



Video Services Forum (VSF) Technical Recommendation TR-07

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



May 26, 2026
TR-07:2026

© 2026 Video Services Forum

This work is licensed under the Creative Commons Attribution-No Derivatives 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nd/4.0/>

or send a letter to Creative Commons, PO Box 1866
Mountain View, CA 94042, USA.



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 IMPLEMENTER 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THE FOREGOING NEGATION OF DAMAGES IS A FUNDAMENTAL ELEMENT OF THE USE OF THE CONTENTS HEREOF, AND THESE CONTENTS WOULD NOT BE PUBLISHED BY VSF WITHOUT SUCH LIMITATIONS.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.

The VSF has published a previous version of this TR known as *VSF TR-07:2022 Transport of JPEG XS Video in MPEG-2 Transport Stream over IP*. This document, *VSF TR-07:2026* adds the following capabilities and clarifications to the original version.

1. **Capability Set and Interop Points for JPEG XS Temporal Differential Coding (TDC).** TDC can optimize quality at a given rate or allow reduced bit rates for some content.
2. **Addition of Interop Points** for some non-broadcast frame rates.
3. **Addition of Traffic Shaping and Delivery Timing Model** in order to provide implementers with clear information for low latency transmission and improved interoperability
4. **Clarifications, Examples, References and Constraints,** including specific supported functionality in order to improve interoperability
5. **Updated normative references** to current versions
6. Consolidation of Rows to provide a more concise Appendix B Capability Set and Interop Points table

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) | 5 |
| 1.1 | Contributors | 5 |
| 1.2 | About the Video Services Forum | 5 |
| 2. | Conformance Notation | 6 |
| 3. | Normative References | 7 |
| 4. | Acronyms | 8 |
| 5. | Definitions | 9 |
| 6. | System Overview/Use Case (Informative) | 9 |
| 7. | Organization and Signaling of a VSF TR-07 Stream | 10 |
| 8. | Capability Sets and Interoperability Points | 11 |
| 9. | Essence Service Components | 12 |
| 9.1 | JPEG XS Video | 12 |
| 9.1.1 | MPEG-2 Transport Stream and JPEG XS Codestream | 12 |
| 9.1.2 | JPEG XS Codestream Restrictions | 13 |
| 9.1.3 | JPEG XS Elementary Stream | 15 |
| 9.1.3.1 | JPEG XS Video descriptor | 15 |
| 9.1.4 | JPEG XS Elementary Stream Header | 16 |
| 9.1.4.1 | Field coding and frame rate | 16 |
| 9.1.4.2 | Maximum Bitrate | 16 |
| 9.1.4.3 | Color Space Specification & Dynamic Range | 16 |
| 9.1.4.4 | Mastering Display Metadata | 17 |
| 9.1.4.5 | JPEG XS Still Pictures | 18 |
| 9.2 | Audio Transport (PCM and Non-PCM signals) | 18 |
| 9.3 | Ancillary Data | 19 |
| 9.3.1 | HANC and VANC data which are excluded from transport | 19 |
| 9.3.2 | Limits on the total amount of ANC data to be transported | 20 |
| 10. | Transport Stream Requirements/Buffer Model | 21 |
| 10.1 | General | 21 |
| 10.2 | Buffer Model from ISO/IEC 21122-2 | 21 |
| 10.3 | Gapped Encoding and Decoding for Low Latency | 21 |
| 10.4 | Low-Latency Gapped Transmission Timing Model | 22 |
| 10.4.1 | Overview and Definitions | 22 |
| 10.4.2 | Low-Latency Gapped Timing Model – Format Specific Parameters | 23 |
| 10.4.3 | Low-Latency Gapped Transmission Timing Model - Video ES to TS | 24 |
| 10.4.4 | Timing of ANC data transmission in the Gapped Model | 25 |

| | | |
|----------|---|----|
| 10.5 | Linear transmission model | 25 |
| 10.5.1 | Linear transmission model - video | 25 |
| 10.5.1.1 | Low Latency Linear model - preserving low latency | 26 |
| 10.5.1.2 | Encoding Implementation Latency | 26 |
| 10.5.1.3 | ANC data transmission in Linear Model | 27 |
| 10.5.1.4 | Audio considerations | 27 |
| 10.5.1.5 | ANC considerations | 27 |
| 10.6 | Transmission Traffic Shape Models | 27 |
| 10.6.1 | Network Compatibility Model | 27 |
| 11. | IP Encapsulation and Forward Error Correction | 28 |
| 12. | Bibliography (Informative) | 29 |
| | Appendix A (Informative) Examples | 30 |
| | Example PMT | 31 |
| | Example PES header | 32 |
| | Appendix B (Informative) Alignment of PES and JXS headers to TS packets | 33 |
| | Appendix C Capabilities and Interop Points | 36 |
| | High 444.12 Profile | 36 |
| 12.1 | TDC 444.12 Profile | 38 |

1. Introduction (Informative)

The VSF created the JPEG XS Activity Group which 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.

As a follow up to the revision work started in February of 2020 and concluding with the publication of TR-07 in April 2022, this activity group was formed to expand and improve this VSF Technical Recommendation. The main focus is on improving interoperability for JPEG-XS coded video over MPEG-2 Transport Stream and including the encapsulation in IP/RTP via ST 2022-2. At the JPEG-XS elementary stream level, the TDC profile will be included in this revision.

1.1 Contributors

Contributors to this document include:

- John Dale, Activity Group Chairman, Media Links, Inc.
- Alain Champenois, Matrox
- Christian Kellner, Imagine
- Eben Jenkins, Leader
- Grant McGilvray, NBC
- John Mailhot, Imagine Communications
- Tim Bruylants, intoPIX SA

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 the VSF Operations Manager (opsmgr@vsf.tv)

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 (2025) | ISO/IEC 13818-1:2025: "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 21122-1:2024 "Information technology — JPEG XS Low-latency Lightweight Image Coding System — Part 1: Core coding system”

ISO/IEC 21122-2:2024 “Information technology — JPEG XS Low-latency Lightweight Image Coding System — Part 2: Profiles and buffer models”

ISO/IEC 21122-3:2024 “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”

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

¹Note: 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.

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”

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 |
| PAT | Program Association Table |
| 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 |
| TDC | Temporal Differential Coding (profile of JPEG XS) |
| TR | Video Services Forum Technical Recommendation ² |
| TS | Transport Stream |
| 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 |

² 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.

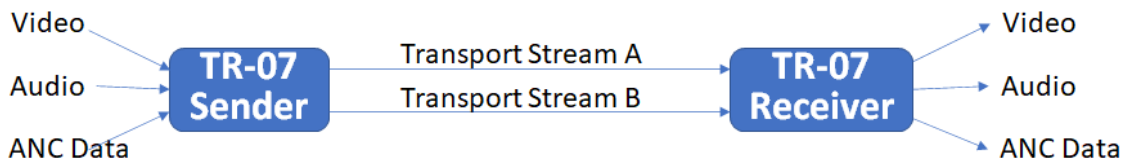


Figure 1 – Example system for transmission of real time television signal over IP

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, at a minimum a compliant transmitter can send a stream which will be processed by the compliant receiver without over or under-flow. The compliant receiver shall receive the specified sender stream without errors. 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 be constructed in compliance with Rec. ITU-T H.222.0.

MPEG-2 Transport Streams shall contain a Program Association Table (PAT) defining a single Program and shall contain a single Program Map Table (PMT). The PAT and PMT entries shall be repeated at least every 500ms seconds.

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, ISO/IEC 21122-2, and ISO/IEC 21122-3 as described in Section 9.1.

The Program Map Table (PMT) for the JPEG XS video stream shall be constructed as defined in Rec. ITU-T H.222.0. An example for such 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 (*private_stream_1*), with a *stream_type* = 0x32 (JPEG XS Video Stream). Further details on the *JXS_video_descriptor* may be found in Section 9.1.

The PES header shall have the *PES_header_data_length* always set to a value of 5. The value of exactly 5 bytes supports only the PTS signalid in the PES header. An example is shown in Appendix A.

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 one SMPTE ST 2038 PES stream.

