SMPTE RP 2047-3:2023
Revision of SMPTE RP 2047-3:2016

SMPTE RECOMMENDED PRACTICE

VC-2 Level 65 Compression of High Definition Video Sources for Use with a Standard Definition Infrastructure

Table of Contents

Approved November 29, 2022

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 Essence.

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 clause 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; then formal languages; then figures; and then any other language forms.

This revision makes updates to improve clarity. It also removes references to “low delay syntax” because the term is not used in the current VC-2 standard.

Introduction

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

The transition from standard definition to high-definition television production is well advanced. Yet many television production facilities still have a substantial legacy of a standard definition infrastructure that is impractical, or uneconomic, to immediately upgrade to a high-definition infrastructure. There is therefore a need to compress high-definition signals so that high-definition production can continue to use a standard definition infrastructure. Standard definition infrastructure components that have the potential to be reused by such compression include cabling, routers, uncompressed video links and uncompressed video recorders.

This Recommended Practice defines constraints on the VC-2 standard (SMPTE ST 2042-1) such that high-definition source video images can be compressed to standard definition data rates. The bit rate of the resulting VC-2 stream is such that it can be carried over a standard definition serial digital interface (SDI) operating at 270Mbit/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.

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.

1 Scope

This Recommended Practice defines constraints on the VC-2 video compression standard such that 10-bit high-definition video signals can be compressed to only occupy the bit rate of 8-bit standard definition signals. Specifically, it defines constraints for compressing interlaced source video having 1920 or 1440 pixels/line by 1080 lines/frame at 25 or 30/1.001 frames/s, or progressive source video having 1280 pixels/line by 720 lines/frame at 50 or 60/1.001 frames/s, using 4:2:2 sampling in all cases. It also defines an application-specific level value corresponding to these constraints.

2 Normative references

The following documents contain provisions that, through reference in this text, constitute provisions of this Recommended Practice. 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 documents 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 274, Television — 1920 x 1080 Image Sample Structure, Digital Representation and Digital Timing Reference Sequences for Multiple Picture Rates

SMPTE ST 296, 1280 × 720 Progressive 4:2:2 and 4:4:4 Image Sample Structure — Analog and Digital Representation and Analog Interface

SMPTE ST 2042-1:2022, VC-2 Video Compression

3 Terms and definitions

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

4 Overview (informative)

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

- a choice of Wavelet transform filters,
- a choice in the number of levels of Wavelet decimation,

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 standard definition infrastructure. It 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 VC2 level number 65 (as indicated below in Clause 5.3.2.2. That is, VC-2 streams conformant to this Recommended Practice ensure compliance by specifying specialized level 65 in the stream. This Recommended Practice does not define how the VC-2 stream is mapped to any specific transport layer. Transport over SDI is specified in SMPTE ST 2047-4.

5 Coding constraints

5.1 Summary (informative)

This application has the following key coding constraints:

1. The video sources are limited to 4:2:2 sampling with the following picture formats:
    - 1920 pixels by 1080 interlaced lines with frame rates of 25 frames/sec or 30/1.001 frames/sec,
    - 1440 pixels by 1080 interlaced lines with frame rates of 25 frames/sec or 30/1.001 frames/sec,
    - 1280 pixels by 720 progressive lines with frame rates of 50 frames/sec or 60/1.001 frames/sec.
2. The codec uses only the low delay profile.
3. The codec uses only the LeGall (5,3) wavelet kernel.
4. The wavelet depth is exactly 3.
5. Slices are kept as small as possible, consistent with the wavelet depth, to minimize latency and implementation complexity.

