

SMPTE RECOMMENDED PRACTICE

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



Table of Contents	Page
Foreword	2
Intellectual Property	2
Introduction.....	2
1 Scope	3
2 Conformance Notation	3
3 Normative References	3
4 General Description	4
4.1 Acronyms and Terms.....	4
5 Coding Constraints.....	4
5.1 Video Source Parameters.....	5
5.2 Coding Parameters	5
5.2.1 Sequence Header	5
5.2.2 Picture Syntax	7
5.2.3 Transform Parameters	7
5.3 Entropy Coding	8
6 Syntax Constraints.....	8
Annex A Byte Values of Stream Components (Informative)	10
A.1 Parse Info 1.....	10
A.2 Sequence Header	11
A.2.1 Parse Parameters.....	11
A.2.2 Base Video Format	12
A.2.3 Source Parameters.....	12
A.2.4 Picture Coding Mode	12
A.2.5 Sequence Header Bytes.....	13
A.3 Parse Info 2.....	13
A.4 Picture Number	14
A.5 Transform Parameters	14
A.6 Transform Data	15
A.7 End of Sequence.....	16
Annex B Filter Coefficients for 1920 to 1440 Conversion (Informative).....	17
B.1 1920 to 1440 Conversion (Down-conversion)	17
B.2 1440 to 1920 Conversion (Up-conversion)	18
Annex C Bibliography (Informative).....	20

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 2047-3 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 Standard. 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

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 may continue to use a standard definition infrastructure. Standard definition infrastructure components that may 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 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.

1 Scope

This recommended practice defines constraints on the VC-2 video compression specification such that 10-bit high definition video signals may 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.

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

Note: All references in this document to other SMPTE documents use the current numbering style (e.g. SMPTE ST 274:2008) although, during a transitional phase, the document as published (printed or PDF) may bear an older designation (such as SMPTE 274M-2008). Documents with the same root number (e.g. 274) and publication year (e.g. 2008) are functionally identical.

The following standards contain provisions which, through reference in this text, constitute provisions of this standard. At the time of publication, the editions indicated were valid. All standards are subject to revision,

and parties to agreements based on this standard are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

SMPTE ST 274:2008, Television — 1920 x 1080 Image Sample Structure, Digital Representation and Digital Timing Reference Sequences for Multiple Picture Rates

SMPTE ST 296:2001, Television — 1280 x 720 Progressive Image Sample Structure — Analog and Digital Representation and Analog Interface

SMPTE ST 2042-1:2009, VC-2 Video Compression

4 General Description

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 and low-delay operation,
- A choice of entropy coding.

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 practice defines coding constraints to provide the interoperability of VC-2 streams coded for compatibility with a standard definition infrastructure. This 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 65 (as indicated in Section 5.2.1.1). That is, VC-2 streams conformant to this practice ensure compliance by specifying specialized level 65 in the stream. This practice does not define how the VC-2 stream is mapped to any specific transport layer.

Note: Such VC-2 stream transport specifications are defined in other SMPTE documents, such as SMPTE ST 2047-4 for the carriage of VC-2 Level 65 Compressed Video over SDI.

4.1 Acronyms and Terms

Lists of acronyms and terms are defined in SMPTE ST 2042-1. They are not replicated here to avoid any divergence of meaning.

5 Coding Constraints

This section defines the VC-2 coding constraints that ensure conformance to this application.

In summary, 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 29.97 frames/sec,
 - 1440 pixels by 1080 interlaced lines with frame rates of 25 frames/sec or 29.97 frames/sec,
 - 1280 pixels by 720 progressive lines with frame rates of 50 frames/sec or 59.94 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.

5.1 Video Source parameters

The video sources shall be:

1. as defined by SMPTE ST 274 for the field rates of 50 fields/sec and 59.94 field/sec only; and
2. as defined by SMPTE ST 296 for the frame rates of 50 frames/sec and 59.94 frames/sec only.

The image sampling shall be:

1. as defined by SMPTE ST 274 systems 5 and 6 only;
2. as defined by SMPTE ST 274 systems 5 and 6 only, in which the horizontal sample rate has been reduced to 1440 pixels; and
3. as defined by SMPTE ST 296 systems 2 and 3 only.

Note: 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 Section 5.1 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 up conversions. 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, Section 13.1.2. 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: 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.

