|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fond-Rec_e | | **International Telecommunication Union** | | |
|  | |  | | |
| **ITU-T** | **H.222.0** | |
| TELECOMMUNICATION STANDARDIZATION SECTOR OF ITU | | (06/2021) |
|  | SERIES H: AUDIOVISUAL AND MULTIMEDIA SYSTEMS  Infrastructure of audiovisual services – Transmission multiplexing and synchronization | | | |
|  | **Information technology – Generic coding of moving pictures and associated audio information: Systems** | | | |
|  | Recommendation ITU‑T H.222.0 | | | |

Logo

Description automatically generated

ITU-T H-SERIES RECOMMENDATIONS

**AUDIOVISUAL AND MULTIMEDIA SYSTEMS**

|  |  |
| --- | --- |
|  |  |
| CHARACTERISTICS OF VISUAL TELEPHONE SYSTEMS | H.100–H.199 |
| INFRASTRUCTURE OF AUDIOVISUAL SERVICES |  |
| General | H.200–H.219 |
| **Transmission multiplexing and synchronization** | **H.220–H.229** |
| Systems aspects | H.230–H.239 |
| Communication procedures | H.240–H.259 |
| Coding of moving video | H.260–H.279 |
| Related systems aspects | H.280–H.299 |
| Systems and terminal equipment for audiovisual services | H.300–H.349 |
| Directory services architecture for audiovisual and multimedia services | H.350–H.359 |
| Quality of service architecture for audiovisual and multimedia services | H.360–H.369 |
| Telepresence, immersive environments, virtual and extended reality | H.420–H.439 |
| Supplementary services for multimedia | H.450–H.499 |
| MOBILITY AND COLLABORATION PROCEDURES |  |
| Overview of Mobility and Collaboration, definitions, protocols and procedures | H.500–H.509 |
| Mobility for H-Series multimedia systems and services | H.510–H.519 |
| Mobile multimedia collaboration applications and services | H.520–H.529 |
| Security for mobile multimedia systems and services | H.530–H.539 |
| Security for mobile multimedia collaboration applications and services | H.540–H.549 |
| VEHICULAR GATEWAYS AND INTELLIGENT TRANSPORTATION SYSTEMS (ITS) |  |
| Architecture for vehicular gateways | H.550–H.559 |
| Vehicular gateway interfaces | H.560–H.569 |
| BROADBAND, TRIPLE-PLAY AND ADVANCED MULTIMEDIA SERVICES |  |
| Broadband multimedia services over VDSL | H.610–H.619 |
| Advanced multimedia services and applications | H.620–H.629 |
| Content delivery and ubiquitous sensor network applications | H.640–H.649 |
| IPTV MULTIMEDIA SERVICES AND APPLICATIONS FOR IPTV |  |
| General aspects | H.700–H.719 |
| IPTV terminal devices | H.720–H.729 |
| IPTV middleware | H.730–H.739 |
| IPTV application event handling | H.740–H.749 |
| IPTV metadata | H.750–H.759 |
| IPTV multimedia application frameworks | H.760–H.769 |
| IPTV service discovery up to consumption | H.770–H.779 |
| Digital Signage | H.780–H.789 |
| E-HEALTH MULTIMEDIA SYSTEMS, SERVICES AND APPLICATIONS |  |
| Personal health systems | H.810–H.819 |
| Interoperability compliance testing of personal health systems (HRN, PAN, LAN, TAN and WAN) | H.820–H.859 |
| Multimedia e-health data exchange services | H.860–H.869 |
| Safe listening | H.870–H.879 |
|  |  |