5.2 Video source parameters

The video source shall be:

- as defined by SMPTE ST 274 for the field rates of 50 fields/sec and 60/1.001 field/sec only

or

- as defined by SMPTE ST 296 for the frame rates of 50 frames/sec and 60/1.001 frames/sec only

The image sampling for a SMPTE ST 274 video source shall be:

- as defined by SMPTE ST 274 systems 5 and 6 only

or

- as defined by SMPTE ST 274 systems 5 and 6 only, in which the horizontal sample rate has been reduced to 1440 pixels

The image sampling for a SMPTE ST 296 video source shall be:

- as defined by SMPTE ST 296 systems 2 and 3 only

NOTE 1 In the second case, the horizontal sampling structure of SMPTE ST 274 systems 5 and 6 is reduced in the ratio 3 output pixels for every 4 input pixels prior to compression. The sampling structure needs to be restored in a complementary fashion after decompression. Thus, the overall system (sample rate changing plus compression) can maintain compatibility with SMPTE ST 274. The provision to support 1440-pixel video is provided for compatibility with other video compression systems that use 1440 pixel video sub-sampling (such as Type D-11, SMPTE ST 367). Failure to support 1440-pixel video compression would result in increased compression artifacts and noise in systems with concatenated compression. To summarize, the text in Clause 5 normatively provides for the sub-sampling of 1920 pixels to 1440 pixels for the purpose of compression coding. Since the sub-sampling is a part of the compression process, there is no externally accessible presence of the 1440-pixel format so there is no requirement to document a standardized method for the down- and upconversions. Filter coefficients suitable for sample rate changing using FIR filters are provided in informative Annex B.

The bit depth shall be limited to the 10 most significant bits.

The video sampling format shall be 4:2:2 as defined by SMPTE ST 274.

Only the active picture area shall be coded. For the 1080-line video formats only, each field of 540 lines shall be padded to 544 lines as part of the compression process, as defined in SMPTE ST 2042-1, Clause 13.2.3. These extra four lines shall be added to the end of the picture and shall be removed as part of the VC-2 decoding process.

NOTE 2 These extra 4 lines are needed to pad the number of lines to be divisible by 8 in order to support the transform depth value of 3.

5.3 Coding parameters

5.3.1 Syntax

The coding syntax structure shall be as defined in SMPTE ST 2042-1.

5.3.2 Sequence header

5.3.2.1 General

The sequence header comprises the parse parameters, base video format, source parameters and picture coding mode and is defined in Clauses 5.3.2.2 through 5.3.2.5.

5.3.2.2 Parse parameters

The major version and minor version numbers shall be as defined in SMPTE ST 2042-1.

The profile value shall be the low delay profile as defined in SMPTE ST 2042-1.

The level value shall be a specialized level value with the value of 65. This level value shall be used only for the VC-2 coding constraints that comply with this Recommended Practice.

5.3.2.3 Base video format

The base video format shall be one of format numbers 9, 10, 11 or 12 (HD 720p60, HD 720p50, HD 1080i60 or HD 1080i50) as defined in SMPTE ST 2042-1.

NOTE 1 ST 2042-1, the base video format names for 30 and 60 frames/sec systems ignore the 1.001 divisor for simplicity. Thus, HD 1080i60 means that the default field rate is 60/1.001 fields/sec, not 60 fields/sec.

NOTE 2 The format is thus defined as follows:

- The picture size is one of 1920 x 1080, 1440 x 1080 or 1280 x 720.
- The sampling format is 4:2:2.
- The scanning is interlaced for 1920 x 1080 and 1440 x 1080 video sources, and progressive for 1280 x 720 video sources.
- The frame rate is 30/1.001 Hz or 25 Hz for 1920 x 1080 and 1440 x 1080 video sources, and 60/1.001 Hz or 50 Hz (for 1280 x 720 video sources).
- The pixel aspect ratio is 1:1 for 1920 x 1080 and 1280 x 720 video sources (aspect ratio index 1) or 4:3 for 1440 x 1080 video sources (aspect ratio index 6, reduced horizontal resolution systems).
- The clean picture area is the same as the picture size, i.e., 1920 x 1080, 1440 x 1080 or 1280 x 720.
- The video depth is 10 bits.
- The color primaries and color matrix are as defined in ITU-R BT.709.
- The transfer function is as defined in ITU-R BT.709.

