SMPTE RP 2047-1:2023
Revision of SMPTE RP 2047-1:2009

# SMPTE RECOMMENDED PRACTICE

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



## 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.

This Recommended Practice defines constraints on the VC-2 standard (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 1.485 Gbps (1.485/1.001 in the case of the 60/1.001 frame rate). This document defines constraints on the coding and syntax, it does not define a mapping to any specific transport layer.

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 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 bitrate that permits carriage over a single-link SDI operating at 1.5 Gb/s. 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 standard. Dated references require that the specific edition cited shall be used as the reference. Undated citations refer to the edition of the referenced document (including any amendments) current at the date of publication of this document. All 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 × 1080 Image Sample Structure, Digital Representation and Digital Timing Reference Sequences for Multiple Picture Rates

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

# 3 Terms and definitions

For the purposes of this document, the terms and definitions given in 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 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

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 ensure the interoperability of VC-2 streams coded with mezzanine compression. 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 VC-2 level number 64, as indicated in Clause 5.3.2.2.

This Recommended Practice does not define how the VC-2 stream is mapped to any particular transport layer. Transport over HD-SDI is specified in SMPTE ST 2047-2.

# 5 Coding constraints

## 5.1 Summary (informative)

This application has the following key coding constraints:

- The video sources are limited to 1920 x 1080p with frame rates of 50 fps, 60/1.001 fps and 60 fps only.
- The codec uses only the Haar filter for simple implementation and to ensure that transformed samples remain bounded within each slice.

## 5.2 Video source parameters

The video sources shall be as defined by SMPTE ST 274 for the frame rates of 50 fps, 60/1.001 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 ST 274.

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.

## 5.3 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.

#### 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 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.

#### 5.3.2.3 Base video format

The base video format index shall be either 13 or 14 (HD 1080p60 or HD 1080p50) 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 Clause 5.3.2.4 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

#### 5.3.2.4 Source parameters

The source parameters part of the VC-2 standard allows 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 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 1080p60) and the frame rate shall be overridden by the frame rate index value 8 (60/1) as defined in SMPTE ST 2042-1:2022, Clause 11.4.6.

#### 5.3.2.5 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.

### 5.3.3 Picture syntax

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

### 5.3.4 Transform parameters

#### 5.3.4.1 Wavelet filter

The wavelet filter shall be “Haar with single shift per level” filter (Wavelet index value = 4) as defined in SMPTE ST 2042-1:2022 Clause 12.4.2.

#### 5.3.4.2 Transform depth

The transform depth shall be 2 as defined in SMPTE ST 2042-1:2022, Clause 12.4.3.

#### 5.3.4.3 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 quantization matrix shall be the default quantization matrix corresponding to the Haar filter with single shift as defined in SMPTE ST 2042-1:2022 Table D.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 ensure interoperability and 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 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 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.

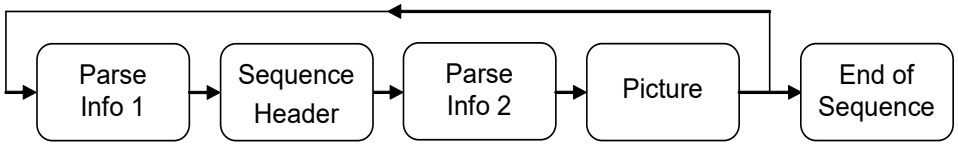


**Figure 1 — Constrained stream syntax.**

The structure of the sequence header and picture are defined in SMPTE ST 2042-1. For clarity, the structure of the picture is shown in Figure 2.

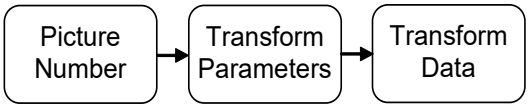


**Figure 2 — picture syntax.**

The string of bytes defining parse info, sequence header, transform parameters and end of sequence is constant for streams that conform to this Recommended Practice. 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 frame 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 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.

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

Video system names use the notation recommended in ITU-R BT.1846.