*For further details, please refer to the list of ITU-T Recommendations.*

|  |
| --- |
| INTERNATIONAL STANDARD ISO/IEC 13818-1  Recommendation ITU-T H.222.0  Information technology – Generic coding of moving pictures and associated  audio information: Systems |
| Summary  This Recommendation | International Standard specifies the system layer of the coding. It was developed in 1994 to principally support the combination and synchronization of video and audio coding methods defined in ISO/IEC 13818 Part 2 (ITU-T H.262) and Part 3. Since 1994, this standard has been extended to support additional video coding specifications (e.g., ISO/IEC 14496-2, ITU-T H.264 | ISO/IEC 14496-10, ITU-T H.265 | ISO/IEC 23008-2 and ITU‑T T.800 | ISO/IEC 15444-1 Annex M JPEG 2000 video), audio coding specifications (e.g., ISO/IEC 13818-7 and ISO/IEC 14496-3), system streams (e.g., ISO/IEC 14496-1 and ISO/IEC 15938-1), ISO/IEC 23009-1 dynamic adaptive streaming over HTTP (DASH), ISO/IEC 13818-11 intellectual property management and protection (IPMP) as well as generic metadata. The system layer supports six basic functions:  1) the synchronization of multiple compressed streams on decoding;  2) the interleaving of multiple compressed streams into a single stream;  3) the initialization of buffering for decoding start up;  4) continuous buffer management;  5) time identification; and  6) multiplexing and signalling of various components in a system stream.  Recommendation ITU-T H.222.0 | ISO/IEC 13818-1 multiplexed bit stream is either a transport stream or a program stream. Both streams are constructed from packetized elementary stream (PES) packets and packets containing other necessary information. Both stream types support multiplexing of video and audio compressed streams from one program with a common time base. The transport stream additionally supports the multiplexing of video and audio compressed streams from multiple programs with independent time bases. For almost error-free environments the program stream is generally more appropriate, supporting software processing of program information. The transport stream is more suitable for use in environments where errors are likely.  Either multiplexed bit stream is constructed in two layers: the outermost layer is the system layer, and the innermost is the compression layer. The system layer provides the functions necessary for using one or more compressed data streams in a system. The video and audio parts of this Specification define the compression coding layer for audio and video data. Coding of other types of data is not defined by this Recommendation | International Standard, but is supported by the system layer provided that the other types of data adhere to the constraints defined in this Recommendation | International Standard.  The 8th edition includes Amd. 1 and Cor. 1 to the 7th edition as well as the following changes: how VVC (ITU-T H.266 | ISO/IEC 23090-3) is carried over MPEG-2 systems; how EVC (ISO/IEC 23094-1) is carried over MPEG-2 systems; how compatible profile sets for MPEG-H 3D Audio (ISO/IEC 23008-3) are signalled in MPEG-2 systems; and an extension of the semantics for the ISO 639 language descriptor. |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **History**   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Edition | | Recommendation | | Approval | Study Group | Unique ID[[1]](#footnote-2)\* | | 1.0 | | ITU-T H.222.0 | | 1995-07-10 | 15 | [11.1002/1000/1071](http://handle.itu.int/11.1002/1000/1071) | | 1.1 | | ITU-T H.222.0 (1995) Amd. 1 | | 1996-11-11 | 16 | [11.1002/1000/3834](http://handle.itu.int/11.1002/1000/3834) | | 1.2 | | ITU-T H.222.0 (1995) Amd. 2 | | 1996-11-11 | 16 | [11.1002/1000/4096](http://handle.itu.int/11.1002/1000/4096) | | 1.3 | | ITU-T H.222.0 (1995) Technical Cor. 1 | | 1998-02-06 | 16 | [11.1002/1000/4532](http://handle.itu.int/11.1002/1000/4532) | | 1.4 | | ITU-T H.222.0 (1995) Amd. 3 | | 1998-02-06 | 16 | [11.1002/1000/4228](http://handle.itu.int/11.1002/1000/4228) | | 1.5 | | ITU-T H.222.0 (1995) Amd. 4 | | 1998-02-06 | 16 | [11.1002/1000/4229](http://handle.itu.int/11.1002/1000/4229) | | 1.6 | | ITU-T H.222.0 (1995) Amd. 5 | | 1999-05-27 | 16 | [11.1002/1000/4498](http://handle.itu.int/11.1002/1000/4498) | | 1.7 | | ITU-T H.222.0 (1995) Amd. 6 | | 1999-05-27 | 16 | [11.1002/1000/4671](http://handle.itu.int/11.1002/1000/4671) | | 2.0 | | ITU-T H.222.0 | | 2000-02-17 | 16 | [11.1002/1000/5142](http://handle.itu.int/11.1002/1000/5142) | | 2.1 | | ITU-T H.222.0 (2000) Technical Cor. 1 | | 2001-03-01 | 16 | [11.1002/1000/5419](http://handle.itu.int/11.1002/1000/5419) | | 2.2 | | ITU-T H.222.0 (2000) Technical Cor. 2 | | 2002-03-29 | 16 | [11.1002/1000/5675](http://handle.itu.int/11.1002/1000/5675) | | 2.3 | | ITU-T H.222.0 (2000) Amd. 1 | | 2002-12-14 | 16 | [11.1002/1000/6190](http://handle.itu.int/11.1002/1000/6190) | | 2.4 | | ITU-T H.222.0 (2000) Amd. 1/Cor. 1 | | 2003-06-29 | 16 | [11.1002/1000/6449](http://handle.itu.int/11.1002/1000/6449) | | 2.5 | | ITU-T H.222.0 (2000) Amd. 2 | | 2003-06-29 | 16 | [11.1002/1000/6363](http://handle.itu.int/11.1002/1000/6363) | | 2.6 | | ITU-T H.222.0 (2000) Amd. 3 | | 2004-03-15 | 16 | [11.1002/1000/7208](http://handle.itu.int/11.1002/1000/7208) | | 2.7 | | ITU-T H.222.0 (2000) Technical Cor. 3 | | 2005-01-08 | 16 | [11.1002/1000/7435](http://handle.itu.int/11.1002/1000/7435) | | 2.8 | | ITU-T H.222.0 (2000) Amd. 4 | | 2005-01-08 | 16 | [11.1002/1000/7436](http://handle.itu.int/11.1002/1000/7436) | | 2.9 | | ITU-T H.222.0 (2000) Amd. 5 | | 2005-01-08 | 16 | [11.1002/1000/7437](http://handle.itu.int/11.1002/1000/7437) | | 2.10 | | ITU-T H.222.0 (2000) Technical Cor. 4 | | 2005-09-13 | 16 | [11.1002/1000/8560](http://handle.itu.int/11.1002/1000/8560) | | 3.0 | | ITU-T H.222.0 | | 2006-05-29 | 16 | [11.1002/1000/8802](http://handle.itu.int/11.1002/1000/8802) | | 3.1 | | ITU-T H.222.0 (2006) Amd. 1 | | 2007-01-13 | 16 | [11.1002/1000/9024](http://handle.itu.int/11.1002/1000/9024) | | 3.2 | | ITU-T H.222.0 (2006) Amd. 2 | | 2007-08-29 | 16 | [11.1002/1000/9214](http://handle.itu.int/11.1002/1000/9214) | | 3.3 | | ITU-T H.222.0 (2006) Cor. 1 | | 2008-06-13 | 16 | [11.1002/1000/9471](http://handle.itu.int/11.1002/1000/9471) | | 3.4 | | ITU-T H.222.0 (2006) Cor. 2 | | 2009-03-16 | 16 | [11.1002/1000/9692](http://handle.itu.int/11.1002/1000/9692) | | 3.5 | | ITU-T H.222.0 (2006) Amd. 3 | | 2009-03-16 | 16 | [11.1002/1000/9691](http://handle.itu.int/11.1002/1000/9691) | | 3.6 | | ITU-T H.222.0 (2006) Cor. 3 | | 2009-12-14 | 16 | [11.1002/1000/10621](http://handle.itu.int/11.1002/1000/10621) | | 3.7 | | ITU-T H.222.0 (2006) Cor. 4 | | 2009-12-14 | 16 | [11.1002/1000/10622](http://handle.itu.int/11.1002/1000/10622) | | 3.8 | | ITU-T H.222.0 (2006) Amd. 4 | | 2009-12-14 | 16 | [11.1002/1000/10623](http://handle.itu.int/11.1002/1000/10623) | | 3.9 | | ITU-T H.222.0 (2006) Amd. 5 | | 2011-05-14 | 16 | [11.1002/1000/11287](http://handle.itu.int/11.1002/1000/11287) | | 3.10 | | ITU-T H.222.0 (2006) Amd. 6 | | 2011-05-14 | 16 | [11.1002/1000/11288](http://handle.itu.int/11.1002/1000/11288) | | 4.0 | | ITU-T H.222.0 | | 2012-06-29 | 16 | [11.1002/1000/11655](http://handle.itu.int/11.1002/1000/11655) | | 4.1 | | ITU-T H.222.0 (2012) Amd. 1 | | 2014-01-13 | 16 | [11.1002/1000/12054](http://handle.itu.int/11.1002/1000/12054) | | 4.2 | | ITU-T H.222.0 (2012) Amd. 2 | | 2014-01-13 | 16 | [11.1002/1000/12055](http://handle.itu.int/11.1002/1000/12055) | | 4.3 | | ITU-T H.222.0 (2012) Amd. 3 | | 2014-01-13 | 16 | [11.1002/1000/12056](http://handle.itu.int/11.1002/1000/12056) | | 4.4 | | ITU-T H.222.0 (2012) Amd. 4 | | 2014-01-13 | 16 | [11.1002/1000/12057](http://handle.itu.int/11.1002/1000/12057) | | 4.5 | | ITU-T H.222.0 (2012) Amd. 5 | | 2014-10-14 | 16 | [11.1002/1000/12306](http://handle.itu.int/11.1002/1000/12306) | | 5.0 | | ITU-T H.222.0 | | 2014-10-14 | 16 | [11.1002/1000/12359](http://handle.itu.int/11.1002/1000/12359) | | 5.1 | | ITU-T H.222.0 (2014) Amd. 1 | | 2015-04-29 | 16 | [11.1002/1000/12452](http://handle.itu.int/11.1002/1000/12452) | | 5.2 | | ITU-T H.222.0 (2014) Amd. 1 Cor. 1 | | 2015-11-29 | 16 | [11.1002/1000/12625](http://handle.itu.int/11.1002/1000/12625) | | 5.3 | | ITU-T H.222.0 (2014) Amd. 2 | | 2015-12-14 | 16 | [11.1002/1000/12632](http://handle.itu.int/11.1002/1000/12632) | | 5.4 | | ITU-T H.222.0 (2014) Amd. 3 | | 2015-12-14 | 16 | [11.1002/1000/12633](http://handle.itu.int/11.1002/1000/12633) | | 5.5 | | ITU-T H.222.0 (2014) Amd. 1 Cor. 2 | | 2016-07-14 | 16 | [11.1002/1000/12899](http://handle.itu.int/11.1002/1000/12899) | | 5.5 | | ITU-T H.222.0 (2014) Cor. 1 | | 2016-07-14 | 16 | [11.1002/1000/12903](http://handle.itu.int/11.1002/1000/12903) | | 5.7 | | ITU-T H.222.0 (2014) Amd. 4 | | 2016-07-14 | 16 | [11.1002/1000/12900](http://handle.itu.int/11.1002/1000/12900) | | 5.8 | | ITU-T H.222.0 (2014) Amd. 5 | | 2016-07-14 | 16 | [11.1002/1000/12901](http://handle.itu.int/11.1002/1000/12901) | | 5.9 | | ITU-T H.222.0 (2014) Amd. 6 | | 2016-07-14 | 16 | [11.1002/1000/12902](http://handle.itu.int/11.1002/1000/12902) | | 5.10 | | ITU-T H.222.0 (2014) Amd. 3 Cor. 1 | | 2017-03-01 | 16 | [11.1002/1000/13184](http://handle.itu.int/11.1002/1000/13184) | | 5.10 | | ITU-T H.222.0 (2014) Cor. 2 | | 2017-03-01 | 16 | [11.1002/1000/13188](http://handle.itu.int/11.1002/1000/13188) | | 5.12 | | ITU-T H.222.0 (2014) Amd. 7 | | 2017-03-01 | 16 | [11.1002/1000/13186](http://handle.itu.int/11.1002/1000/13186) | | 5.13 | | ITU-T H.222.0 (2014) Amd. 8 | | 2017-03-01 | 16 | [11.1002/1000/13187](http://handle.itu.int/11.1002/1000/13187) | | 6.0 | | ITU-T H.222.0 | | 2017-03-01 | 16 | [11.1002/1000/13269](http://handle.itu.int/11.1002/1000/13269) | | 6.1 | | ITU-T H.222.0 (2017) Amd. 1 | | 2017-12-14 | 16 | [11.1002/1000/13431](http://handle.itu.int/11.1002/1000/13431) | | 7.0 | | ITU-T H.222.0 | | 2018-08-29 | 16 | [11.1002/1000/13664](http://handle.itu.int/11.1002/1000/13664) | | 7.1 | | ITU-T H.222.0 (2018) Cor. 1 | | 2019-11-29 | 16 | [11.1002/1000/14106](http://handle.itu.int/11.1002/1000/14106) | | 7.2 | | ITU-T H.222.0 (2018) Amd. 1 | | 2019-11-29 | 16 | [11.1002/1000/14105](http://handle.itu.int/11.1002/1000/14105) | | 8.0 | | ITU-T H.222.0 | | 2021-06-13 | 16 | [11.1002/1000/14658](http://handle.itu.int/11.1002/1000/14658) | | 8.1 | ITU-T H.222.0 (2021) Cor. 1 | | 2022-12-14 | 16 | [11.1002/1000/15172](http://handle.itu.int/11.1002/1000/15172" \o "Click to download the respective PDF version) | | 8.2 | ITU-T H.222.0 (2021) Amd. 1 | | 2022-12-14 | 16 | [11.1002/1000/15171](http://handle.itu.int/11.1002/1000/15171" \o "Click to download the respective PDF version) | | 9.0 | ITU-T H.222.0 (V9) | |  | 16 |  | |

