-time discontinuity occurs. The **discontinuity_indicator** bit may also set to 1 in Transport Stream packets of the same **PCR_PID** prior to the packet which contains the new system time-base PCR. In this case, once the **discontinuity_indicator** has been set to 1, in all Transport Stream packets of the same **PCR_PID** up to and including the Transport Stream packet which contains the first PCR of the new system time-base. After the occurrence of a system time-base discontinuity, no fewer than two PCRs for the new system time-base shall be received before another system time-base discontinuity can occur. Further, except when trick mode status is true, data from no more than two system time-bases shall be present in the set of T-STD buffers for one program at any time.

Prior to the occurrence of a system time-base discontinuity, the first byte of a Transport Stream packet which contains a PTS or DTS which refers to the new system time-base shall not arrive at the input of the T-STD. After the occurrence of a system time-base discontinuity, the first byte of a Transport Stream octet which contains a PTS or DTS which refers to the previous system time-base shall not arrive at the input of the T-STD.

A **continuity_counter** discontinuity shall be indicated by the use of the **discontinuity_indicator** in any Transport Stream packet. When the discontinuity state is true in any Transport Stream packet of a PID not designated as a **PCR_PID**, the **continuity_counter** in that packet may be discontinuous with respect to the previous Transport Stream packet of the same PID. When the continuity state is true in a Transport Stream packet of a PID that is designated as a **PCR_PID**, the **continuity_counter** may only be discontinuous in the packet in which a system time-base discontinuity occurs. A continuity counter discontinuity point occurs when the discontinuity state is true in a Transport Stream packet and the **continuity_counter** in the same packet is discontinuous with respect to the previous Transport Stream packet of the same PID. A continuity counter discontinuity point shall occur at most one time from the initiation of the discontinuity state until the conclusion of the discontinuity state. Furthermore, for all PIDs that are not designated as **PCR_PIDs**, when the **discontinuity_indicator** is set to 1 in a packet of a specific PID, the **discontinuity_indicator** may be set to 1 in the next Transport Stream packet of that same PID, but shall not be set to 1 in three consecutive Transport Stream packets of that same PID.

For the purpose of this clause, a VC-4 elementary stream access point is defined as follows:

- The first byte of a VC-4 sequence header if there is no sequence start code preceding the sequence header.
- The first byte of the sequence start code if a sequence start code immediately precedes the sequence header.

After continuity counter discontinuity is a Transport Stream packet which is designated as containing VC-4 data is received, the first byte of elementary stream data in a Transport Stream packet of the same PID shall be the first byte of a VC-4 elementary stream access point or a VC-4 end-of-sequence coded followed by an access point. Each Transport Stream packet which contains elementary stream data with a PID not designated as a **PCR_PID**, and in which a continuity counter discontinuity point occurs, and in which a PTS or DTS occurs, shall arrive at the input of the T-STD after the system time-base discontinuity for the associated program occurs. In the case where the discontinuity state is true, if two consecutive Transport Stream packets of the same PID occur which have the same **continuity_counter** value and have **adaptation_field_control** values set to '01b' or '11b', the second packet may be discarded. A Transport Stream shall not be constructed in such a way that discarding such a packet will cause the loss of PES packet payload data or PSI data.

After the occurrence of a **discontinuity_indicator** set to 1 in a Transport Stream packet which contains PSI information, a single discontinuity in the **version_number** of PSI section may occur. At the occurrence of such a discontinuity, a version of the **TS_program_map_section()** of the appropriate program shall be sent with **section_length** == 13 and the **current_next_indicator** == 1, such that there are no program descriptors and no elementary streams described. This shall then be followed by a version of the **TS_program_map_section** for each affected program with the **version_number** incremented by one and the **current_next_indicator** == 1, containing a complete program definition. This indicates a version change in PSI data.

6.2.8 Random Access Indicator

For VC-4 elementary streams, the **random_access_indicator** field shall adopt the following semantics:

random_access_indicator – The **random_access_indicator** is a 1-bit field that indicates that the current Transport Stream packet, and possibly subsequent Transport Stream packets with the same PID, contains some information to aid random access at this point. Specially, 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 a VC-4 elementary stream access point as defined in the semantics of the **discontinuity_indicator** field. In addition, a presentation timestamp shall be present in the PES packet containing the first picture following the VC-4 elementary stream access point.

In a scalable video sub-stream set or service, the random access point of the lower-layer elementary stream shall be the random access point of the upper-layer elementary stream. Therefore, each **hierarchy_descriptor()** is defined for each VC-4 elementary stream and a random access point of the lower-layer elementary stream, which is referred by the **hierarchy_embedded_layer_index** of the **hierarchy_descriptor()** for a certain elementary stream, shall be also a random access point of the elementary stream associated with this **hierarchy_descriptor()**.

6.2.9 Elementary Stream Priority Indicator

For VC-4 elementary streams, the **elementary_stream_priority_indicator** field shall adopt the following semantics:

elementary_stream_priority_indicator – The **elementary_stream_priority_indicator** is a 1-bit field. It indicates, among packet with the same PID, the priority of the elementary stream data carried within the payload of the Transport Stream packet. A '1' indicates that the payload has a higher priority than the payloads of other Transport Stream packets. In the case of a VC-4 elementary stream, this field may be set to 1 only if the payload contains one or more bytes from an intra-coded data. A value of 0 indicates that the payload has the same priority as all other packets which do not have this bit set to 1.

6.2.10 Splice Countdown

For VC-4 elementary streams, the **splice_countdown** field shall adopt the following semantics:

splice_countdown – The **splice_countdown** is an 8-bit field representing a value which may be positive or negative. A positive value specifies the remaining number of Transport Stream packets, of the same PID, following the associated Transport Stream packet until a splice point is reached. Duplicate Transport Stream packets and Transport Stream packets which only contain adaptation fields are excluded. The splicing point is located immediately after the last byte of the Transport Stream packet in which the associated **splice_countdown** field reaches zero. In the Transport Stream packet where the **splice_countdown** reaches zero, the last data byte of the Transport Stream packet payload shall be the last byte of a coded audio frame or a coded picture. In the case of VC-4 elementary streams, the corresponding access unit may or may not be terminated by an end-of-sequence start code. Transport Stream packets with the same PID, which follow, may contain data from a different elementary stream of the same type.

The payload of the next Transport Stream packet of the same PID (duplicated packets and packets without payload being excluded) shall commence with the first byte of a PES packet. In the case of a VC-4 elementary stream, the PES packet payload shall commence with an access point as defined in the semantics of the **discontinuity_indicator** field or with a VC-4 end-of-sequence start code followed by an access point. Thus, the previous coded picture aligns with the packet boundary, or is padded to make this so. Subsequent to the splicing point, the countdown field may also be present. When the **splice_countdown** is a negative number whose value is minus n (-n), it indicates that the associated Transport Stream packet is the n-th packet following the splicing point (duplicate packets and packets without payload being excluded).