5.3.2.4 Source parameters

The source parameters allow the video parameters predefined by the base video format to be overridden with custom values. To comply with this document, the predefined video parameters shall not be overridden except in the following case:

- The frame width to be 1440 as specified in SMPTE ST 2042-1:2022, Clause 11.4.3.
- The pixel aspect ratio index to be 6 (4:3 reduced horizontal resolution systems) as specified in SMPTE ST 2042-1:2022, Clause 11.4.7.
- The clean area to be 1440 x 1080 with zero offset from top left as specified in SMPTE ST 2042-1, Clause 11.4.8.

5.3.2.5 Picture coding mode

Picture coding mode shall be 1 for the interlaced source video formats (base video format numbers 11 and 12) and shall be 0 for the progressive source video formats (base video format numbers 9 and 10).

NOTE This means that interlaced source fields are coded separately to minimize latency, whereas progressive source frames are coded as a whole.

5.3.3 Picture syntax

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

As per SMPTE ST 2042-1, the picture number increments by one for each successive picture.

NOTE According to this Recommended Practice, pictures are fields for interlaced video sources and frames for progressive video sources.

5.3.4 Transform parameters

5.3.4.1 Wavelet filter

The wavelet filter shall be “LeGall (5,3)” filter (Wavelet index value = 1) as defined in SMPTE ST 2042-1:2022, Table 15.

5.3.4.2 Transform depth

The discrete wavelet transform (DWT) depth, as defined in SMPTE ST 2042-1:2022, Clause 12.4.3, shall be 3.

5.3.4.3 Slice coding parameters

The number of slices horizontally per picture (slices x) shall be the number of horizontal pixels divided by 16, i.e., there are 120, 90 or 80 horizontal slices for 1920, 1440 or 1280 horizontal pixel source video respectively.

NOTE 1 This means that each individual slice notionally corresponds to 16 input pixels for luminance (given the ripple effects of the wavelet filter).

The number of slices vertically per picture (slices y) shall be the number of vertical lines divided by 8 rounded up to the nearest integer, i.e., the number of vertical slices per picture is 68 for 1080-line interlaced source video and 90 for 720-line progressive source video.

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 vertical slices is defined per picture, which is a field for interlaced source video and a frame for progressive source video. Each field of 1080-line source video is padded to 544 lines within the compression process.

The number of bytes allocated to each slice shall be 349920 bytes divided by the number of slices per picture for source video with the 30/1.001-Hz frame rate, and 414720 bytes divided by the number of slices per picture for source video with the 25-Hz frame rate. The number of bytes per slice is not a whole number but can be precisely specified as a rational number by defining bytes per slice numerator and bytes per slice denominator as specified in SMPTE ST 2042-1.

NOTE 3 The total number of bytes per picture, 349920 or 414720, is the number of video samples in a field of standard definition video at the corresponding frame rate. So, for 30/1.001 frames/s standard definition, 349920 is derived by multiplying the number of pixels horizontally, by the number of lines vertically, by 2 (to allow for both luma and color difference components), i.e., 720 x 486 x 2. Similarly, for 25 frames/s standard definition, 414720 is the product 720 x 576 x 2. Using these byte count values ensures that the compressed video can be mapped onto the standard 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.

Table 1 defines the slice coding parameter values for each combination of picture size and frame rate.

Table 1 — Slice coding parameters.

Video Format	**1920 x 1080 (interlaced)**		**1440 x 1080 (interlaced)**		**1280 x 720 (progressive)**	
Frame Rate	**30/1.001 Hz**	**25 Hz**	**30/1.001 Hz**	**25 Hz**	**60/1.001 Hz**	**50 Hz**
Picture Rate	**60/1.001 Hz**	**50 Hz**	**60/1.001 Hz**	**50 Hz**	**60/1.001 Hz**	**50 Hz**
Horizontal Slices	120	120	90	90	80	80
Vertical Slices	68	68	68	68	90	90
Total Slices	8160	8160	6120	6120	7200	7200
Total Bytes per picture	349920	414720	349920	414720	349920	414720
Bytes per slice Numerator	729	864	972	1152	243	288
Bytes per slice Denominator	17	17	17	17	5	5

5.4 Entropy coding

All entropy coded data shall comply with requirements of the low delay profile.

