Video Services Forum (VSF)
Technical Recommendation TR-01

Transport of JPEG 2000 Broadcast Profile video in MPEG-2 TS over IP

April 15, 2013

VSF_TR-01_2013-04-15

http://www.videoservicesforum.org

INTELLECTUAL PROPERTY RIGHTS

VSF DRAWS ATTENTION TO THE POSSIBILITY THAT THE PRACTICE OR IMPLEMENTATION OF THIS RECOMMENDATION MAY INVOLVE THE USE OF A CLAIMED INTELLECTUAL PROPERTY RIGHT. VSF TAKES NO POSITION CONCERNING THE EVIDENCE, VALIDITY OR APPLICABILITY OF CLAIMED INTELLECTUAL PROPERTY RIGHTS, WHETHER ASSERTED BY VSF MEMBERS OR OTHERS OUTSIDE OF THE RECOMMENDATION DEVELOPMENT PROCESS. AS OF THE DATE OF APPROVAL OF THIS RECOMMENDATION, VSF HAD NOT RECEIVED NOTICE OF INTELLECTUAL PROPERTY, PROTECTED BY PATENTS, WHICH MAY BE REQUIRED TO IMPLEMENT THIS RECOMMENDATION.

LIMITATION OF LIABILITY

VSF SHALL NOT BE LIABLE FOR ANY AND ALL DAMAGES, DIRECT OR INDIRECT, ARISING FROM OR RELATING TO ANY USE OF THE CONTENTS CONTAINED HEREIN, INCLUDING WITHOUT LIMITATION ANY AND ALL INDIRECT, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES (INCLUDING DAMAGES FOR LOSS OF BUSINESS, LOSS OF PROFITS, LITIGATION, OR THE LIKE), WHETHER BASED UPON BREACH OF CONTRACT, BREACH OF WARRANTY, TORT (INCLUDING NEGLIGENCE), PRODUCT LIABILITY OR OTHERWISE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THE FOREGOING NEGATION OF DAMAGES IS A FUNDAMENTAL ELEMENT OF THE USE OF THE CONTENTS HEREOF, AND THESE CONTENTS WOULD NOT BE PUBLISHED BY VSF WITHOUT SUCH LIMITATIONS.

Executive Summary

JPEG 2000 is used in broadcast transmission applications for cost-effective and high quality real-time transport of television video signals over IP networks. While there is a clear demand among end-users and service providers for interoperability between equipment vendors, the existing standards provide a wide range of implementation choices that has made it difficult to achieve interoperability. This VSF Technical Recommendation (TR) defines profiles for streaming of JPEG 2000 Broadcast Profile video, audio and ancillary data in MPEG-2 Transport Stream over IP with optional Forward Error Correction.

Objectives:

- Utilize existing industry standards, constraining them when necessary, to create a VSF Technical Recommendation that provides specific Profiles for interoperable real time JPEG2000 video, audio and ancillary data streams over IP.
- Enable/Facilitate interoperability between products from different equipment manufacturers.

Table of Contents

1. Introduction

The J2K Activity Group's goal was to utilize existing industry standards, constraining them when necessary, to create a VSF Technical Recommendation that provides specific Profiles for interoperable real time streaming of JPEG2000 video, audio and ancillary data in MPEG-2 TS over IP.

In order to address this gap, the Video Services Forum, Inc. (VSF) authorized Activity Group No. 2012-1, titled JPEG2000 Broadcast Profile Technical Recommendation Group, in June 2012.

Recognizing that feedback from interoperability testing and actual use in the field may warrant clarification or improvement of this document, the VSF is committed to a program of maintenance for its technical recommendations, including periodic reviews of its technical recommendations. The first review of this document is planned for approximately one year after its initial publication date.

The Activity Group chair persons are:

John Dale III
Media Links Inc.
jdale@medialinks.com

Inge Hillestad
Nevion
ihillestad@nevion.com

1.1 Contributors

John Dale, Activity Group Co-Chairman, Media Links, Inc.
Inge Hillestad, Activity Group Co-Chairman, Nevion
Mike Bany, DVBLink
Nelson Botsford, Harris
Richard Bullock, Ericsson
John Clark, Artel
Jean-Marie Cloquet, Barco Silex
Jack Douglass, PacketStorm
Brad Gilmer, Gilmer & Associates, Inc.
Igor Komir, Analog Devices
Francois Mace, intoPIX s.a
John Mailhot, Harris
Robert Ruhl, Operations Manager, VSF
Patrick Waddell, Harmonic Inc.
John Wagner, Ericsson

1.2 About the Video Services Forum

The Video Services Forum, Inc. (www.videoservicesforum.org) is an international association dedicated to video transport technologies, interoperability, quality metrics and education. The VSF is composed of service providers, users and manufacturers. The organization's activities include:

- providing forums to identify issues involving the development, engineering, installation, testing and maintenance of audio and video services;
- exchanging non-proprietary information to promote the development of video transport service technology and to foster resolution of issues common to the video services industry;
- identification of video services applications and educational services utilizing video transport services;
- promoting interoperability and encouraging technical standards for national and international standards bodies.