It indicates, among packet with the same PID, the priority of the elementary stream data carried within the payload of the Transport Stream packet. A '1' indicates that the payload has a higher priority than the payloads of other Transport Stream packets. In the case of a VC-4 elementary stream, this field may be set to 1 only if the payload contains one or more bytes from an intra-coded data. A value of 0 indicates that the payload has the same priority as all other packets which do not have this bit set to 1. For the definition of a VC-4 elementary stream access point, see the semantics of the **discontinuity_indicator** field.

6.2.11 Seamless Splice Flag

For VC-4 elementary streams, the **seamless_splice_flag** field shall adopt the following semantics:

seamless_splice_flag – This is a 1-bit flag which when set 1, indicates that the **splice_type** and **DTS_next_AU** fields are present. A value of 0 indicates that neither **splice_type** nor **DTS_next_AU** fields are present. This field shall not be set to 1 in Transport Stream packets in which the **splicing_point_flag** is not set to 1. Once it is set to 1 in a Transport Stream packet in which the **splice_countdown** is positive, it shall be set to 1 in all the subsequent Transport Stream packets of the same PID that have the **splicing_point_flag** set to 1, until the packet in which the **splice_countdown** reaches zero (including this packet). When this flag is set, and if the elementary stream carried in this PID is a VC-4 elementary stream, it shall fulfill the constraints indicated by the **splice_type** value.

6.2.12 Splice Type

For VC-4 elementary streams, the **splice_type** field shall adopt the following semantics:

splice_type – This is a 4-bit field. From the first occurrence of this field onwards, it shall have same value in all the subsequent Transport Stream packets of the same PID in which it is present, until the packet in which the **splice_countdown** reaches zero (including this packet). If the elementary stream carried in that PID is a VC-4 elementary stream, the value of this field shall always be set to '0000'.

6.3 T-STD Buffer Model for VC-4 Streams

6.3.1 Transport Stream System Target Decoder

Delivery of VC-4 elementary streams in MPEG-2 Transport Streams shall be governed by a Transport Stream System Target Decoder model (T-STD) as defined in ISO/IEC 13818-1.

In this model, there is a set of received video sub-streams in an ISO/IEC 13818-1 program which provides scalable video service, the dependencies among those video sub-streams are signaled in the ***hierarchy_descriptor()***, one video sub-stream is the base layer stream and at least one of video sub-streams are SMPTE=LVE elementary streams. The T-STD model for VC-4 streams is illustrated in Figure 1 and Table 9 describes the elements and the notations used for the block diagram. Here, the operations for the buffers of the base layer elementary stream are controlled by the buffering rules which the base layer elementary stream shall follow. This document describes the buffering of the VC-4 streams not the base layer elementary stream. All Transport Stream packets are demultiplexed by the PID values which are described in ***TS_program_map_section()***.

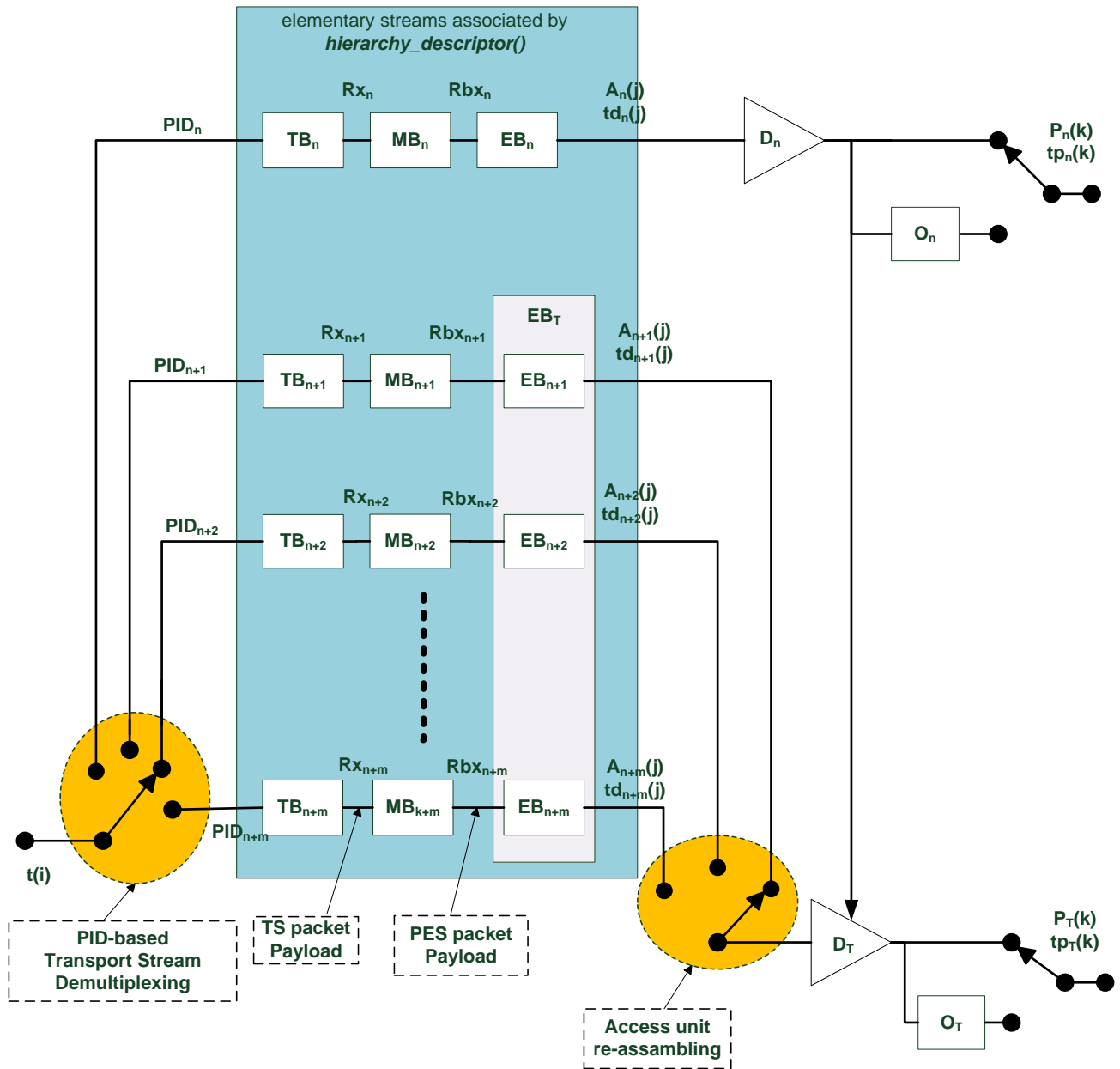


Figure 1 – Transport Stream System Target Decoder for VC-4 streams

Table 9 – Elements and notations for T-STD model

ES _n	A received elementary stream associated to the base layer specified by other coding standards.
ES _{n+m}	A received VC-4 elementary stream associated to the layer_id value equal to 'm'
TB _n	The transport buffer for an elementary stream 'n' from the base layer
TB _{n+m}	The transport buffer for a VC-4 elementary stream 'n+m'
MB _n	The multiplexing buffer for an elementary stream 'n' from the base layer
MB _{n+m}	The multiplexing buffer for a VC-4 elementary stream 'n+m'

