SMPTE RP 2047-5:2017

SMPTE RECOMMENDED PRACTICE

VC-2 Level 66 Compression of Ultra-high Definition Video Sources for use with a High Definition Infrastructure

Table of Contents

3 Barker Ave. White Plains, NY 10601 +1 914 761 1100

Approved
July 1, 2017

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 its Standards Operations Manual. This SMPTE Engineering Document was prepared by Technology Committee 10E.

Intellectual Property

At the time of publication no notice had been received by SMPTE claiming patent rights essential to the implementation of this Engineering Document. However, attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. SMPTE shall not be held responsible for identifying any or all such patent rights.

Introduction

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

The transition from high definition to ultra-high definition television production is beginning, but many television production facilities still have a substantial legacy of high definition infrastructure that is impractical, or uneconomic, to immediately upgrade to an ultra-high definition infrastructure. There is therefore a need to compress ultra-high definition signals so that ultra-high definition production can continue to use a high definition infrastructure. High definition infrastructure components that can be re-used by such compression include cabling, routers, uncompressed video links and uncompressed video recorders.

This recommended practice defines constraints on the VC-2 specification (SMPTE ST 2042-1) such that ultra-high definition source video images can be compressed to high definition data rates. The bit rate of the

resulting VC-2 stream is such that it can be carried over a high definition serial digital interface (SDI) operating at 2.970 Gbit/s. This document defines constraints on the coding and syntax. It does not define a mapping to any specific transport layer. The constraints defined herein are designed to support low latency, low complexity implementations suitable for television production environments.

1 Scope

This recommended practice defines constraints on the VC-2 video compression specification such that 10-bit ultra-high definition video signals can be compressed to only occupy the bit rate of 8-bit high definition signals. Specifically, it defines constraints for compressing progressive source video having 3840 pixels/line by 2160 lines/frame at 50 or 60/1.001 frames/s using 4:2:2 sampling.

This document defines a VC-2 bit-stream for video compressed according to this practice. It also defines an application specific level for VC-2 compression.

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.

A conformant implementation according to this document is one that includes all mandatory provisions ("shall") and, if implemented, all recommended provisions ("should") as described. A conformant implementation need not implement optional provisions ("may") and need not implement them as described.

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 that, through reference in this text, constitute provisions of this standard. Dated references require that the specific edition cited shall be used as the reference. Undated citations refer to the edition of the referenced document (including any amendments) current at the date of publication of this document. All standards are subject to revision, and users of this engineering document are encouraged to investigate the possibility of applying the most recent edition of any undated reference.

SMPTE ST 2042-1:2017, VC-2 Video Compression

SMPTE ST 2036-1:2014, Ultra High Definition Television – Image Parameter Values for Program Production

4 Terms and Definitions

For the purposes of this document, the terms and definitions of SMPTE ST 2042-1 apply.

5 Overview (informative)

The VC-2 compression standard defines a compression system that uses Wavelet coding as the basis for image de-correlation with a number of compression tools including:

- A choice of Wavelet transform filters,
- A choice in the number of levels of Wavelet decimation,
- A choice between standard, low-delay, and high-quality operation,
- A choice of entropy coding if using standard operation.

The VC-2 compression standard defines various coding parameters that are needed by a decoder to successfully decode and present the output video signal. This Recommended Practice defines coding constraints to provide the interoperability of VC-2 streams coded for compatibility with a high definition infrastructure. This Recommended Practice also constrains the stream syntax to ensure that coded streams can be mapped in an interoperable fashion onto a transport layer.

The constraints defined here are signaled in the stream as the specialized VC-2 level number 66 (as indicated below in section 7.3.1.2). That is, VC-2 streams conformant to this practice ensure compliance by specifying specialized level 66 in the stream. This practice does not define how the VC-2 stream is mapped to any specific transport layer.

Note: VC-2 stream transport specifications are defined in other SMPTE documents.

6 General

The VC-2 stream shall be compliant with SMPTE ST 2042-1 and the constraints specified below.

7 Coding Constraints

7.1 Overview (informative)

The key coding constraints in the following sub-sections are as follows:

1. The video sources are limited to 4:2:2 sampling with the following picture formats:
    - 3840 pixels by 2160 progressive lines with frame rates of 50 frames/sec, or 60/1.001 frames/sec.
2. The codec uses only the High Quality Profile.
3. The codec uses only the LeGall (5,3) wavelet kernel.
4. The wavelet depth is exactly 3.

7.2 Video Source parameters

The video sources and image sampling shall be as defined by SMPTE ST 2036-1 for either the 3840 x 2160/50/P system or the 3840 x 2160/59.94/P system.

The system shall be a 10-bit system.

The color encoding shall be $Y'C'_BC'_R$ 4:2:2 as defined by SMPTE ST 2036-1.

