Video Services Forum (VSF)
Technical Recommendation TR-07

Transport of JPEG XS Video in MPEG-2 Transport Stream over IP

April 20, 2022
TR-07:2022

INTELLECTUAL PROPERTY RIGHTS

RECIPIENTS OF THIS DOCUMENT ARE REQUESTED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT CLAIMS OR OTHER INTELLECTUAL PROPERTY RIGHTS OF WHICH THEY MAY BE AWARE THAT MIGHT BE INFRINGED BY ANY IMPLEMENTATION OF THE RECOMMENDATION SET FORTH IN THIS DOCUMENT, AND TO PROVIDE SUPPORTING DOCUMENTATION.

THIS RECOMMENDATION IS BEING OFFERED WITHOUT ANY WARRANTY WHATSOEVER, AND IN PARTICULAR, ANY WARRANTY OF NONINFRINGEMENT IS EXPRESSLY DISCLAIMED. ANY USE OF THIS RECOMMENDATION SHALL BE MADE ENTIRELY AT THE IMPLEMENTER'S OWN RISK, AND NEITHER THE FORUM, NOR ANY OF ITS MEMBERS OR SUBMITTERS, SHALL HAVE ANY LIABILITY WHATSOEVER TO ANY MPLEMENTER OR THIRD PARTY FOR ANY DAMAGES OF ANY NATURE WHATSOEVER, DIRECTLY OR INDIRECTLY, ARISING FROM THE USE OF 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THE FOREGOING NEGATION OF DAMAGES IS A FUNDAMENTAL ELEMENT OF THE USE OF THE CONTENTS HEREOF, AND THESE CONTENTS WOULD NOT BE PUBLISHED BY VSF WITHOUT SUCH LIMITATIONS.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

The JPEG XS compression method is used in low latency transmission applications for cost-effective, high quality, real-time transport of television video signals over IP networks. The term “XS” is meant to convey the “extra small”, “extra speed” nature of this compression method. This VSF Technical Recommendation (TR), defines profiles for streaming of JPEG XS video, and establishes an interoperable method for transporting that compressed video along with associated audio and ancillary data in an MPEG-2 Transport Stream. This TR also defines an optional Forward Error Correction scheme.

The primary use case addressed by this TR is the transport of video, audio, and ancillary data in wide area network (WAN) applications.

Recipients of this document are invited to submit technical comments. The VSF also requests that recipients notify us of any relevant patent claims or other intellectual property rights of which they may be aware, that might be infringed by any implementation of the Recommendation set forth in this document, and to provide supporting documentation.

Table of Contents

1. Introduction (Informative)

In February of 2020, the VSF created the JPEG XS Activity Group. This group was formed to develop an interoperable method for the transport of low-latency JPEG XS compressed video for WAN & LAN transport. The group was also charged with defining interoperability points for the compression and transmission of High Definition and Ultra-High Definition formats. This document addresses MPEG-2 Transport Stream encapsulation in SMPTE(c) ST 2022-2 IP, and also addresses potential system interoperability issues with other media components within the MPEG-2 Transport Stream.

1.1 Contributors

Contributors to this document include:

- John Dale, Activity Group Chairman, Media Links, Inc.
- Tim Bruylants, intoPIX SA
- Antonin Descampe, intoPIX SA
- Brad Gilmer, Gilmer & Associates, Inc.
- Andrew Krupiczka, Disney
- Jean-Baptiste Lorent, intoPIX SA
- John Mailhot, Imagine Communications
- John Schilberg, Utah Scientific

1.2 About the Video Services Forum

The Video Services Forum, Inc. (www.vsf.tv)) 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
+1 929-279-1995
bob.ruhl1@verizon.net.

2. Conformance Notation

Normative text is text that describes elements of the design that are indispensable or that contain 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

AES: AES3:2009 (r2019), “Digital input-output interfacing — Serial transmission format for two-channel linearly-represented digital audio data”

ANSI/CTA-861-H (2021) “A DTV Profile for Uncompressed High Speed Digital Interfaces”

ANSI/SCTE 127 2007 “Carriage of Vertical Blanking Interval (VBI) Data in North American Digital Television Bitstreams”

ETSI EN 301 775 “Digital Video Broadcasting (DVB); Specification for the carriage of Vertical Blanking Information (VBI) data in DVB bitstreams”

Rec. ITU-R BT.2020-2 “Parameter values for ultra-high definition television systems for production and international programme exchange”

Rec. ITU-R BT.2100-2 “Image parameter values for high dynamic range television for use in production and international programme exchange”

Rec. ITU-T H.222.0 (2021) | ISO/IEC 13818-1:2021: "Information Technology - Generic Coding of moving pictures and associated audio information - Part 1: Systems"[1]

Rec. ITU-T H.273 (2021) | ISO/IEC 23091-2:2021: "Information Technology - Coding-independent code points - Part 2: Video"

ISO/IEC 21122-1:2022[1] "Information technology — JPEG XS Low-latency Lightweight Image Coding System — Part 1: Core coding system”

ISO/IEC 21122-2:2022[2] “Information technology — JPEG XS low-latency lightweight image coding system — Part 2: Profiles and buffer models”

ISO/IEC 21122-3:2022 “Information technology — JPEG XS low-latency lightweight image coding system — Part 3: Transport and container formats”

SMPTE ST 299-2:2010 “Extension of the 24-Bit Digital Audio Format to 32 Channels for 3Gb/s Bit – Serial Interfaces”

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

SMPTE ST 337:2015 “Television - Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface"

SMPTE ST 338:2016 “Format for Non-PCM Audio and Data in an AES3 – Data Types”.