EB_n	The elementary stream buffer for an elementary stream 'n' from the base layer
EB_{n+m}	The elementary stream buffer for a VC-4 elementary stream 'n+m'
EB_T	The stream buffer for all enhancement VC-4 elementary streams
TBS_n	The size of buffer TB_n
TBS_{n+m}	The size of buffer TB_{n+m}
MBS_n	The size of buffer MB_n
MBS_{n+m}	The size of buffer MB_{n+m}
EBS_n	The size of buffer EB_n
EBS_{n+m}	The size of buffer EB_{n+m}
EBS_T	The size of buffer EB_T
D_n	The decoder for a video elementary stream 'n' associated to the base layer
D_T	The decoder for a VC-4 elementary stream 'n+m'
O_n	The re-order buffer for a video elementary stream 'n' associated to the base layer
O_T	The re-order buffer according to the profile and the level associated to the highest LAER_ID for the presentation of the scalable video service using VC-4 codec
Rx_n	The rate at which data are removed from TB_n
Rx_{n+m}	The rate at which data are removed from TB_{n+m}
Rbx_n	The rate at which PES packet payload data are removed from MB_n when the leak method is used
Rbx_{n+m}	The rate at which PES packet payload data are removed from MB_{n+m} when the leak method is used
$A_n(j)$	The j^{th} access unit in video elementary stream 'n' from the base layer. $A_n(j)$ is indexed in decoding order.
$A_{n+m}(j)$	The j^{th} access unit in VC-4 elementary stream 'n+m'. $A_{n+m}(j)$ is indexed in decoding order.
$td_n(j)$	The decoding time, measured in seconds, in the system target decoder of the j^{th} access unit in video elementary stream 'n' from the base layer
$td_{n+m}(j)$	The decoding time, measured in seconds, in the system target decoder of the j^{th} access unit in VC-4 elementary stream 'n+m'
$P_n(k)$	The k^{th} presentation unit in video elementary stream 'n' from the base layer. $P_n(k)$ results from decoding $A_n(j)$. $P_n(k)$ is indexed in presentation order.
$P_T(k)$	The k^{th} presentation unit in the highest layer. $P_T(k)$ results from decoding $A_{n+1}(j)$, $A_{n+2}(j)$, ... $A_{n+m}(j)$. $P_T(k)$ is indexed in presentation order.
$tp_n(k)$	The presentation time, measured in seconds, in the system target decoder of the k^{th} presentation unit in video elementary stream 'n' from the base layer
$tp_T(k)$	The presentation time, measured in seconds, in the system target decoder of the k^{th} presentation unit in the highest layer
$t(i)$	The time in seconds that at which the i^{th} byte of the Transport Stream enters the T-STD

6.3.2 Mode of Operation

The main model constrains impose that buffer TB_{n+m} , buffer MB_{n+m} and buffer EB_{n+m} shall not overflow. Furthermore, buffer EB_{n+m} shall not underflow except when the value of the **HRD_PARM_FLAG** field in the sequence header of the VC-4 elementary stream is equal to 0 in which case the codec operates in variable delay mode. In this mode, the HRD shall wait until a complete video frame has arrived at the buffer before decoding the frame. As a result, the delay is minimized for a given frame, but the end-to-end delay is not constraint. The detail of this mode is described in Annex C.7 of SMPTE ST 421.

6.3.3 Operation of Buffer TB_{n+m}

Complete Transport Stream packets containing data from the 'n+m'-th VC-4 elementary stream are passed to the transport buffer for stream 'n+m', TB_{n+m} . This includes duplicate Transport Stream packets and packets with no payload. Transfer of the i-th byte from the system target decoder input to TB_{n+m} is instantaneous, so that the i-th byte enters the buffer for steam 'n+m', of size TBS_{n+m} , at time $t(i)$.

All bytes that enter the buffer TB_{n+m} are removed at the rate Rx_{n+m} specified below. Bytes which are part of the PES packets or its content are delivered to the multiplexing buffer MB_{n+m} . Other bytes (for example, bytes from the header of the Transport Stream packet) are not delivered to the multiplexing buffer MB_{n+m} and may be used to control the system. Duplicate Transport Stream packets are not delivered to MB_{n+m} .

The size of buffer TB_{n+m} , known as TBS_{n+m} , shall be fixed to 512 bytes. The TB_{n+m} shall be emptied as follows:

- Where there is no data in TB_{n+m}
 $Rx_{n+m} = 0$
- Otherwise
 $Rx_{n+m} = 1.2 \times R_{\max}[\text{profile, level}]$

where $R_{\max}[\text{profile, level}]$ is specified in Annex D.2 of SMPTE ST 2058-1.

6.3.4 Operation of Buffer MB_{n+m}

The multiplexing buffer size MBS_{n+m} is defined as follows:

$$MBS_{n+m} = BS_{\text{mux}} + BS_{\text{oh}} + B_{\max}[\text{profile, level}] - \text{vbv_buffer_size}$$

where BS_{oh} , PES packet overhead buffering is defined as:

$$BS_{\text{oh}} = (1/750) \text{ second} \times R_{\max}[\text{profile, level}]$$

and BS_{mux} , additional multiplex buffering is defined as:

$$BS_{\text{mux}} = 0.004 \text{ seconds} \times R_{\max}[\text{profile, level}]$$

and $B_{\max}[\text{profile, level}]$ is defined in Annex D.2 of SMPTE ST 2058-1, **vbv_buffer_size** is the size of the smallest buffer among the buffers defined with **HRD_BUFFER[]** and **BUFFER_SIZE_EXPONENT[]** parameters in the sequence header of the VC-4 elementary stream.