NOTE Only interleaved Exp-Golomb coding is used.

6 Syntax constraints

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

- A sequence header shall be included in the bitstream prior to each compressed frame.
- The VC-2 stream shall not contain any auxiliary or padding data units.
- The number of bytes allocated to each slice, “slice bytes numerator” and “slice bytes denominator” shall be signaled as defined in the last two rows of Table 1. That is, for each format in Table 1, the “slice bytes numerator” and “slice bytes denominator” values shall be the lowest integer values that do not contain a common factor.
- Bits which are otherwise not defined shall be assigned the value zero.

  NOTE 1 Such bits are used to pad partially filled bytes to occupy a whole byte.

The format of the bitstream with these syntax constraints is shown in Figure 1.

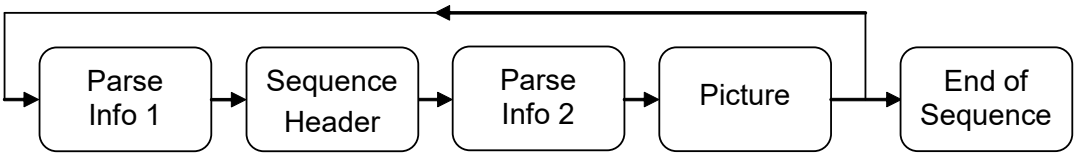

Figure 1 — Constrained stream syntax.

The structure of the sequence header and picture shall be as defined in SMPTE ST 2042-1. For clarity the structure of the picture syntax is shown in Figure 2.

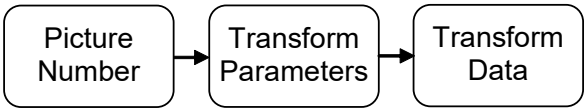

Figure 2 — Picture syntax.

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

Only the picture number and the transform data can vary between pictures in a coded bitstream. The picture number increments by one for each successive picture and the transform data depends on the picture content of each coded frame.

Each component of the bitstream shown in Figure 1 and Figure 2 shall occupy a whole number of bytes.

NOTE 2 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 concatenating the separate components, in the correct order, without the need to shift bits within a byte.

Annex A (Informative) Byte values of stream components

A.1 General

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

This Recommended Practice specifies the compression of the six video formats defined in Clause 5. For brevity in this annex these formats are referred to by the names defined in Table A.1. These names use the notation recommended in ITU-R BT.1868.

Table A.1 — Video format names.

Format name	Frame height	Width	Frame rate (Hz)	Scanning
1920x1080/59.94/I	1080	1920	30/1.001	interlace
1920x1080/50/I	1080	1920	25	interlace
1440x1080/59.94/I	1080	1440	30/1.001	interlace
1440x1080/50/I	1080	1440	25	interlace
1280x720/59.94/P	720	1280	60/1.001	progressive
1280x720/50/P	720	1280	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 rightmost 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 as listed in Table A.2

Table A.2 — Next parse offset.

Format name	Next parse offset (decimal)	Next parse offset (hex string)	Number of bytes
1920x1080/59.94/I	18	0x00000012	4
1920x1080/50/I	18	0x00000012	4
1440x1080/59.94/I	29	0x0000001D	4
1440x1080/50/I	29	0x0000001D	4
1280x720/59.94/P	18	0x00000012	4
1280x720/50/P	18	0x00000012	4

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 in the sequence (i.e., there is no previous parse info) and as indicated in Table A.3 for all other parse infos.

Table A.3 — Previous parse offset.

Format name	Previous parse offset (decimal)	Previous parse offset (hex string)	Number of bytes
1920x1080/59.94/I	349945	0x000556F9	4
1920x1080/50/I	414745	0x00065419	4
1440x1080/59.94/I	349945	0x000556F9	4
1440x1080/50/I	414746	0x0006541A	4
1280x720/59.94/P	349944	0x000556F8	4
1280x720/50/P	414745	0x00065419	4

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 are listed in Table A.4, except for the first parse info in a sequence in which the last 4 bytes (previous parse offset) are set to zero.

Table A.4 — Parse info 1 byte sequence.

