SMPTE ST 2047-2:2010

SMPTE STANDARD

Carriage of VC-2 Compressed Video over HD-SDI

Table of Contents

Approved
January 12. 2010

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 ST 2047-2 was prepared by Technology Committee 32NF.

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

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

The VC-2 compression standard defines a generic intra-frame compression system that can handle all defined video formats and more. A companion recommended practice defines constraints on the VC-2 specification such that source video images of 1920 × 1080 pixels using 4:2:2 sampling at 50, 60, and 60/1.001 frames per second can be compressed by a ratio of approximately 5 to 2 resulting in a compressed bit-rate that can be carried over a single-link HD SDI operating at 1.5 Gb/s.

This standard defines the provisions for mapping constrained VC-2 bit-streams onto the 1.5 Gb/s HD SDI link and describes a method by which additional bits are added to ensure that the resultant signal can be viewed on a HD SDI monitor, albeit in a somewhat noisy form.

A key feature of this specification is that the transport of compressed data in this way is backward compatible with installed equipment. The compressed video is mapped entirely into active picture. The signal on the link therefore comprises a valid, viewable, video signal. This ensures compatibility with existing equipment such routing and monitoring apparatus.

An important feature of this mapping process is that it can be implemented in a practical form using currently available firmware and hardware, and, in the course of time, could be implemented within new chip designs to ensure the minimum impact on hardware power requirements and physical size.

To aid readability, the serial digital bit rates are abbreviated throughout this document to 1.5 Gb/s in the cases of the bit rates 1.485 Gb/s (for 50 and 60 frames per second) and 1.485/1.001 Gb/s (for 60/1.001 frames per second), and to 3 Gb/s in the cases of the bit rates 2.97 Gb/s (for 50 and 60 frames per second) and 2.97/1.001 Gb/s (for 60/1.001 frames per second).

1 Scope

This standard defines a means of mapping VC-2 data onto a single HD SDI link, based upon the SMPTE 292 data structure, to support the transport of 1920 x 1080/50/P, 1920 x 1080/59.94/P (60/1.001), and 1920 x 1080/60/P systems with 4:2:2 $Y'C'_BC'_R$ color encoding. As these systems require a raw uncompressed data capacity twice that of SMPTE 292, VC-2 mezzanine compression as defined by SMPTE RP 2047-1 is required to reduce the data rate. This link supports the carriage of embedded audio, ancillary data and payload identification.

A signal coded in accordance with this standard incorporates a noisy but recognizable picture that can be viewed using a standard SMPTE 292 video monitor.

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 labelled as "Informative" or individual paragraphs that start with "Note:"

The keywords "shall" and "shall not" indicate requirements strictly to be followed in order to conform to the document and from which no deviation is permitted.

The keywords, "should" and "should not" indicate that, among several possibilities, one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required; or that (in the negative form) a certain possibility or course of action is deprecated but not prohibited.

The keywords "may" and "need not" indicate courses of action permissible within the limits of the document.
The keyword “reserved” indicates a provision that is not defined at this time, shall not be used, and may be defined in the future. The keyword “forbidden” indicates “reserved” and in addition indicates that the provision will never be defined in the future.

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

Unless otherwise specified the order of precedence of the types of normative information in this document shall be as follows. Normative prose shall be the authoritative definition. Tables shall be next, followed by formal languages, then figures, and then any other language forms.

3 Normative References

The following standards contain provisions 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 12M-2-2008, Television — Transmission of Time Code in the Ancillary Data Space

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

SMPTE 291M-2006, Television — Ancillary Data Packet and Space Formatting

SMPTE 292-2008, 1.5 Gb/s Signal/Data Serial Interface

SMPTE 299-2009, 24-Bit Digital Audio Format for SMPTE 292 Bit-Serial Interface

SMPTE 352M-2002, Television (Dynamic) — Video Payload Identification for Digital Interfaces

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

SMPTE RP 2047-1:2009, VC-2 Mezzanine Level Compression of 1080P High Definition Video Sources

4 General Description

4.1 Source Image Formats

The data for this interface shall be derived from the source image formats identified in Table 1.

Table 1 – Source image formats

SMPTE 274 System No.	System nomenclature	Frame rate (Hz)	Scanning format	Signal format sampling structure/pixel depth
1	1920 × 1080/60/P	60	Progressive	4:2:2 ($Y'C'_BC'_R$) / 10-bit
2	1920 x 1080/59.94/P	60/1.001	Progressive	4:2:2 ($Y'C'_BC'_R$) / 10-bit
3	1920 x 1080/50/P	50	Progressive	4:2:2 ($Y'C'_BC'_R$) / 10-bit

4.2 Timing Reference Codes

Timing Reference codes shall be in conformance with those specified in SMPTE 274M and SMPTE 292 for the corresponding interlaced system (1920 x 1080/60/I, 1920 x 1080/59.94/I or 1920 x 1080/50/I).