|  |
| --- |
| Keywords  Audiovisual content, multimedia multiplexing, MPEG-2 system, multiplexed bit stream, program stream, transport stream. |

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of tele­com­mu­ni­ca­tions, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU‑T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents/software copyrights, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the appropriate ITU-T databases available via the ITU-T website at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2021

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

**CONTENTS**

| Page |
| --- |
| [SECTION 1 – GENERAL 1](#_Toc75201044)  [1.1 Scope 1](#_Toc75201045)  [1.2 Normative references 1](#_Toc75201046)  [SECTION 2 – TECHNICAL ELEMENTS 4](#_Toc75201047)  [2.1 Definitions 4](#_Toc75201048)  [2.2 Symbols and abbreviations 12](#_Toc75201049)  [2.3 Method of describing bit stream syntax 14](#_Toc75201050)  [2.4 Transport stream bitstream requirements 15](#_Toc75201051)  [2.5 Program stream bitstream requirements 63](#_Toc75201052)  [2.6 Program and program element descriptors 76](#_Toc75201053)  [2.7 Restrictions on the multiplexed stream semantics 157](#_Toc75201054)  [2.8 Compatibility with ISO/IEC 11172 161](#_Toc75201055)  [2.9 Registration of copyright identifiers 161](#_Toc75201056)  [2.10 Registration of private data format 162](#_Toc75201057)  [2.11 Carriage of ISO/IEC 14496 data 162](#_Toc75201058)  [2.12 Carriage of metadata 174](#_Toc75201059)  [2.13 Carriage of ISO 15938 data 182](#_Toc75201060)  [2.14 Carriage of Rec. ITU-T H.264 | ISO/IEC 14496-10 video 183](#_Toc75201061)  [2.15 Carriage of ISO/IEC 14496-17 text streams 199](#_Toc75201062)  [2.16 Carriage of auxiliary video streams 200](#_Toc75201063)  [2.17 Carriage of HEVC 201](#_Toc75201064)  [2.18 Carriage of green access units 215](#_Toc75201065)  [2.19 Carriage of ISO/IEC 23008-3 MPEG-H 3D audio data 217](#_Toc75201066)  [2.20 Carriage of Quality Access Units in MPEG-2 sections 219](#_Toc75201067)  [2.21 Carriage of sample variants 220](#_Toc75201068)  [2.22 Carriage of Media Orchestration Access Units 221](#_Toc75201069)  [2.23 Carriage of VVC 221](#_Toc75201070)  [2.24 Carriage of EVC 226](#_Toc75201071)  [Annex A CRC decoder model 230](#_Toc75201072)  [A.1 CRC decoder model 230](#_Toc75201073)  [Annex B Digital storage medium command and control (DSM-CC) 231](#_Toc75201074)  [B.1 Introduction 231](#_Toc75201075)  [B.2 General elements 232](#_Toc75201076)  [B.3 Technical elements 234](#_Toc75201077)  [Annex C Program-specific information 240](#_Toc75201078)  [C.1 Explanation of program-specific information in transport streams 240](#_Toc75201079)  [C.2 Introduction 240](#_Toc75201080)  [C.3 Functional mechanism 240](#_Toc75201081)  [C.4 The mapping of sections into transport stream packets 241](#_Toc75201082)  [C.5 Repetition rates and random access 241](#_Toc75201083)  [C.6 What is a program? 242](#_Toc75201084)  [C.7 Allocation of program\_number 242](#_Toc75201085)  [C.8 Usage of PSI in a typical system 242](#_Toc75201086)  [C.9 The relationships of PSI structures 243](#_Toc75201087)  [C.10 Bandwidth utilization and signal acquisition time 245](#_Toc75201088)  [Annex D Systems timing model and application implications of this Recommendation | International Standard 248](#_Toc75201089)  [D.1 Introduction 248](#_Toc75201090)  [Annex E Data transmission applications 257](#_Toc75201091)  [E.1 General considerations 257](#_Toc75201092)  [E.2 Suggestion 257](#_Toc75201093)  [Annex F Graphics of syntax for this Recommendation | International Standard 258](#_Toc75201094)  [F.1 Introduction 258](#_Toc75201095)  [Annex G General information 262](#_Toc75201096)  [G.1 General information 262](#_Toc75201097)  [Annex H Private data 263](#_Toc75201098)  [H.1 Private data 263](#_Toc75201099)  [Annex I Systems conformance and real-time interface 264](#_Toc75201100)  [I.1 Systems conformance and real-time interface 264](#_Toc75201101)  [Annex J Interfacing jitter-inducing networks to MPEG-2 decoders 265](#_Toc75201102)  [J.1 Introduction 265](#_Toc75201103)  [J.2 Network compliance models 265](#_Toc75201104)  [J.3 Network specification for jitter smoothing 266](#_Toc75201105)  [J.4 Example decoder implementations 267](#_Toc75201106)  [Annex K Splicing transport streams 268](#_Toc75201107)  [K.1 Introduction 268](#_Toc75201108)  [K.2 The different types of splicing point 268](#_Toc75201109)  [K.3 Decoder behaviour on splices 269](#_Toc75201110)  [Annex L Registration procedure (see 2.9) 271](#_Toc75201111)  [L.1 Procedure for the request of a Registered Identifier (RID) 271](#_Toc75201112)  [L.2 Responsibilities of the Registration Authority 271](#_Toc75201113)  [L.3 Responsibilities of parties requesting an RID 271](#_Toc75201114)  [L.4 Appeal procedure for denied applications 271](#_Toc75201115)  [Annex M Registration application form (see 2.9) 273](#_Toc75201116)  [M.1 Contact information of organization requesting a Registered Identifier (RID) 273](#_Toc75201117)  [M.2 Statement of an intention to apply the assigned RID 273](#_Toc75201118)  [M.3 Date of intended implementation of the RID 273](#_Toc75201119)  [M.4 Authorized representative 273](#_Toc75201120)  [M.5 For official use only of the Registration Authority 273](#_Toc75201121)  [Annex N Registration Authority Diagram of administration structure (see 2.9) 274](#_Toc75201122)  [Annex O Registration procedure (see 2.10) 275](#_Toc75201123)  [O.1 Procedure for the request of an RID 275](#_Toc75201124)  [O.2 Responsibilities of the Registration Authority 275](#_Toc75201125)  [O.3 Contact information for the Registration Authority 275](#_Toc75201126)  [O.4 Responsibilities of parties requesting an RID 275](#_Toc75201127)  [O.5 Appeal procedure for denied applications 275](#_Toc75201128)  [Annex P Registration application form 277](#_Toc75201129)  [P.1 Contact information of organization requesting an RID 277](#_Toc75201130)  [P.2 Request for a specific RID 277](#_Toc75201131)  [P.3 Short description of RID that is in use and date system that was implemented 277](#_Toc75201132)  [P.4 Statement of an intention to apply the assigned RID 277](#_Toc75201133)  [P.5 Date of intended implementation of the RID 277](#_Toc75201134)  [P.6 Authorized representative 277](#_Toc75201135)  [P.7 For official use of the Registration Authority 277](#_Toc75201136)  [Annex Q T-STD and P-STD buffer models for ISO/IEC 13818-7 ADTS 278](#_Toc75201137)  [Q.1 Introduction 278](#_Toc75201138)  [Q.2 Leak rate from transport buffer 278](#_Toc75201139)  [Q.3 Buffer size 278](#_Toc75201140)  [Q.4 Conclusion 279](#_Toc75201141)  [Annex R Carriage of ISO/IEC 14496 scenes in Rec. ITU-T H.222.0 | ISO/IEC 13818-1 281](#_Toc75201142)  [R.1 Content access procedure for ISO/IEC 14496 program components within a program stream 281](#_Toc75201143)  [R.2 Content access procedure for ISO/IEC 14496 program components within a transport stream 282](#_Toc75201144)  [Annex S Carriage of JPEG 2000 part 1 video over MPEG-2 transport streams 286](#_Toc75201145)  [S.1 Introduction 286](#_Toc75201146)  [S.2 J2K video access unit, J2K video elementary stream, J2K video sequence and J2K still picture 286](#_Toc75201147)  [S.3 Optional J2K block mode for high resolution support 286](#_Toc75201148)  [S.4 Optional J2K stripe mode for Ultra-Low Latency 287](#_Toc75201149)  [S.5 Elementary stream header (elsm) and mapping to PES packets 287](#_Toc75201150)  [S.6 J2K transport constraints 290](#_Toc75201151)  [S.7 Interpretation of flags in adaptation and PES headers for J2K video elementary streams 291](#_Toc75201152)  [S.8 T-STD extension for J2K video elementary streams 291](#_Toc75201153)  [Annex T MIME type for MPEG-2 transport streams 294](#_Toc75201154)  [T.1 Introduction 294](#_Toc75201155)  [T.2 MIME type and subtype 294](#_Toc75201156)  [T.3 Security considerations 295](#_Toc75201157)  [T.4 Parameters 295](#_Toc75201158)  [Annex U Carriage of timeline and external media information over MPEG-2 transport streams 297](#_Toc75201159)  [U.1 Introduction 297](#_Toc75201160)  [U.2 TEMI access unit and TEMI elementary stream 298](#_Toc75201161)  [U.3 AF descriptors 299](#_Toc75201162)  [Annex V Transport of HEVC tiles 308](#_Toc75201163)  [V.1 Introduction 308](#_Toc75201164)  [V.2 HEVC tile substream identification example 309](#_Toc75201165)  [V.3 Subregion layout example 309](#_Toc75201166)  [Annex W Carriage of JPEG XS part 1 video over MPEG-2 Transport Streams 311](#_Toc75201167)  [W.1 Introduction 311](#_Toc75201168)  [W.2 JPEG XS video access unit, JPEG XS video elementary stream, JPEG XS video sequence and JPEG XS still picture 311](#_Toc75201169)  [W.3 Elementary stream header (jxes) and mapping to PES packets 311](#_Toc75201170)  [W.4 JPEG XS transport constraints 312](#_Toc75201171)  [W.5 Interpretation of flags in adaptation field and PES packet for JPEG XS video elementary streams 313](#_Toc75201172)  [W.6 T-STD extension for JPEG XS video elementary streams 313](#_Toc75201173) |