Only the active picture area shall be coded.

7.3 Coding parameters

7.3.1 Profile and syntax

7.3.1.1 General

The Sequence Header comprises the parse parameters, base video format, source parameters and picture coding mode and is defined as follows.

7.3.1.2 Parse parameters

The profile value shall be the High Quality Profile defined in SMPTE ST 2042-1.

The level value shall be a specialized level value with the value of 66. This level value shall be used only for the VC-2 coding constraints that comply with this practice.

7.3.1.3 Base video format

The base video format shall be one of format numbers 17 (UHDTV 4K-60) or 18 (UHDTV 4K-50) as defined in SMPTE ST 2042-1.

Note 1: Per SMPTE ST 2042-1, the format names for 60 frames/sec systems ignore the 1.001 divisor for simplicity. Thus UHDTV 4K-60 means that the default frame rate is 60/1.001 frames/sec, not 60 frame/sec.

Note 2: Per SMPTE ST 2042-1, the format names for UHDTV refer to 4K rather than UHDTV1, although the default line width is 3840 pixels, not 4096.

7.3.1.4 Source parameters

The source parameters allow the video parameters pre-defined by the base video format to be overridden with custom values. To comply with this document, the pre-defined video parameters shall not be overridden.

Note: The format is thus defined as follows:

- The picture size is 3840 x 2160.
- The scanning is progressive.
- The frame rate is 60/1.001 Hz or 50 Hz.
- The pixel aspect ratio is 1:1.
- The clean picture area is the same as the picture size, i.e. 3840 x 2160.
- The video depth parameters luma_depth and color_diff_depth are both equal to 10 (bits).
- The color primaries and color matrix are as defined in Recommendation ITU-R BT.2020.
- The transfer function is as defined in Recommendation ITU-R BT.2020.

7.3.1.5 Picture Coding Mode

Picture coding mode shall be 0.

Note: This means that progressive source frames are coded as a whole.

7.3.2 Picture syntax

The base video format and source parameters shall be static values over the VC-2 sequence. The syntax constraints (section 5) define that each picture is preceded by a sequence header. Thus, in the initialization process, all the sequence and source parameters can be accessed for each picture in the sequence.

Per SMPTE ST 2042-1, the picture header value shall be an incrementing number for each successive picture.

Note: According to this practice, pictures are frames for progressive video sources.

7.3.3 Transform parameters

7.3.3.1 Wavelet filter

The wavelet filter shall be “LeGall (5,3)” filter (Wavelet index value = 1) as defined in SMPTE ST 2042-1:2017 Table 12.1.

7.3.3.2 Transform depth

The discrete wavelet transform (DWT) depth, as defined in SMPTE ST 2042-1:2017 section 12.4.3, shall be 3.

Note: VC-2 requires that the number of decimation stages be equal in both horizontal and vertical axes.

7.3.3.3 Slice Coding Parameters

The number of slices horizontally per picture (slices x) shall be 120.

Note 1: this means that each individual slice notionally corresponds to 32 input pixels for luminance (given the ripple effects of the wavelet filter).

The number of slices vertically per picture (slices y) shall be 270.

Note 2: This means that each individual slice notionally corresponds to 8 input picture lines (given the ripple effects of the wavelet filter).

The number of bytes allocated to each slice shall be 128 bytes. The slice prefix bytes shall be 0 and the slice size scalar shall be 1 as defined in SMPTE ST 2042-1.

Note 3: The total number of bytes per picture, 4147200, is the number of video samples in a frame of 1920 x 1080 video. Using these byte count values ensures that the compressed video can be mapped onto the high definition infrastructure.

The quantization matrix shall be the default quantization matrix corresponding to the LeGall (5,3) kernel as defined in SMPTE ST 2042-1:2017 Annex D, table D.2.

Table 4.1 documents the slice coding parameter values for each combination of picture size and frame rate.

Table 7-1: Slice Coding Parameters

Picture Size	3840 x 2160	
Frame Rate	60/1.001 Hz	50 Hz
Horizontal Slices	120	
Vertical Slices	270	
Total Slices	32400	
Total Bytes per picture	4147200	
Slice Prefix Bytes	0	
Slice Size Scalar	1	

7.4 Entropy coding

Only interleaved Exp-Golomb coding is used, as per the provisions for the High Quality Profile in ST 2042-1.

8 Syntax Constraints

The syntax of the VC-2 bit-stream shall be constrained as defined in this section to provide interoperability and to facilitate mapping to transport layers.

- A sequence header shall be included in the bit-stream prior to each compressed frame.
- The VC-2 stream shall not contain any auxiliary or padding data units.
- Bits which are otherwise not defined shall be assigned the value zero.

  Note: Such bits are used to pad partially filled bytes to occupy a whole byte.