The VSF is an association incorporated under the Not For Profit Corporation Law of the State of New York. Membership is open to businesses, public sector organizations and individuals worldwide. For more information on the Video Services Forum, contact Bob Ruhl, Operations Manager, Video Services Forum, +1 856 627 6672, bob.ruhl1@verizon.net.

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

[1] ITU-T H.222.0: "Information Technology - Generic Coding of moving pictures and associated audio information: Systems" - and ITU-T Recommendation H.222.0 / ISO/IEC 13818-1 /Amd 5:2011: "Transport of JPEG 2000 Part 1 (ITU-T T.800 | ISO/IEC 15444-1) video over ITU-T H.222.0 | ISO/IEC 13818-1".

[2] ITU-T T.800: "Information technology - JPEG 2000 image coding system: Core coding system" – and ITU-T T.800|ISO/IEC 15444-1:2004/Amd 3:2010: "Profiles for broadcast applications".

[3] SMPTE ST 302M-2007: "Television - Mapping of AES3 Data into MPEG-2 Transport Stream".

[4] SMPTE ST 337-2008: “Television - Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface".

[5] SMPTE ST 338:2010 “Format for Non-PCM Audio and Data in an AES3 – Data Types”.

[6] SMPTE ST 2022-1:2007 “Forward Error Correction for Real-Time Video/Audio Transport Over IP Networks”

[7] SMPTE ST 2022-2:2007 “Unidirectional Transport of Constant Bit Rate MPEG-2 Transport Streams on IP Networks”

[8] SMPTE ST 2038:2008 “Carriage of Ancillary Data Packets in MPEG-2 Transport Stream”

[9] SMPTE ST 2063:2012 “Stereoscopic 3D Full Resolution Contribution Link Based on MPEG-2 TS”

[10] AES: AES3:2009, “Digital input-output interfacing — Serial transmission format for two-channel linearly-represented digital audio data”

[11] ETSI EN 301 775 "Digital Video Broadcasting (DVB); Specification for the carriage of Vertical Blanking Information (VBI) data in DVB bitstreams"

[12] ANSI/SCTE 127 2007 "Carriage of Vertical Blanking Interval (VBI) Data in North American Digital Television Bitstreams"

4. Acronyms

3G	3 Gbit/s (Serial Digital Interface)
AES	Audio Engineering Society
AG	Activity Group
ATSC	Advanced Television Systems Committee
ES	Elementary Stream
ETSI	European Telecommunications Standards Institute
FEC	Forward Error Correction
HD	High Definition
IEC	International Electrotechnical Commission
IP	Internet Protocol
IRD	Integrated Receiver/Decoder
ISO	International Organization for Standardization
ITU	International Telecommunication Union
J2K	JPEG 2000
JPEG	Joint Photographic Experts Group
MPEG	Moving Picture Experts Group
PCR	Program Clock Reference
PES	Packetized Elementary Stream
PID	Packet Identifier
PTS	Presentation Time Stamp
RFC	Request for Comments
RTP	Real-time Transport Protocol
S3D	Stereoscopic 3D
SD	Standard Definition
SDI	Serial Digital Interface
SMPTE	Society of Motion Picture Television Engineers
TR	Video Services Forum Technical Recommendation[1]
TS	Transport Stream
TV	Television
UDP	User Datagram Protocol
VSF	Video Services Forum

[1] Note that the term Technical Recommendation is also used by other other organizations such as ETSI.

5. Definitions

Interoperability — An end user or service provider can transport a signal using devices from different manufacturers that state compliance with this Technical Recommendation with the expectation that they will successfully achieve their business objective.

6. System Overview (Informative)

An end-user or service provider of broadcast transmission services can utilize devices that implement this Technical Recommendation (TR) for unidirectional transport of various SDI television video signals over IP. Television video signals include a subset of the formats that can be carried in SD-SDI (ST 259), HD-SDI (ST 292) and 3G-SDI (ST 424).

