

SMPTE RECOMMENDED PRACTICE

VC-4 Bitstream Transport Encodings



Page 1 of 32 pages

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
registration_descriptor() {		
descriptor_tag	8	0x05
descriptor_length	8	uimbsf
format_identifier	32	0x56432D34
for(i = 0; i < K; i++) {		
subdescriptor()	N*8	uimbsf
}		
}		

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
buffer_size_subdescriptor() {		
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.