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

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

SMPTE ST 2038:2021 “Carriage of Ancillary Data Packets in MPEG-2 Transport Stream”

---

[1] We anticipate that, in 2022, Rec. ITU-T H.222.0 (2021) will be amended by Rec. ITU-T H.222.0 (2021)/AMD1 | ISO/IEC 13818-1:2021/AMD1: "Information technology Generic coding of moving pictures and associated audio information Part 1: Systems Amendment 1: Carriage of LCEVC in MPEG-2 TS and other improvements". This amendment addresses some errors in MPEG-2 TS with respect to carriage of JPEG XS.

[2] Implementers should note that AMD1 for ISO/IEC 21122-2:2022 will contain fixes and also defines a 4bpp sublevel.

SMPTE ST 2086:2018 “Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images”

SMPTE ST 2108-1:2018 “HDR/WCG Metadata Packing and Signaling in the Vertical Ancillary Data Space”

SMPTE ST 2108-2:2019 “Vertical Ancillary Data Mapping of KLV Formatted HDR/WCG Metadata”

Note: Joint ITU and ISO/IEC documents refer to exactly the same standard text, and may share the same title, however in reference 6 there is an exception, the titles are different, but the standard text is identical. Titles listed are from ISO/IEC documents.

4. Acronyms

AES	Audio Engineering Society
ANC	Ancillary Data
bpp	Bits per pixel
ES	Elementary Stream
FEC	Forward Error Correction
HD	High Definition
HDR	High Dynamic Range
HLG	Hybrid Log Gamma
IEC	International Electrotechnical Commission
IP	Internet Protocol
ISO	International Organization for Standardization
ITU	International Telecommunication Union
JPEG	Joint Photographic Experts Group
JPEG XS	Joint Photographic Experts Group 21122 Coding Standard
jxes	JPEG XS Elementary Stream
MPEG	Moving Picture Experts Group
PCM	Pulse-code Modulation
PCR	Program Clock Reference
PES	Packetized Elementary Stream
PID	Packet Identifier
PMT	Program Map Table
PQ	Perceptual Quantization
PTS	Presentation Time Stamp
RTP	Real-time Transport Protocol
SDR	Standard Dynamic Range
SDI	Serial Digital Interface
SMPTE	Society of Motion Picture and Television Engineers
TR	Video Services Forum Technical Recommendation[3]
UHD1	Ultra high resolution with a resolution of 3840 × 2160, which is found in ITU-R BT 2020
UHD2	Ultra high resolution with a resolution of 7680 × 4320, which is found in ITU-R BT 2020
YCbCr	Luminance Component, Blue-Difference and Red-Difference Chroma Components

[3] Note that the term Technical Recommendation is also used by other organizations such as European Technical Standards Institute (ETSI).

5. Definitions

Codestream	Compressed image data representation that includes all necessary data to allow (visually lossless or lossy) reconstruction of the sample values of a digital image.
JPEG XS video elementary stream	Video elementary stream consisting of a succession of JPEG XS Video Access Units.
JPEG XS Video Access Unit	The JPEG XS codestream or multiple JPEG XS codestreams, as defined in ISO/IEC 21122-1, comprising a decodable image, preceded by a JPEG XS elementary stream header.
JPEG XS video sequence	A JPEG XS video elementary stream where all the access units have the same profile, level, and sublevel (as defined in ISO/IEC 21122-2), JPEG XS video access unit coding parameters, and video parameters.
JPEG XS video	The JPEG XS video descriptor definition is provided in Rec. ITU-T H.222.0 section 2.6.127
Receiver	Element within a device that terminates one RTP stream from the Network.
Sender	Element within a device that originates one RTP stream into the Network.

6. System Overview/Use Case (Informative)

An end-user or service provider of broadcast transmission services may utilize devices that implement this Technical Recommendation for the unidirectional transport of real time television signals over an IP network between two facilities. The signals may include video, audio, and SMPTE 291-formatted Ancillary Data packets.

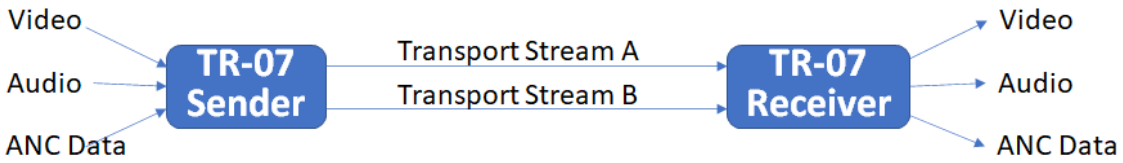

Figure 1: Example system for transmission of real time television signal over IP

As shown in Figure 1 above, the Sender has video, audio, and Ancillary Data inputs. These inputs are presumed to be time-aligned at their presentation to the Sender; the video is compressed using a JPEG XS compression engine. The JPEG XS codestream is multiplexed into an MPEG-2 Transport Stream together with the audio and ancillary data by the TR-07 Sender. The system defined in this TR supports transparent pass-through of linear PCM and non-PCM audio formatted using SMPTE ST 302, and transparent pass-through of ancillary data formatted using SMPTE ST 2038. The Sender encapsulates the transport stream into an RTP stream in accordance with SMPTE ST 2022-2 and transmits this stream (potentially redundantly, as shown in the figure above) using Internet Protocol to a TR-07 Receiver. The Receiver de-encapsulates the RTP/IP stream, de-