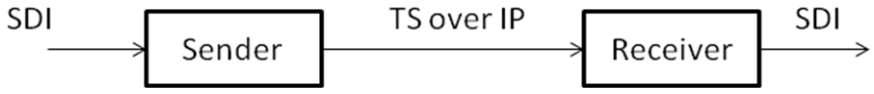

Figure 1: Example system for transmission of SDI signal over IP

As shown in the figure above, a sending device ("sender") with an input SDI video interface extracts the active video, audio and ancillary data components from the input SDI signal. The active video is compressed with JPEG 2000 image compression in order to save bandwidth on the network while providing high image quality. The encoded JPEG 2000 image is multiplexed into an MPEG-2 Transport Stream (TS) together with the associated audio and ancillary data. The system supports transparent pass-through of linear PCM and non-PCM audio, and transparent pass-through of ancillary data. The TS is encapsulated in an RTP stream and transmitted over IP to a receiving device ("receiver"). The receiver will de-encapsulate the RTP/IP stream, de-multiplex the TS, perform JPEG 2000 decoding, and place the video together with associated audio and ancillary data into the output SDI signal.

This TR specifies the syntax and semantics of the signal between the sender and the receiver, and in so doing places constraints on the behavior of the sender; it also specifies some minimum requirements for the receiver. These requirements of the sender and receiver are needed for interoperability between sender and receiver devices.

7. Interoperability Profiles

This TR defines a set of interoperability points, which shall be as defined in Table 1.

Table 1: Video formats supported by the different Profiles

Profile: **Source Video**	1 SD	2 HD	3 3G	4 S3D-HD	5 S3D-3G
270 Mb/s SDTV	576i/25 480i/29.97				
1.5 Gb/s HDTV		720p/50 720p/59.94 1080i/25 1080i/29.97		720p/50 720p/59.94 1080i/25 1080i/29.97	
3 Gb/s HDTV			1080p/50 1080p/59.94		1080p/50 1080p/59.94
	Video format is given as active lines, scanning (interlaced or progressive) and frame rate ("59.94" is equivalent to "60/1.001", while "29.97" is equivalent to "30/1.001")				

Sender and receivers that claim compliance to this TR shall support at least one of the profiles in Table 1, and shall state in their published documentation which profile, or profiles, are supported by a compliant device, system or service.

Note: In Table 1, the following abbreviations are used:

- SD – transport of SDTV JPEG 2000 in TS over IP
- HD – transport of HDTV JPEG 2000 in TS over IP
- 3G – transport of 3 Gb/s HDTV JPEG 2000 in TS over IP
- S3D-HD – transport of Stereoscopic 3D HDTV JPEG 2000 in TS over IP
- S3D- 3G – transport of Stereoscopic 3D 3 Gbit/s HDTV JPEG 2000 in TS over IP

Senders and receivers should also support the video formats given in Table 2.

Table 2: Video formats optionally supported by the different Profiles

Profile: **Source Video**	1 SD	2 HD	3 3G	4 S3D-HD	5 S3D-3G
1.5 Gb/s HDTV		1080p/23.98 1080p/24 1080p/25		1080p/23.98 1080p/24 1080p/25	
	Video format is given as active lines, scanning (interlaced or progressive) and frame rate ("23.98" is equivalent to "24/1.001")				

8. Service Components and Mapping to MPEG-2 TS

8.1 JPEG 2000 Video

The JPEG 2000 video elementary stream (ES) is a concatenation of JPEG 2000 access units. An access unit contains the compressed image representation of one video frame, and consists of an ES header followed by either one or two JPEG 2000 codestreams, depending on whether the video signal is progressive or interlaced. The JPEG 2000 codestream(s) shall be constructed according to section 8.1.1.

8.1.1 JPEG 2000 Codestream Restrictions

Senders creating compressed JPEG 2000 codestreams shall comply with the "*Broadcast Contribution Single Tile Profile*" as specified in ITU-T T.800 Amd 3 [2]. Therefore, each single field or frame shall be encoded as a single tile. Senders shall not use the "*Broadcast Contribution Multi-Tile Profile*" or the "*Broadcast Contribution Multi-Tile Reversible Profiles*".

Receivers shall, at a minimum, be able to decode compressed JPEG 2000 codestreams if those codestreams comply with the "*Broadcast Contribution Single Tile Profile*" as specified in ITU-T T.800 Amd 3 [2].

For each of the Profiles in this Technical Recommendation, Senders shall create codestreams that comply with the J2K Broadcast Profile and Level as stated in Table 3 of this TR.

Note: Profile and Level are terms defined in ITU-T T.800 [2].

Table 3: J2K Broadcast Profile, Level and associated codestream limitations for the defined profiles for interoperability

Profile: **JPEG 2000 Video**	1 SD	2 HD	3 3G	4 S3D-HD	5 S3D-3G
J2K Broadcast Profile	Broadcast Contribution Single Tile Profile				
J2K Broadcast Profile Level	Level 1	Level 2	Level 4	Level 2	Level 4
Capability Rsiz parameter	0x0101	0x0102	0x0104	0x0102	0x0104
Max J2K ES codestream bit rate	200	200	400	200	400
Max J2K ES bit rate (for S3D)	-	-	-	400	800
Min J2K ES codestream bit rate	25	75	100	150	200
	Bit rates given in Mbit/s				

In Table 3, Max J2K ES codestream bit rate, Max J2K ES bit rate (for S3D) and Min J2K ES codestream bit rate provide a maximum and minimum JPEG2000 codestream bit rate that a compliant device shall support for that specific profile.