**List of Tables**

| Page |
| --- |
| [Table 2-1 – Transport stream 26](#_Toc130989888)  [Table 2-2 – Transport packet of this Recommendation | International Standard 26](#_Toc130989889)  [Table 2-3 – PID table 27](#_Toc130989890)  [Table 2-4 – Scrambling control values 27](#_Toc130989891)  [Table 2-5 – Adaptation field control values 28](#_Toc130989892)  [Table 2-6 – Transport stream adaptation field 28](#_Toc130989893)  [Table 2-7 – Splice parameters Table 1 Simple Profile Main Level, Main Profile Main Level, SNR Profile Main Level (both layers), Spatial Profile High-1440 Level (base layer), High Profile Main Level (middle + base layers), Multi-view Profile Main Level (base layer) Video 35](#_Toc130989894)  [Table 2-8 – Splice parameters Table 2 Main Profile Low Level, SNR Profile Low Level (both layers), High Profile Main Level (base layer), Multi-view Profile Low Level (base layer) Video 36](#_Toc130989895)  [Table 2-9 – Splice parameters Table 3 Main Profile High-1440 Level, Spatial Profile High-1440 Level (all layers), High Profile High-1440 Level (middle + base layers), Multi-view Profile High-1440 Level (base layer) Video 36](#_Toc130989896)  [Table 2-10 – Splice parameters Table 4 Main Profile High Level, High Profile High-1440 Level (all layers), High Profile High Level (middle + base layers), Multi-view Profile High Level (base layer) Video 36](#_Toc130989897)  [Table 2-11 – Splice parameters Table 5 SNR Profile Low Level (base layer) Video 36](#_Toc130989898)  [Table 2-12 – Splice parameters Table 6 SNR Profile Main Level (base layer) Video 37](#_Toc130989899)  [Table 2-13 – Splice parameters Table 7 Spatial Profile High-1440 Level (middle + base layers) Video 37](#_Toc130989900)  [Table 2-14 – Splice parameters Table 8 High Profile Main Level (all layers), High Profile High-1440 Level (base layer) Video 37](#_Toc130989901)  [Table 2-15 – Splice parameters Table 9 High Profile High Level (base layer), Multi-view Profile Main Level (both layers) Video 37](#_Toc130989902)  [Table 2-16 – Splice parameters Table 10 High Profile High Level (all layers), Multi-view Profile High-1440 Level (both layers) Video 38](#_Toc130989903)  [Table 2-17 – Splice parameters Table 11 4:2:2 Profile Main Level Video 38](#_Toc130989904)  [Table 2-18 – Splice parameters Table 12 Multi-view Profile Low Level (both layers) Video 38](#_Toc130989905)  [Table 2-19 – Splice parameters Table 13 Multi-view Profile High Level (both layers) Video 38](#_Toc130989906)  [Table 2-20 – Splice parameters Table 14 4:2:2 Profile High Level Video 39](#_Toc130989907)  [Table 2-21 – PES packet 39](#_Toc130989908)  [Table 2-22 – Stream\_id assignments 42](#_Toc130989909)  [Table 2-23 – PES scrambling control values 43](#_Toc130989910)  [Table 2-24 – Trick mode control values 48](#_Toc130989911)  [Table 2-25 – Field\_id field control values 49](#_Toc130989912)  [Table 2-26 – Coefficient selection values 49](#_Toc130989913)  [Table 2-27 – Stream\_id\_extension assignments 51](#_Toc130989914)  [Table 2-28 – Program-specific information 52](#_Toc130989915)  [Table 2-29 – Program-specific information pointer 54](#_Toc130989916)  [Table 2-30 – Program association section 54](#_Toc130989917)  [Table 2-31 – table\_id assignment values 55](#_Toc130989918)  [Table 2-32 – Conditional access section 56](#_Toc130989919)  [Table 2-33 – Transport stream program map section 57](#_Toc130989920)  [Table 2-34 – Stream type assignments 58](#_Toc130989921)  [Table 2-35 – Private section 61](#_Toc130989922)  [Table 2-36 – The transport stream description table 62](#_Toc130989923)  [Table 2-37 – Program stream 68](#_Toc130989924)  [Table 2-38 – Program stream pack 68](#_Toc130989925)  [Table 2-39 – Program stream pack header 68](#_Toc130989926)  [Table 2-40 – Program stream system header 69](#_Toc130989927)  [Table 2-41 – Program stream map 72](#_Toc130989928)  [Table 2-42 – Program stream directory packet 74](#_Toc130989929)  [Table 2-43 – Intra\_coded indicator 76](#_Toc130989930)  [Table 2-44 – Coding\_parameters indicator 76](#_Toc130989931)  [Table 2-45 – Program and program element descriptors 77](#_Toc130989932)  [Table 2-46 – Video stream descriptor 78](#_Toc130989933)  [Table 2-47 – Frame rate code 78](#_Toc130989934)  [Table 2-48 – Audio stream descriptor 79](#_Toc130989935)  [Table 2-49 – Hierarchy descriptor 80](#_Toc130989936)  [Table 2-50 – Hierarchy\_type field values 81](#_Toc130989937)  [Table 2-51 – Registration descriptor 81](#_Toc130989938)  [Table 2-52 – Data stream alignment descriptor 82](#_Toc130989939)  [Table 2-53 – Video stream alignment values 82](#_Toc130989940)  [Table 2-54 – AVC video stream alignment values 83](#_Toc130989941)  [Table 2-55 – HEVC video stream alignment values 83](#_Toc130989942)  [Table 2-56 – Audio