multiplexes the MPEG-2 Transport Stream, decodes the JPEG XS codestream, and places the video together with its associated audio and ancillary data onto the output. MPEG-2 Presentation Timestamps are used to preserve the end-to-end timing relationships of the video, audio, and ancillary data items.

The target end to end transmission latency, (less network and route mile delay), for the real time transmission of all essence components including video, audio and ancillary data is approximately one tenth of a video frame duration.

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 constraints and requirements are needed for interoperability.

Receivers under this TR-07 are expected to make the relevant signal metadata (colorimetry, transfer characteristic, frame rate, sampling structure, image dimensions, etc.) available to downstream devices using the technical standards appropriate to the interfaces provided.

7. Organization and Signaling of a VSF TR-07 Stream

MPEG-2 Transport Streams shall contain a single Program and a single Program Map Table (PMT).

MPEG-2 Transport Streams shall be constructed in compliance with Rec. ITU-T H.222.0.

MPEG-2 Transport Streams shall be of a constant bitrate utilizing null packet stuffing, if required, to maintain a constant bitrate.

The Program Clock Reference (PCR) shall be carried on an otherwise empty PID utilizing adaptation layer stuffing to fill the remainder of the packet.

The PCR shall not be part of any of the PES streams.

Video signals shall be compressed using a JPEG XS codec that conforms to ISO/IEC 21122-1 as described in Section 9.1.

The Program Map Table for the JPEG XS video stream shall be constructed as defined in Rec. ITU-T H.222.0. A typical PMT is provided in Appendix A.

The PMT Entry for JPEG XS Video under this TR shall contain the *JXS_video_descriptor* (*extension_descriptor_tag* == 0x14) defined in Rec. ITU-T H.222.0. Mastering Display Metadata may be present in this descriptor. JPEG XS PES streams are carried using *stream_id* = 0xBD, with a *stream_type* = 0x32. Further details on the *JXS_video_descriptor* may be found in Section 9.1.

Audio signals shall be organized into Transport Stream packets in accordance with SMPTE ST 302, as described in Section 9.2.

The Program Map Table shall contain no more than four (4) audio PES streams.

Each SMPTE ST 302 PES stream shall be signaled using *stream_id* = 0xBD, with a *stream_type* = 0x06 (Private Data), and the specific descriptors defined in ST 302.

Ancillary Data signals shall be organized into Transport Stream in accordance with SMPTE ST 2038, as described in Section 9.3.

The Program Map Table shall contain no more than one SMPTE ST 2038 PES stream.

The SMPTE ST 2038 PES stream shall be signaled using *stream_id* = 0xBD, with a *stream_type* = 0x06 (Private Data), and the specific descriptors defined in SMPTE ST 2038.

8. Capability Sets and Interoperability Points

Implementations which claim conformance with this document shall consist of one or more Senders and/or Receivers that support one or more of the Capability Sets described in the sections below.

Receivers compliant with this document shall conform to one or more of the following Capability Sets as further detailed in Appendix B *Capability Sets and Interoperability Points*.

Capability Set A
- HD video capability
- JPEG XS 2k-1 video level capability with only those formats specified in Appendix B
- Audio level capability as described in Section 9.2
- ANC capability as described in Section 9.3

Capability Set B
- UHD1 video capability
- JPEG XS 4k-2 video level capability with only those formats specified in Appendix B
- Audio level capability as described Section 9.2
- ANC capability as described in Section 9.3

Capability Set C
- UHD2 video capability with only those formats specified in Appendix B
- JPEG XS 8k-2 video level capability with only those formats specified in Appendix B
- Audio level capability as described in Section 9.2
- ANC capability as described in Section 9.3

Note: Capability Sets define the features and capacities which can appear in a bitstream, and the capability expected in a compliant receiver. Appendix B contains a table of interoperability points, including notes of the encompassing capability set. The interoperability points reflect typical television production industry use-cases and could be used during testing events to ensure reasonable coverage of common formats and features used in contribution applications at the time of this publication.

9. Essence Service Components

This section establishes specific restrictions for JPEG XS video, audio, and metadata in order to improve interoperability between Senders and Receivers from different implementers.

9.1 JPEG XS Video

9.1.1 MPEG-2 Transport Stream and JPEG XS Codestream (Informative)

The following is an informative summary of the organization of a TR-07 JPEG XS video stream within a MPEG-2 Transport Stream, according to Rec. ITU-T H.222.0. During this discussion, it may be useful to refer to Figure 2 below.