For purposes of interoperability compliance testing, senders and receivers shall support the maximum and minimum codestream bit rates as stated in Table 3.

Receivers should support all codestream bit rates in-between the minimum and maximum bit rates, and should support codestream bit rates lower than the minimum bit rate given in Table 3. It is expected that Senders support a variety of codestream bit rates in between the minimum and maximum bit rates given in Table 3.

Senders shall use the Rsiz parameter, located in the SIZ marker segment of the JPEG 2000 codestream to identify both the "*Broadcast Contribution Single Tile Profile*" and the associated Level of the codestream. Senders shall ensure that the JPEG 2000 video ES payload constructed from the codestreams shall not exceed the bit rate limitations noted as "*Max J2K ES codestream bit rate*" in Table 3. Furthermore, in the case of the Stereoscopic 3D profiles (as indicated in Table 3 for Levels with the prefix "S3D"), Senders shall ensure that the individual codestreams for each image of the image pair shall conform to the "*Max J2K ES codestream bit rate*". Senders shall also ensure that the combined codestream bitrate for both image pairs shall not exceed the limitation defined as "*Max TS J2K bit rate (for S3D)*".

JPEG 2000 codestreams created by Senders that are compliant with this TR shall set the values of SIZ marker segment in the JPEG 2000 codestream as follows:

- Senders shall set the number of components equal to 3 by setting the following parameter to the value indicated; Csiz=3
- Senders shall set the component sub-sampling mode to signal 4:2:2 video format by setting the following parameters to the values indicated; XRsiz1=1, XRsiz2=2, XRsiz3=2
- Senders shall set the bit depth to indicate 10 bit resolution by setting the following parameters to the values indicated; Ssiz1=9, Ssiz2=9, Ssiz3=9

The order of components in the JPEG 2000 codestream is indicated by the broadcast color specification (see section 8.1.2.5). Under this recommendation, for YCbCr digital representation, Senders shall use the following order of components in the JPEG 2000 codestream:; Y (component index shall be set to 0), Cb (component index shall be set to 1), Cr (component index shall be set to 2).

Senders shall ensure that the codeblock size used in the JPEG 2000 codestream shall be coherent amongst all components, all decomposition levels and all subbands. Senders shall use a single codeblock size value for the image. Moreover, Senders shall limit the set of possible values for the codeblock size to either 32x32 (xcb=ycb=5) or 128x32 (xcb=7; ycb=5). Optionally, a Sender may support other code block sizes, such as 64x64 (xcb=ycb=6).

Senders shall observe the following constraints regarding marker segments:

- As provisioned in ITU-T T.800 Amd 3 [2], TLM marker segments shall be used.
- COC markers segments shall not be used since (due to other constraints) they do not bring any other relevant information than what is already provided by the COD marker segments.

- QCC marker segment(s) may be used, since utilizing different quantization parameters for the different components may improve rate allocation.
- PLM, PLT, SOP and EPH markers and marker segments shall not be present.

Senders supporting stereoscopic operation shall comply with SMPTE ST 2063[9] (see section 8.5), by ensuring that the sets of JPEG 2000 encoding parameters used for the stereoscopic image pairs in Profiles "S3D-HD" and " S3D-3G" are strictly identical.

8.1.2 JPEG 2000 Elementary Stream

Senders shall construct JPEG 2000 video elementary streams according to Annex M of ITU-T T.800 Amd 3 [2].

Senders shall encapsulate the JPEG 2000 elementary stream (ES) in an MPEG-2 Transport Stream according to ITU-T H.222.0 Amd 5 [1], and shall set the following parameters accordingly: stream_id = private_stream_1 and stream_type = '0x21'. Senders shall include the JPEG 2000 video descriptor in the Program Map Table (PMT) for the JPEG 2000 service component.

8.1.2.1 JPEG 2000 ES Header

Senders shall construct the JPEG 2000 ES header as specified in Table S.1 of ITU-T H.222.0 Amd 5 [1], including the order of the ES header boxes.

Note: Due to the fact that some syntax elements of the ES header are not always present, a Receiver is advised to parse the JPEG 2000 video descriptor (Located in the PMT of the Transport Stream) in order to determine the value of interlaced_video prior to parsing the JPEG 2000 ES header.

