

# SMPTE RECOMMENDED PRACTICE

## VC-2 Mezzanine Level Compression of 1080P High Definition Video Sources



<b>Table of Contents</b>	<b>Page</b>
Foreword .....	3
Intellectual Property .....	3
Introduction.....	3
1 Scope .....	4
2 Conformance Notation .....	4
3 Normative References .....	4
4 General Description .....	5
4.1 Acronyms and Terms .....	5
5 Coding Constraints.....	5
5.1 Video Source Parameters.....	5
5.2 Coding Parameters .....	5
5.2.1 Sequence Header .....	6
5.2.1.1 Parse parameters .....	6
5.2.1.2 Base video format .....	6
5.2.1.3 Source parameters .....	6
5.2.1.4 Picture Coding Mode .....	6
5.2.2 Picture Syntax.....	6
5.2.3 Transform Parameters .....	7
5.2.3.1 Wavelet filter .....	7
5.2.3.2 Transform depth.....	7
5.2.3.3 Slice Coding Parameters .....	7
5.3 Entropy Coding .....	7
6 Syntax Constraints .....	7

- Annex A Byte Values of Stream Components (Informative) ..... 9
  - A.1 Parse Info 1 ..... 9
  - A.2 Sequence Header ..... 9
    - A.2.1 Parse Parameters ..... 9
    - A.2.2 Base Video Format ..... 10
    - A.2.3 Source Parameters ..... 10
    - A.2.4 Picture Coding Mode ..... 10
    - A.2.5 Sequence Header Bytes ..... 10
  - A.3 Parse Info 2 ..... 11
  - A.4 Picture Number ..... 11
  - A.5 Transform Parameters ..... 11
  - A.6 Transform Data ..... 12
  - A.7 End of Sequence ..... 12
- Annex B Bibliography (Informative) ..... 13

## 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 2047-1 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 Recommended Practice. 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 recommended practice defines constraints on the VC-2 specification (SMPTE ST 2042-1) such that source video images of 1920 x 1080 pixels using 4:2:2 sampling at 50, 60 and 60/1.001 frames per second will be compressed by a factor of approximately 2.5 to 1. The bit rate of the resulting VC-2 stream is such that it can be carried over a single-link serial digital interface (SDI) operating at 1.485Gbps (1.485/1.001 in the case of the 60/1.001 frame rate). This documents defines constraints on the coding and syntax, it does not define a mapping to any specific transport layer.

## 1 Scope

This recommended practice defines constraints on the VC-2 video compression specification to provide a mezzanine level compression coding. It defines coding constraints such that source video of 1920 x 1080 pixels using 4:2:2 sampling at 50, 60 and 60/1.001 frames per second can be compressed by a factor of approximately 2.5 to 1 resulting in a compressed bit-rate that permits carriage over a single-link SDI operating at 1.5 Gb/s.

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 Reference

The following standards contain provisions which, through reference in this text, constitute provisions of this recommended practice. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this recommended practice are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

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

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

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

## 4 General Description

The VC-2 compression standard defines a compression system that uses Wavelet coding as the basis for image de-correlation.

The VC-2 compression standard defines 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 ensure the interoperability of VC-2 streams coded with mezzanine compression. It also constrains the stream syntax to ensure that coded streams may be mapped in an interoperable fashion onto a transport layer. The constraints defined here are signalled in the stream, as the specialized VC-2 level number 64 as indicated below in Section 5.2.1.1. That is, VC-2 streams conformant to this practice shall ensure compliance by specifying specialized level 64 in the stream. This practice does not define how the VC-2 stream is mapped to any particular transport layer.

Note: Such VC-2 stream transport specifications may be defined in other SMPTE documents such as SMPTE ST 2047-2 for the carriage of VC-2 Compressed Video over HD-SDI.