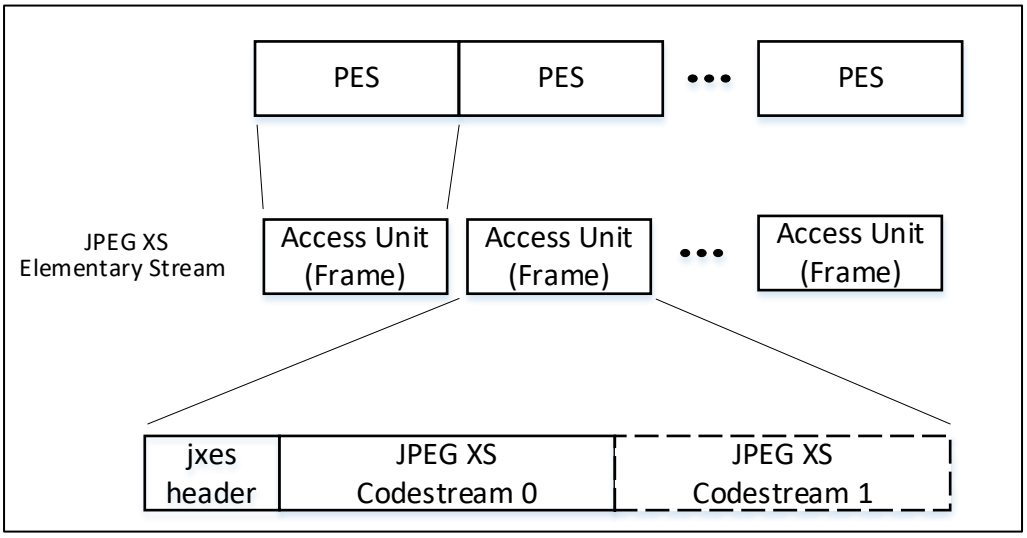

Figure 2: Structure and order of JPEG XS Video Access Units

The following signaling elements and headers are used to transport JPEG XS streams in MPEG-2 Transport Stream packets.

- A JPEG XS video descriptor (JXS_video_descriptor) that includes the description of a *JPEG XS video elementary stream*. This descriptor is included for each *JPEG XS video elementary stream* component in the PMT with *stream_type* equal to 0x32, as described in Rec. ITU-T H.222.0.
- A JPEG XS video elementary stream consisting of a succession of JPEG XS Video Access Units, each of them embedded in a Packetized Elementary Stream (PES). Each such access unit includes a JPEG XS Elementary Stream Header (*jxes_header*) followed by one (if progressive video stream) or two (if interlaced video stream) JPEG XS codestreams as illustrated in Figure 2 above.[4]

[4] Implementers should note that table 2-132 of Rec. ITU-T H.222.0 defining the JXS_video_descriptor contains an error. This error will be addressed in ISO/IEC 13818-1 (2021)/AMD1 by removing the *descriptor_tag* and *descriptor_length* fields from the table. We anticipate AMD1 will be published in 2022.

9.1.2 JPEG XS Codestream Restrictions

JPEG XS Codestreams emitted by Senders that are compliant with this TR shall conform to the "High444.12" profile specified in ISO/IEC 21122-2, with the following constraints:

- **Number of components, chroma sampling format and alpha channel**: if there is an alpha channel, it shall be transported in its own codestream. The allowed number of components and chroma sampling formats are listed in Table 1 below, and the corresponding parameters shall be set accordingly.

Allowed Configurations	**N_c value (= number of components)**	**$s_x[c]$ and $s_y[c]$ values (= horizontal and vertical sampling factors)**
3 color components	$N_c = 3$	$s_x[0] = 1$ $s_x[1] = s_x[2] = 2$ $s_y[0] = s_y[1] = s_y[2] = 1$ (*4:2:2 chroma sampling*)
Alpha channel only	$N_c = 1$	$s_x[0] = s_y[0] = 1$ (*4:0:0 chroma sampling*)

Table 1. Number of components and chroma sampling

- **Color transformation**: Cpih shall be set to 0 (no color transformation). This implies that the 3 components of the image shall use the YCbCr digital representation. Moreover, Senders shall use the following order of components in the JPEG XS 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).

- **Input bit depth**: B[c] shall be set to 10 for all values of [c] (i.e. all components).

- **Number of horizontal wavelet transformations**: $N_{L,x}$ shall be set to 5

- **Number of vertical wavelet transformations**: $N_{L,y}$ shall be set to 2

- **Quantizer type**: Qpih shall be set to 1 (uniform quantizer)

- **Level**: shall be set to either 2k-1, 4k-2 or 8k-2, as required by the targeted use case (see Capability Sets and Interoperability Points defined in Appendix B).

  Note: Level defines constraints on the maximum dimensions and framerate of the images in the uncompressed domain. The lower bound of this level is just above the maximum value of the preceding level. Levels are defined ISO/IEC 21122-2. Implementers should note that Part 2 defines a 4K-1 level, but that level is not used in this document.