4.3 Ancillary Data

Ancillary data, if present, shall be mapped into the blanking area in conformance with SMPTE 291M.

4.4 Audio Data

Audio data, if present, shall be mapped into the ancillary data space in conformance with SMPTE 299.

4.5 Time Code

Time code, if present, shall be mapped into the ancillary data space in conformance with SMPTE 12M-2.

4.6 Video Payload Identifier

The payload identifier shall be present and shall be mapped into the ancillary data space in conformance with SMPTE 352M.

The first byte of the payload identifier shall be used to identify the combination of video payload format and digital interface transport. For an interface in accordance with this standard Byte 1 shall be set to $B0_h$.

Bytes 2 and 3 of the payload identifier shall be set in accordance with the picture rate, sampling structure, and aspect ratio etc of the viewable image being carried on the interface as shown in Table 2.

Byte 4 of the payload identifier shall be used to identify the compressed payload. The meaning of byte 4 shall depend on the values of bytes 2 and 3 as shown in Table 3.

Table 2 – Payload Identifier Definition for SMPTE 352M

<table>
<tr><th>Bits</th><th>Byte 1</th><th>Byte 2</th><th>Byte 3</th><th>Byte 4</th></tr>
<tr><td>Bit 7</td><td>1</td><td>0 (Interlaced transport)</td><td>1
Image aspect ratio: 16:9</td><td>0</td></tr>
<tr><td>Bit 6</td><td>0</td><td>0 (Interlaced picture)</td><td>Reserved</td><td>0</td></tr>
<tr><td>Bit 5</td><td>1</td><td rowspan="2">00 (VC-2 low-delay syntax compression)</td><td>Reserved</td><td>0</td></tr>
<tr><td>Bit 4</td><td>1</td><td>Reserved</td><td>0</td></tr>
<tr><td>Bit 3</td><td>0</td><td rowspan="4">Picture Rate
(see SMPTE 352M Table 2)</td><td rowspan="4">0h
(Sampling structure: 4:2:2 Y/Cb/Cr)</td><td>0</td></tr>
<tr><td>Bit 2</td><td>0</td><td>0</td></tr>
<tr><td>Bit 1</td><td>0</td><td>0</td></tr>
<tr><td>Bit 0</td><td>0</td><td>1</td></tr>
</table>

Table 3 – Compressed Payload Formats

Byte 2	Byte 3	Byte 4	SMPTE 274M System no.	System nomenclature
00001011 (Picture rate = 60Hz)	10000000	00000001	1	1920 × 1080p @ 60
00001010h (Picture rate = 60/1.001 Hz)	10000000	00000001	2	1920 x 1080p @ 59.94
00001001h (Picture rate = 50Hz)	10000000	00000001	3	1920 x 1080p @ 50

Byte 2, bits 4 and 5 shall be set to 00 to indicate that VC-2 low delay syntax is used for the video compression data.part of the signal.

Byte 4 shall be set to 01h to indicate the carriage of the VC-2 mezzanine compressed payload and the viewable video data according to this standard.

4.7 Viewable Content

A signal coded in accordance with this interface will produce a recognizable version of the compressed payload (although noisy and coarsely quantised) when decoded and displayed by a SMPTE 292 compliant receiving interface, as if it were a standard 1920 x 1080/60/I, 1920 x 1080/59.94/I, or 1920 x 1080/50/I interlaced signal.

5 VC-2 Mezzanine Bit-stream Mapping

This section defines which data from a VC-2 stream according to SMPTE RP 2047-1 is mapped into the SDI signal.

A compressed VC-2 stream comprises data and metadata. The metadata describes the parameters of the coded signal and the way in which it is encoded. The data describes the details of the specific video signal that has been compressed. When a VC-2 stream is transported over an SDI link, according to this specification, the stream metadata shall be inferred from the payload identifier (see Section 4.6). Therefore, it is only the data from the VC-2 stream that is mapped to the SDI link.

The signal data part of a VC-2 stream shall comprise the transform data part as defined in SMPTE ST 2042-1, VC-2 Video Compression. The transform data part of the VC-2 stream is defined in Sections 12.4 and 12.5 of the VC-2 specification.

The transform data for each VC-2 picture shall be mapped into exactly one interlaced picture area within the SDI signal. No metadata from the VC-2 stream shall be mapped into the SDI signal (except by inference from the payload identifier). Consecutive VC-2 pictures shall be mapped to consecutive SDI fields.