Format name	Parse info 1 byte string	Number of bytes
1920x1080/59.94/I	0x42424344000000012000556F9	13
1920x1080/50/I	0x42424344000000001200065419	13
1440x1080/59.94/I	0x4242434400000001D000556F9	13
1440x1080/50/I	0x4242434400000001D0006541A	13
1280x720/59.94/P	0x42424344000000012000556F8	13
1280x720/50/P	0x42424344000000001200065419	13

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 in Clauses A.3.2 through A.3.5, and defined in SMPTE ST 2042-1.

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

A.3.2 Parse parameters

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

- Version Number Major: Value = 1, VLC = 0b001, Bits = 3
- Version Number Minor: Value = 0, VLC = 0b1, Bits = 1
- Profile: Value = 0 (low delay), VLC = 0b1, Bits = 1
- Level: Value = 65, VLC = 0x0009, Bits = 13

Concatenated parse parameters (18 bits) = 0x0E009

A.3.3 Base video format

The base video format index defines default values for source parameters and picture coding mode. Table A.5 lists the values for each format.

Table A.5 — Base video format index.

Format name	Base video format index	Base video format (VLC Hex string)	Number of bits
1920x1080/59.94/I	11	0x21	7
1920x1080/50/I	12	0x23	7
1440x1080/59.94/I	11	0x21	7
1440x1080/50/I	12	0x23	7
1280x720/59.94/P	9	0x09	7
1280x720/50/P	10	0x0B	7

A.3.4 Source parameters

Source parameters describe the source video format, i.e., the formats listed in Table A.1. The bit strings for each format are listed in Table A.6.

- Custom dimensions flag:
    - 1920x1080 & 1280x720: Value = False, Code = 0b0, Bits = 1
    - 1440x1080: Value = True, Code = 0b1, Bits = 1
        - frame width: Value = 1440, VLC = 0x028803, Bits = 21
        - frame height: Value = 1080, VLC = 0x000A83, Bits = 21
- Custom color difference format flag: Value = False, Code = 0b0, Bits = 1
- Custom scan format flag: Value = False, Code = 0b0, Bits = 1
- Custom frame rate flag: Value = False, Code = 0b0, Bits = 1
- Custom pixel aspect ratio flag:
    - 1920x1080 & 1280x720: Value = False, Code = 0b0, Bits = 1
    - 1440x1080: Value = True, Code = 0b1, Bits = 1
        - Custom pixel aspect ratio: Value = 6, VLC = 0xB, Bits = 5
- Custom clean area flag:
    - 1920x1080 & 1280x720: Value = False, Code = 0b0, Bits = 1
    - 1440x1080: Value = True, Code = 0b1, Bits = 1
        - Clean Width: Value = 1440, VLC = 0x028803, Bits = 21
        - Clean Height: Value = 1080, VLC = 0x000A83, Bits = 21
        - Left Offset: Value = 0, VLC = 0b1, Bits = 1
        - Top Offset: Value = 0, VLC = 0b1, Bits = 1
- Custom signal range flag: Value = False, Code = 0b0, Bits = 1
- Custom color specification flag: Value = False, Code = 0b0, Bits = 1

Table A.6 — Source parameters.

Format name	Source parameters bit string	Number of bits
1920x1080/59.94/I	0x00	8
1920x1080/50/I	0x00	8
1440x1080/59.94/I	0x8A200C015062B8A200C015078	99
1440x1080/50/I	0x8A200C015062B8A200C015078	99
1280x720/59.94/P	0x00	8
1280x720/50/P	0x00	8

A.3.5 Picture coding mode

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

- Picture coding mode 1080I: Value = 1 (interlaced), VLC = 0b001, Bits = 3
- Picture coding mode 720P: Value = 0 (progressive), VLC = 0b1, Bits = 1

A.3.6 Sequence header bytes

This clause enumerates the byte sequence obtained by concatenating the component parts enumerated in Clauses A.3.2 through A.3.5 and padding with zero (LSBs) to produce a whole number of bytes.