- **Sublevel**: encoders shall set Sublevel to Sublev3bpp or Sublev4bpp. When the bpp is less than or equal to 3bpp, Sublev3bpp shall be used. When the bpp is above 3bpp up to 4bpp, Sublev4bpp shall be used. Implementations compliant with this TR shall not set

Sublevel to any other values.[5]

- Compression bit rate shall be constrained to a maximum of 4bpp (see Appendix B - *Capability Sets and Interoperability Points*).

  Sublev4bpp shall be signaled setting the binary value of *Plev* field to 'xxxx xxxx 0000 0110'.
  (Note: *Plev* value to be published in ISO/IEC 21122-2:2022/AMD1)

- SCHAR: shall be set to 0.

9.1.3 JPEG XS Elementary Stream

JPEG XS streams emitted by Senders that are compliant with this TR shall conform to ISO/IEC 21122-1 and 21122-2.

These streams shall be encapsulated in an MPEG-2 Transport Stream according to Rec. ITU-T H.222.0 with the following additional constraints:

In the PES Header, the JPEG XS video streams (*stream_type* = 0x32) are carried using the same PES packet syntax as *private_stream_1*, *stream_id* shall be set to 0xBD

Note: Except for the *tcod* field, all fields from the *jxes_header* are also found in the JXS video descriptor in the Program Map Table. This is done to allow for a greater flexibility in implementations, at a cost of a negligible overhead.

Values found in the JXS video descriptor and in the *jxes_header* shall be consistent. If inconsistent values are found, values from the *jxes_header* shall take precedence.

Senders shall ensure that the JPEG XS Video Access Unit contains the correct number of contiguous JPEG XS codestreams, depending on whether the video stream is progressive or interlaced. The Video Access Unit shall contain a single codestream in the progressive case and shall contain two code streams, one codestream per field, in the interlace case. The video access unit always represents a frame.

Senders shall process and transmit codestreams in raster order to achieve minimum latency. Receivers should not have to handle out of order codestream components

9.1.3.1 JPEG XS Video descriptor

Senders shall construct the *JPEG XS video descriptor* as specified in Clause 2.6.127 of Rec. ITU-T H.222.0.

Senders shall include the *JPEG XS video descriptor* in the Program Map Table for the JPEG XS service component.

9.1.4 JPEG XS Elementary Stream Header

[5] Sublev4bpp will be specified in ISO/IEC 21122-2:2022/AMD1.

Senders shall construct the *jxes_header* as specified in Table W-1 of Rec. ITU-T H.222.0.

9.1.4.1 Field coding and frame rate

Field coding and frame rate shall be set by the *frat* field, whose semantics are defined in ISO/IEC 21122-3.

Senders shall set the *frat* field in the JPEG XS video descriptor and in each JPEG XS elementary stream header as follows:

For interlaced signals, fields shall be transmitted in temporal order of appearance in all cases. The first line of the first transmitted field shall be located temporally above the first line of the second field (top-field-first). The *frat* field in the header and descriptors allows signaling of bottom-field-first, however this bottom-field-first mode shall not be used. The *Interlace_Mode* of the *frat* field shall be set to '1'.

For progressive signals, the *Interlace_Mode* of the *frat* field shall be set to '0'.

9.1.4.2 Maximum Bitrate

Senders shall set the maximum bitrate in the *brat* field of the JPEG XS video descriptor and in each JPEG XS elementary stream header according to the semantics defined in ISO/IEC 21122-3.

9.1.4.3 Color Space Specification & Dynamic Range

Color space information shall be specified in the JPEG XS video descriptor and in each JPEG XS elementary stream header using *color_primaries*, *transfer_characteristics*, *matrix_coefficients*, and *video_full_range_*(three 8-bit fields and one 1-bit field) as described in Rec. ITU-T H.222.0. These fields shall be coded according to the semantics with the same name defined in Rec. ITU-T H.273.

Note: Table 2. below summarizes the signaling code values a sender might employ for commonly used color spaces. This information may be found in ISO/IEC 21122-3 and is provided here for the convenience of implementers.

Color space	Color primaries code	Transfer characteristics code	Matrix coefficients code	Video full range flag	Notes
Rec. ITU-R BT.709-6	1	1	1	0	BT 709 SDR
Rec. ITU-R BT.2020-2	9	14 (10 bit)	9 (non-constant luminance)	0	Wide Color Gamut SDR
Rec. ITU-R BT.2020-2	9	15 (12 bit)	9 (non-constant luminance)	0	Wide Color Gamut SDR
Rec. ITU-R BT.2020-2	9	14 (10 bit)	10 (constant luminance)	0	Wide Color Gamut SDR
Rec. ITU-R	9	15 (12 bit)	10 (constant	0	Wide Color

BT.2020-2			luminance)		Gamut SDR
Rec. ITU-R BT.2100-2	9	16	9 (Y'CbCr)	0	Wide Color Gamut PQ HDR
Rec. ITU-R BT.2100-2	9	18	9 (Y'CbCr)	0	Wide Color Gamut HLG/HDR

Table 2 (Informative) – Selected examples of color space specification

9.1.4.4 Mastering Display Metadata

Note: Mastering Display Color Volume is described in SMPTE ST 2086. This Metadata transport is defined in SMPTE ST 2108-1 and ST 2108-2. Mastering Display Metadata is fixed length and described in ST 2108-1.