The Program Map Table shall contain no more than eight (8) 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 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 AT TDC

- Same as Capability A but also including the TDC profile

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 BT TDC

- Same as Capability B + the TDC profile is allowed

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

Capability Set CT TDC

- Same as Capability C + the TDC profile is allowed

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. It is not an exhaustive list.

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

Figure 2 below is a 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.

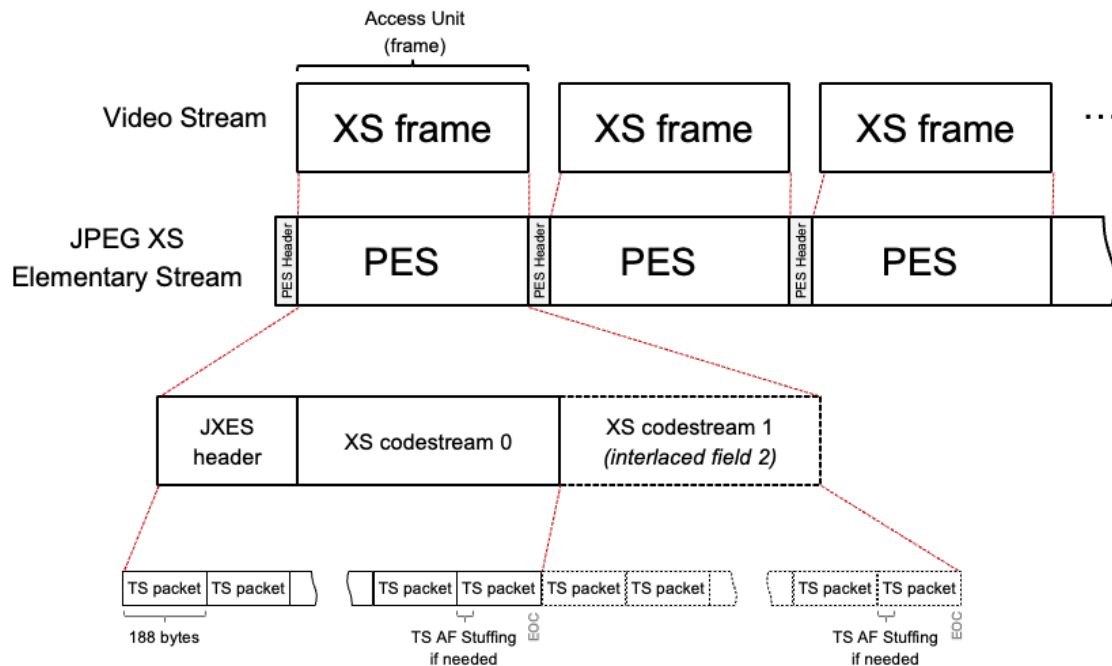


Figure 2 – Structure and order of JPEG XS Video Access Units

Note that this figure shows only the JPEG XS video content. An actual MPEG2-TS will also contain audio, ancillary data, PAT/PMT etc. in separate TS packets.

Figure 2. Explanation and Requirements:

The EOC (0xFF11) marker of codestream 0, and codestream 1 when present, shall be the last 2 bytes of the TS packet (using TS Adaptation Field stuffing if needed). For interlace video, the start of codestream 1 shall start in the next TS packet carrying the JPEG XS video data.

JXS PES shall not start with adaptation field

For interlace, data bytes from codestream 0 and codestream 1 shall not be included in the same TS packet

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.

9.1.2 JPEG XS Codestream Restrictions

JPEG XS codestreams generated under this recommendation shall contain a constant number of bits per frame, as determined by the BPP and other codec settings. For the case of interlaced signals, the frame includes the codestreams of both fields.

JPEG XS Codestreams emitted by Senders that are compliant with this TR shall conform to either the “High444.12” or the “TDC444.12” profile as specified in ISO/IEC 21122-1, ISO/IEC 21122-2, and ISO/IEC 21122-3, with the following parameters, settings and changes:

JPEG XS Codestreams emitted by Senders that are compliant with this TR shall be compliant with

- **Number of components, chroma sampling format and alpha channel:** If there is an alpha channel, it shall be transported in its own codestream.

Codestream parameters and the JPEG XS Video Information box parameters shall be set according to one of the allowed configurations shown in Table 1. below.

- **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).

| Allowed Configurations | N_c value (= number of components) | $s_x[c]$ and $s_y[c]$ values (= horizontal and vertical sampling factors) | Profile | Ppih | Cpih | SCHAR |
|---|---|--|----------------|--------|------------------|----------------|
| 3 color components 4:2:0 YCbCr sampling | $N_c = 3$ | $s_x[0] = s_y[0] = 1$ $s_x[1] = s_x[2] = 2$ $s_y[1] = s_y[2] = 2$ | High 420.12 | 0x4240 | 0 | 3 |
| | | | TDC 444.12 | 0x4a45 | | |
| 3 color components 4:2:2 YCbCr sampling | $N_c = 3$ | $s_x[0] = 1$ $s_x[1] = s_x[2] = 2$ $s_y[0] = s_y[1] = s_y[2] = 1$ | High 444.12 | 0x4a40 | 0 | 0 |
| | | | TDC 444.12 | 0x4a45 | | |
| 3 color components 4:4:4 YCbCr sampling | $N_c = 3$ | $s_x[0] = s_x[1] = s_x[2] = 1$ $s_y[0] = s_y[1] = s_y[2] = 1$ | High 444.12 | 0x4a40 | 0 (No RCT) | 1 |
| | | | TDC 444.12 | 0x4a45 | | |
| 3 color components 4:4:4 RGB sampling | $N_c = 3$ | $s_x[0] = s_x[1] = s_x[2] = 1$ $s_y[0] = s_y[1] = s_y[2] = 1$ | High 444.12 | 0x4a40 | 1 (RCT) | 2 |
| | | | TDC 444.12 | 0x4a45 | | |
| alpha channel only 4:0:0 sampling | $N_c = 1$ | $s_x[0] = s_y[0] = 1$ | High 444.12 | 0x4a40 | 0 | 0 ³ |

Table 1. Number of components and chroma sampling.

- **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:** Q_{pih} 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.
- **FBB level:** when employing the TDC 444.12 profile, encoders shall set FBB level to either $Fbblev8bpp^4$ or $Fbblev12bpp$ for levels 1K-1, 2k-1 and 4k-2, and shall set FBB level to $Fbblev8bpp$ for 8k-2 level.

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

³ Note that no value is specified in ISO/IEC 21122-3 for 4:0:0 sampling, however, TR-07 includes the value of '0' for alpha channel 4:0:0 sampling, as well as for 4:2:2 YCbCr sampling.

⁴ $Fbblev8bpp$ will be defined by 21122-2:2024 AMD1 (which is planned for publication in 2026).

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.

While the *tcod* field (as defined in ISO/IEC 21122-3:2024) can be used for time code, it is not mandatory for the receiver to use values of the *tcod* field (senders may set it to all 0's). It is recommended to use ANC data as described in section 9.3 of this document to convey timecodes and other information in applications of this recommendation.

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 codestreams, one codestream per field, in the interlace case. The video access unit always represents a frame.

Special Requirements for interlaced video streams:

In each JPEG XS Video Access Unit the second JPEG XS codestream shall also start in a new TS packet. This shall be achieved by padding (stuffing) the previous TS packet using the Adaptation Field when necessary, so that the beginning of the second video box shall be aligned to the start of a TS packet.

Likewise, the JXS PES should not start with an adaptation field, instead any padding required to achieve TS alignment should be done in the last packet of the prior PES.

Appendix B shows details regarding the allowed/required adaptation field stuffing for these cases.

Combining the requirements above, bytes from codestream 0 and codestream 1 shall not be included in the same TS packet under any circumstance.

Senders shall process and transmit codestreams in raster order to achieve minimum latency. Receivers under this recommendation are not expected 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

Senders shall construct the *jxes_header* as specified in Clause W.3 of Rec. ITU-T H.222.0. An example of this header is shown in Appendix A.

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 BT.2020-2 | 9 | 15 (12 bit) | 10 (constant luminance) | 0 | Wide Color 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.