The transfer of data from MB_{n+m} to EB_{n+m} shall be governed by a leak method. Use of a leak method shall be signaled via one of the following two methods:

- There is no MPEG-2 **STD_descriptor()** present in the inner descriptor loop of the MPEG-2 program element in the **TS_program_map_section()** corresponding to the VC-4 elementary stream.
- An MPEG-2 **STD_descriptor()** is present in the inner descriptor loop of the MPGE-2 program element in the **TS_program_map_section()** corresponding to the VC-4 elementary stream and the **leak_valid** flag has the value 1.

The leak method transfers data from MB_{n+m} to EB_{n+m} using the leak rate R_{bxn} where

$$R_{bx_n} = R_{\max}[\text{profile, level}]$$

If there is PES packet payload data in MB_{n+m} and the buffer EB_{n+m} is not full, the PES packet payload is transferred from MB_{n+m} to EB_{n+m} at a rate equal to $R_{bx_{n+m}}$. If EB_{n+m} is full, data are not removed from MB_{n+m} . When a byte of data is transferred from MB_{n+m} to EB_{n+m} , all PES packet header bytes that are in MB_{n+m} and immediately precede that byte, are instantaneously removed and discarded. When there is no PES packet payload data present in MB_{n+m} , no data is removed from MB_{n+m} . All data that enters MB_{n+m} leaves it. All PES packet payload data byte except any stuffing byte appearing before a start code enter EB_{n+m} instantaneously upon leaving MB_{n+m} . This means that only access unit bytes enter the video elementary stream buffer. Any stuffing bytes before a start code do not enter EB_{n+m} and are discarded but may be used to control the system.

6.3.5 Operation of Buffer EB_{n+m}

There is exactly only one elementary stream buffer EB_{n+m} for each VC-4 elementary stream in the set of received VC-4 elementary streams where each VC-4 stream buffer EB_{n+m} in the set of received VC-4 elementary streams is allocated within EB_T . Even though the size EBS_{n+m} of individual EB_{n+m} is not constrained, the sum of the sizes EBS_{n+m} is constrained as follows:

$$EBS_T = \sum_m (EBS_{n+m})$$

The default size of the elementary stream buffer EB_{n+m} , EBS_{n+m} shall be the $B_{\max}[\text{profile, level}]$ associated with the profile and level of the VC-4 elementary stream as specified in SMPTE ST 2058-1. The profile and level may be identified by the **profile** and **level** fields in the **profile_level_subdescriptor()**.

With the assumption that the incoming service delivery rate R is known (for example, by means of the **ES_rate_flag** and **ES_rate** fields in the PES packet header), a receiver may opt to use a smaller elementary stream buffer for its internal representation of the hypothetical reference decoder. However, the size of the elementary stream buffer shall always be equal or the minimum buffer value EBS_{\min} specified by the generalized hypothetical reference decoder for rate R . See Annex C.2 and Annex C.4 of SMPTE ST 421. In this case, the size of the elementary stream buffer may be computed from the values of **BIT_RATE_EXPONENT**, **BUFFER_SIZE_EXPONENT**, **HRD_RATE[]** and **HRD_BUFFER[]** fields listed in the sequence header of the VC-4 elementary stream, if present. The buffer size $EBS_{n+m}[k]$ for the k -th buffer model in the sequence header may be computed from its respective **HRD_BUFFER[k]** and **BUFFER_SIZE_EXPONENT** fields as follows:

$$EBS_{n+m}[k] = (\text{HRD_BUFFER}[k] + 1) \ll (\text{BUFFER_SIZE_EXPONENT} + 4) \text{ (in units of bits)}$$

and the associated rate $R[k]$ may be computed from the **HRD_RATE[k]** and the **BIT_RATE_EXPONENT** fields as follows:

$$R[k] = (\text{HRD_RATE}[k] + 1) \ll (\text{BIT_RATE_EXPONENT} + 6) \text{ (in bits/sec)}$$

The value EBS_{n+m} selected for the elementary stream buffer EB_{n+m} shall not exceed the value of $B_{\max}[\text{profile, level}]$ associated with the profile and level for the VC-4 elementary stream as specified by the **profile** and **level** fields in the **profile_level_subdescriptor()**.

6.3.6 Decoding and Presentation

Each VC-4 access unit is reassembled to make one single bitstream whose configuration is defined in SMPTE ST 2058-1. Each access $A_{n+m}(j)$ unit within the re-ordered bitstream shall be extracted from buffer EB_T at time $td_{n+m}(j)$ and is decoded instantaneously by decoder D_T . All decoded access units which have the same presentation time $tp_T(k)$ shall be used to make the presentation unit $P_T(k)$ and may be delayed in a re-order buffer O_T before being presented at the output of the T-STD. Here, the decoded output of the base layer is also used to make the presentation unit. The procedure to generate the presentation unit is defined in SMPTE ST 2058-1.

Re-order buffers are used only in the case of VC-4 element streams where some access units are not carried in presentation order

Decoding time $td_{n+m}(j+1)$, $td_{n+m}(j+2)$, ... of access units without encoded DTS or PTS field which directly follow access unit 'j' in the same PES packet may be derived from the frame or field rate information in the elementary stream.

For any access unit associated with one or more **HRD_FULL[]** values (the first access unit after sequence header or entry point), the decoding time shall be the instant of the System Time Clock corresponding to the time when all buffers the corresponding layer and all hierarchy-embedded VC-4 layers are filled up to the each buffer fullness defined from each buffer fullness decided in each elementary stream.

6.3.7 Decoding of the First Access Unit following an Sequence Start Code

To minimize start-up latency and with the assumption that the incoming service data rate R is known (for example, by means of the **ES_rate_flag** and **ES_rate** fields in the PES packet header), a decoder may opt to use the **HRD_RATE[]**, the **HRD_BUFFER[]** and **HRD_FULL[]** fields listed in the sequence header of the VC-4 elementary stream for determining when to initiate the decoding of the first video access unit following sequence header. However, the value of the elementary stream buffer fullness used in this situation shall always be equal or greater than the minimum fullness value F_{min} specified by the generalized hypothetical reference decoder for data rate R . See Annex C.2 and Annex C.4 of SMPTE ST 421. Here, since each VC-4 elementary stream defines **HRD_RATE[]**, the **HRD_BUFFER[]** and **HRD_FULL[]** fields individually, the initiation time of the each layer is after the corresponding layer and its lower layers achieve each buffer fullness.

6.3.8 Rate and Buffer Sizes

The rate and buffer size for the different profiles and levels supported by VC-4 are listed in Annex D.2 of SMPTE ST 2058-1.