stream alignment values 83](#_Toc130989943)  [Table 2-57 – VVC video stream alignment values 84](#_Toc130989944)  [Table 2-58 – EVC video stream alignment values 84](#_Toc130989945)  [Table 2-59 – Target background grid descriptor 85](#_Toc130989946)  [Table 2-60 – Video window descriptor 85](#_Toc130989947)  [Table 2-61 – Conditional access descriptor 86](#_Toc130989948)  [Table 2-62 – ISO 639 language descriptor 86](#_Toc130989949)  [Table 2-63 – Audio type values 87](#_Toc130989950)  [Table 2-64 – System clock descriptor 88](#_Toc130989951)  [Table 2-65 – Multiplex buffer utilization descriptor 88](#_Toc130989952)  [Table 2-66 – Copyright descriptor 89](#_Toc130989953)  [Table 2-67 – Maximum bitrate descriptor 89](#_Toc130989954)  [Table 2-68 – Private data indicator descriptor 90](#_Toc130989955)  [Table 2-69 – Smoothing buffer descriptor 90](#_Toc130989956)  [Table 2-70 – STD descriptor 91](#_Toc130989957)  [Table 2-71 – IBP descriptor 91](#_Toc130989958)  [Table 2-72 – MPEG-4 video descriptor 92](#_Toc130989959)  [Table 2-73 – MPEG-4 audio descriptor 92](#_Toc130989960)  [Table 2-75 – IOD descriptor 95](#_Toc130989961)  [Table 2-76 – SL descriptor 95](#_Toc130989962)  [Table 2-77 – FMC descriptor 96](#_Toc130989963)  [Table 2-78 – External\_ES\_ID descriptor 96](#_Toc130989964)  [Table 2-79 – Muxcode descriptor 97](#_Toc130989965)  [Table 2-80 – M4MuxBufferSize descriptor 97](#_Toc130989966)  [Table 2-81 – MultiplexBuffer descriptor 98](#_Toc130989967)  [Table 2-82 – M4MuxTiming descriptor 98](#_Toc130989968)  [Table 2-83 – Content labelling descriptor 99](#_Toc130989969)  [Table 2-84 – Metadata\_application\_format 99](#_Toc130989970)  [Table 2-85 – Content\_time\_base\_indicator values 100](#_Toc130989971)  [Table 2-86 – Metadata pointer descriptor 101](#_Toc130989972)  [Table 2-87 – Metadata format values 101](#_Toc130989973)  [Table 2-88 – MPEG\_carriage\_flags 102](#_Toc130989974)  [Table 2-89 – Metadata descriptor 103](#_Toc130989975)  [Table 2-90 – decoder\_config\_flags 104](#_Toc130989976)  [Table 2-91 – Metadata STD descriptor 105](#_Toc130989977)  [Table 2-92 – AVC video descriptor 105](#_Toc130989978)  [Table 2-93 – AVC timing and HRD descriptor 107](#_Toc130989979)  [Table 2-94 – MPEG-2 AAC\_audio\_descriptor 108](#_Toc130989980)  [Table 2-95 – MPEG-2\_AAC\_additional\_information field values 109](#_Toc130989981)  [Table 2-96 – MPEG-4 text descriptor 109](#_Toc130989982)  [Table 2-97 – MPEG-4 audio extension descriptor 109](#_Toc130989983)  [Table 2-98 – Auxiliary video stream descriptor 110](#_Toc130989984)  [Table 2-99 – SVC extension descriptor 111](#_Toc130989985)  [Table 2-100 – MVC extension descriptor 112](#_Toc130989986)  [Table 2-101 – J2K video descriptor 113](#_Toc130989987)  [Table 2-102 – Example frame rates based on DEN\_frame\_rate and NUM\_frame\_rate values 115](#_Toc130989988)  [Table 2-103 – MVC operation point descriptor 117](#_Toc130989989)  [Table 2-104 – MPEG2\_stereoscopic\_video\_format\_descriptor syntax 118](#_Toc130989990)  [Table 2-105 – Stereoscopic\_program\_info\_descriptor syntax 118](#_Toc130989991)  [Table 2-106 – Stereoscopic\_service\_type values 119](#_Toc130989992)  [Table 2-107 – Stereoscopic\_video\_info\_descriptor syntax 119](#_Toc130989993)  [Table 2-108 – Upsampling factor values 120](#_Toc130989994)  [Table 2-109 – Extension descriptor 120](#_Toc130989995)  [Table 2-110 – Extension descriptor tag values 123](#_Toc130989996)  [Table 2-111 – Transport\_profile\_descriptor syntax 124](#_Toc130989997)  [Table 2-112 – Transport\_profile values 124](#_Toc130989998)  [Table 2-113 – HEVC video descriptor 125](#_Toc130989999)  [Table 2-114 – Semantics of HDR\_WGC\_idc 127](#_Toc130990000)  [Table 2-115 – HEVC timing and HRD descriptor 127](#_Toc130990001)  [Table 2-116 – Adaptation field extension descriptor 128](#_Toc130990002)  [Table 2-117 – HEVC operation point descriptor 129](#_Toc130990003)  [Table 2-118 – HEVC hierarchy extension descriptor 131](#_Toc130990004)  [Table 2-119 – Semantics of extension dimension bits 131](#_Toc130990005)  [Table 2-120 – Green extension descriptor 132](#_Toc130990006)  [Table 2-121 – MPEG-H 3D audio descriptor 133](#_Toc130990007)  [Table 2-122 – MPEG-H 3D audio config descriptor 133](#_Toc130990008)  [Table 2-123 – MPEG-H 3D audio scene descriptor 134](#_Toc130990009)  [Table 2-124 – MPEG-H 3D audio text label descriptor 137](#_Toc130990010)  [Table 2-125 – MPEG-H 3D audio multi-stream descriptor 139](#_Toc130990011)  [Table 2-126 – MPEG-H 3D audio DRC and Loudness descriptor() 140](#_Toc130990012)  [Table 2-127 – MPEG-H 3D audio command descriptor 142](#_Toc130990013)  [Table 2-128 – Quality extension descriptor 143](#_Toc130990014)  [Table 2-129 – Virtual segmentation descriptor 144](#_Toc130990015)  [Table 2-130 – HEVC tile substream descriptor 145](#_Toc130990016)  [Table 2-131 — HEVC subregion descriptor 146](#_Toc130990017)  [Table 2-132 – JPEG XS video descriptor 147](#_Toc130990018)  [Table 2-133 – VVC video descriptor 150](#_Toc130990019)  [Table 2-134 – Semantics of HDR\_WGC\_idc 151](#_Toc130990020)  [Table 2-135 – SDR widely used video property combinations 152](#_Toc130990021)  [Table 2-136 – WCG widely used video property combinations 152](#_Toc130990022)  [Table 2-137 – HDR/WCG widely used video property combinations 152](#_Toc130990023)  [Table 2-138 – No Indication 153](#_Toc130990024)  [Table 2-139 – VVC timing and HRD descriptor 153](#_Toc130990025)  [Table 2-140 – EVC video descriptor 154](#_Toc130990026)  [Table 2-141 – EVC timing and HRD descriptor 156](#_Toc130990027)  [Table 2-142 – LCEVC video descriptor 157](#_Toc130990028)  [Table 2-143 – LCEVC linkage descriptor 157](#_Toc130990029)  [Table 2-144 – Media service kind descriptor 158](#_Toc130990030)  [Table 2-145 – media\_description\_flag 159](#_Toc130990031)  [Table 2-146 – Media type indicator 159](#_Toc130990032)  [Table 2-147 – ID\_length\_code 159](#_Toc130990033)  [Table 2-148 – ID\_type 159](#_Toc130990034)  [Table 2-149 – configuration type values 160](#_Toc130990035)  [Table 2-150 – lang\_len\_indicator 160](#_Toc130990036)  [Table 2-151 – Media service type values 160](#_Toc130990037)  [Table 2-152 – Carriage of individual ISO/IEC 14496 streams in Rec. ITU-T H.222.0 | ISO/IEC 13818-1 170](#_Toc130990038)  [Table 2-153 – Section syntax for transport of ISO/IEC 14496 stream 175](#_Toc130990039)  [Table 2-154 – ISO/IEC defined options for carriage of an ISO/IEC 14496 scene and associated streams in Rec. ITU-T H.222.0 | ISO/IEC 13818-1 179](#_Toc130990040)  [Table 2-155 – Metadata Access Unit Wrapper 185](#_Toc130990041)  [Table 2-156 – Metadata AU cell 185](#_Toc130990042)  [Table 2-157 – Cell fragment indication 185](#_Toc130990043)  [Table 2-158 – Section syntax for transport of metadata 186](#_Toc130990044)  [Table 2-159 – Section fragment indication 187](#_Toc130990045)  [Table 2-160 – View and dependency representation delimiter NAL unit 196](#_Toc130990046)  [Table 2-161 – Implied hierarchy\_layer\_index if no hierarchy descriptors are used 216](#_Toc130990047)  [Table 2-162 – Green access unit section syntax 223](#_Toc130990048)  [Table 2-163 – Green access unit 223](#_Toc130990049)  [Table 2-164 – Quality Access Unit 226](#_Toc130990050)  [Table B.1 – DSM-CC syntax 243](#_Toc130990051)  [Table B.2 – Command\_id assigned values 243](#_Toc130990052)  [Table B.3 – DSM-CC control 244](#_Toc130990053)  [Table B.4 – Select mode assigned values 245](#_Toc130990054)  [Table B.5 – DSM-CC Acknowledgement 246](#_Toc130990055)  [Table B.6 – Time code 247](#_Toc130990056)  [Table C.1 – Composite\_descriptor 253](#_Toc130990057)  [Table C.2 – Sub-descriptor 253](#_Toc130990058)  [Table C.3 – Program association table bandwidth usage (bit/s) Number of programs per transport stream 254](#_Toc130990059)  [Table C.4 – Program map table bandwidth usage (bit/s) Number of programs per transport stream 254](#_Toc130990060)  [Table D.1 – Re-multiplexing strategy 261](#_Toc130990061)  [Table E.1 – PES packet header example 265](#_Toc130990062)  [Table S.1 – J2K Access unit elementary stream header 296](#_Toc130990063)  [Table S.2 – Operating levels and maximum buffer size for JPEG 2000 broadcast profiles (from Table A.49 in Rec. ITU-T T.800 (2015) | ISO/IEC 15444-1:2016) 301](#_Toc130990064)  [Table T.1 – 'codecs' parameter values for some specific stream\_type values 303](#_Toc130990065)  [Table U.1 – Variable field length notation example 305](#_Toc130990066)  [Table U.1](#_Toc130990067)*[bis](#_Toc130990067)* [– Table U.1 in equivalent full notation 306](#_Toc130990067)  [Table U.2 – TEMI access unit 306](#_Toc130990068)  [Table U.3 – AF descriptor tags 307](#_Toc130990069)  [Table U.4 – TEMI location descriptor 308](#_Toc130990070)  [Table U.5 – TEMI URL scheme types 308](#_Toc130990071)  [Table U.6 – TEMI service types 309](#_Toc130990072)  [Table U.7 – TEMI base URL descriptor 309](#_Toc130990073)  [Table U.8 – TEMI timeline descriptor 310](#_Toc130990074)  [Table U.9 – TEMI MPEG-H\_3dAudio\_extStreamID descriptor 312](#_Toc130990075)  [Table U.10 – Boundary descriptor 313](#_Toc130990076)  [Table U.11 – sequence\_number\_length\_code interpretation 313](#_Toc130990077)  [Table U.12 – Labelling Descriptor 314](#_Toc130990078)  [Table U.13 – label\_length\_code interpretation 314](#_Toc130990079)  [Table U.14 – label\_type values 314](#_Toc130990080)  [Table U.15 – HEVC tile substream af\_descriptor 315](#_Toc130990081)  [Table W.1 – JPEG XS Access unit elementary stream header (jxes header) 319](#_Toc130990082) |