8.1.2.2 Field Coding

For interlaced video signals, Senders shall set the interlaced_video flag in the JPEG 2000 video descriptor to '1', they shall ensure that the Field Coding box is present in the JPEG 2000 ES header, and they shall ensure that the JPEG 2000 access unit contains two JPEG 2000 codestreams. Senders shall transmit fields in temporal order, and they shall endure that the codestream corresponding to the field with the top-most line is located first in the access unit. In the Field Coding box, Senders shall ensure that the following parameters are set: Fic = 2, Fio = 1.

For progressive video signals, Senders shall set the interlaced_video flag in the JPEG 2000 video descriptor to '0', they shall ensure that the Field Coding box is not present in the JPEG 2000 ES header, and they shall ensure that the JPEG 2000 access unit contains one JPEG 2000 codestream.

8.1.2.3 Maximum Bitrate and Codestream size

In accordance with ITU-T H.222.0 [1], the interlaced_video flag in the JPEG 2000 video descriptor shall be used to determine whether the AUF_2 field in the Maximum Bit Rate box of the JPEG 2000 ES Header is present or not

- If interlaced_video = '0', then AUF_2 shall not be present
- If interlaced_video = '1', then AUF_2 shall be present

Note: A receiver can use the values indicated in the Maximum Bitrate box to determine the size of the JPEG 2000 codestream(s). Alternatively, a receiver can use the tile-part length as indicated in the TLM marker segment, or the value of the Psot parameter in the SOT marker segment in the JPEG 2000 main header (note that the latter may be zero).

Note: Implementers of this Technical Recommendation are advised that, since the exact codestream sizes for both fields of an interlaced frame (AUF_1 and AUF_2) needs to be known prior to transmission of an access unit, an additional field period of buffering latency can be expected on the encoder side in addition to the actual encoding latency.

8.1.2.4 Frame rate

The frame rate of the JPEG 2000 video ES is signaled using the NUM and DEN parameters of the Frame Rate box located in the ES header, as well as in the JPEG 2000 video descriptor. For the supported video formats, Senders shall signal the frame rate of the JPEG 2000 video ES using the values shown in Table 4 below.

Table 4: Signaling of supported frame rates

Frame rate	Interlaced (I) or Progressive (P)	Numerator NUM	Denominator DEN
25	I	25	1
30/1.001	I	30 000	1001
50	P	50	1
60/1.001	P	60 000	1001

8.1.2.5 Broadcast color

Senders shall signal the Broadcast Color Specification Code (located in the Broadcast Color Box in the ES header) using the values shown in Table 5 below.

Table 5: Supported color space for the defined Profiles

Profile:	1 SD	2 HD	3 3G	4 S3D-HD	5 S3D-3G
Broadcast Color Specification Code	0x02	0x03	0x03	0x03	0x03
	Notes: the value"0x02" indicates use of Rec. ITU-R BT 601-6, and the value "0x03" indicates Rec. ITU-R BT 709-5.				

8.1.2.6 JPEG 2000 Still Pictures

Senders shall ensure that the JPEG 2000 video ES does not contain any JPEG 2000 still pictures. Senders shall set the still_mode field in the JPEG 2000 video descriptor to '0'.

8.2 Audio

Senders shall ensure that audio signals shall be 48 kHz sampling rate, and shall be synchronous to the video pixel clock. Senders shall organize audio signals into monaural channel-pairs and shall ensure that they are formatted in accordance with AES3 [10]. These AES3 signals shall consist of PCM audio samples, or of non-PCM signals which may be compressed audio.

8.2.1 Audio Transport (PCM and Non-PCM signals)

SMPTE ST 302 shall be used for the transport of all audio signals. Under this technical recommendation, ST 302 is further constrained such that Senders shall transmit exactly one AES3 signal pair per ST 302 audio service, and shall use only the 20-bits-per-sample mode. As stated in ST 302, Senders shall mark each audio service with MPEG-2 Presentation Time Stamps (PTS) corresponding to video frames in the source video format.

Table 6: Number of AES3 stereo channel pairs supported by the different Profiles

Profile: **Audio**	1 SD	2 HD	3 3G	4 S3D-HD	5 S3D-3G
Total AES3 stereo channel pairs including Linear PCM audio and non-PCM signals	4	8	8	8	8

Receivers under this technical recommendation shall be capable of simultaneously receiving and processing the number of AES3 stereo channel pairs shown in Table 6 though senders may send more. A mechanism should be incorporated in receiving equipment to allow selection of the audio signals to be processed from amongst those sent. A transmitter should assign audio services to ascending MPEG-2 transport stream PID values such that the first audio service has the lowest PID and the last audio service has the highest PID. Similarly, unless overridden by the user, the receiver should receive the first audio service from the lowest value audio PID and last audio service from the highest PID.