The format of the bit-stream with the above syntax constraints will therefore be as indicated below.

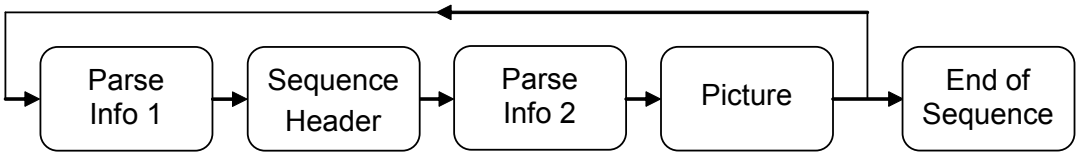

Figure 1: Constrained Stream Syntax

The structure of the picture syntax is as indicated below.

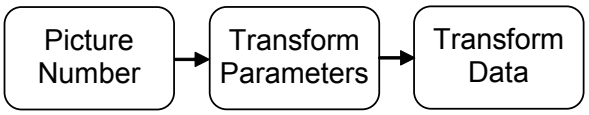

Figure 2: Picture Syntax

The string of bytes defining parse info, sequence header, transform parameters and end of sequence shall be constant within a sequence for streams that conform to this practice. The actual values of the bytes, given the constraints defined in this practice, are defined in SMPTE ST 2042-1 and are enumerated in Annex A of this practice.

Only the picture number and the transform data can vary between pictures in a coded bit-stream (see section 4.2.2 above). The picture number increments for each successive frame and the transform data depends on the picture content of each coded frame.

Each component of the bit stream shown above shall occupy a whole number of bytes.

Note: Each picture thus comprises a picture header (parse info, sequence header, picture number, transform parameters) followed by the transform data. The picture header is constant within the stream except for the picture number, which occupies 4 bytes. Thus the components of the stream can be assembled by juxtaposing the separate components (in the correct order) without the need to shift bits within a byte.

Annex A Byte Values of Stream Components (Informative)

A.1 General

This annex enumerates the specific bytes used as components of a compliant bit-stream. The byte values enumerated here can be derived unambiguously from SMPTE ST 2042-1 plus the constraints defined in preceding sections of this Recommended Practice.

In this annex, numbered tables (i.e. Table A-x) define concatenated byte sequences rather than individual byte string values.

This practice specifies the compression of the two video formats defined in section 4. For brevity in this annex these formats are referred to by the names defined in the table below.

Format Name	Frame Height	Width	Frame Rate (Hz)	Scanning
3840x2160/60	2160	3840	59.94	progressive
3840x2160/50	2160	3840	50	progressive

In this annex the notation ‘0x’ represents a hexadecimal literal and ‘0b’ represents a binary literal. Variable length coding is used to code many components of the picture header. Variable length coded components, in general, occupy a whole number of bytes. In this annex, a hexadecimal string plus the number of bits, is used to define such components in a concise way. The numbers of bits are the right most bits in the hexadecimal number.

A.2 Parse Info 1

Parse info 1 comprises a parse info prefix, parse code, next parse offset and previous parse offset.

- Parse info Prefix is 0x42424344, 4 bytes
- Parse code is 0x00, 1 byte (sequence header)
- Next parse offset is 0x00000012, 4 bytes (an offset of 18)

  Next parse offset is the size of this parse info plus the size of the sequence header.
- Previous parse offset is 0x00000000 (4 bytes) for the first parse info 1 in the sequence (i.e. there is no previous parse info) and 0x003F4817 (4 bytes) for all other parse info 1s.

  Previous parse offset is the sum of the number of bytes for picture (picture number + transform parameters + transform data) plus the number of bytes for parse info 2 (13 bytes).

Parse info 1 therefore occupies 13 bytes, whose values is 0x42424344000000012003F4817 , except for the first parse info in a sequence in which the last 4 bytes (Previous Parse Offset) are set to zero.

A.3 Sequence Header

A.3.1 General

The sequence header comprises parse parameters, base video format, source parameters and picture coding mode. Each of these components can be further divided as outlined below and defined in SMPTE ST 2042-1

Each of the values in the sequence header is coded using the variable length code (abbreviated VLC below) defined in SMPTE ST 2042-1.

A.3.2 Parse Parameters

The parse parameters indicate the revision number of the codec specification, plus the profile and levels of the coded stream.

- Version Number Major: Value = 2, VLC = 0b011, Bits = 3
- Version Number Minor: Value = 0, VLC = 0b1, Bits = 1
- Profile: Value = 3 (high quality), VLC = 0b00001, Bits = 5
- Level: Value = 66, VLC = 0x000B, Bits = 13

Concatenated Parse Parameters (22 bits) = 0x1C200B

A.3.3 Base Video Format

The base video format defines default values for source parameters and picture coding mode.