When video Mastering Display Metadata is present, Senders must include a SMPTE ST 2038 ancillary data packet for Mastering Display Metadata. Additionally, 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.

In the case that both are utilized, senders shall keep the metadata consistent and avoid conflict. Receivers under this recommendation which re-generate SDI or output SMPTE 2110-40 ANC streams shall forward the ANC metadata as transmitted in the SMPTE ST 2038 stream.

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, for all Audio Conformance Levels as listed in Table 3 below, a single PID with 8 audio channels, (4 AES 3 channel pairs) as a base requirement.

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.

| Audio Conformance Level (ACL) | 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 8 audio PIDs (up to 64 audio channels total) |

Table 3: Audio Conformance Levels

Note that the minimal requirement for 1 PID with 8 audio channels, (4 AES channel pairs) applies to all ACLs as a base requirement.

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.

For transmission of time code, senders shall use 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 | <i>Number of 10-bit words of ANC Data payload per second (2096 per frame @ 50Hz) (1748 per frame @ 59.94Hz)</i> | 104,800 |
| | <i>Transport Maximum bit rate (Rmax) (bits/second)</i> | 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 (B_n) as described in Table 4 above.

The transport buffer TB_n 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 (R_{xn}) at 1.2 times R_{max} 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. Transport Stream Requirements/Buffer Model

10.1 General

This section specifies a parametric model for the delivery timing characteristics of JPEG XS video elementary streams and ANC data streams as they exist inside the produced Transport Stream as it leaves the transmission interface of the Sender.

This model builds on the Buffer model of the JPEG XS specification ISO/IEC 21122-2. However a practical receiver needs to also accommodate the network-induced IP packet jitter, TS packet jitter and overall latency which accumulate along the path from sender to receiver. Design of receivers is outside the scope of this standard.

The purpose of this model is to ensure that the encoder's stream is constructed such that any compliant receiver can receive and decode the stream with minimum latency while avoiding over- and under-flow of the receiver elementary stream buffer. The Gapped model allows for a lower overall latency, while the linear model may allow for lower-rate transmissions at the same BPP value.

10.2 Buffer Model from ISO/IEC 21122-2

The ISO/IEC 21122-2 standard documents a buffer model for the elementary stream encoder and decoder as shown in Figure 3 (taken from Figure 1 of the 2024 version of ISO/IEC 21122-2 and duplicated below).

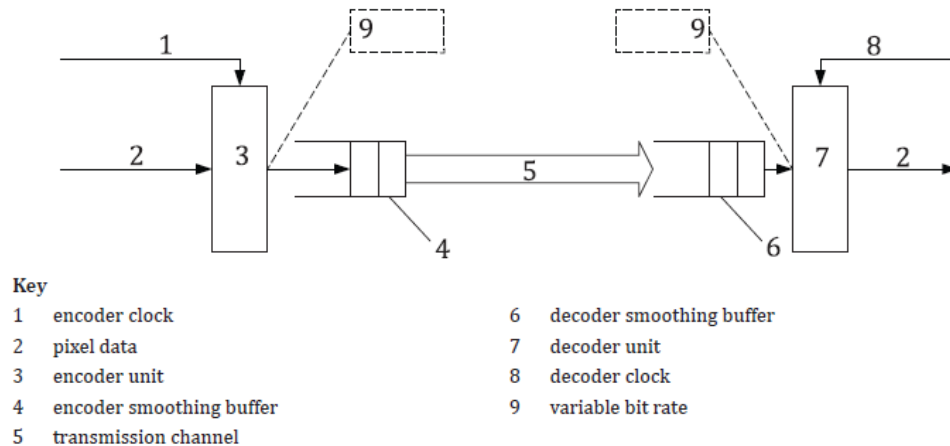


Figure 3 – Buffer model from ISO/IEC 21122-2

This recommendation (VSF TR-07) sets requirements on the transmission channel (item 5 in the figure above, including the encapsulation of the elementary stream into transport stream packets) such that the buffer model of ISO/IEC 21122-2 is not violated by timing changes caused by this transport specification.

10.3 Gapped Encoding and Decoding for Low Latency

The buffer model of 21122-2 assumes that pixels are delivered uniformly throughout the frame period, and that bits are also transmitted through the transmission channel at a constant rate throughout the frame period. In low-latency applications where the video input and output signals are SMPTE ST 2110-20 or SDI, this is not the case – instead pixels are presented to the encoder in a gapped manner in SDI and ST 2110 gapped mode, and the encoder performs its function in an accelerated manner during the active video period. Then the transmission channel must deliver a

higher instantaneous bit rate of elementary stream bits during the active video period to account for this difference. Likewise the decoder must consume the elementary stream faster during the active video period in order to decode the video boxes during the active video period, in order to output a gapped pixel stream in a low latency manner.

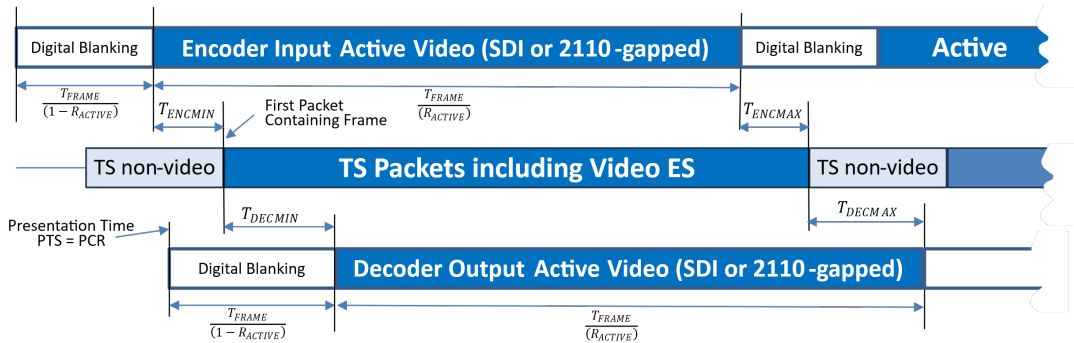


Figure 4 – Gapped encoding timeline, Progressive video

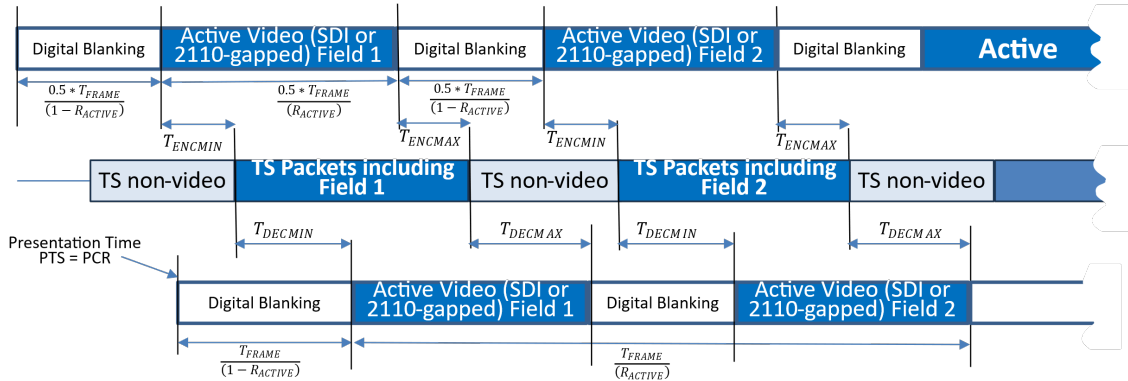


Figure 5 – Gapped encoding timeline, Interlaced video

10.4 Low-Latency Gapped Transmission Timing Model

10.4.1 Overview and Definitions

The gapped encoding and packing timeline, illustrated in Figure 4 for progressive-scan video, consolidates the encoding and transmission of video elementary stream into a period of time shorter than the frame time T_{FRAME} loosely approximating the delivery of active video in the SDI signal, including a gap in the transmission corresponding to the vertical blanking time of SDI. A similar expectation on the reception and decoding of the elementary stream and regeneration of the pixel data is required in order to produce a continuous gapped output signal aligned with SDI or ST2110-20 timing. The progressive scan signal has one gap between each frame of active video.