At the MPEG-2 Transport Stream layer as described in ITU-T H.222.0 senders may specify the Mastering Display Metadata in the JPEG XS video descriptor.

Alternatively, Senders may utilize SMPTE ST 2038 ancillary data packets for Mastering Display Metadata.

In the case that both are utilized, senders shall keep the metadata consistent and avoid conflict.

When Senders specify the Mastering Display Metadata in the *JPEG XS video descriptor*. Senders shall set the *mdm_flag* to '1', as described in Rec. ITU-T H.222.0.

Mastering Display Metadata shall be specified using the following fields:

- *X_c0, Y_c0, X_c1, Y_c1, X_c2, Y_c2, X_wp, Y_wp, L_max* and *L_min*, as defined in SMPTE ST 2086:2018 "Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images"

- *MaxFALL* and *MaxCLL*, as defined in ANSI/CTA 861-H

If the Mastering Display Metadata is unknown or not included in the video input, when the stream is generated, then the Mastering Display Metadata shall not be included in the in the JPEG XS video descriptor, and the *mdm_flag* shall be set to '0

9.1.4.5 JPEG XS Still Pictures

Rec. ITU-T H.222.0 includes still picture mode. Senders shall set the still_mode field in the *JPEG XS video descriptor* to '0'. Senders compliant with this TR shall not enable still picture mode.

9.2 Audio Transport (PCM and Non-PCM signals)

Audio signals shall be sampled at a rate of 48 kHz, using a sampling clock which shall be synchronous with the Program Clock Reference (PCR) timebase.

Audio signals shall be formatted in accordance with AES3 and may contain PCM audio samples or non-PCM signals.

Audio signals shall be formatted for transport in accordance with SMPTE ST 302. As noted in SMPTE ST 302, the audio is organized into multiple Packetized Elementary Streams (PES), and each PES shall contain no more than 8 channels (four AES3 signals) of audio.

Senders and Receivers shall support, at a minimum, the Capability Set A requirements defined in Table 3 below.

Receivers shall be capable of simultaneously receiving and processing the number of AES3 signals (channel pairs) shown in Table 3 below. Receivers should be capable of ignoring additional audio PMT entries and PIDs beyond those they can process.

Capability Set	Bit Depth, Sampling, PIDs and AES Channel Pairs per PID
A	24 Bits, 48 kHz 1-4 AES3 (channel pairs) per PID up to 4 audio PIDs (up to 16 audio channels total)
B	24 Bits, 48 kHz 1-4 AES3 (channel pairs) per PID up to 4 audio PIDs (up to 32 audio channels total)
C	24 Bits, 48 kHz 1-4 AES3 (channel pairs) per PID up to 4 audio PIDs (up to 32 audio channels total)

Table 3: Audio Capability Sets

Receivers should incorporate a selection mechanism that allows the user to choose which audio signals (channel pair) to process, from amongst those sent.

The order of the streams, as listed in the PMT, shall constitute the ordering of the streams presented to the user in any user presentation or selection mechanism.

Senders should assign ascending MPEG-2 transport stream PID values to SMPTE ST 302 audio elementary streams such that the first audio stream within the PMT has the lowest PID and the last audio stream has the highest PID.

Unless overridden by the user, by default, the Receiver shall receive and process audio streams, and assign the channels inside of them in the order they appear within the PMT.

Senders shall mark each audio PES with a Presentation Time Stamp (PTS) corresponding to a video frame in the source video as required by SMPTE ST 302.

Upon receipt and processing of SMPTE ST 302 audio streams, Receivers shall synchronize the audio streams such that upon presentation, audio samples from Access Units containing the same PTS value shall be synchronized to the video within +/- 2 ms.

Audio Access Units with the same PTS value shall be aligned with each other in a sample-phase-accurate manner.

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 referred to SMPTE RDD 19.

9.3 Ancillary Data

Subject to the exceptions and limits noted below, Senders and Receivers shall support the transport of SMPTE ST 291-1 data using the method specified in SMPTE ST 2038.

Note: Most SDI signals include Horizontal Ancillary (HANC) and Vertical Ancillary (VANC) data packets formatted in accordance with SMPTE ST 291-1. ANC data packets of this format are also transported via SMPTE ST 2110-40.

Note: Since video metadata is included in the video descriptor, access unit elementary header, and potentially within SMPTE ST 2038 ANC packets, it may be possible that a contradiction could occur between this metadata.

Implementers shall follow best standard practices as indicated in this document to avoid duplication of metadata. If duplication is unavoidable, then metadata shall be consistent in each location with no difference. If a receiver detects a conflict, the video descriptor or access unit elementary header shall prevail.

9.3.1 HANC and VANC data which are excluded from transport

Section 9.2 above describes audio encapsulation and transport. Senders shall not use HANC or VANC for the encapsulation of audio.

The Embedded Audio Control Packet defined in SMPTE ST 299-1 should not be transmitted by Senders and shall be ignored by Receivers if present in the SMPTE ST 2038 payload.

Receivers shall generate a locally correct Embedded Audio Control Packet based on their specific configuration if they are creating an SDI output.