5.2 Coding Parameters

The codec shall use only the low delay syntax. The coding syntax structure shall be as defined in SMPTE ST 2042-1.

5.2.1 Sequence Header

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

5.2.1.1 Parse parameters

The major version and minor version numbers shall be as defined in SMPTE ST 2042-1, Section 11.1.

The profile value shall be the low delay profile as defined in SMPTE ST 2042-1, Annex D.1.1.

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 practice.

5.2.1.2 Base video format

The base video format shall be one of format numbers 9, 10, 11 or 12 (HD 720P-60, HD 720P-50, HD 1080I-60 or HD-1080I-50) as defined in SMPTE ST 2042-1.

Video formats with 1440 horizontal pixels use the base video format 11 or 12 and override the frame width to be 1440 using the source parameter values as defined in Section 5.2.1.3.

Notes:

1. Per SMPTE ST 2042-1, the format names for 30/60 frames/sec systems ignore the 1.001 divisor for simplicity. Thus HD 1080I-60 means that the default field rate is 60/1.001 fields/sec, not 60 fields/sec.

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 30/1.001Hz or 25Hz for 1920 x 1080 and 1440 x 1080 video sources, and 60/1.001Hz or 50Hz (for 1280 x 720 video sources).
- The pixel aspect ratio is 1:1 (aspect ratio index 1, SMPTE ST 2042-1, Table 11.4) or 4:3 for 1440 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.

5.2.1.3 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 with the single exception as follows:

- Source video formats with 1440 horizontal pixels shall use base video format 11 or 12 and shall override the frame width to be 1440 as specified in SMPTE ST 2042-1, Section 11.3.2.

5.2.1.4 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.

5.2.2 Picture Syntax

The base video format and source parameters shall be static values over the VC-2 sequence. The syntax constraints (Section 6) 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 fields for interlaced video sources and frames for progressive video sources.

5.2.3 Transform Parameters

The transform parameters shall be as defined below.

5.2.3.1 Wavelet filter

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

5.2.3.2 Transform depth

The discrete wavelet transform (DWT) depth, as defined in SMPTE ST 2042-1, Section 12.3.2, shall be 3.

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

5.2.3.3 Slice Coding Parameters

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

Note: 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 vertically lines divided by 8 rounded up to the nearest integer. That is the number of vertical slices per picture shall be 68 for 1080 line interlaced source video and 90 for 720 line progressive source video.

Notes:

1. This means that each individual slice notionally corresponds to 8 input picture lines (given the ripple effects of the wavelet filter).
2. The number of vertical slices is defined per picture, which is a field for interlaced source video and a frame for progressive source video.
3. 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 29.97 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 may 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: 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 29.97 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, Annex E, Table E.2.

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		1440 x 1080		1280 x 720	
Frame Rate	29.97 Hz	25 Hz	29.97 Hz	25 Hz	29.97 Hz	25 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

5.3 Entropy Coding

All entropy coded data shall comply with requirements of the low-delay profile. Thus the entropy coding methods will be as defined in SMPTE ST 2042-1, Section 9.1.2.1 and Annex A.3.

Note: Only interleaved Exp-Golomb coding is used in the low-delay syntax. There is no arithmetic coding.

6 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.
- 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: 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 in Figure 1.

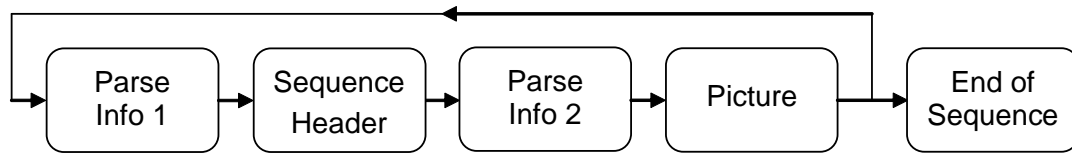


Figure 1 – Constrained Stream Syntax

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 as indicated in Figure 2.

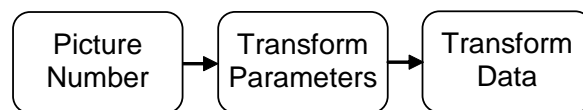


Figure 2 – Picture Syntax

The string of bytes defining parse info, sequence header, transform parameters and end of sequence shall be constant within a sequence 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 5.2.2). 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)

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 practice.