Interlaced scan signals (and PsF signals) have two gaps, one after each field (or segment in the case of PsF) as shown in Figure 5.

The following variables are used in the equations below:

T_{FRAME} is the time period between consecutive frames of video at the prevailing frame rate

N_{ESBITS} is the number of video elementary stream bits per frame of video

T_{PTSPCR} a time point corresponding to the Presentation Timestamp (PTS) of the video frame, when its value is equal to the clock defined by the Program Clock References (PCR) of the transport stream

T_{ENCMIN} is the smallest time delay between when the first pixel of video enters the encoding process, until elementary stream bits can begin to be withdrawn from the encoder smoothing buffer shown in Figure 3. This value is a function of the prevailing video format

T_{ENCMAX} is the longest time delay between when the last pixel of video enters the encoding process, until elementary stream bits cease to egress from the encoder smoothing buffer shown in Figure 3. This value is a function of the prevailing video format.

T_{DECMIN} is the smallest time delay between when the first elementary stream bit of video bitstream enters the decoder smoothing buffer shown in Figure 3, until pixel data can begin to egress from the decoding process. This value is a function of the prevailing video format.

T_{DECMAX} is the longest time delay between when the last elementary stream bit of video bitstream enters the decoder smoothing buffer shown in Figure 3, until pixel data ceases to egress from the decoding process. This value is a function of the prevailing video format.

R_{ACTIVE} is the ratio of active time to total time within the frame period, a function of the prevailing video format

$T_{ENCBUF} = T_{ENCMIN} = T_{ENCMAX}$ The encoder smoothing buffer and algorithmic latency of ISO/IEC 21122-2 is managed by the codec such that the delay from pixel input to bitstream output is constant throughout the frame (field) period.

$T_{DECBUF} = T_{DECMIN} = T_{DECMAX}$ The decoder smoothing buffer and algorithmic latency of ISO/IEC 21122-2 is managed by the codec such that the delay from bitstream input to pixel output is constant throughout the frame (field) period.

10.4.2 Low-Latency Gapped Timing Model – Format Specific Parameters

For progressively scanned images (but excluding Progressive segmented Frame (PsF) images):

$$R_{ACTIVE} = \left(\frac{1080}{1125} \right)$$

$$T_{ENCBUF} = (\text{constrained in ISO/IEC 21122-2})$$

$$T_{DECBUF} = (\text{constrained in ISO/IEC 21122-2})$$

Note: These values are a function of pixel size, encoding rate, frame rate, and other factors. This gapped model defines the inter-frame gap and R_{ACTIVE} value for all (non-PsF) progressive formats, including 720p, 1080p, 2160p, and 4320p.

Note: for 1080p video at 3BPP 4:2:2, with $N_{L,y}=2$, T_{ENCBUF} is approximately $(12/1125)*T_{FRAME}$, and T_{DECBUF} is approximately twice T_{ENCBUF} , but practical encoders and decoders often use a larger buffer to account for other sources of jitter and implementation constraints.

For interlaced images and Progressive segmented Frame (PsF) images, T_{ENCBUF} and T_{DECBUF} are as constrained in ISO/IEC 21122-2, and R_{ACTIVE} is defined in Table 1:

| System | R_{ACTIVE} |
|--|----------------------------------|
| 525 line interlaced system as specified in Recommendation ITU-R BT.656-5 | $\left(\frac{487}{525}\right)$ |
| 625 line interlaced system as specified in Recommendation ITU-R BT.656-5 | $\left(\frac{576}{625}\right)$ |
| 1125 line interlaced systems and Progressive segmented Frame (PsF) systems as specified in Recommendation ITU-R BT.709-6 | $\left(\frac{1080}{1125}\right)$ |

Table 5 – Ratio of active to total time for interlaced systems

10.4.3 Low-Latency Gapped Transmission Timing Model - Video ES to TS

This specification requires a constant bit rate transmission of the full transport stream. Due to the gapped nature of the availability of the video elementary stream bits in this low-latency model, transport stream level “stuffing” with null packets and/or other payloads is required in order to maintain the TS rate during the gaps. In order to manage overall latency, this specification also models (and sets some limits on) the packing of the ES bits into TS packets, and the unpacking of TS back into ES bits, as shown in Figure 6.

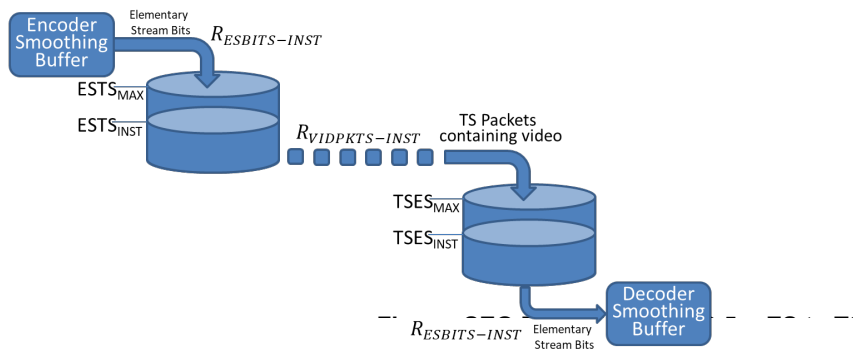


Figure 6 – Gapped encoding timeline, Interlaced video

$R_{ESBITS-INST}$ is the instantaneous bit rate of the Video Elementary Stream at egress from the encoder smoothing buffer. This value is a function of the prevailing video format and the BPP of the encoding. The model also provides video elementary stream bits to the decoder smoothing buffer at this rate during the active period.

Note: the primary necessity is that the elementary stream bits must be delivered to the decoder smoothing buffer before the decoding process requires them.

$R_{VIDPKTS-INST}$ is the instantaneous rate of TS packets containing video during the active period.

$ESTS_{INST}$ is the instantaneous backlog of Elementary Stream bits waiting to be packaged into TS packets.

$ESTS_{MAX}$ is the maximum amount of ES-TS backlog. This shall be limited to $184*8*2$ bits in the gapped model.

$TSES_{INST}$ is the instantaneous backlog of ES bits from received TS packets, waiting to be drained into the decoder smoothing buffer.

$TSES_{MAX}$ is the maximum amount of ES bits from received TS packets waiting to be delivered to the decoder smoothing buffer. This shall be limited to $184*8*2$ bits in the gapped model.

10.4.4 Timing of ANC data transmission in the Gapped Model

The timing diagrams of figures 2 and 3 above show the gapped timing of input video and digital blanking. VANC data arrives during the digital blanking. In order to preserve the low-latency nature of the transmission, any VANC data must be packed into TS packets and multiplexed into the output stream expediently. TENCBUF should be considered an upper bound for the delay in encapsulation and multiplexing of ANC data input to TS output on the gapped mode encoder.

N.n Timing of Audio Transmission

Notwithstanding the gapped or other models of transmission for the Video or ANC data, Audio data must be sent in a smooth consistent manner, and must be delivered before its presentation time (PTS) requires it. Within any 5ms window, the short-term average audio packet delivery rate should approximate the long-term average audio packet delivery rate.

10.5 Linear transmission model

10.5.1 Linear transmission model - video

Regardless of the actual arrival pattern of the pixels to the encoder, in the Linear transmission model the encoding process assumes a timing model in which the entire frame of pixels is encoded over the course of the entire frame time without gaps. In this case the buffer model of ISO/IEC 21122-2 applies directly, and the packing of the elementary stream into TS packets should still be undertaken expediently in order to avoid violation of the decoding buffer within that model.

The ES-TS packing limits of section (#ref) apply to the linear transmission model.

10.5.1.1 Low Latency Linear model - preserving low latency

In practical systems with a gapped pixel arrival, the start of the encoding process is likely offset by at least the duration of the digital blanking period even if it is then carried out in a linear manner, as shown in Figure 7. The minimum achievable latency in this type of system is the length of a blanking period, plus TENCBUF plus TDECBUF, since it must be ensured that the last pixel is decoded before its time to be output. Note that this requires careful specification of the PTS to ensure that the presentation is adequately delayed.