Format Name	Base Video Format Index	Base Video Format (VLC Hex string)	Number of Bits
3840x2160/60	17	0x009	9
3840x2160/50	18	0x00B	9

A.3.4 Source Parameters

Source parameters describe the source video format, i.e. 2160P 50 or 59.94. The source parameters correspond to the default parameters defined in SMPTE ST 2042-1 and so are not over-ridden.

- Custom dimensions flag Value = False, VLC Code = 0b0, Bits = 1
- Custom color difference format flag: Value = False, VLC Code = 0b0, Bits = 1
- Custom scan format flag: Value = False, VLC Code = 0b0, Bits = 1
- Custom frame rate flag: Value = False, VLC Code = 0b0, Bits = 1
- Custom pixel aspect ratio flag: Value = False, VLC Code = 0b0, Bits = 1
- Custom clean area flag: Value = False, VLC Code = 0b0, Bits = 1
- Custom signal range flag: Value = False, VLC Code = 0b0, Bits = 1
- Custom color specification flag: Value = False, VLC Code = 0b0, Bits = 1

The source parameters bit stream is 0x00 (8 bits).

A.3.5 Picture Coding Mode

The picture coding mode indicates whether the frames are coded as fields or frames. In this practice all pictures are coded as frames not fields (i.e. the source video is progressive not interlaced).

- Value = 0 (progressive), VLC = 0b1, Bits = 1

This value is 0x1 (1 bits).

A.3.6 Sequence Header Bytes

This section enumerates the byte sequence obtained by concatenating the component parts enumerated above and padding with zero (LSBs) to produce a whole number of bytes.

Format Name	Sequence Header (Byte String)	Number of Bytes
3840x2160/60	0x70802C1200	5
3840x2160/50	0x70802C1600	5

A.4 Parse Info 2

Parse info 2 comprises parse info prefix, parse code, next parse offset and previous parse offset.

- Parse info Prefix is 0x42424344, 4 bytes
- Parse code is 0xE8, 1 byte (high quality picture)
- Next parse offset is 0x003F4817, 4 bytes (an offset of 4147223)
- Previous Parse Offset is 0x00000012, 4 bytes (an offset of 18)

Next parse offset is the size of the sequence header plus the size of parse info 1.

Overall Parse info 2 occupies 13 bytes which are the concatenation of the components above in this sub section. The value of this header is 0x42424344E8003F481700000012 (13 bytes).

A.5 Picture Number

The picture number is located in the picture header as defined in SMPTE ST 2042-1. The value of the first picture number in the sequence is unconstrained. Subsequent pictures increment by one per frame.

Picture number is an unsigned 32 bit integer in which the most significant bit is the first bit in the stream.

A.6 Transform Parameters

The transform parameters describe the details of the wavelet transform that is used and the way the transform coefficients are coded. Like the sequence header these parameters are coded using the VLC defined in SMPTE ST 2042-1.

- Wavelet Filter Index: Value = 1 (LeGall (5,3)) VLC = 0b001 Bits = 3
- DWT Depth: Value = 3 (3 level transform) VLC = 0b00001 Bits = 5
- Slices X: Value = 120 VLC = 0x0A83 Bits = 13
- Slices Y: Value = 270 VLC = 0x000AB Bits = 17
- Slice Prefix Bytes: Value = 0 VLC = 0b1 Bits = 1
- Slice Size Scalar: Value = 1 VLC = 0b001 Bits = 3
- Custom Quantization Matrix: Value = False VLC= 0b0 Bits = 1

Concatenating the VLC above, padding with zeros to produce a whole number of bytes and converting to hexadecimal gives the byte sequence for the transform parameters: 0x21541802AE40 (6 bytes).

A.7 Transform Data

Transform data is picture dependent and varies frame by frame. The actual values are calculated as defined in SMPTE ST 2042-1.

With the constrained parameters defined in this practice, the transform data for each frame occupies precisely 4147200 bytes.

A.8 End of Sequence

End of sequence is a parse info comprising a parse info prefix, parse code, next parse offset and previous parse offset.

- Parse info Prefix is 0x42424344, 4 bytes
- Parse code is 0x10, 1 byte (end of sequence)
- Next parse offset is 0x00000000 (no next parse offset)
- Previous parse offset is 0x003F4817, 4 bytes.

Previous parse offset is the sum of the number of bytes for picture data (picture number, transform parameters, transform data) plus the number of bytes for parse info 2 (13 bytes).

End of sequence therefore occupies 13 bytes which are 0x42424344100000000003F4817.

Bibliography

SMPTE ST 2042-2:2009, VC-2 Level Definitions

Recommendation ITU-R BT.2020-2 (10/2015), Parameter values for ultra-high definition television systems for production and international programme exchange