Upon receipt and processing of multiple ST 302 audio services, receivers shall synchronize the audio services such that, upon presentation, audio samples from access units containing the same PTS value shall be emitted synchronously (phase aligned) with each other.

8.2.2 PCM (Uncompressed) Audio (Informative)

The number of 48 kHz audio samples corresponding to a given frame of video is not an integer in some frame rates. Therefore the size of the PES packets for carrying said audio samples will vary. Senders and Receivers shall support 48 kHz audio at all supported frame rates, without altering the number of audio samples per frame.

8.2.3 Non-PCM (Pre-compressed Audio) Signals

There is a common use-case of packaging compressed audio bitstreams into an AES3 signal for transport, documented in SMPTE ST 337. Senders and Receivers shall take care to ensure that non-PCM audio signals covering the gamut of options identified in SMPTE (see SMPTE ST 337 and ST 338) are transported in such a manner as to not disrupt their contents. Sample-Rate-Conversion operations shall not be performed on non-PCM audio signals.

The following pre-compressed audio formats, as documented in SMPTE 338, shall be supported:

- AC-3 audio data (data_type = 1) per SMPTE ST 340.
- MPEG 2 AAC in ADTS audio data (data_type = 7) per SMPTE ST 2041-2.
- MPEG-4 AAC data in ADTS or LATM/LOAS (data_type = 10) per SMPTE ST 2041-3.
- MPEG-4 HE AAC data in ADTS or LATM/LOAS (data_type = 11) per SMPTE ST 2041-3.
- E-AC-3 audio data (data_type = 16) per SMPTE ST 340.
- Dolby E audio data (data_type = 28).

Note: Other formats may be supported as standardized by SMPTE.

8.2.4 Maintenance of A/V Sync

Receivers shall use the PTS values in both audio and video PES to maintain A/V synchronization within +/- 2 ms when outputting video on the recovered stream time base.

Note: When synchronizing the output video to a local (GenLock) time base, A/V synchronization error of +- ½ video frame time can be expected. Implementers are referred to Rec. ITU-R BT.1359-1 “Relative Timing of Sound and Vision for Broadcasting”, as well as CEA-CEB20 “A/V Synchronization Processing Recommended Practice” for additional guidance on this topic. For the specific case of Dolby E, implementers are also referred to SMPTE RDD 19.

8.3 Ancillary Data

Most SDI signals (both SD and HD) include Horizontal Ancillary (HANC) and Vertical Ancillary (VANC) data packets formatted in accordance with SMPTE ST 291-1. Subject to the exceptions and limits noted below, HANC and VANC data shall be transported from sender to receiver using the method specified in SMPTE ST 2038.

HANC and VANC data with other standardized transports in MPEG-2 TS or J2k video PES may also be transported using those methods, although implementers may provide controls to permit the user to choose whether an alternate transport method is used or not.

Note that VANC and HANC packets can appear on any line, and the method specified for transport herein allows for them to be placed back into the line and space from which they came.

8.3.1 HANC and VANC data which are excluded from transport

Although embedded audio is formatted as HANC data, Senders shall use the method identified in section 8.2 for the transport of all audio signals; Senders shall not use the methods in this section for audio.

The Embedded Audio Control Packet defined in SMPTE ST 299-1 should not be transmitted by senders, and should be ignored by receivers. Receivers shall generate a locally correct Embedded Audio Control Packet based on their specific configuration.

EDH, CRC, and Line Number information, while present in the ancillary data spaces, is not formatted as ANC packets under SMPTE ST 291-1 and therefore Senders shall not send this data.

8.3.2 Limits on the total amount of ANC data to be transported

If all of the available ancillary data spaces were packed full of well-formed packets, a very significant amount of transport bandwidth might be required to transport all of the data. In practice, while the use of ANC data to carry information is expected to increase during the lifetime of this TR, this document places practical limits on the amount of data to be transported in order to foster interoperability in balance with reasonable implementation.

For each of the profiles defined, Table 7 indicates the MINIMUM number of 10-bit words of ANC data, carried using SMPTE ST 2038, which will be supported by compliant senders and receivers. Senders shall be able to create streams which meet these minimum bit rates, and Receivers shall be able to output ANC data streams which meet these bit rates. Senders and Receivers which are compliant with this TR may support the transport of higher amounts of ANC data than those specified in Table 7.