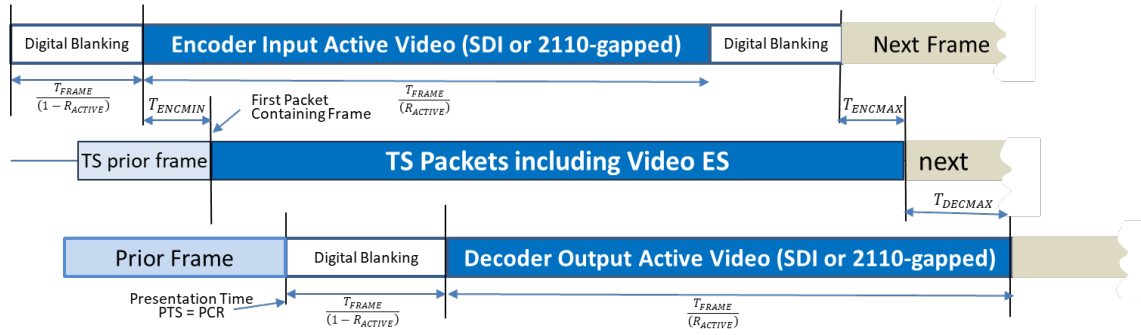


Figure 7 – Linear encoding timeline

Similar considerations apply for the interlaced case even in a linear model.

Even in the case that the input to the encoder (or output of the decoder) is delivered in a non-gapped manner, the relative timing of the first pixel of the encoded video relates to the appearance of the first bit of the frame in the bitstream, and likewise the last bit of the frame in the bitstream relates to the timing of the last pixel of decoded video - Figure 8 shows the same linear arrangement with linear pixel arrive to the encoder, and linear pixels departing from the decoder.

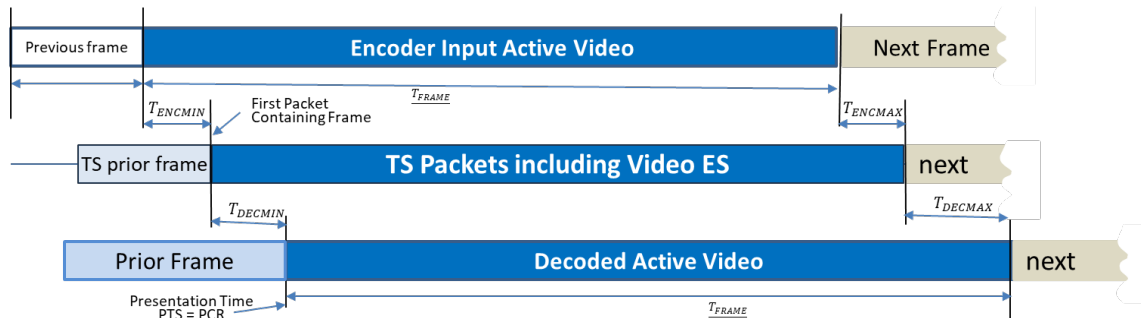


Figure 8 – Linear encoding timeline

10.5.1.2 Encoding Implementation Latency

Encoding operations can take more time or exhibit more variability and more latency depending on their specific implementation. Such encoder implementations must still produce an output that allows a low-latency decoder to decode the signal. This may be accomplished through extra buffering on the encoder and delay of the PTS, but such implementations are outside the scope of this specification. The requirement of this specification is that the output Transport Stream be constructed such that the decoder is able to complete decoding of the picture without starving for bits (buffer underflow). The Presentation Timestamp (PTS) and Program Clock Reference (PCR) are the only communication from encoder to decoder about the time/size of these buffers.

10.5.1.3 ANC data transmission in Linear Model

The ANC data of the linear model must still arrive to the decoder in a timely manner for presentation, ahead of the decoded video frame. Prompt encapsulation in the linear model as noted in section (above for gapped) will ensure that the ANC data is available well before the first pixel of video data output.

10.5.1.4 Audio considerations

In the Linear model, Audio data must be sent in a smooth consistent manner, and must be delivered before its presentation time (PTS) requires it. Within any 5ms window, the short-term average audio packet delivery rate should approximate the long-term average audio packet delivery rate.

10.5.1.5 ANC considerations

In the Linear model, Ancillary data must be sent in a smooth consistent manner, and should be delivered in line with the frame/field it was originally tied to. To optimize latency of the end-to-end system, ANC data should arrive before the beginning of the related video field/frame.

10.6 Transmission Traffic Shape Models

10.6.1 Network Compatibility Model

In order to avoid loss in transport due to burst affects inside switching and routing infrastructure, the stream of RTP packets generated by this specification must be transmitted onto the network in a reasonably smooth manner. The leaky-buffer model shown in Figure 9 applies, where the drain rate is determined by the RTP packet rate, and CMAX has a value of 4 packets.

Transport Stream RTP Senders shall ensure that their sequence of actual transmission instants, as measured on their network egress interface, passes the Network Compatibility Model shown in Figure 9 at all times and in all operating configurations. The model is tested at the output of the Sender, prior to any network-induced delivery impairments.

Packets from the sender shall enter a leaky bucket of infinite capacity at the instant they are emitted from the sender. The bucket drains an RTP packet every T_{DRAIN} seconds, if a packet is available. C_{INST} represents the instantaneous number of packets in the bucket at any time. The value of C_{INST} shall never exceed 4 RTP packets.

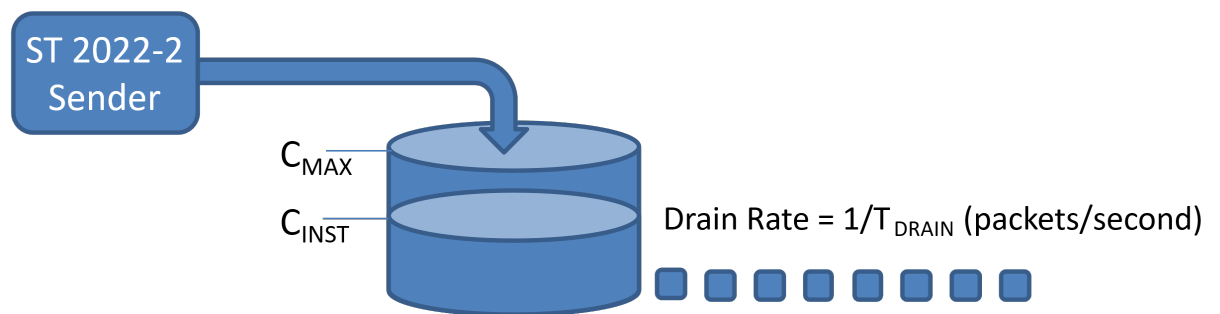


Figure 9 – Network compatibility model

The parameters of the Network Compatibility Model are as follows:

| | |
|----------------------|--|
| R_{NOMINAL} | the long-term average packet rate generated by a sender, in units of packets per second. |
| β | a scaling factor applied to R_{NOMINAL} . In this specification, the value shall be $\beta=1.1$ |
| T_{DRAIN} | the time interval (in seconds) between packets being drained in the Network Compatibility buffer model. $T_{\text{DRAIN}} = (1 / R_{\text{NOMINAL}}) * (1 / \beta)$ |
| C_{INST} | the instantaneous fullness (in packets) of the bucket at any time |

11. IP Encapsulation and Forward Error Correction

The first step in encapsulation is JPEG XS, audio and ancillary ES to MPEG2TS which has been described in section 9.1.1 of this document. This MPEG2TS encapsulation is then wrapped in ST 2022-2 RTP encapsulation as described below with optional ST 2022-1 FEC.

It should be noted that the MPEG2TS layer, defined in this document can be used with other ARQ based encapsulation protocols. However, a receiver compliant to this TR-07 document does not need to support such protocols.

When SMPTE ST 2022-2 is used to transport the streams under this recommendation, the following provisions shall apply. Encapsulation of the Transport Stream defined in this recommendation in other forms of transport is possible, but the details of such encapsulations are outside the scope of this recommendation.

Senders shall map MPEG-2 Transport Stream for transport over an IP network in accordance with SMPTE ST 2022-2. Notwithstanding the provision of ST 2022-2 supporting 204 byte TS packets, all streams generated or consumed under this recommendation shall use 188 byte TS packets.

