

SMPTE RECOMMENDED PRACTICE

VC-4 Bitstream Storage in the ISO Base Media File Format



Page 1 of 14 pages

Table of Contents	Page
Foreword	2
Intellectual Property	2
Introduction.....	2
1 Scope	3
2 Conformance Notation	3
3 Normative References	4
4 Notation	4
4.1 Abbreviations	4
4.2 Definition of Terminology	4
5 File Identification	5
6 Track Definition	5
7 Layer Definition	6
7.1 Layer Table Box	6
7.2 Layer Information Box	6
8 Video Stream Definition	8
8.1 Base Track	8
8.2 VC-4 Track	8
Annex A Bibliography (Informative)	14

Table of Tables

Table 1 – Scalability field	7
-----------------------------------	---

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 2058-4 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 Recommended Practice.

The purpose of this specification is to define a minimum set of rules for the storage file format of a layered video stream derived from the ISO/IEC 14496-12 ISO based media file format. In particular, this specification is aimed to provide a storage file format encapsulating both base layer video stream from other video coding standards and enhancement layer video streams from the SMPTE Layered Video Extension (VC-4) coding where the video coding standard for the base layer is such as SMPTE ST 421 (SMPTE VC-1), ITU-T H.262 (MPEG-2), ISO/IEC 14496-2 (MPEG-4 Part2) or ITU-T H.264 (MPEG-4 AVC), etc.

1 Scope

The purpose of this specification is to define the storage file format for SMPTE Layered Video Extension (VC-4) video streams which extends a file format that is compliant with the ISO Base Media File Format. The VC-4 video streams can be dependent on any base layer video streams such as SMPTE ST 421 (SMPTE VC-1), ITU-T H.262 (MPEG-2), ISO/IEC 14496-2 (MPEG-4 Part 2) or ITU-T H.264 (MPEG-4 AVC), etc. The storage format for VC-4 shall be extended from a particular file format which encapsulates a base layer video stream on which the VC-4 shall be dependent. The following file formats derived from ISO Base Media File format shall be extensible from this specification but are not limited to these:

- MP4 file format (ISO/IEC 14496-14)
- AVC file format (ISO/IEC 14496-15)
- SMPTE VC-1 file format (SMPTE RP 2025)

The file format for storage of VC-4 content uses the existing capability of the ISO base media file format. In addition, the following new extensions to support new features of VC-4 codec are specified:

- **VC-4 Sample:** a newly defined sample including a subset of enhancement layers or all enhancement layers of a picture.
- **VC-4 Extractor:** a structure to enable efficient extraction of bitstream segment units from samples in other tracks instead of containing duplicate media data.

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

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

ISO/IEC 14496-12:2008, Information Technology – Coding of Audio-Visual Objects – Part 12: ISO Base Media File Format

4 Notation

4.1 Abbreviations

This section provides a list of acronyms used in this document.

EBS	Encapsulated Bitstream Segment
HRD	Hypothetical Reference Decoder

4.2 Definition of Terminology

4.2.1 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, is used in reconstructing a new picture when integrating with the residual picture decoded from the enhancement layer stream. The synchronization of the base layer picture and other enhancement layer streams is expected to be signaled in the system layer.

4.2.2 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.

4.2.3 Enhancement Layer

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

4.2.4 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.

4.2.5 Entry Point Sample

A sample which includes sequence header and is considered as the new starting point of the sequence. The subsequent coded samples do not use the previously coded samples as the reference sample before this entry point sample. It also means the current entry point sample does not use any reference samples for motion compensation.

5 File Identification

A file type box, as defined in the ISO base media file format (ISO/IEC 14496-12), shall be present in conforming files.

`major_brand`: This specification does not define the use of 'vc-4' as a major brand.

`minor_version`: Major brand is not defined.

`compatible_brands`: Files that conform to this specification shall have 'vc-4' in the compatible brands list. The brand used for base layer shall be in the compatible brands list, thus at least the base layer bitstream can be decoded by a player that supports the brand for the base layer but is non-conformant to this specification.

6 Track Definition

A base layer bitstream and an enhancement layer bitstream shall be separately represented by a base track and one or more VC-4 tracks, respectively. There shall be exactly one base track and at least one VC-4 track in a VC-4 file. In the terminology of the ISO base media file format (ISO/IEC 14496-12), a base track and a VC-4 track are video tracks. Therefore they use:

- A `handler_type` of 'vide' in the Handler Reference Box.
- A Video Media Header Box in the Media Information Box.

Additionally, a VC-4 track shall use a box called **VC4SampleEntry** in the Sample Description Box (this box is derived from the `VisualSampleEntry` defined in ISO/IEC 14496-12).

A VC-4 track shall contain a subset of enhancement layers or all enhancement layers. When a VC-4 file has multiple enhancement layers, each enhancement layer may be contained separately in different VC-4 tracks, or each enhancement layer may be contained in multiple VC-4 tracks.

Each track shall be identified by the `track_ID` in the Track Header Box.

The base track and VC-4 tracks shall be linked to each other by means of track reference boxes, defined in Section 8.3.3 of ISO/IEC 14496-12. Each VC-4 track shall have one Track Reference Box. The following reference types are defined for VC-4 file format:

- 'vbas' shall be used by VC-4 tracks to reference the base track. Every VC-4 track shall have this reference type in its track reference box.
- 'vext' shall be used by VC-4 extractors to indicate the track from which the media data shall be copied (see Section 8.2.3).
- 'vdep' shall be used by VC-4 tracks to reference the VC-4 tracks which are required to decode samples in the current track. These referenced VC-4 tracks shall have enhancement layers referenced by the layers in the current track. The referenced track IDs in 'vdep' shall be arranged in the ascending order.

The visual samples in the base track and VC-4 tracks shall be temporally aligned in the decoding timeline and mapped at the same decoding time.

For every VC-4 track, its `Track_in_preview` flag in the flags of the Track Header Box shall be set to 0.