Table 7: Amount of ANC data transported using ST 2038 to be supported by the different Profiles

Profile: / Ancillary Data	1 SD	2 HD	3 3G	4 S3D-HD	5 S3D-3G
Number of 10-bit words of ANC Data to be transported per second	104800	104800	104800	104800	104800
Worst-Case PES bit rate for ANC Data (bits/field)	26106	26106	26106	26106	26106
ES Buffer Size (Bn) (bits)	4 * 26106	4 * 26106	4 * 26106	4 * 26106	4 * 26106
Transport Maximum bit rate (Rmax) (bits/second)	2500000	2500000	2500000	2500000	2500000

Note: in Table 7, 104800 10-bit words per second is equivalent to 8 ANC data packets per frame, each having 7 header words and 255 User Data Words, at 50 frames per second. Since the size of ANC data packets is variable, more than 8 packets can of course be transported.

Note: users should be aware that encapsulation of ST 2038 is in TS packets (188 bytes). The resulting TS bit rate can be substantially higher than the underlying ANC data rate. For example, a 2-byte payload of CEA 608 data becomes 188 bytes in the TS layer (ignoring the header overhead).

The ANC data transmitted by Senders shall be compliant with the T-STD model as specified in ISO/IEC 13818-1 [1], using an elementary stream buffer size (Bn) as articulated in Table 7. The transport buffer TBn for the ANC service is specified to be 512 bytes in ISO/IEC 13818-1 [1]. For ST 2038 ANC data, the transport buffer shall be drained (Rxn) at 1.2 times [Rmax] as per Table 7.

Note: the PES bit rate calculation in Table 7 assumes 20 lines of ANC data per field (frame in progressive formats), and 10 separate ANC data packets per each line for the purpose of the ST 2038 overhead estimation (14 bytes PES header, plus 4 bytes per each packet header, 10 packets per line). A worst-case field rate of 60 fields per second is used for this calculation.

Note: The Transport Maximum bit rate calculation assumes the worst-case PES rate, and adds TS header overhead, plus an average 92 bytes stuffing per PES for PES alignment as indicated in ST 2038. The calculated value of 2449560 is rounded up to 2500000 for the purposes of this TR.

8.3.3 Prioritization of certain ANC signals (Informative)

Senders might want to provide a mechanism for filtering what types of ancillary data types that should be sent. In case of over-subscription, Senders might also wish to provide prioritization for the ANC data types being sent. In no particular order, examples of important ANC data types include:

Closed Captions (SMPTE ST 334-1)
Time Code (SMPTE ST 12-2)
AFD/Bar data (SMPTE ST 2016-3)
Audio metadata (SMPTE ST 2020-2, SMPTE ST 2020-3)
ANSI/SCTE 104 Messages (SMPTE ST 2010)
DVB/SCTE VBI data (SMPTE ST 2031, OP-47, SMPTE RDD 8)

8.4 Vertical Blanking Interval Data for Standard-Definition Signals

While both High-Definition and Standard-Definition systems routinely carry associated information as ANC data packets, Standard-Definition (SD) digital signals may carry important digital data representations of analogue waveforms, which are band-limited to permit a compatible analog conversion, and are located within the Vertical Blanking Interval. Transport of these VBI signals is expected as part of the use case for transporting SD signals and is described in this section.

Certain VBI signals have well-known transport mappings which are required under this standard; in addition a pass-through mechanism is defined for transport of arbitrary waveform VBI data.

8.4.1 VBI signals with specific transport mappings

Senders shall support the transport of VBI data using the method specified in ETSI EN 301 775 [11] and ANSI/SCTE 127 [12] subject to the exceptions and limits noted below. Receivers shall recreate VBI data in accordance with the ETSI standard as well. Senders and Receivers may transport VBI data using other methods specified in standards for MPEG-2 TS or J2k video PES transport. Implementers may provide controls to permit the user to choose whether an alternate transport method is used or not.

Senders shall support the transport of VBI data formatted in accordance with CEA-608B (often found on line 21 of field 1 and field 2 in 525-line systems) using the method described in ETSI EN 301 775 [11], and Receivers shall output VBI data which has been transported in accordance with the method described in the ETSI standard.

Senders shall support the transport of VBI data formatted as teletext (including subtitles in teletext) using the method described in ETSI EN 301 775 [11], and Receivers shall output VBI data formatted as teletext (including subtitles in teletext) which has been transported in accordance with the method described in the ETSI standard.

Senders shall support the transport of Wide Screen Signaling (WSS) data in VBI data using the method described in ETSI EN 301 775 [11], and Receivers shall output WSS data which has been transported in accordance with the method described in the ETSI standard.

8.4.2 Arbitrary sample-based transport mapping

In addition to the well-known signals described above, there are additional waveforms defined in ETSI EN 301 755 [11] and in extensions defined in SCTE 127 [12], as well as proprietary or unknown waveforms in the VBI area which might need to be transported. While transporting the entire VBI region using the method below could be onerous in terms of bit rate, transport service providers, in coordination with their customers, might choose to select some lines of the VBI to transport using the sample-based method below.