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 TS packets per IP datagram.

12. Bibliography (Informative)

- CEA-CEB20 “A/V Synchronization Processing Recommended Practice”
 - o 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) Examples

This appendix contains annotated examples of a Program Map Table and a PES header.

Program Map Table example

| | 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 bits) current/next (1 bit) |
| | 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) |

Table 6. - Program MAP Table

Example PMT

Below is output from a widely available WireShark Dissector showing an example PMT:

MPEG2 Program Map Table

```

Table ID: Program Map Table (PMT) (0x02)
1... .. = Syntax indicator: 1
.011 ... .. = Reserved: 0x3
.... 0000 0111 1111 = Length: 127
Program Number: 0x0001
11.. ... = Reserved: 0x3
..01 101. = Version Number: 0x0d
.... ...1 = Current/Next Indicator: Currently Applicable
Section Number: 0
Last Section Number: 0
111. .... .. = Reserved: 0x7
...0 0001 0000 0000 = PCR PID: 0x0100
1111 .... .. = Reserved: 0xf
.... 0000 0000 0000 = Program Info Length: 0
Stream PID=0x0065
Stream type: JPEG XS video stream conforming to ISO/IEC 21122-2 (0x32)
111. .... .. = Reserved: 0x7
...0 0000 0110 0101 = Elementary PID: 0x0065
1111 .... .. = Reserved: 0xf
.... 0000 0010 0000 = ES Info Length: 32
Descriptor Tag=0x3f
Descriptor Tag: Extension Descriptor (ISO) (0x3f)
Descriptor Length: 30
Descriptor Tag Extension: JXS Video Descriptor (0x14)
JXS Video Descriptor
Descriptor version: 0x00
Horizontal size (width): 1920
Vertical size (height): 1080
Max bit rate (brat): 400 Mbps
Frame Rate (frat): 0x0200003c (59.94 fps Progressive)
00.. .... .. = Interlace Mode: Progressive (0)
..00 0010 .... .. = Frame rate denominator: 1.001 (2)
.... .... 0000 0000 .... .. = Reserved: 0
.... .... .... 0000 0000 0011 1100 = Frame rate numerator: 60
Sample characteristics (schar): 0x8090 (10-bit 4:2:2 YCbCr)
1... .... .. = Valid flag: 1
.... .... 1010 .... = Bit depth: 10 (9)

```

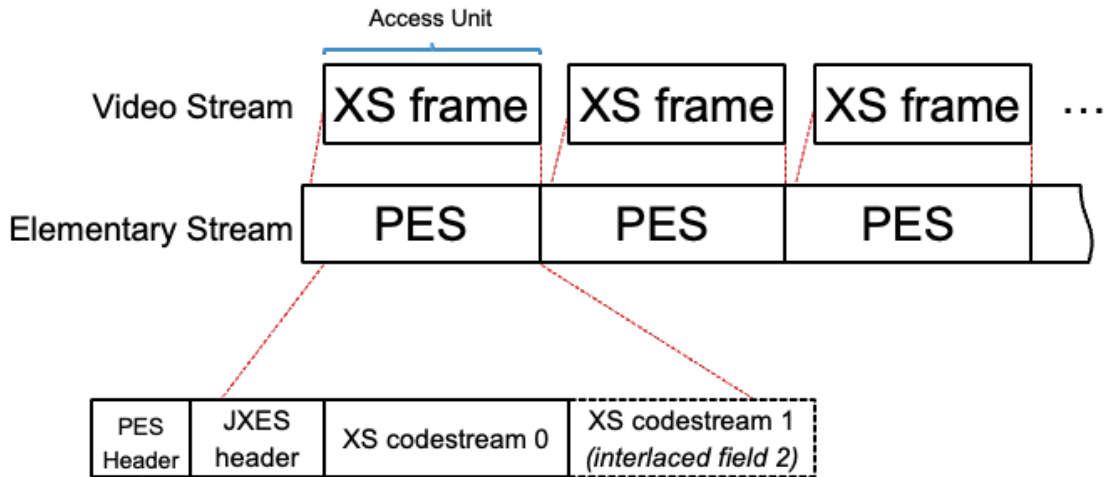
```

    .... 0000 = Sample struct: 4:2:2 YCbCr (0)
    XS profile (Ppjh): High 444.12 (0x4a40)
    XS Plev: 0x1008
    0001 00.. .... = Level: 2k-1 / bayer4k-1 (0x04)
    .... 0..0 1000 = Sublevel: Sublev6bpp (0x08)
    Max bit rate (brat): 3 MB
    Buffer model type: 2
    Color primaries: 1
    Transfer characteristics: 1
    Matrix coefficients: 1
    Video full range flag: 0
    Flags: 0x00
    Stream PID=0x006e <redacted>
    Stream PID=0x00c8 <redacted>
    Stream PID=0x00c9 <redacted>
    Stream PID=0x00ca <redacted>
    Stream PID=0x00cb <redacted>

```

Example PES header

The diagram below details other aspects of the PES header.



```

Packet_start_code_prefix = 0x000001 (24 bits)
Stream_id = 0xBD (8 bits, private_stream_1)
PES_packet_len = 0x0000 (16 bits)
16 bits set to 0b10000100 0b10000000
PES_header_data_length = 5 (8 bits)
PTS_DTS == '10' which defines PTS in 5 bytes

```

Figure 10. - PES Header Details

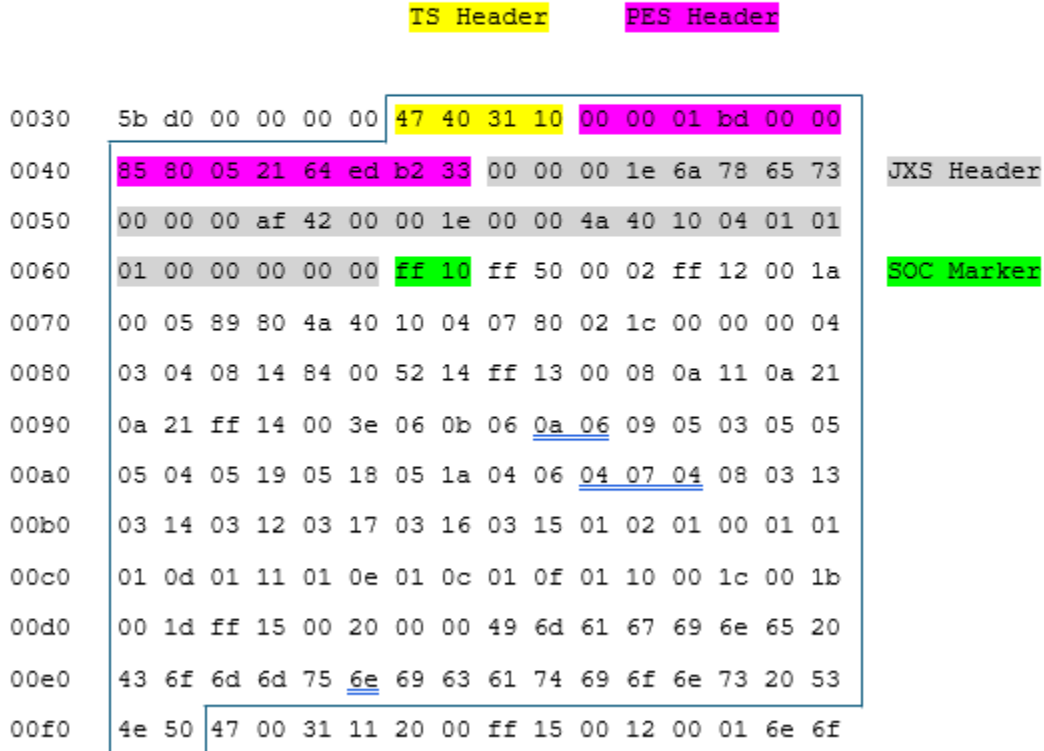
Appendix B (Informative) Alignment of PES and JXS headers to TS packets

Note: Content of headers, shown in examples below, can vary. The purpose of the examples is to show the location or presence of certain fields in a TS packet.