7 MPEG-2 Program Stream Encoding

7.1 Signaling of VC-4 Elementary Streams

The stream type value 0xEB as defined in Section 6.1.1 and the usage of the registration descriptor and sub-descriptors as defined as defined in Section 6.1.2 through Section 6.1.6 shall also be applicable to the carriage of VC-4 elementary streams in an MPEG-2 Program Stream. The hierarchy descriptor also shall be used to specify the dependency information and scalability information between the current elementary stream and the embedded lower layer in an MPEG-2 Program Stream. In case of an MPEG-2 Program Stream, these fields and descriptors are used in the Program Stream Map (PSM) instead of the Program Map Table. The **hierarchy_descriptor()** of ISO/IEC 13818-1:2007 and ISO/IEC 13818-1:2007/Amd3:2009 also shall be present in the inner descriptor loop of the MPEG-2 program element listed in the **program_stream_map()** corresponding to the VC-4 elementary stream like Section 6.1.7.

7.2 Encapsulation of VC-4 Elementary Streams

Encapsulation of a VC-4 elementary stream in an MPEG-2 Program Stream shall follow the same rules defined for MPEG-2 Transport Streams in Section 6.2.

7.3 P-STD Buffer Model for VC-4 Streams

7.3.1 Program Stream Target Decoder

Delivery of VC-4 elementary streams in MPEG-2 Program Streams shall be governed by a Program Stream System Target Decoder model (P-STD) as defined in ISO/IEC 13818-1.

In this model, there is a set of received video sub-streams in an ISO/IEC 13818-1 program which provides scalable video service, the dependencies among those video sub-streams are signaled in the ***hierarchy_descriptor()***, one video sub-stream is the base layer stream and at least one of video sub-streams are SMPTE=LVE elementary streams. The P-STD model for VC-4 streams is illustrated in Figure 2 and Table 10 describes the elements and the notations used for the block diagram. Here, the operations for the buffers of the base layer elementary stream are controlled by the buffering rules which the base layer elementary stream shall follow. This document describes the buffering of the VC-4 streams not the base layer elementary stream. All Program Stream packets are demultiplexed by the stream_id.