### 4.1 Acronyms and of 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

The Mezzanine compression process has the following key coding constraints:

- a) The codec uses only the low delay syntax.
- b) The video sources are limited to 1920 x 1080p with frame rates of 50 fps, 59.94 fps and 60 fps only.
- c) The codec uses only the Haar filter for simple implementation and to ensure that transformed samples remain bounded within each slice.

### 5.1 Video Source Parameters

The video sources shall be as defined by SMPTE 274M for the frame rates of 50 fps, 59.94 fps and 60 fps only.

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

The video sampling format is 4:2:2 as defined by SMPTE 274M.

Only the active picture area of 1920 x 1080 pixels for the luminance channels and 960 x 1080 pixels for each of the color difference channels shall be coded.

### 5.2 Coding Parameters

The codec shall use only the low delay syntax as defined by SMPTE ST 2042-1 (VC-2 Video Compression).

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.

### 5.2.1.1 Parse parameters

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

The profile value shall be Low Delay Profile as defined in SMPTE ST 2042-1.

The level value shall be a specialized level value with the value of 64. This level value shall be used only for the VC-2 coding constraints that comply with this recommended practice.

### 5.2.1.2 Base video format

The base video format shall be either format number 13 or 14 (HD 1080P-60 or HD 1080P-50) as defined in SMPTE ST 2042-1.

Note: The format is thus defined as follows:

- The picture size is 1920 x 1080
- The sampling format is 4:2:2
- The scanning is progressive
- The frame rate is 60/1.001 or 50 (see below for 60-Hz frame rate)
- The pixel aspect ratio is 1:1
- The clean picture area is the full 1920 x 1080 pixels
- 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 part of the VC-2 specification allows 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 that follows:

For a video source with a 60-Hz frame rate (rather than 60/1.001 Hz), the base video format shall be 13 (HD 1080P-60) and the frame rate shall be overridden by the frame rate index value 8 (60/1) as defined in SMPTE ST 2042-1, Section 11.3.5.

### 5.2.1.4 Picture Coding Mode

Picture coding mode shall be 0, which means that VC-2 pictures are frames and thus the wavelet transform operates on each whole frame.

## 5.2.2 Picture Syntax

The base video format and source parameters shall be static values over the VC-2 sequence. Thus, in the initialization process, all the sequence and source parameters can be accessed for each picture in the sequence.

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

### 5.2.3 Transform Parameters

The transform parameters shall be as defined below.

#### 5.2.3.1 Wavelet filter

The wavelet filter shall be “Haar with single shift per level” filter (Wavelet index value = 4) as defined in SMPTE ST 2042-1.

#### 5.2.3.2 Transform depth

The transform depth shall be 2 as defined in SMPTE ST 2042-1.

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 frame (slices x) shall be 120. This means that each slice corresponds to 16 input pixels for luminance.

The number of slices vertically per frame (slices y) shall be 270. This means that each slice corresponds to 4 input picture lines.

Note: Therefore each picture contains a total of 32,400 slices.

The number of bytes allocated to each slice (slice bytes) shall be 64. That is “slice bytes numerator” divided by slice bytes denominator shall be 64, where slice bytes numerator and denominator are as defined in SMPTE ST 2042-1.

The quantisation matrix shall be the default quantisation matrix corresponding to the Haar filter with single shift as defined in SMPTE ST 2042-1, Annex E, Table E.5.

### 5.3 Entropy Coding

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

Note: Only interleaved Exp-Golomb coding is used in the low-delay syntax.

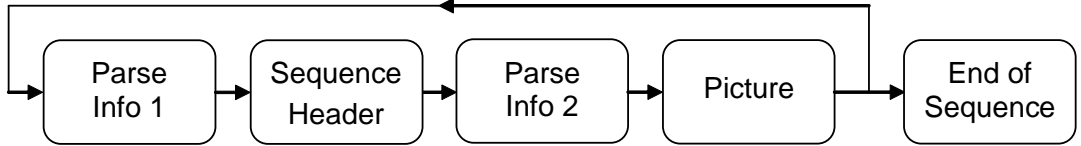
## 6 Syntax Constraints

The syntax of the VC-2 bit-stream shall be constrained as defined in this section to ensure interoperability and 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 shall be signaled with “slice bytes numerator” = 64 and “slice bytes denominator” = 1
- Bits which are otherwise undefined shall be assigned the value zero.