First TS packet in JPEG-XS PES: codestream0

Progressive formats: Start of Frame

Interlaced formats: Start of Field1



SOC: Start of Codestream

Figure 11. - SOC Marker

First TS packet: codestream1 (interlaced formats only)

Start of Field2

TS Header SOC Marker

```

0260  00 00 00 00 00 00 00 00 00 da 11 47 00 31 15 ff 10
0270  ff 50 00 02 ff 12 00 1a 00 05 89 80 4a 40 10 04
0280  07 80 02 1c 00 00 00 04 03 04 08 14 84 00 52 14
0290  ff 13 00 08 0a 11 0a 21 0a 21 ff 14 00 3e 06 0b
02a0  06 0a 06 09 05 03 05 05 05 04 05 19 05 18 05 1a
02b0  04 06 04 07 04 08 03 13 03 14 03 12 03 17 03 16
02c0  03 15 01 02 01 00 01 01 01 0d 01 11 01 0e 01 0c
02d0  01 0f 01 10 00 1c 00 1b 00 1d ff 15 00 20 00 00
02e0  49 6d 61 67 69 6e 65 20 43 6f 6d 6d 75 6e 69 63
02f0  61 74 69 6f 6e 73 20 53 4e 50 20 00 ff 15 00 12
0300  00 01 6e 6f 74 20 73 70 65 63 69 66 69 65 64 00
0310  ff 20 00 04 00 00 00 0a 70 07 10 aa aa aa aa aa
0320  aa aa a0 81 a9 03 47 00 31 16 c8 58 aa aa aa ab

```

SOC: Start of Codestream

Figure 12. - SOC Marker

Appendix C Capabilities and Interop Points
High 444.12 Profile

| Interop Points | Capability Set | Buffer Model Gapped Linear (F4) | Conformance Level | Format & Frame Rate (F1) | Sampling Points | Pixels per Second | Min rate (Mbps) | Min Bpp | Max Rate (Mbps) | Max Bpp | Bit Depth | Color Sampling | Color Space (F3) | Audio Conformance Level (F5) | JPEG XS Profile | | | Reference Uncompressed Video BitRate (Mbps) |
|----------------|----------------|---------------------------------|-------------------|--------------------------|-----------------|-------------------|-----------------|---------|-----------------|---------|-----------|----------------|---------------------|------------------------------|-----------------|-------|---------------|---|
| | | | | | | | | | | | | | | | Profile | Level | Sublevel (F2) | |
| 1 | A | | FHD | 720px1280/59 | 921,600 | 55,240,759 | 83 | 1.5 | 221 | 4.0 | 10bit | 4:2:2 | Rec. ITU-R BT.709-6 | ACL-A | High 444.12 | 1K-1 | Sublev4bpp | 1,105 |
| 2 | A | | FHD | 720px1280/50 | 921,600 | 46,080,000 | 69 | 1.5 | 184 | 4.0 | 10bit | 4:2:2 | Rec. ITU-R BT.709-6 | ACL-A | High 444.12 | 1K-1 | Sublev4bpp | 922 |
| 3 | A | | FHD | 1080ix1920/29 | 2,073,600 | 62,145,854 | 93 | 1.5 | 249 | 4.0 | 10bit | 4:2:2 | Rec. ITU-R BT.709-6 | ACL-A | High 444.12 | 2K-1 | Sublev4bpp | 1,243 |
| 4 | A | | FHD | 1080ix1920/25 | 2,073,600 | 51,840,000 | 78 | 1.5 | 207 | 4.0 | 10bit | 4:2:2 | Rec. ITU-R BT.709-6 | ACL-A | High 444.12 | 2K-1 | Sublev4bpp | 1,037 |
| 5a,b,c | A | | FHD | 1080px1920/59 | 2,073,600 | 124,291,708 | 186 | 1.5 | 497 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-A,B,C | High 444.12 | 2K-1 | Sublev4bpp | 2,486 |
| 6a,b,c | A | | FHD | 1080px1920/50 | 2,073,600 | 103,680,000 | 156 | 1.5 | 415 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-A,B,C | High 444.12 | 2K-1 | Sublev4bpp | 2,074 |
| 11a.b.c | A | | FHD | 1080px1920/23.98 | 2,073,600 | 49,716,683 | 75 | 1.5 | 199 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-A,B,C | High 444.12 | 2K-1 | Sublev4bpp | 994 |
| 12a.b.c | A | | FHD | 1080px1920/24 | 2,073,600 | 49,766,400 | 75 | 1.5 | 199 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-A,B,C | High 444.12 | 2K-1 | Sublev4bpp | 995 |
| 13a.b.c | A | | FHD | 1080px1920/29.97 | 2,073,600 | 62,145,854 | 93 | 1.5 | 249 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-A,B,C | High 444.12 | 2K-1 | Sublev4bpp | 1,243 |
| 14a.b.c | A | | FHD | 1080px1920/30 | 2,073,600 | 62,208,000 | 93 | 1.5 | 249 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-A,B,C | High 444.12 | 2K-1 | Sublev4bpp | 1,244 |
| 15a.b.c | A | | FHD | 1080px1920/60 | 2,073,600 | 124,416,000 | 187 | 1.5 | 498 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-A,B,C | High 444.12 | 2K-1 | Sublev4bpp | 2,488 |
| 7a.b.c | B | | UHD1 | 2160px3840/59 | 8,294,400 | 497,166,833 | 746 | 1.5 | 1989 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | High 444.12 | 4K-2 | Sublev4bpp | 9,943 |
| 8a.b.c | B | | UHD1 | 2160px3840/50 | 8,294,400 | 414,720,000 | 622 | 1.5 | 1,659 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | High 444.12 | 4K-2 | Sublev4bpp | 8,294 |
| 16a.b.c | B | | UHD1 | 2160px3840/23.98 | 8,294,400 | 198,866,733 | 298 | 1.5 | 795 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | High 444.12 | 4K-2 | Sublev4bpp | 9,943 |
| 17a.b.c | B | | UHD1 | 2160px3840/24 | 8,294,400 | 199,065,600 | 299 | 1.5 | 796 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | High 444.12 | 4K-2 | Sublev4bpp | 8,294 |
| 18a.b.c | B | | UHD1 | 2160px3840/29.97 | 8,294,400 | 248,583,417 | 373 | 1.5 | 994 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | High 444.12 | 4K-2 | Sublev4bpp | 8,294 |
| 19a.b.c | B | | UHD1 | 2160px3840/30 | 8,294,400 | 248,832,000 | 373 | 1.5 | 995 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | High 444.12 | 4K-2 | Sublev4bpp | 8,294 |
| 20a.b.c | B | | UHD1 | 2160px3840/60 | 8,294,400 | 497,664,000 | 746 | 1.5 | 1,991 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | High 444.12 | 4K-2 | Sublev4bpp | 8,294 |
| 9a.b.c | C | | UHD2 | 4320px7680/59 | 33,177,600 | 1,988,667,333 | 2,983 | 2 | 7,955 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | High 444.12 | 8K-2 | Sublev4bpp | 39,773 |
| 10a.b.c | C | | UHD2 | 4320px7680/50 | 33,177,600 | 1,658,880,000 | 2,488 | 1.5 | 6,636 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | High 444.12 | 8K-2 | Sublev4bpp | 33,178 |



Notes:

F1 "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 $-30/1.001$, 23.98 is equivalent to $24/1001$ ")

F2 Sublevel: shall be set to Sublev3bpp or Sublev4bpp when the bpp is less than or equal to 3bpp, and shall be set to Sublev4bpp when the bpp exceeds 3bpp. Implementations compliant with this TR shall not set Sublevel to any other values.

F3 Color Space: a=BT.709-6, b=BT.2100-2 (PQ), c=BT.2100-2 (HLG)

F4 ES/TS Buffer Model: G=Gapped, L=Linear

F5 Note that the minimal requirement for 1 PID with 8 audio channels, (4 AES channel pairs) applies to all ACLs as a base requirement.