EDH, CRC, and Line Number information shall not appear in SMPTE ST 2038 streams. While present in the ancillary data spaces of SDI, this information shall not be formatted as ANC packets according to SMPTE ST 291-1.

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

For each of the Capability Sets defined, Table 4 below indicates the maximum number of 10-bit words of SMPTE ST 291-1 ANC data, carried using SMPTE ST 2038, which shall be supported by compliant Receivers.

Senders shall not exceed the maximum data rates in Table 4 below.

Receivers shall be able to process ANC data streams that conform to the bit rates listed in Table 4

below.

Capability Set	Description	Number of 10-bit words/sec
Capability Set A, B & C	*Number of 10-bit words of ANC Data payload per second (2096 per frame @ 50Hz) (1748 per frame @ 59.94Hz)*	104,800
	Transport Maximum bit rate (Rmax) (bits/second)	2,500,000

Table 4 – Amount of ANC Data Transported Using ST 2038 to be Supported by the Different Capability Sets

Note: In Table 4, 104800 10-bit words per second is equivalent to 8 full-sized 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.

The ANC data transmitted by Senders shall be compliant with the T-STD model as specified in Rec. ITU-T H.222.0, using an elementary stream buffer size (*Bn*) as described in Table 4 above.

The transport buffer *TBn* for the ANC service shall be 512 bytes as specified in Rec. ITU-T H.222.0.

For ST 2038 ANC data, the transport buffer shall be drained (*Rxn*) at 1.2 times *Rmax* as per Table 4 above.

Note: Users should note that ST 2038 ANC data is encapsulated in 188 byte MPEG-2 Transport Stream packets. The resulting transport stream 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 transport stream layer (ignoring the header overhead).

10. IP Encapsulation and Forward Error Correction

Senders shall map MPEG-2 Transport Stream for transport over an IP network in accordance with SMPTE ST 2022-2.

If FEC is implemented, Senders shall construct the FEC stream in accordance with SMPTE ST 2022-1.

Receivers shall be able to accept IP streams that are compliant with SMPTE ST 2022-2.

If FEC is implemented, Receivers shall be able to accept and process FEC streams constructed in accordance with SMPTE ST 2022-1.

Although SMPTE 2022-2 allows for 1, 4 and 7 MPEG-2 Transport Stream packets per IP datagram, Senders that are compliant with this TR shall emit streams with 7 TS packets per IP datagram and shall not emit streams with 1 or 4 transport stream per datagram.

11. Bibliography (Informative)

- CEA-CEB20 “A/V Synchronization Processing Recommended Practice”
    - CEA-608-B-2000 “Line 21 data services” Rec. ITU-R BT 709-6 “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 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 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 Asynchronous 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”

Appendix A (Informative) Program Map Table Example

This appendix contains an annotated example of a Program Map Table.

	value	bits	contents
PMT Header	0x02	8	table_id
	1011<PMT-length>	16	length of the PMT section
	<program num>	16	program number
	11<version><cn>	8	version (5) current/next (1)
	0x0000	16	section #, last section #
	111<PCR-PID>	16	what PID carries the PCR
	1111<pgm-info-length>	16	length of the program descriptors which follow
	program-level descriptors if any	8*n	variable length, byte aligned
JPEG XS Video Stream	0x32	8	stream_type (0x32 == JPEG XS)
	111<pid-value>	16	which PID carries this PES
	1111<ES-info length>	16	length of the elementary descriptors which follow if any
	0x3F<len1>0x14<len2>00 <horizontal_size16> <vertical_size16> <brat32><frat32> <other ES information>		Extension Descriptor (0x3F) containing JPEG XS Video Descriptor (0x14)
	other descriptors		other descriptors
SMPTE ST 302 Audio Stream (PMT could contain several of these)	0x06	8	stream_type (0x06 == private data)
	111<pid-value>	16	which PID carries this PES
	1111<ES-info-length>	16	length of the elementary descriptors which follow
	0x05<len>0x42535344	8*6	registration descriptor (0x05) for SMPTE 302 audio. 0x42535344 =”BSSD”
	other descriptors	8*n	other descriptors

SMPTE ST 2038 ANC Data Stream	0x06	8	stream_type (0x06 == private data)
	111<pid-value>	16	which PID carries this PES
	1111<ES-info-length>	16	length of the descriptors which follow
	0x05<len>0x56414E43	8*6	registration descriptor (0x05) for SMPTE 2038 VANC data. 0x56414E43 = “VANC”
	0xC4<len><descriptor>	var	anc_data_descriptor (mandatory, but contents is optional)