For example, for 1080x1920I60 the concatenated bits are:
0b00111000,00000010,01010000,10000000,0001, padding with 4 zero bits on the right and expressing in hexadecimal gives the value in Table A.7.

Table A.7 — Sequence header.

Format name	Sequence header (byte string)	Number of bytes
1920x1080/59.94/I	0x3802508010	5
1920x1080/50/I	0x3802518010	5
1440x1080/59.94/I	0x380250C5100600A8315C5100600A83C2	16
1440x1080/50/I	0x380251C5100600A8315C5100600A83C2	16
1280x720/59.94/P	0x3802448040	5
1280x720/50/P	0x3802458040	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 0xC8, 1 byte (low delay picture)
- Next parse offset is as indicated in Table A.8:

Table A.8 — Next parse offset.

Format name	Next parse offset (decimal)	Next parse offset (hex string)	Number of bytes
1920x1080/59.94/I	349945	0x000556F9	4
1920x1080/50/I	414745	0x00065419	4
1440x1080/59.94/I	349945	0x000556F9	4
1440x1080/50/I	414746	0x0006541A	4
1280x720/59.94/P	349944	0x000556F8	4
1280x720/50/P	414745	0x00065419	4

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

- Previous parse offset is as indicated in Table A.9:

Table A.9 — Previous parse offset.

Format name	Previous parse offset (decimal)	Previous parse offset (hex string)	Number of bytes
1920x1080/59.94/I	18	0x00000012	4
1920x1080/50/I	18	0x00000012	4
1440x1080/59.94/I	29	0x0000001D	4
1440x1080/50/I	29	0x0000001D	4
1280x720/59.94/P	18	0x00000012	4
1280x720/50/P	18	0x00000012	4

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 parse info prefix, parse code, next parse offset and previous parse offset. Table A.10 lists the byte sequences for each format.

Table A.10 — Parse info 2 byte sequence.

Format name	Parse info 2 bytes	Number of bytes
1920x1080/59.94/I	0x42424344C8000556F900000012	13
1920x1080/50/I	0x42424344C80006541900000012	13
1440x1080/59.94/I	0x42424344C8000556F90000001D	13
1440x1080/50/I	0x42424344C80006541A0000001D	13
1280x720/59.94/P	0x42424344C8000556F800000012	13
1280x720/50/P	0x42424344C80006541900000012	13

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 picture numbers increment by one for each successive picture.

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
- Slice parameters – per Table A.11
- Custom Quantization Matrix: Value = False, Code = 0b0, Bits = 1

Table A.11 — Slice parameters.

Format name	Slices X	Slices Y	Slice bytes numerator	Slice bytes denominator
	VLC / bits	VLC / bits	VLC / bits	VLC / bits
1920x1080/59.94/I	120	68	729	17
	0x0A83 / 13	0x0023 / 13	0x0A289 / 19	0x009 / 9
1920x1080/50/I	120	68	864	17
	0x0A83 / 13	0x0023 / 13	0x22803 / 19	0x009 / 9
1440x1080/59.94/I	90	68	972	17
	0x028B / 13	0x0023 / 13	0x2A0A3 / 19	0x009 / 9
1440x1080/50/I	90	68	1152	17
	0x028B / 13	0x0023 / 13	0x008003 / 21	0x009 / 9
1280x720/59.94/P	80	90	243	5
	0x0203 / 13	0x028B / 13	0x2A21 /15	0x09 / 5
1280x720/50/P	80	90	288	5
	0x0203 / 13	0x028B / 13	0x00803 /17	0x09 / 5

Concatenating the VLC for wavelet filter index, DWT depth and slice parameters, padding with zeros to produce a whole number of bytes and converting to hexadecimal gives the byte sequence for the transform parameters. Table A.12 lists the byte sequences for each format.

Table A.12 — Transform parameters byte sequence.

Format name	Transform parameters bytes	Number of bytes
1920x1080/59.94/I	0x21541808C5144824	8
1920x1080/50/I	0x21541808D1401824	8
1440x1080/59.94/I	0x21145808D5051824	8
1440x1080/50/I	0x21145808C100060900	9
1280x720/59.94/P	0x211018A2D510A4	7
1280x720/50/P	0x211018A2C1006900	8

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 Recommended Practice, the transform data for each frame occupies precisely 349920 bytes (all three 60Hz formats) or 414720 (all three 50Hz formats).