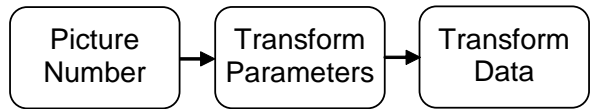
Note: Such bits are used to pad partially filled bytes to occupy a whole byte.

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 sequence header and picture are defined in SMPTE ST 2042-1. For clarity the structure of the picture is as indicated below.



**Figure 2 – Picture Syntax**

The string of bytes defining, parse info, sequence header, transform parameters and end of sequence are constant 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: 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.

In this annex the notation 0x..... represents a hexadecimal literal and 0b..... represents a binary literal.

### 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 0x42, 0x42, 0x43, 0x44
- Parse code is 0x00 (sequence header)
- Next parse offset is
  - 0x00, 0x00, 0x00, 0x12 for 1080P50/59.94 (i.e., 18 bytes until the next parse info)
  - 0x00, 0x00, 0x00, 0x13 for 1080P60 (i.e., 19 bytes until the next parse info)
- Previous parse offset is
  - 0x00, 0x00, 0x00, 0x00 for the first frame (i.e., there is no previous parse info)
  - 0x00, 0x1F, 0xA4, 0x18 for other frames

Note: Previous parse offset is the sum of the number of bytes for picture (2073611 bytes) plus the number of bytes for parse info 2 (13 bytes); i.e., a total of 2073624 bytes.

Parse info 1 therefore occupies 13 bytes which are:

- First Frame:  
0x42, 0x42, 0x43, 0x44, 0x00, 0x00, 0x00, 0x00, 0x12, 0x00, 0x00, 0x00, 0x00
- Other frames at 1080P50/59.94:  
0x42, 0x42, 0x43, 0x44, 0x00, 0x00, 0x00, 0x00, 0x12, 0x00, 0x1F, 0xA4, 0x18
- Other frames at 1080P60:  
0x42, 0x42, 0x43, 0x44, 0x00, 0x00, 0x00, 0x00, 0x13, 0x00, 0x1F, 0xA4, 0x18

### 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 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 = 64,            | VLC = 0b0000000000011, | Bits = 13 |

Total number of bits = 18

### A.2.2 Base Video Format

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

- 1080P59.94 & 60: Value = 13 (HD 1080P-60), VLC = 0b0101001, Bits = 7
- 1080P50: Value = 14 (HD 1080P-50), VLC = 0b0101011, Bits = 7

Total number of bits = 7

### A.2.3 Source Parameters

Source parameters describe the source video format; i.e., 1080P 50, 59.94 or 60. 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: Value = False, Code = 0b0, Bits = 1
- 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
  - 1080P50/59.94: Value = False, Code = 0b0, Bits = 1
  - 1080P60: Value = True, Code = 0b1, Bits = 1
    - frame rate index: Value = 8, VLC = 0b0000011, Bits = 7
- Custom dimensions flag: Value = False, Code = 0b0, Bits = 1
- Custom dimensions flag: Value = False, Code = 0b0, Bits = 1
- Custom dimensions flag: Value = False, Code = 0b0, Bits = 1
- Custom dimensions flag: Value = False, Code = 0b0, Bits = 1

Total number of bits:

1080P50/59.94	= 8
1080P60	= 15

### 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: Value = 0, VLC = 0b1, Bits = 1

Total number of bits = 1

### A.2.5 Sequence Header Bytes

This section enumerates the byte sequence obtained by concatenating the component parts enumerated above.