12.1 TDC 444.12 Profile

| Interop Points | Capability Set | Buffer Model Gapped Linear (F4) | Conformance Level | Format & Frame Rate (F1) | Sampling Points | Pixels per Second | Min rate (Mbps) | Min Bpp | Max Rate (Mbps) | Max Bpp | Bit Depth | Color Sampling | Color Space (F3) | Audio Conformance Level (F5) | JPEG XS Profile | | | Minimum Decoder FBB Level | Reference Uncompressed Video BitRate (Mbps) |
|----------------|----------------|---------------------------------|-------------------|--------------------------|-----------------|-------------------|-----------------|---------|-----------------|---------|-----------|----------------|---------------------|------------------------------|------------------------|-------|---------------|---------------------------|---|
| | | | | | | | | | | | | | | | Profile | Level | Sublevel (F2) | | |
| 1 | AT | | FHD | 720px1280/59 | 921,600 | 55,240,759 | 55 | 1.0 | 221 | 4.0 | 10bit | 4:2:2 | Rec. ITU-R BT.709-6 | ACL-A | TDC 444.12 High 444.12 | 1K-1 | Sublev4bpp | Fbblev12 bpp | 1,105 |
| 2 | AT | | FHD | 720px1280/50 | 921,600 | 46,080,000 | 46 | 1.0 | 184 | 4.0 | 10bit | 4:2:2 | Rec. ITU-R BT.709-6 | ACL-A | TDC 444.12 High 444.12 | 1K-1 | Sublev4bpp | Fbblev12 bpp | 922 |
| 3 | AT | | FHD | 1080ix1920/29 | 2,073,600 | 62,145,854 | 62 | 1.0 | 249 | 4.0 | 10bit | 4:2:2 | Rec. ITU-R BT.709-6 | ACL-A | TDC 444.12 High 444.12 | 2K-1 | Sublev4bpp | Fbblev12 bpp | 1,243 |
| 4 | AT | | FHD | 1080ix1920/25 | 2,073,600 | 51,840,000 | 52 | 1.0 | 207 | 4.0 | 10bit | 4:2:2 | Rec. ITU-R BT.709-6 | ACL-A | TDC 444.12 High 444.12 | 2K-1 | Sublev4bpp | Fbblev12 bpp | 1,037 |
| 5a,b,c | AT | | FHD | 1080px1920/59 | 2,073,600 | 124,291,708 | 124 | 1.0 | 497 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-A,B,C | TDC 444.12 High 444.12 | 2K-1 | Sublev4bpp | Fbblev12 bpp | 2,486 |
| 6a,b,c | AT | | FHD | 1080px1920/50 | 2,073,600 | 103,680,000 | 104 | 1.0 | 415 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-A,B,C | TDC 444.12 High 444.12 | 2K-1 | Sublev4bpp | Fbblev12 bpp | 2,074 |
| 11a.b.c | AT | | FHD | 1080px1920/23.98 | 2,073,600 | 49,716,683 | 50 | 1.0 | 199 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-A,B,C | TDC 444.12 High 444.12 | 2K-1 | Sublev4bpp | Fbblev12 bpp | 994 |
| 12a.b.c | AT | | FHD | 1080px1920/24 | 2,073,600 | 49,766,400 | 50 | 1.0 | 199 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-A,B,C | TDC 444.12 High 444.12 | 2K-1 | Sublev4bpp | Fbblev12 bpp | 995 |
| 13a.b.c | AT | | FHD | 1080px1920/29.97 | 2,073,600 | 62,145,854 | 62 | 1.0 | 249 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-A,B,C | TDC 444.12 High 444.12 | 2K-1 | Sublev4bpp | Fbblev12 bpp | 1,243 |
| 14a.b.c | AT | | FHD | 1080px1920/30 | 2,073,600 | 62,208,000 | 62 | 1.0 | 249 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-A,B,C | TDC 444.12 High 444.12 | 2K-1 | Sublev4bpp | Fbblev12 bpp | 1,244 |
| 15a.b.c | AT | | FHD | 1080px1920/60 | 2,073,600 | 124,416,000 | 124 | 1.0 | 498 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-A,B,C | TDC 444.12 High 444.12 | 2K-1 | Sublev4bpp | Fbblev12 bpp | 2,488 |
| 7a.b.c | BT | | UHD1 | 2160px3840/59 | 8,294,400 | 497,166,833 | 497 | 1.0 | 1989 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | TDC 444.12 High 444.12 | 4K-2 | Sublev4bpp | Fbblev12 bpp | 9,943 |
| 8a.b.c | BT | | UHD1 | 2160px3840/50 | 8,294,400 | 414,720,000 | 415 | 1.0 | 1,659 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | TDC 444.12 High 444.12 | 4K-2 | Sublev4bpp | Fbblev12 bpp | 8,294 |
| 16a.b.c | BT | | UHD1 | 2160px3840/23.98 | 8,294,400 | 198,866,733 | 199 | 1.0 | 795 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | TDC 444.12 High 444.12 | 4K-2 | Sublev4bpp | Fbblev12 bpp | 9,943 |



VSF TR-07:2026

| Interop Points | Capability Set | Buffer Model Gapped Linear (F4) | Conformance Level | Format & Frame Rate (F1) | Sampling Points | Pixels per Second | Min rate (Mbps) | Min Bpp | Max Rate (Mbps) | Max Bpp | Bit Depth | Color Sampling | Color Space (F3) | Audio Conformance Level (F5) | JPEG XS Profile | | | Minimum Decoder FBB Level | Reference Uncompressed Video BitRate, in Mbps |
|----------------|----------------|---------------------------------|-------------------|--------------------------|-----------------|-------------------|-----------------|---------|-----------------|---------|-----------|----------------|------------------|------------------------------|---------------------------|-------|---------------|---------------------------|---|
| | | | | | | | | | | | | | | | Profile | Level | Sublevel (F2) | | |
| 17a.b.c | BT | | UHD1 | 2160px3840/24 | 8,294,400 | 199,065,600 | 199 | 1.0 | 796 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | TDC 444.12 High 444.12 | 4K-2 | Sublev4bpp | Fbblev12 bpp | 8,294 |
| 18a.b.c | BT | | UHD1 | 2160px3840/29.97 | 8,294,400 | 248,583,417 | 249 | 1.0 | 994 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | TDC 444.12 High 444.12 | 4K-2 | Sublev4bpp | Fbblev12 bpp | 8,294 |
| 19a.b.c | BT | | UHD1 | 2160px3840/30 | 8,294,400 | 248,832,000 | 249 | 1.0 | 995 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | TDC 444.12 High 444.12 | 4K-2 | Sublev4bpp | Fbblev12 bpp | 8,294 |
| 20a.b.c | BT | | UHD1 | 2160px3840/60 | 8,294,400 | 497,664,000 | 498 | 1.0 | 1,991 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B,C | TDC 444.12 High 444.12 | 4K-2 | Sublev4bpp | Fbblev12 bpp | 8,294 |
| 9a.b.c | CT | | UHD2 | 4320px7680/59 | 33,177,600 | 1,988,667,333 | 1,989 | 1.0 | 7,955 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B.C | TDC 444.12 High 444.12 | 8K-2 | Sublev4bpp | Fbblev8b pp | 39,773 |
| 10a.b.c | CT | | UHD2 | 4320px7680/50 | 33,177,600 | 1,658,880,000 | 1,659 | 1.0 | 6,636 | 4.0 | 10bit | 4:2:2 | Footnote 3 | ACL-B.C | TDC 444.12 High 444.12 | 8K-2 | Sublev4bpp | Fbblev8b pp | 33,178 |

Notes:

F1 "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, 23.98 is equivalent to 24/1001"

F2 Sublevel: shall be set to Sublev3bpp or Sublev4bpp when the bpp is less than or equal to 3bpp, and shall be set to Sublev4bpp when the bpp exceeds 3bpp. Implementations compliant with this TR shall not set Sublevel to any other values.

F3 Color Space: a=BT.709-6, b=BT.2100-2 (PQ), c=BT.2100-2 (HLG)

F4 Buffer Model: G=Gapped, L=Linear Narrow, W= Linear Wide

F5 Note that the minimal requirement for 1 PID with 8 audio channels, (4 AES channel pairs) applies to all ACLs as a base requirement.

F6 Frame buffer Level for TDC can be either 8bpp or 12bpp, these are maximums, they support lower value