Appendix B – Capabilities and Interop Points

Interop Points	Capability Set	Format & Frame Rate*	Pixels per Frame	Max Coding Efficiency Mbps	Max Coding Efficiency Bpp	Min Coding Efficiency Mbps	Min Coding Efficiency Bpp	Bit Depth	Color Sampling	Color Space	Audio Capability Set	JPEG XS Profile			Reference Uncompressed Mbps Video, Mbps
												Profile	Level	Sublevel	
1	A	720px1280/59	55,240,759	92	1.67	221	4.00	10bit	4:2:2	Rec. ITU-R BT.709-6	A	High 444.12	1k-1	Sublev4bpp	1105
2	A	720px1280/50	46,080,000	77	1.67	184	3.99	10bit	4:2:2	Rec. ITU-R BT.709-6	A	High 444.12	1k-1	Sublev4bpp	922
3	A	1080ix1920/29	62,145,854	104	1.67	249	4.01	10bit	4:2:2	Rec. ITU-R BT.709-6	A	High 444.12	2k-1	Sublev4bpp	1243
4	A	1080ix1920/25	51,840,000	86	1.66	207	3.99	10bit	4:2:2	Rec. ITU-R BT.709-6	A	High 444.12	2k-1	Sublev4bpp	1037
5a	A	1080px1920/59	124,291,708	207	1.67	497	4.00	10bit	4:2:2	Rec. ITU-R BT.709-6	A	High 444.12	2k-1	Sublev4bpp	2486
5b	A	1080px1920/59	124,291,708	207	1.67	497	4.00	10bit	4:2:2	Rec. ITU-R BT.2100-2 (PQ)	A	High 444.12	2k-1	Sublev4bpp	2486
5c	A	1080px1920/59	124,291,708	207	1.67	497	4.00	10bit	4:2:2	Rec. ITU-R BT.2100-2 (HLG)	A	High 444.12	2k-1	Sublev4bpp	2486
6a	A	1080px1920/50	103,680,000	173	1.67	415	4.00	10bit	4:2:2	Rec. ITU-R BT.709-6	A	High 444.12	2k-1	Sublev4bpp	2074
6b	A	1080px1920/50	103,680,000	173	1.67	415	4.00	10bit	4:2:2	Rec. ITU-R BT.2100-2 (PQ)	A	High 444.12	2k-1	Sublev4bpp	2074
6c	A	1080px1920/50	103,680,000	173	1.67	415	4.00	10bit	4:2:2	Rec. ITU-R BT.2100-2 (HLG)	A	High 444.12	2k-1	Sublev4bpp	2074
7a	B	2160px3840/59	497,166,833	829	1.67	1989	4.00	10bit	4:2:2	Rec. ITU-R BT.709-6	B	High 444.12	4k-2	Sublev4bpp	9,943
7b	B	2160px3840/59	497,166,833	829	1.67	1989	4.00	10bit	4:2:2	Rec. ITU-R BT.2100-2 (PQ)	B	High 444.12	4k-2	Sublev4bpp	9,943
7c	B	2160px3840/59	497,166,833	829	1.67	1989	4.00	10bit	4:2:2	Rec. ITU-R BT.2100-2 (HLG)	B	High 444.12	4k-2	Sublev4bpp	9,943
8a	B	2160px3840/50	414,720,000	691	1.67	1659	4.00	10bit	4:2:2	Rec. ITU-R BT.709-6	B	High 444.12	4k-2	Sublev4bpp	8,294
8b	B	2160px3840/50	414,720,000	691	1.67	1659	4.00	10bit	4:2:2	Rec. ITU-R BT.2100-2 (PQ)	B	High 444.12	4k-2	Sublev4bpp	8,294

Interop Points	Capability Set	Format & Frame Rate*	Pixels per Frame	Max Coding Efficiency Mbps	Max Coding Efficiency Bpp	Min Coding Efficiency Mbps	Min Coding Efficiency Bpp	Bit Depth	Color Sampling	Color Space	Audio Capability Set	JPEG XS Profile			Reference Uncompressed Mbps Video, Mbps
												Profile	Level	Sublevel	
8c	B	2160px3840/50	414,720,000	691	1.67	1659	4.00	10bit	4:2:2	Rec. ITU-R BT.2100-2 (HLG)	B	High 444.12	4k-2	Sublev4bpp	8,294
9a	C	4320px7680/59	1,988,667,333	3314	1.67	7955	4.00	10bit	4:2:2	Rec. ITU-R BT.709-6	C	High 444.12	8k-2	Sublev4bpp	39,773
9b	C	4320px7680/59	1,988,667,333	3314	1.67	7955	4.00	10bit	4:2:2	Rec. ITU-R BT.2100-2 (PQ)	C	High 444.12	8k-2	Sublev4bpp	39,773
9c	C	4320px7680/59	1,988,667,333	3314	1.67	7955	4.00	10bit	4:2:2	Rec. ITU-R BT.2100-2 (HLG)	C	High 444.12	8k-2	Sublev4bpp	39,773
10a	C	4320px7680/50	1,658,880,000	2765	1.67	6636	4.00	10bit	4:2:2	Rec. ITU-R BT.709-6	C	High 444.12	8k-2	Sublev4bpp	33,178
10b	C	4320px7680/50	1,658,880,000	2765	1.67	6636	4.00	10bit	4:2:2	Rec. ITU-R BT.2100-2 (PQ)	C	High 444.12	8k-2	Sublev4bpp	33,178
10c	C	4320px7680/50	1,658,880,000	2765	1.67	6636	4.00	10bit	4:2:2	Rec. ITU-R BT.2100-2 (HLG)	C	High 444.12	8k-2	Sublev4bpp	33,178

* Video format is given as active lines, scanning (interlaced or progressive) and frame rate (—59.9 is equivalent to —60/1.001 while —29.97 is equivalent to —30/1.001