These figures are calculated by multiplying the number of slices by the slice bytes numerator and dividing by the slice bytes denominator.

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 listed in Table A.13

Table A.13 — Previous parse offset.

Format name	Previous parse offset (decimal)	Previous parse offset (hexadecimal)
1920x1080/59.94/I	349945	0x000556F9
1920x1080/50/I	414745	0x00065419
1440x1080/59.94/I	349945	0x000556F9
1440x1080/50/I	414746	0x0006541A
1280x720/59.94/P	349944	0x000556F8
1280x720/50/P	414745	0x00065419

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 listed for each format in Table A.14.

Table A.14 — End of sequence bytes.

Format name	End of sequence bytes
1920x1080/59.94/I	0x42424344100000000000556F9
1920x1080/50/I	0x42424344100000000000065419
1440x1080/59.94/I	0x42424344100000000000556F9
1440x1080/50/I	0x4242434410000000000006541A
1280x720/59.94/P	0x42424344100000000000556F8
1280x720/50/P	0x42424344100000000000065419

Annex B (Informative) Filter coefficients for conversion between 1920 and 1440 pixels

B.1 General

This Recommended Practice provides VC-2 coding parameters etc. for both 1920-pixel and 1440-pixel HDTV formats. The 1440-pixel format is added to support the use of other compression techniques that use 1440-pixel formats. Repeated conversion between 1920- and 1440-pixel formats can increase the overall compression loss and compressing a 1920-pixel format that only contains information for a 1440-pixel format unnecessarily increases compression artifacts. Compression loss in such concatenated compression systems is minimized by only compressing the 1440-pixel format throughout a concatenated compression system and up-converting to 1920 pixels only for display or processing.

This Recommended Practice does not require the use of any particular method for conversion between 1920- and 1440-pixel formats. However, for implementation convenience, this annex provides a set of filter coefficients that can be used in a linear sample rate converter to achieve the desired conversion. Details of such sample rate conversion techniques, and their optimizations, are widely known and will only briefly be described here. The purpose of this annex is to provide a set of finite impulse response filter coefficients that implementers are free to use.

The downconversion and upconversion filters are applied equally to all color components noting that there is half the number of color-difference samples per horizontal scan. Thus, the luminance component is converted between 1920 and 1440 pixels and the color-difference components are converted between 960 and 720 pixels. All references to 1920/1440 conversion for the luminance component apply equally to the 960/720 conversion for the color-difference components.

B.2 1920 to 1440 Conversion (downconversion)

Conversion between 1920 and 1440 pixels requires changing the horizontal sampling rate by a factor of 4:3. Conceptually, in a linear converter, this is achieved by first up-sampling by a factor of 3 (i.e., inserting 2 zero value samples between each input sample), followed by filtering using a finite impulse response filter, and finally down-sampling by a factor of 4 (i.e., extracting only every 4th filtered sample).

This clause offers the filter coefficients listed in Table B.1 for down-converting from 1920 to 1440 luminance pixels and 960 to 720 color-difference pixels.

Table B.1 — Downconversion filter coefficients.

Coefficient number	1	2	3	4	5	6	7	8	9	10	11	12
Coefficient value	60	25	-30	-141	-174	-31	174	364	297	-4	-395	-648
Coefficient number	13	14	15	16	17	18	19	20	21	22	23	24
Coefficient value	-540	24	939	1957	2746	3042	2746	1957	939	24	-540	-648
Coefficient number	25	26	27	28	29	30	31	32	33	34	35	
Coefficient value	-395	-4	297	364	174	-31	-174	-141	-30	25	60	

There are 35 filter coefficients in Table B.1. They are sampled at 3 times the horizontal sampling rate for 1920-pixel video, i.e., they are sampled at the 3 times up-sampling rate. A well-known optimization of the downconversion process arises because only about 1/3 of the filter taps operate on non-zero sample values (because of the 3:1 up-sampling). Therefore, the downconversion process, using the filter coefficients in Table B.1, can be accomplished using a 12-tap filter with (3 sets of) variable coefficients.