Note: 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 six video formats defined in Section 5. 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
1920x1080/60/I	1080	1920	29.97	interlace
1920x1080/50/I	1080	1920	25	interlace
1440x1080/60/I	1080	1440	29.97	interlace
1440x1080/50/I	1080	1440	25	interlace
1280x720/60/P	720	1280	59.94	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 right most bits in the hexadecimal number.

A.1 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

Format Name	Next Parse Offset (Decimal)	Next Parse Offset (Hex string)	Number of Bytes
1920x1080/60/I	18	0x00000012	4
1920x1080/50/I	18	0x00000012	4
1440x1080/60/I	23	0x00000017	4
1440x1080/50/I	23	0x00000017	4
1280x720/60/P	18	0x00000012	4
1280x720/50/P	18	0x00000012	4

Note: 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 below for all other parse infos.

Format Name	Previous Parse Offset (Decimal)	Previous Parse Offset (Hex string)	Number of Bytes
1920x1080/60/I	349945	0x000556F9	4
1920x1080/50/I	414745	0x00065419	4
1440x1080/60/I	349945	0x000556F9	4
1440x1080/50/I	414746	0x0006541A	4
1280x720/60/P	349944	0x000556F8	4
1280x720/50/P	414745	0x00065419	4

Note: 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 below, except for the first parse info in a sequence in which the last 4 bytes (Previous Parse Offset) are set to zero.

Table A.1 – Parse Info 1 Byte Sequence

Format Name	Parse Info 1 Byte String	Number of Bytes
1920x1080/60/I	0x424243440000000012000556F9	13
1920x1080/50/I	0x42424344000000001200065419	13
1440x1080/60/I	0x424243440000000017000556F9	13
1440x1080/50/I	0x4242434400000000170006541A	13
1280x720/60/P	0x424243440000000012000556F8	13
1280x720/50/P	0x42424344000000001200065419	13

A.2 Sequence Header

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.2.1 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 = 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.2.2 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
1920x1080/60/I	11	0x21	7
1920x1080/50/I	12	0x23	7
1440x1080/60/I	11	0x21	7
1440x1080/50/I	12	0x23	7
1280x720/60/P	9	0x09	7
1280x720/50/P	10	0x0B	7

A.2.3 Source Parameters

Source parameters describe the source video format; i.e., 1080I 50, 59.94 or 720P 50 or 59.94. With the exception of the frame rate of 1080P60 the source parameters correspond to the default parameters defined in SMPTE ST 2042-1 and so are not over-ridden.

- Custom dimensions flag
 - 1920x1080 & 1280x720: Value = False, VLC Code = 0b0, Bits = 1
 - 1440x1080: Value = True, VLC Code = 0b1, Bits = 1
 - frame width Value = 1440, VLC Code = 0x028803 Bits = 21
 - frame height Value = 1080, VLC Code = 0x000A83 Bits = 21
- 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

Format Name	Source Parameters Bit String	Number of Bits
1920x1080/60/I	0x00	8
1920x1080/50/I	0x00	8
1440x1080/60/I	0x2288030054180	50
1440x1080/50/I	0x2288030054180	50
1280x720/60/P	0x00	8
1280x720/50/P	0x00	8

A.2.4 Picture Coding Mode

The picture coding mode indicates whether the frames are coded as fields or frames. In this 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.2.5 Sequence Header Bytes

This section enumerates the byte sequence obtained by concatenating the component parts enumerated above and padding with zero (LSBs) 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 below.

Table A.2 – Sequence Header

Format Name	Sequence Header (Byte String)	Number of Bytes
1920x1080/60/I	0x3802508010	5
1920x1080/50/I	0x3802508010	5
1440x1080/60/I	0x380250C5100600A83004	10
1440x1080/50/I	0x380250C5100600A83004	10
1280x720/60/P	0x3802508040	5
1280x720/50/P	0x3802508040	5

A.3 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 syntax picture)
- Next parse offset is as indicated in the table below:

Format Name	Next Parse Offset (Decimal)	Next Parse Offset (Hex string)	Number of Bytes
1920x1080/60/I	349945	0x000556F9	4
1920x1080/50/I	414745	0x00065419	4
1440x1080/60/I	349945	0x000556F9	4
1440x1080/50/I	414746	0x0006541A	4
1280x720/60/P	349944	0x000556F8	4
1280x720/50/P	414745	0x00065419	4

Note: 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), see below.