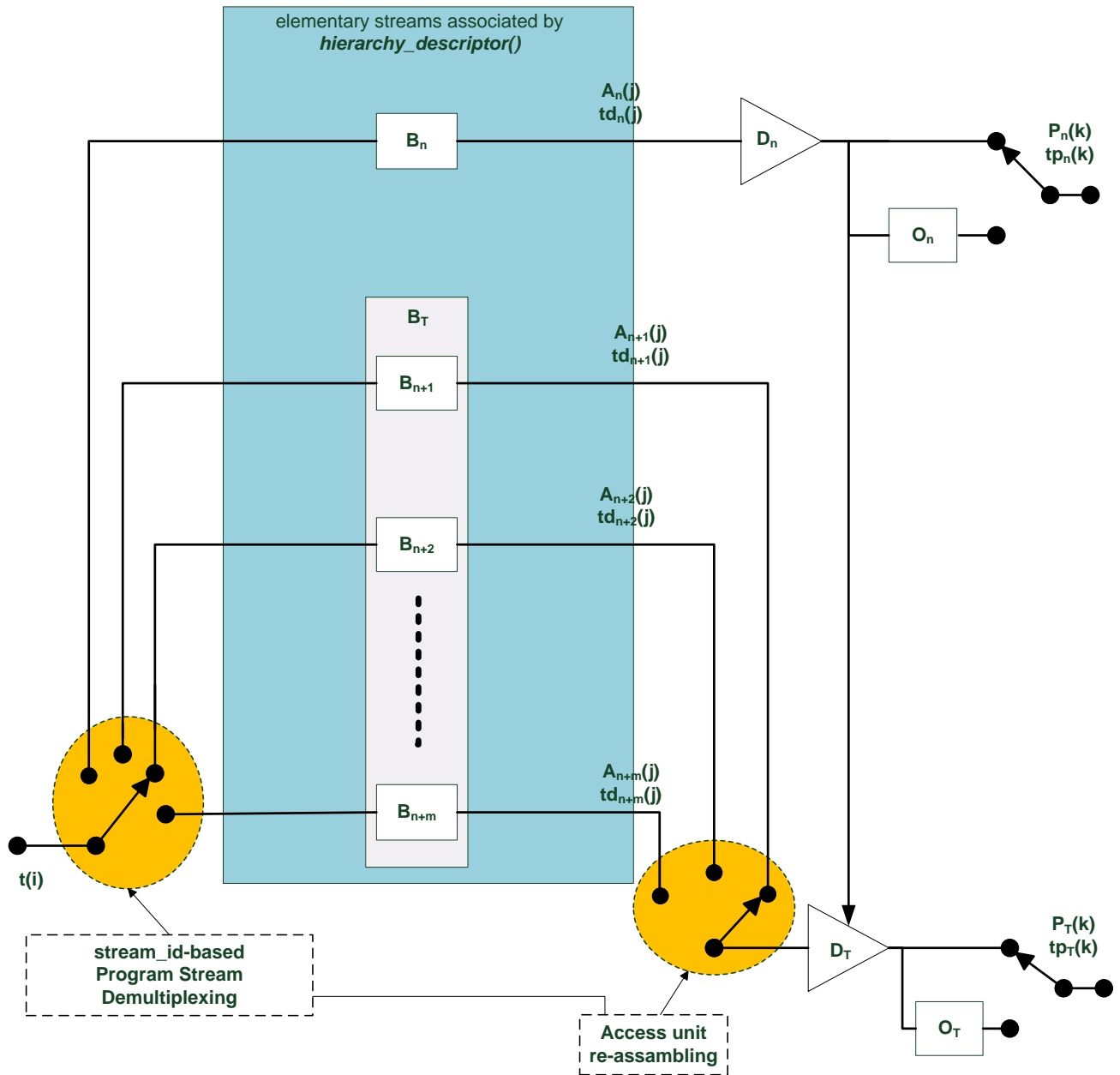


Figure 2 – Program Stream System Target Decoder for VC-4 streams

Table 10 – Elements and notations for P-STD model

ES_n	A received elementary stream associated to the base layer specified by other coding standards.
ES_{n+m}	A received VC-4 elementary stream associated to the layer_id value equal to 'm'
B_n	The input buffer for an elementary stream 'n' from the base layer
B_{n+m}	The input buffer for a VC-4 elementary stream 'n+m'
B_T	The input buffer for all enhancement VC-4 elementary streams
EBS_n	The size of buffer B_n
BS_{n+m}	The size of buffer B_{n+m}
BS_T	The size of buffer B_T
D_n	The decoder for a video elementary stream 'n' associated to the base layer
D_T	The decoder for a VC-4 elementary stream 'n+m'
O_n	The re-order buffer for a video elementary stream 'n' associated to the base layer
O_T	The re-order buffer according to the profile and the level associated to the highest LAER_ID for the presentation of the scalable video service using VC-4 codec
$A_n(j)$	The j^{th} access unit in video elementary stream 'n' from the base layer. $A_n(j)$ is indexed in decoding order.
$A_{n+m}(j)$	The j^{th} access unit in VC-4 elementary stream 'n+m'. $A_{n+m}(j)$ is indexed in decoding order.
$td_n(j)$	The decoding time, measured in seconds, in the system target decoder of the j^{th} access unit in video elementary stream 'n' from the base layer
$td_{n+m}(j)$	The decoding time, measured in seconds, in the system target decoder of the j^{th} access unit in VC-4 elementary stream 'n+m'
$P_n(k)$	The k^{th} presentation unit in video elementary stream 'n' from the base layer. $P_n(k)$ results from decoding $A_n(j)$. $P_n(k)$ is indexed in presentation order.
$P_T(k)$	The k^{th} presentation unit in the highest layer. $P_T(k)$ results from decoding $A_{n+1}(j)$, $A_{n+2}(j)$, ... $A_{n+m}(j)$. $P_T(k)$ is indexed in presentation order.
$tp_n(k)$	The presentation time, measured in seconds, in the system target decoder of the k^{th} presentation unit in video elementary stream 'n' from the base layer
$tp_T(k)$	The presentation time, measured in seconds, in the system target decoder of the k^{th} presentation unit in the highest layer
$t(i)$	The time in seconds that at which the i^{th} byte of the Transport Stream enters the T-STD

7.3.2 Mode of Operation

The input buffers B_{n+1} through B_{n+m} in the P-STD model shall not overflow. Furthermore, they shall not underflow except when the value of the **HRD_PARM_FLAG** field in the sequence header of the VC-4 elementary stream is equal to 0 in which case the codec operates in variable delay mode. In this mode, the HRD shall wait until a complete video frame has arrived at the buffer before decoding the frame. As a result, the delay is minimized for a given frame, but the end-to-end delay is not constraint. The detail of this mode is described in Annex C.7 of SMPTE ST 421.

7.3.3 Operation of Buffer B_{n+m}

There is exactly only one elementary stream buffer B_{n+m} for each VC-4 elementary stream in the set of received VC-4 elementary streams where each VC-4 stream buffer B_{n+m} in the set of received VC-4 elementary streams is allocated within B_T . Even though the size BS_{n+m} of individual B_{n+m} is not constrained, the sum of the sizes BS_{n+m} is constrained as follows:

$$BS_T = \sum_m (BS_{n+m})$$

Data enter the P-STD at the rate specified by the value of the field **program_mux_rate** in the pack header. The PES packet data byte from the elementary stream 'n+m' are passed to the input buffer B_{n+m} . Transfer of byte l from the system target decoder input to B_{n+m} is instantaneous, so that byte l enters the buffer for stream 'n+m' of size BS_{n+m} , at the time $t(l)$. Bytes present in the pack header, system headers, Program Stream Map and PES packet headers of the Program Stream do not enter B_{n+m} and may be used to control the system. Any stuffing bytes before a start code do not enter B_{n+m} . This means that only data access unit bytes enter the video elementary stream buffer.

When present in the PES header (value of **P-STD_buffer_flag** equal to 1) of the 'n+m'-th VC-4 elementary stream, the value of the **P-STD_buffer_scale** field shall always be set to 1. And the value of the **P-STD_buffer_size** field shall specify the size of the corresponding input video buffer BS_{n+m} . In particular, the size of the buffer BS_{n+m} shall always be equal to the sum of the elementary stream buffer size for the 'n+m'-th VC-4 stream and the value BS_{oh} as defined in the Program Stream Target Decoder of ISO/IEC 13818-1. The size of the VC-4 elementary stream buffer is either signaled explicitly in the VC-4 sequence header or implicitly. In the latter case, it is equal to the $B_{max}[profile, level]$ value associated with the profile and level of the 'n+m'-th VC-4 elementary stream.

With the assumption that the incoming service delivery rate R is known (for example, by means of the **ES_rate_flag** and **ES_rate** fields in the PES packet header), a receiver may opt to use a smaller elementary stream buffer for its internal representation of the hypothetical reference decoder. However, the size of the input buffer shall always be equal or greater than the minimum buffer value BS_{min} specified by the generalized hypothetical reference decoder for rate R . See Annex C.2 and Annex C.4 of SMPTE ST 421. In this case, the size of the elementary stream buffer may be computed from the values of **BIT_RATE_EXPONENT**, **BUFFER_SIZE_EXPONENT**, **HRD_RATE[]** and **HRD_BUFFER[]** fields listed in the sequence header of the VC-4 elementary stream, if present. The buffer size $BS_{n+m}[k]$ for the k -th buffer model in the sequence header may be computed from its respective **HRD_BUFFER[k]** and **BUFFER_SIZE_EXPONENT** fields as follows:

$$BS_{n+m}[k] = BS_{oh} + (HRD_BUFFER[k] + 1) \ll (BUFFER_SIZE_EXPONENT + 4) \text{ (in units of bits)}$$

and the associated rate $R[k]$ may be computed from the **HRD_RATE[k]** and the **BIT_RATE_EXPONENT** fields as follows:

$$R[k] = (HRD_RATE[k] + 1) \ll (BIT_RATE_EXPONENT + 6) \text{ (in bits/sec)}$$

The value BS_{n+m} selected for the input buffer B_{n+m} shall not exceed the value of $B_{max}[profile, level] + BS_{oh}$ associated with the profile and level for the VC-4 elementary stream as specified by the **profile** and **level** fields in the **profile_level_subdescriptor()**.

7.3.4 Decoding and Presentation

Decoding and presentation rules are identical to those described in Section 6.3.6 except that EB_n , EB_{n+m} and EB_T are replaced with B_n , B_{n+m} and B_T respectively.

7.3.5 Decoding of the First Access Unit following an Entry Point Start Code

Use of the information of the **HRD_RATE[]**, the **HRD_BUFFER[]** and **HRD_FULL[]** fields of the VC-4 elementary stream is similar to that described in Section 6.3.7.

7.3.6 Buffer Sizes

The values of $B_{\max}[\text{profile, level}]$ are defined in Annex D.2 of SMPTE ST 2058-1.