The complete VC-2 stream (with the exception of VC-2 “Picture Numbers” defined in Section 11.1 of the VC-2 specification) can be reconstructed from the “Transform Data” present on the SDI and the metadata inferred from the payload identifier.

VC-2 “Picture Numbers” provide unique identifiers for pictures within a sequence. Picture numbers within a VC-2 sequence are required to increment by one for each successive picture within a given sequence. Between sequences they are unconstrained and so picture numbers, starting from any arbitrary value, may be inserted in a decoded VC-2 stream while retaining compliance with the specification.

Decoding apparatus may decode the VC-2 stream implied as in the previous paragraph to produce an uncompressed video signal.

Note: The coding parameters, defined in SMPTE RP 2047-1, VC-2 Compression of 1080P High Definition Video Sources for transport over 1.5 Gb/s SDI, ensure that the compressed “Transform Data”, plus the bits added for the viewable picture occupy precisely the space allocated within the SDI signal for the active picture.

6 Interface Data Format

6.1 Compressed Data Mapping

The VC-2 mezzanine compressed video data shall be transported in the least-significant 8 bits of the (10-bit) video words as shown in Table 4. The first video word shall contain the first byte of the VC-2 “Transform Data”. Subsequent bytes of the “Transform Data” shall be transported in the 8 least-significant bits of consecutive video words, irrespective of whether the video words correspond to Cb, Y or Cr data. No compressed data shall be transported other than as the low 8 bits of video words.

Table 4 – Mapping of compressed video data to interface words

Word	Bit number									
	9 (MSB)	8	7	6	5	4	3	2	1	0 (LSB)
XYZ (C)	1	F	0	0	P3	P2	P1	P0	0	0
XYZ (Y)	1	F	0	0	P3	P2	P1	P0	0	0
Cb data			Byte 0 of VC-2 Transform Data							
Y data			Byte 1 of VC-2 Transform Data							
Cr data			Byte 2 of VC-2 Transform Data							
Y data			Byte 3 of VC-2 Transform Data							
Cb data			Byte 4 of VC-2 Transform Data							
Y data			Byte 5 of VC-2 Transform Data							
Cr data			Byte 6 of VC-2 Transform Data							
Y data			Byte 7 of VC-2 Transform Data							
•			•							
•			•							
•			•							

(MSB – Most Significant Bit, LSB – Least Significant Bit, MS – Most Significant, LS – Least Significant)

The shaded region, the 2 most-significant bits of each video word, contains the viewable picture information as described in Section 6.2 below.

Note: Since this specification uses the low delay syntax in accordance with SMPTE RP 2047-1, the 64 bytes of coefficient data for each slice are placed byte-sequentially along each line in the active video part of the SDI. In the absence of the viewable picture information, the coded slice data would appear visually as an array of 30 columns in the active picture area.

6.2 Viewable Picture

The viewable picture shall represent the contents of the compressed video. The two most-significant bits of the 10-bit interface words shall be used to carry a viewable version of the input picture, so that if the SDI signal is viewed as if it were standard SDI video the content will be recognizable, although noisy and coarsely quantised. This makes it possible to view the signal for the purpose of identification and to give confidence that the coder is working.

The encoding method used to generate the viewable image in the top 2 most-significant bits of each video word is not defined in this specification. Whatever encoding method is used, the video words shall not contain the prohibited synchronization codes as specified in SMPTE 274M and SMPTE 292 (000_h to 003_h and $3FC_h$ to $3FF_h$).

6.3 Creating a Viewable Picture

Generating the viewable picture may be achieved as follows:

a) A version of the compressed signal is generated corresponding to the format of the SDI signal. This is the unquantized viewable signal. For example, to generate a 1080i signal from a 1080P signal (where the field rate of the former corresponds to the frame rate of the latter), alternate lines may be taken from the progressive frame. The phase of the line pairing must be different on even fields and odd fields in order to achieve the required interlaced structure. Line averaging or other filtering operations may also be used to generate the interlaced signal.

b) A video signal, corresponding to the compressed data formatted as in the SDI signal with the two most-significant bits set to zero, is subtracted from the unquantized viewable signal. A two dimensional halftone dither signal is added to this difference signal. The result must be limited to a valid 10-bit range. The dither may consist of a repeating 8x8 pattern as follows:

Table 5 – 2D halftone dither

0	128	32	160	8	136	40	168
192	64	224	96	200	72	232	104
48	176	16	144	56	184	24	152
240	112	208	80	248	120	216	88
12	140	44	172	4	132	36	164
204	76	236	108	196	68	228	100
60	188	28	156	52	180	20	148
252	124	220	92	244	116	212	84

The dither is added independently to the luminance, Cb color-difference and Cr color-difference components, where the row selected is determined by the line number evaluated modulo-8, and the column selected is determined by the pixel number (e.g., 0-1919 for luminance, 0-959 for color-difference) evaluated modulo-8.