## 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 0x42, 0x42, 0x43, 0x44
- Parse code is 0x00 (sequence header)
- Next parse offset is:
    - 0x00, 0x00, 0x00, 0x12 for frame rates of 50 fps and 60/1.001 fps (i.e., 18 bytes until the next parse info)
    - 0x00, 0x00, 0x00, 0x13 for frame rate of 60 fps (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

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 frame rates of 50 fps and 60/1.001 fps:
  0x42, 0x42, 0x43, 0x44, 0x00, 0x00, 0x00, 0x00, 0x12, 0x00, 0x1F, 0xA4, 0x18
- Other frames at frame rate of 60 fps:
  0x42, 0x42, 0x43, 0x44, 0x00, 0x00, 0x00, 0x00, 0x13, 0x00, 0x1F, 0xA4, 0x18

## 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 defined in SMPTE ST 2042-1 and as outlined in A.3.2, A.3.3, A.3.4 and A.3.5.

Each of the values in the sequence header is coded using the variable length code (abbreviated VLC) defined in 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 = 64, VLC = 0b0000000000011, Bits = 13

Total number of bits = 18

### A.3.3 Base video format

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

- 1080/59.94/P: Value = 13 (HD 1080p60), VLC = 0b0101001, Bits = 7
- 1080/60/P: Value = 13 (HD 1080p60), VLC = 0b0101001, Bits = 7
- 1080/50/P: Value = 14 (HD 1080p50), VLC = 0b0101011, Bits = 7

Total number of bits = 7

### A.3.4 Source parameters

Source parameters describe the source video format; i.e., 1080/50/P, 1080/59.94/P or 1080/60/P. With the exception of the frame rate of 1080/60/P, the source parameters correspond to the default parameters defined in SMPTE ST 2042-1 and so are not overridden.

- 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:
    - 1080/50/P: Value = False, Code = 0b0, Bits = 1
    - 1080/59.94/P: Value = False, Code = 0b0, Bits = 1
    - 1080/60/P: 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:

| | |
|---|---|
| 1080/50/P | = 8 |
| 1080/59.94/P | = 8 |
| 1080/60/P | = 15 |

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

Total number of bits = 1

### A.3.6 Sequence header bytes

This clause enumerates the byte sequence obtained by concatenating the component parts enumerated in A.3.2, A.3.3, A.3.4 and A.3.5.

Concatenating the VLCs and codes gives:

- 1080/59.94/P: 0b00111000,00000001,11010100,10000000,01
- 1080/50/P: 0b00111000,00000001,11010101,10000000,01
- 1080/60/P: 0b00111000,00000001,11010100,10001000,00110000,1

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

Sequence Header (1080/59.94/P): 0x38, 0x01, 0xD4, 0x80, 0x40

Sequence Header (1080/50/P): 0x38, 0x01, 0xD5, 0x80, 0x40

Sequence Header (1080/60/P): 0x38, 0x01, 0xD4, 0x88, 0x30, 0x80

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

  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 is:
  - 0x00, 0x00, 0x00, 0x12 for 1080/50/P and 1080/59.94/P (i.e., 18 bytes back to the previous parse info)
  - 0x00, 0x00, 0x00, 0x13 for 1080/60/P (i.e., 19 bytes back to the previous parse info)

  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:

- 1080/50/P & 1080/59.94/P:

  0x42, 0x42, 0x43, 0x44, 0xC8, 0x00, 0x1F, 0xA4, 0x18, 0x00, 0x00, 0x00, 0x12

- 1080/60/P:

  0x42, 0x42, 0x43, 0x44, 0xC8, 0x00, 0x1F, 0xA4, 0x18, 0x00, 0x00, 0x00, 0x13

## A.5 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 picture numbers increment by one for each successive frame.

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

## A.6 Transform parameters

The transform parameters describe the details of the wavelet transform that is used and the way that 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.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 2073600 bytes.

This figure is calculated as 32400 slices x 64 bytes and is 2/5 the number of bits in an uncompressed frame.

## 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 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

  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

# Bibliography

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

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