Annex A Graphics of Syntax for this SMPTE Recommended Practice (Informative)

A.1 Introduction

This annex is an informative annex presenting graphically the Transport Stream and Program Stream syntax for VC-4 elementary stream syntax. In order to produce clear drawings, not all the fields have been fully described or represented. Reserved fields may be omitted or indicated by area with no detail. Field lengths are indicated in units of bits. Section F.0.1. (Transport Stream Syntax), F.0.3 (Program Association Section), F.0.4 (CA Section), F.0.6 (Private Section) and F.0.7 (Program Stream) of Annex F in ISO/IEC 13818-1 remain unchanged and are not reproduced herein.

A.2 PES Packet

See Figure A.1.

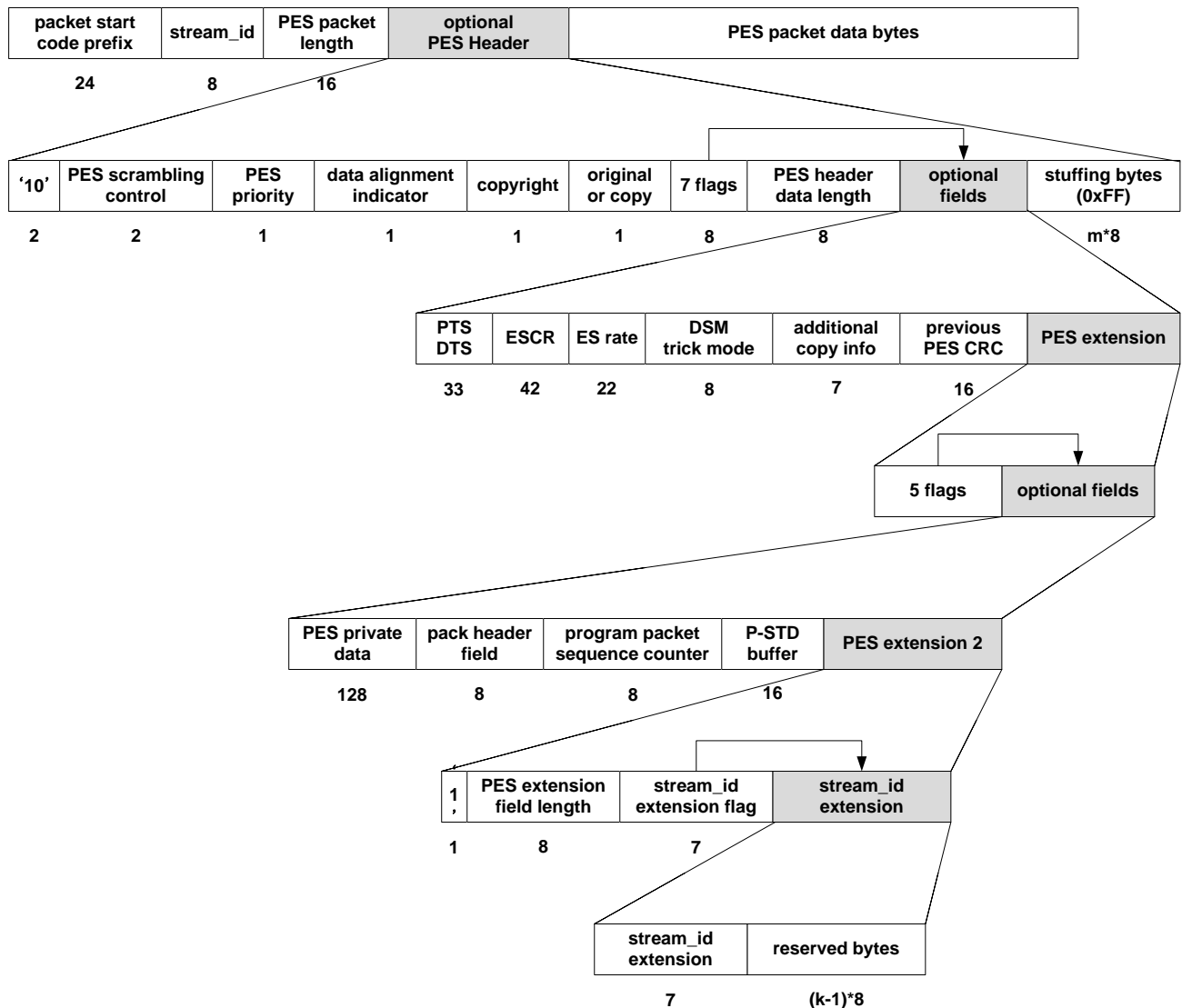


Figure A.1 – PES packet syntax diagram for SMPTE elementary stream

A.3 TS Program Map Section

See Figure A.2.