c) The least-significant 8-bits of the resulting video are discarded and replaced by the 8-bit values of the compressed data.

d) If the final 10-bit value corresponds to a prohibited synchronization code as specified in SMPTE 274M and SMPTE 292 (000_h to 003_h and $3FC_h$ to $3FF_h$), the most-significant two bits shall be modified as follows:

00 000000 xx is modified to 01 000000 xx

11 111111 yy is modified to 10 111111 yy

Annex A Effect of Halftone Dither (Informative)

The effect of the halftone dither can be seen from the following illustrations (magnified so the individual pixels are visible):

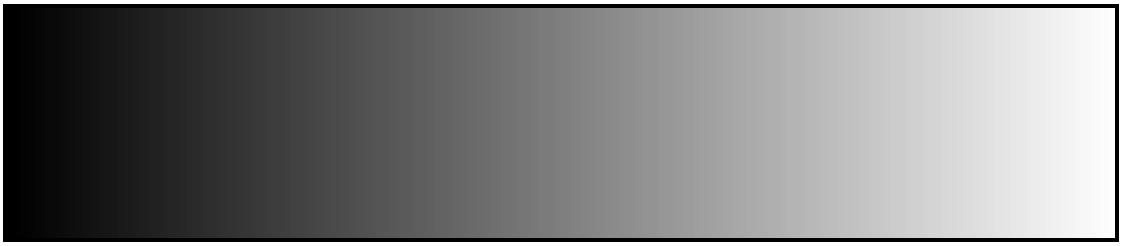

Figure A.1 – Original linear ramp

Figure A.2 – Ramp quantized to two bits

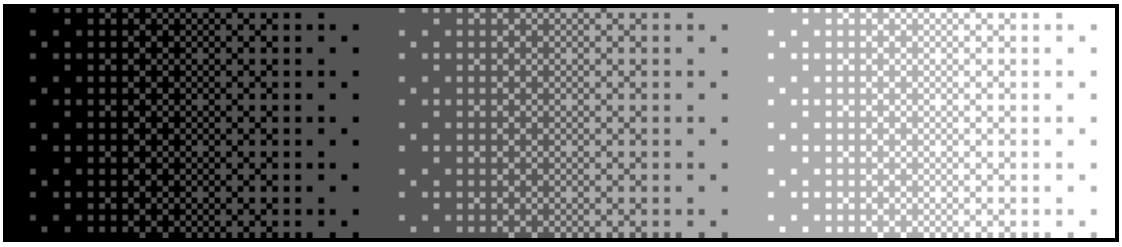

Figure A.3 – Ramp quantized to two bits with 2D halftone dither

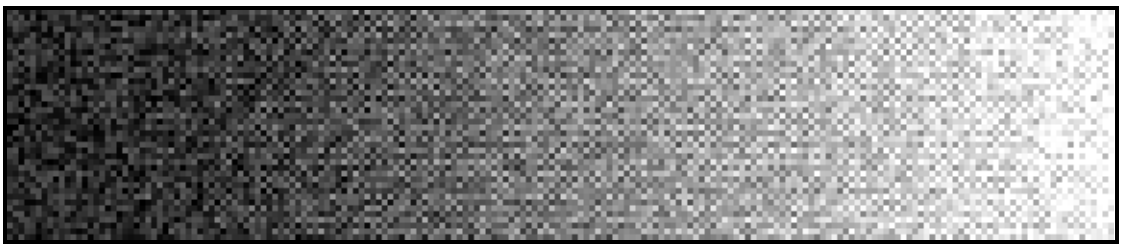

Figure A.4 – Quantized ramp with dither and random data in LSBs

Annex B Characteristics of Compression Method (Informative)

The compression method described in this standard has the following characteristics that make it particularly suitable for the intended application:

- Low delay (latency). The coder and decoder each have a delay of approximately 4 TV lines, giving a total delay through the interface of only about 8 TV lines (at the progressive source image format) or approximately 120 µs for 1080/60/P and 150 µs for 1080/50/P.

- The compressed signal is easily recognizable when displayed as 1080i, as if it were a standard SMPTE 292 signal. This is achieved by allocating the 2 most-significant bits of the interface video words to carry a viewable picture. By using a 2D halftone dither to restore dynamic range a monitoring quality picture is obtained.

- The compression and decompression process is lossless for some easy-to-code material and low-loss for most typical material.

- The loss occurs principally in the first generation. Subsequent generations (assuming the picture has not been moved or processed) rapidly approach a steady state where the loss does not increase any further.

- The algorithm is be simple to implement in hardware and requires little storage.

Annex C 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 RP 291-2009, Assigned Ancillary Identification Codes