Senders shall support the user selectable transport of information represented in VBI data utilizing the "Data field for monochrome 4:2:2 samples" method described in ETSI EN 301 775 section 4.9 [11], and Receivers shall output the information in accordance with the method described in the ETSI standard.

8.4.3 Limits on the transport of VBI signals

The following limits are applied to the transport of VBI signals.
Senders shall be required to format for transport, and Receivers shall be required to output, up to and including:

- two lines per field of CEA-608 data
- 12 lines per field of Teletext (including subtitles in teletext)
- 1 line per field of Wide-Screen Signaling (WSS)
- 2 lines per field of Monochrome 4:2:2 samples

8.5 Stereoscopic 3D

Senders shall format Stereoscopic 3D image pairs for transport in compliance with SMPTE ST 2063 [9].

9. IP Encapsulation and Forward Error Correction

Senders shall map the MPEG-2 TS for transport over an IP network in accordance with SMPTE ST 2022-2 [7]. If FEC is used, Senders shall construct the FEC stream in accordance with SMPTE ST 2022-1 [6].

Receivers shall be able to accept IP streams which are compliant with SMPTE ST 2022-2 [7], and if it is present, Receivers shall be able to accept and process FEC streams constructed in accordance with SMPTE ST 2022-1 [6].

Although SMPTE 2022-2 [7] allows for 1, 4 and 7 TS packets per IP datagram, in order to be compliant with this TR, Senders and Receivers shall support 7 TS packets per IP datagram. Senders and Receivers may support 1 and 4 TS packets per IP datagram, if desired.

10. Bibliography (Informative)

CEA-CEB20 “A/V Synchronization Processing Recommended Practice”
CEA-608-B-2000 “Line 21 data services”
Rec. ITU-R BT 601-6 “Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios”
Rec. ITU-R BT 709-5 “Parameter values for the HDTV standards for production and international programme exchange”
Rec. ITU-R BT.1359-1 “Relative Timing of Sound and Vision for Broadcasting”
SMPTE RDD 8:2008 “Storage and Distribution of Teletext Subtitles and VBI Data for High-Definition Television”
SMPTE RDD 19:2011 “Guidelines on the Use of Dolby® E with Video Signals at Frame Rates Greater than 30 Hz”
SMPTE ST 12-2:2008 “Television – Transmission of Time Code in the Ancillary Data Space”
SMPTE ST 259:2008 “Television – SDTV Digital Signal/Data – Serial Digital Interface”
SMPTE ST 291-1:2011 “Ancillary Data Packet and Space Formatting”
SMPTE RP 291-2:2013 “Ancillary Data Space use – 4:2:2 SDTV and HDTV component systems and 4:2:2 2048 x 1080 Production Image formats”
SMPTE ST 292-1:2011 “1.5 Gb/s Signal/Data Serial Interface”
SMPTE ST 334-1:2007 “Vertical Ancillary Data Mapping of Caption Data and Other Related Data”
SMPTE ST 340:2008 “Format for Non-PCM Audio and Data in AES3 – ATSC A/52B Digital Audio compression Standard for AC-3 Data Types”
SMPTE ST 372:2011 “Dual Link 1/5 Gb/s Digital Interface for 1920 x 1080 and 2048 x 1080 Picture Formats”
SMPTE ST 424:2006 “Television – 3 Gb/s Signal/Data Serial Interface”
SMPTE ST 425:2011 “Source Image Format and Ancillary Data Mapping for the 3 Gb/s Interface”
SMPTE ST 2010:2008 Vertical Ancillary Data Mapping of ANSI/SCTE 104 Messages
SMPTE ST 2016-3:2009 “Vertical Ancillary Data Mapping of Active Format Description and Bar Data”
SMPTE ST 2020-1:2008 “Format of Audio Metadata and Description of the Asychronous Serial Bitstream Transport"
SMPTE ST 2020-2:2008 “Vertical Ancillary Data Mapping of Audio Metadata - Method A”
SMPTE ST 2020-3-2008 “Vertical Ancillary Data Mapping of Audio Metadata - Method B”
SMPTE ST 2031-2:2007 “Carriage of DVB/SCTE VBI Data in VANC”
SMPTE ST 2041-2:2010 “Format for Non-PCM Audio in AES3 - MPEG-2 AAC and HE AAC Audio in ADTS”SMPTE ST 2041-3:2010 “Format for Non-PCM Audio and Data in AES3 - MPEG-4 AAC and HE AAC Compressed Digital Audio in ADTS and LATM/LOAS Wrappers”
ANSI/SCTE 127 2007 “Carriage of Vertical Blanking Interval (VBI) Data in North American Digital Television Bitstreams”