- Previous Parse Offset is as indicated in the table below:

Format Name	Previous Parse Offset (Decimal)	Previous Parse Offset (Hex string)	Number of Bytes
1920x1080/60/I	18	0x00000012	4
1920x1080/50/I	18	0x00000012	4
1440x1080/60/I	23	0x00000017	4
1440x1080/50/I	23	0x00000017	4
1280x720/60/P	18	0x00000012	4
1280x720/50/P	18	0x00000012	4

Note: 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 subsection.

Table A.3 – Parse Info 2 Byte Sequence

Format Name	Parse Info 2 Bytes	Number of Bytes
1920x1080/60/I	0x42424344C8000556F900000012	13
1920x1080/50/I	0x42424344C80006541900000012	13
1440x1080/60/I	0x42424344C8000556F900000017	13
1440x1080/50/I	0x42424344C80006541A00000017	13
1280x720/60/P	0x42424344C8000556F800000012	13
1280x720/50/P	0x42424344C80006541900000012	13

A.4 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 shall 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.5 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 the table below:

Format Name	Slices X	Slices Y	Slice Bytes Numerator	Slice Bytes Denominator
	VLC / Bits	VLC / Bits	VLC / Bits	VLC / Bits
1920x1080/60/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
1440x1080/60/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
1280x720/60/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

- Custom Quantization Matrix: Value = False Code = 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.

Table A.4 – Transform Parameters Byte Sequence

Format Name	Transform Parameters Bytes	Number of Bytes
1920x1080/60/I	0x21541808C5144824	8
1920x1080/50/I	0x21541808D1401824	8
1440x1080/60/I	0x21541808D5051824	8
1440x1080/50/I	0x21541808C100060900	9
1280x720/60/P	0x211018A2D510A4	7
1280x720/50/P	0x211018A2C1006900	8

A.6 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 349920 bytes (all three 60Hz formats) or 414720 (all three 50Hz formats).

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

A.7 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:

Format Name	Previous Parse Offset (Decimal)	Previous Parse Offset (Hexadecimal)
1920x1080/60/I	349945	0x000556F9
1920x1080/50/I	414745	0x00065419
1440x1080/60/I	349945	0x000556F9
1440x1080/50/I	414746	0x0006541A
1280x720/60/P	349944	0x000556F8
1280x720/50/P	414745	0x00065419

Note: 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:

Table A.5 – End Of Sequence Bytes

Format Name	End of Sequence Bytes
1920x1080/60/I	0x424243441000000000000556F9
1920x1080/50/I	0x42424344100000000000065419
1440x1080/60/I	0x424243441000000000000556F9
1440x1080/50/I	0x4242434410000000000006541A
1280x720/60/P	0x424243441000000000000556F8
1280x720/50/P	0x42424344100000000000065419

Annex B Filter Coefficients for Conversion between 1920 and 1440 Pixels (Informative)

This 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 pixels 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 pixels format throughout a concatenated compression system and up-converting to 1920 pixels only for display or processing.

This 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 down-conversion and up-conversion 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 below to 1920/1440 conversion for the luminance component apply equally to the 960/720 conversion for the color-difference components.

B.1 1920 to 1440 Conversion (Down-conversion)

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 section offers the following filter coefficients for down-converting from 1920 to 1440 luminance pixels and 960 to 720 color-difference pixels.

Table B.1 – Down-conversion 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 the table above. 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 down-conversion 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 down-conversion process, using the filter coefficients above, 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 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 as required 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.2 1440 to 1920 Conversion (Up-conversion)

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 section offers the following filter coefficients for up-converting from 1440 to 1920 luminance pixels and 720 to 960 color-difference pixels.

Annex B.2 – Up-conversion 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 the table above. 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 up-conversion 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 up-conversion process, using the filter coefficients above, 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 need 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 as required 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.

Annex C Bibliography (Informative)

SMPTE ST 2042-2:2009, 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 (12/07), 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

ITU-R BT.709-5 (04/02), Parameter Values for the HDTV Standards for Production and International Programme Exchange