Concatenating the VLCs and codes above gives:

- 1080P59.94:           0b00111000,00000001,11010100,10000000,01
- 1080P50:             0b00111000,00000001,11010101,10000000,01
- 1080P60:             0b00111000,00000001,11010100,10001000,00110000,1

Padding with zeros to produce a whole number of bytes, and converting to hexadecimal gives:

Sequence Header (1080P59.94):	0x38, 0x01, 0xD4, 0x80, 0x40
Sequence Header (1080P50):	0x38, 0x01, 0xD5, 0x80, 0x40
Sequence Header (1080P60):	0x38, 0x01, 0xD4, 0x88, 0x30, 0x80

### 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 0x42, 0x42, 0x43, 0x44
- Parse code is 0xC8 (low delay syntax picture)
- Next parse offset is:    0x00, 0x1F, 0xA4, 0x18

Note: This is the sum of the number of bytes for parse info 2 (13 bytes) plus the number of bytes for the picture (2073611 bytes); i.e., a total of 2073624 bytes.

- Previous Parse Offset:
  - 0x00, 0x00, 0x00, 0x12 for 1080P50/59.94 (i.e., 18 bytes back to the previous parse info)
  - 0x00, 0x00, 0x00, 0x13 for 1080P60 (i.e., 19 bytes back to the previous parse info)

Note: This is the sum of the number of bytes for sequence header (5 or 6 bytes) plus the number of bytes for parse info 1 (13 bytes); i.e., a total of 18 or 19 bytes.

Parse info 2 therefore occupies 13 bytes which are:

- 1080P50 & 59.94:  
0x42, 0x42, 0x43, 0x44, 0xC8, 0x00, 0x1F, 0xA4, 0x18, 0x00, 0x00, 0x00, 0x12
- 1080P60:  
0x42, 0x42, 0x43, 0x44, 0xC8, 0x00, 0x1F, 0xA4, 0x18, 0x00, 0x00, 0x00, 0x13

### A.4 Picture Number

The picture number is the picture header 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 = 4 (Haar with single shift)      VLC = 0b00011      Bits = 5
- DWT Depth:      Value = 2 (2 level transform)      VLC = 0b011      Bits = 3
- Slice Parameters
  - Slices X:      Value = 120 (horizontal slices)      VLC = 0b0101010000011      Bits = 13
  - Slices Y:      Value = 270 (vertical slices)      VLC = 0b00000000010101011      Bits = 17
  - Slice Bytes Numerator:      Value = 64      VLC = 0b0000000000011      Bits = 13
  - Slice Bytes Denominator:      Value = 1      VLC = 0b001      Bits = 3
- Custom Quantization Matrix:      Value = False      Code = 0b0      Bits = 1

Total number of bits = 55

Concatenated bits:      0b00011011, 01010100,00011000,00000010,10101100,00000000,0110010

The transform parameters occupy 7 bytes which are:

0x1B, 0x54, 0x18, 0x02, 0xAC, 0x00, 0x64

### 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 2073600 bytes.

Note: this figure is calculated as 32,400 slices x 64 bytes and is 2/5 the number of bits in an uncompressed frame.

### 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 0x42, 0x42, 0x43, 0x44
- Parse code is 0x10 (end of sequence)
- Next parse offset is 0x00, 0x00, 0x00, 0x00 (no next parse offset)
- Previous parse offset is 0x00, 0x1F, 0xA4, 0x18

Note: Previous parse offset is the sum of the number of bytes for picture (2073611 bytes) plus the number of bytes for parse info 2 (13 bytes); i.e., a total of 2073624 bytes.

End of sequence therefore occupies 13 bytes which are:

0x42, 0x42, 0x43, 0x44, 0x10, 0x00, 0x00, 0x00, 0x00, 0x00, 0x1F, 0xA4, 0x18

**Annex B Bibliography (Informative)**

SMPTE 372-2009, Dual Link 1.5 Gb/s Digital Interface for 1920 x 1080 and 2048 x 1080 Picture Formats

SMPTE 424M-2006, Television – 3 Gb/s Signal/Data Serial Interface

SMPTE ST 2047-2:2010, Carriage of VC-2 Compressed Video over HD-SDI