The coefficients are signed 13-bit coefficients i.e., in the range −4096 to +4095. The coefficients are scaled such that unity is defined as 4096, i.e., after filtering the result needs to be divided by 4096 to give the correct magnitude. For a 10-bit input signal, 23-bit precision needs to be maintained during filtering to prevent overflow.

At the left and the right edge of the image the filter's region of support extends beyond the edge of the image. It is recommended that the extreme edge pixel values are replicated to provide inputs to the filter (rather than, for example, those inputs being set to zero). By replicating the edge pixels "ringing" artifacts from the filtering process are minimized at the edge of the picture. Pixel replication needs to be performed prior to up-sampling.

B.3 1440 to 1920 Conversion (upconversion)

Conversion between 1440 and 1920 pixels requires increasing the horizontal sampling rate by a factor of 4:3. Conceptually, in a linear converter, this is achieved by first up-sampling by a factor of 4 (i.e., inserting 3 zero value samples between each input sample), followed by filtering using a finite impulse response filter, and finally down-sampling by a factor of 3 (i.e., extracting only every 3rd filtered sample).

This clause offers the filter coefficients listed in Table B.2 for up-converting from 1440 to 1920 luminance pixels and 720 to 960 color-difference pixels.

Table B.2 — Upconversion filter coefficients.

Coefficient Number	1	2	3	4	5	6	7	8	9	10	11	12
Coefficient Value	-12	21	57	61	64	-21	-157	-214	-202	-65	177	365
Coefficient Number	13	14	15	16	17	18	19	20	21	22	23	24
Coefficient Value	417	267	-68	-459	-728	-704	-283	513	1539	2550	3292	3565
Coefficient Number	25	26	27	28	29	30	31	32	33	34	35	36
Coefficient Value	3292	2550	1539	513	-283	-704	-728	-459	-68	267	417	365
Coefficient Number	37	38	39	40	41	42	43	44	45	46	47	
Coefficient Value	177	-65	-202	-214	-157	-21	64	60	57	21	-12	

There are 47 filter coefficients in Table B.2. They are sampled at 4 times the horizontal sampling rate for 1440-pixel video, i.e., they are at the sampling rate after 4 times up-sampling. A well-known optimization of the upconversion process arises because only about 1/4 of the filter taps operate on non-zero sample values (because of the 4:1 up-sampling). Therefore, the upconversion process, using the filter coefficients in Table B.2, can be accomplished using a 12-tap filter with (4 sets of) variable coefficients.

The coefficients are signed 13-bit coefficients i.e., in the range −4096 to +4095. The coefficients are scaled such that unity is defined as 4096, i.e., after filtering the result needs to be divided by 4096 to give the correct magnitude. For a 10-bit input signal, 23-bit precision needs to be maintained during filtering to prevent partial sums “wrapping round”.

At the left and the right edge of the image the filter’s region of support extends beyond the edge of the image. It is recommended that the extreme edge pixel values are replicated to provide inputs to the filter (rather than, for example, those inputs being set to zero). By replicating the edge pixels “ringing” artifacts from the filtering process are minimized at the edge of the picture. Pixel replication needs to be performed prior to up-sampling.

Bibliography

SMPTE ST 367:2002, For Television – Type D-11 Picture Compression and Data Stream Format

SMPTE ST 2042-2:2017, VC-2 Level Definitions

SMPTE ST 2047-4:2011, Carriage of Level 65 VC-2 Compressed Video Over the SDTV SDI

Recommendation ITU-R BT.656-5 (12/2007), Interfaces for Digital Component Video Signals in 525-line and 625-line Television Systems Operating at the 4:2:2 Level of Recommendation ITU-R BT.601

Recommendation ITU-R BT.709-6 (06/2015), Parameter Values for the HDTV Standards for Production and International Programme Exchange

Recommendation ITU-R BT.1846-0 (10/2008), Notations for video systems