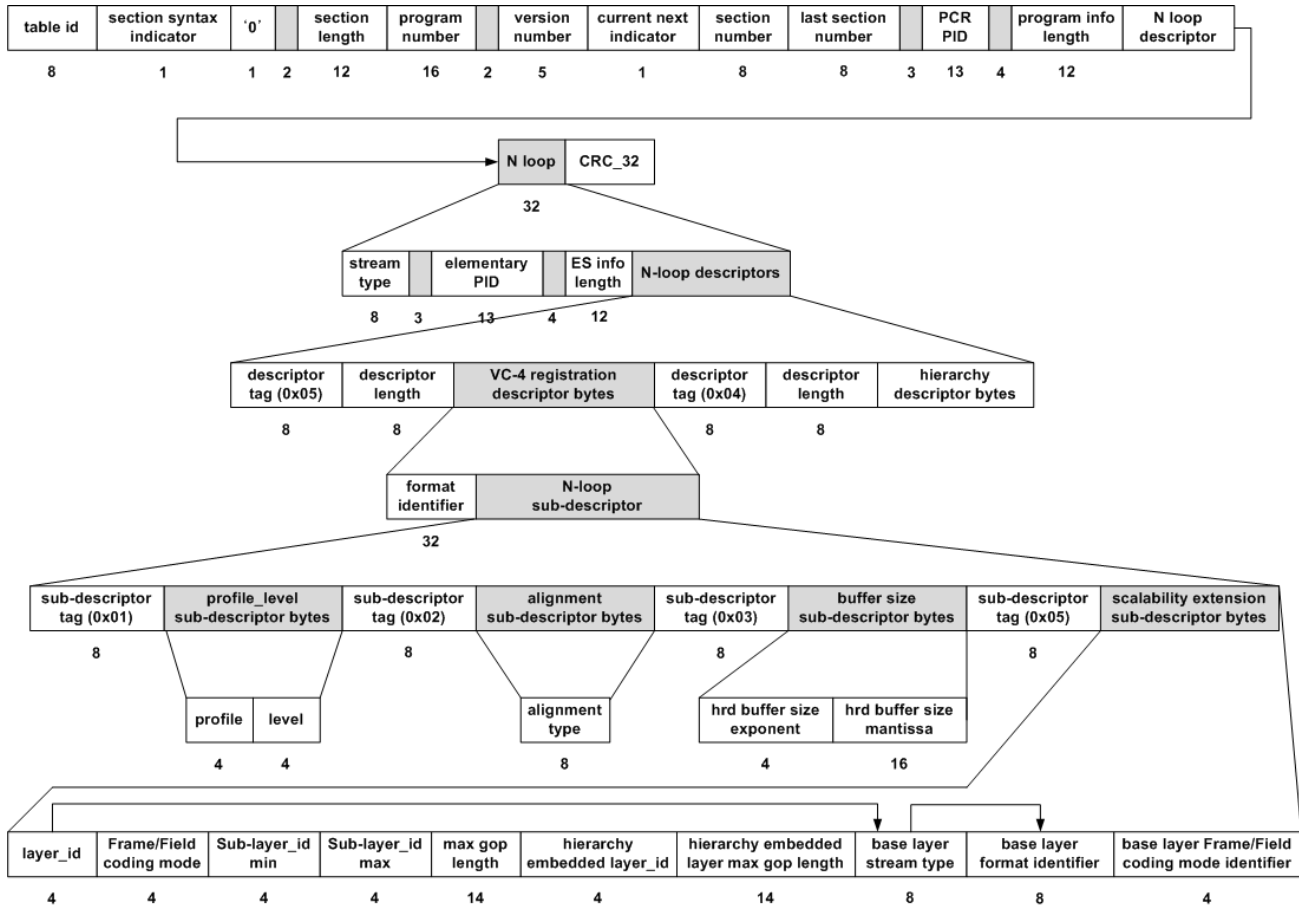


Figure A.2 – TS Program Map Section diagram for VC-4 elementary stream

A.4 Program Stream Map

See Figure A.3.

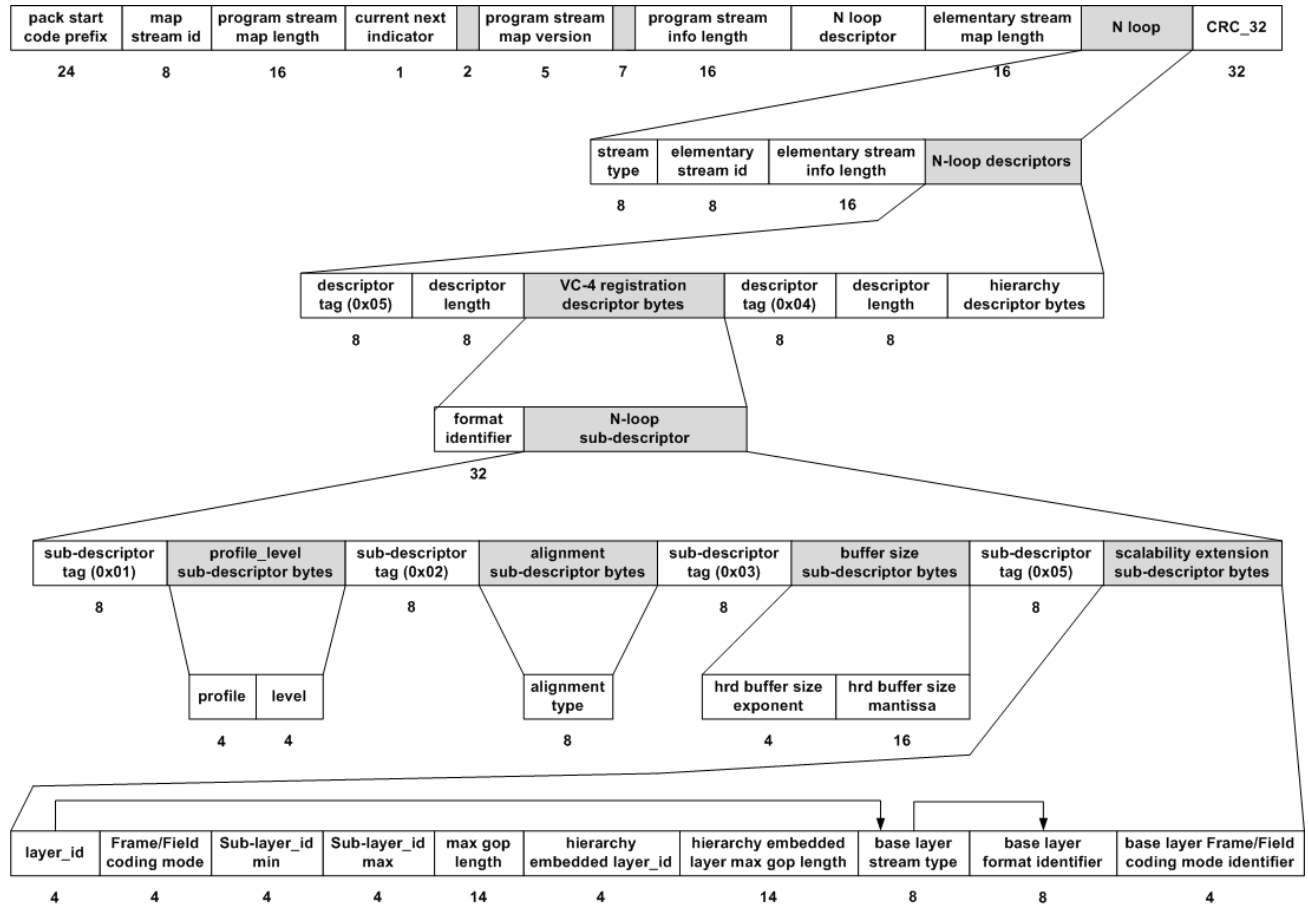


Figure A.3 – Program Stream Map diagram for VC-4 elementary stream

Annex B Bibliography (Informative)

ISO/IEC 14496-1:2004, Information Technology – Coding of Audio-Visual Objects — Part 1: Systems

ISO/IEC 14496-2:2004, Information Technology — Coding of Audio-Visual Objects — Part 2: Visual

Recommendation ITU-T H.262, ISO/IEC 13818-2, Information Technology — Generic Coding of Moving Pictures and Associated Audio Information: Video, February 2000

Recommendation ITU-T H.264, Advanced Video Coding for Generic Audiovisual Services, November 2007.

J. Ribas-Corbera, P.A. Chou, and S.L. Regunathan, "A Generalized Hypothetical Reference Decoder for H.264/AVC," IEEE Transactions on Circuits and Systems for Video Technology, August 2003