**List of Figures**

| Page |
| --- |
| [Figure Intro. 1 – Simplified overview of the scope of this Recommendation | International Standard xiv](#_Toc130991386)  [Figure Intro. 2 – Prototypical transport demultiplexing and decoding example xvi](#_Toc130991387)  [Figure Intro. 3 – Prototypical transport multiplexing example xvi](#_Toc130991388)  [Figure Intro. 4 – Prototypical transport stream to program stream conversion xvi](#_Toc130991389)  [Figure Intro. 5 – Prototypical decoder for program streams xvii](#_Toc130991390)  [Figure 2-1 – Transport stream system target decoder notation 16](#_Toc130991391)  [Figure 2-2 – Program stream system target decoder notation 63](#_Toc130991392)  [Figure 2-3 – Target background grid descriptor display area 84](#_Toc130991393)  [Figure 2-4 – Media Service Kind Descriptor semantics at program level 163](#_Toc130991394)  [Figure 2-5 – –Media Program Kind Descriptor semantics at elementary stream level 164](#_Toc130991395)  [Figure 2-6 – T-STD model extensions for individual ISO/IEC 14496 elementary streams 171](#_Toc130991396)  [Figure 2-7 – T-STD model for ISO/IEC 14496 content 176](#_Toc130991397)  [Figure 2-8 – P-STD model for ISO/IEC 14496 Systems stream 179](#_Toc130991398)  [Figure 2-9 – Timing model for delivery of content and metadata 182](#_Toc130991399)  [Figure 2-10 – Delivery of metadata in PES packets 183](#_Toc130991400)  [Figure 2-11 – Metadata signalling and referencing 189](#_Toc130991401)  [Figure 2-12 – Metadata decoding in the STD 189](#_Toc130991402)  [Figure 2-13 – T-STD model extensions for Rec. ITU-T H.264 | ISO/IEC 14496-10 video 193](#_Toc130991403)  [Figure 2-14 – P-STD model extensions for Rec. ITU-T H.264 | ISO/IEC 14496-10 video 195](#_Toc130991404)  [Figure 2-15 – T-STD model extensions for Rec. ITU-T H.264 | ISO/IEC 14496-10 Video with scalable video sub-bitstreams 196](#_Toc130991405)  [Figure 2-16 – P-STD model extensions for Rec. ITU-T H.264 | ISO/IEC 14496-10 Video with scalable video sub-bitstreams 199](#_Toc130991406)  [Figure 2-17 – T-STD model extensions for Rec. ITU-T H.264 | ISO/IEC 14496-10 Video with MVC video sub-bitstreams 201](#_Toc130991407)  [Figure 2-18 – P-STD model extensions for Rec. ITU-T H.264 | ISO/IEC 14496-10 Video with MVC video sub-bitstreams 205](#_Toc130991408)  [Figure 2-19 – T-STD model extensions for ISO/IEC 14496-17 text streams 207](#_Toc130991409)  [Figure 2-20 – T-STD model extensions for single layer HEVC 210](#_Toc130991410)  [Figure 2-21 – T-STD model extensions for layered transport of HEVC temporal video subsets 212](#_Toc130991411)  [Figure 2-22 – T-STD model extensions for bitstream-partition-specific CPB operation 214](#_Toc130991412)  [Figure 2-23 – T-STD model extensions for transport of HEVC tiles through individual ESs 218](#_Toc130991413)  [Figure 2-24 – T-STD model extensions for transport of HEVC tiles in a common ES using AF descriptors 220](#_Toc130991414)  [Figure 2-25 – T-STD model extension for transport of HEVC tiles in a common ES ignoring AF descriptors 222](#_Toc130991415)  [Figure 2-26 – T-STD model extension for green access units 224](#_Toc130991416)  [Figure 2-27 – Transport stream system target decoder for multiple audio elementary streams 226](#_Toc130991417)  [Figure 2-28 – Quality Access Unit decoder processing model 227](#_Toc130991418)  [Figure 2-29 – T-STD model extensions for single layer VVC 229](#_Toc130991419)  [Figure 2-30 – T-STD model extensions for layered transport of VVC temporal video subsets 231](#_Toc130991420)  [Figure 2-31 – T-STD model extensions for single layer EVC 234](#_Toc130991421)  [Figure A.1 – 32-bit CRC decoder model 238](#_Toc130991422)  [Figure B.1 – Configuration of DSM-CC application 241](#_Toc130991423)  [Figure B.2 – BSM-CC bitstream decoded as a stand-alone bitstream 241](#_Toc130991424)  [Figure B.3 – DSM-CC bitstream decoded as part of the system bitstream 242](#_Toc130991425)  [Figure C.1 – Program and network mapping relationships 251](#_Toc130991426)  [Figure D.1 – Constant delay model 256](#_Toc130991427)  [Figure D.2 – STC recovery using PLL 260](#_Toc130991428)  [Figure F.1 – Transport stream syntax diagram 266](#_Toc130991429)  [Figure F.2 – PES packet syntax diagram 267](#_Toc130991430)  [Figure F.3 – Program association section diagram 267](#_Toc130991431)  [Figure F.4 – Conditional access section diagram 267](#_Toc130991432)  [Figure F.5 – TS program map section diagram 268](#_Toc130991433)  [Figure F.6 – Private section diagram 268](#_Toc130991434)  [Figure F.7 – Program stream diagram 269](#_Toc130991435)  [Figure F.8 – Program stream map diagram 269](#_Toc130991436)  [Figure J.1 – Sending system streams over a jitter-inducing network 274](#_Toc130991437)  [Figure J.2 – Jitter-smoothing using network-layer timestamps 274](#_Toc130991438)  [Figure J.3 – Integrated dejittering and MPEG-2 decoding 275](#_Toc130991439)  [Figure R.1 – Example of ISO/IEC 14496 content in a program stream 290](#_Toc130991440)  [Figure R.2 – Example of ISO/IEC 14496 content in a transport stream 291](#_Toc130991441)  [Figure R.3 – Usage of MPEG-4 in a transport stream with BIFS scene referring to native PES 292](#_Toc130991442)  [Figure R.4 – Usage of MPEG-4 in a transport stream with an ODUpdate\_descriptor carrying an image ObjectDescriptor in the PMT 293](#_Toc130991443)  [Figure S.1 – Structure and order of JPEG 2000 access units 298](#_Toc130991444)  [Figure S.2 – T-STD model extensions for J2K Video 299](#_Toc130991445)  [Figure U.1 – Stream partitioning into 2 and 5 second segments 313](#_Toc130991446)  [Figure V.1 – Illustration of HEVC tiled encoding of panoramic content beyond UHD 316](#_Toc130991447)  [Figure V.2 – Example of HEVC tile substream identification 317](#_Toc130991448)  [Figure V.3 – Example of subregion layout for a 3 x 3 RoI 317](#_Toc130991449)  [Figure W.1 – Structure and order of JPEG XS access units 320](#_Toc130991450)  [Figure W.2 – T-STD model extensions for JPEG XS Video 322](#_Toc130991451) |

Introduction

The systems part of this Recommendation | International Standard addresses the combining of one or more elementary streams of video and audio, as well as other data, into single or multiple streams which are suitable for storage or transmission. Systems coding follows the syntactical and semantic rules imposed by this Specification and provides information to enable synchronized decoding of decoder buffers over a wide range of retrieval or receipt conditions.

System coding shall be specified in two forms: the transport stream and the program stream. Each is optimized for a different set of applications. Both the transport stream and program stream defined in this Recommendation | International Standard provide coding syntax which is necessary and sufficient to synchronize the decoding and presentation of the video and audio information, while ensuring that data buffers in the decoders do not overflow or underflow. Information is coded in the syntax using time stamps concerning the decoding and presentation of coded audio and visual data and time stamps concerning the delivery of the data stream itself. Both stream definitions are packet-oriented multiplexes.

The basic multiplexing approach for single video and audio elementary streams is illustrated in Figure Intro. 1. The video and audio data is encoded as described in Rec. ITU-T H.262 | ISO/IEC 13818-2 and ISO/IEC 13818-3. The resulting compressed elementary streams are packetized to produce PES packets.Information needed to use PES packets independently of either transport streams or program streams may be added when PES packets are formed. This information is not needed and need not be added when PES packets are further combined with system level information to form transport streams or program streams. This systems standard covers those processes to the right of the vertical dashed line.



Figure Intro. 1 – Simplified overview of the scope of this Recommendation | International Standard

The program stream is analogous and similar to the ISO/IEC 11172 systems layer. It results from combining one or more streams of PES packets, which have a common time base, into a single stream.

For applications that require the elementary streams that comprise a single program to be in separate streams that are not multiplexed, the elementary streams can also be encoded as separate program streams, one per elementary stream, with a common time base. In this case the values encoded in the SCR fields of the various streams shall be consistent.

Like the single program stream, all elementary streams can be decoded with synchronization.

The program stream is designed for use in relatively error-free environments and is suitable for applications which may involve software processing of system information such as interactive multi-media applications. Program stream packets may be of variable and relatively great length.

The transport stream combines one or more programs with one or more independent time bases into a single stream. PES packets made up of elementary streams that form a program share a common timebase. The transport stream is designed for use in environments where errors are likely, such as storage or transmission in lossy or noisy media. Transport stream packets are 188 bytes in length.

Program and transport streams are designed for different applications and their definitions do not strictly follow a layered model. It is possible and reasonable to convert from one to the other; however, one is not a subset or superset of the other. In particular, extracting the contents of a program from a transport stream and creating a valid program stream is possible and is accomplished through the common interchange format of PES packets, but not all of the fields needed in a program stream are contained within the transport stream; some must be derived. The transport stream may be used to span a range of layers in a layered model, and is designed for efficiency and ease of implementation in high bandwidth applications.

The scope of syntactical and semantic rules set forth in the systems specification differs: the syntactical rules apply to systems layer coding only, and do not extend to the compression layer coding of the video and audio specifications; by contrast, the semantic rules apply to the combined stream in its entirety.

The systems specification does not specify the architecture or implementation of encoders or decoders, nor those of multiplexors or demultiplexors. However, bit stream properties do impose functional and performance requirements on encoders, decoders, multiplexors and demultiplexors. For instance, encoders must meet minimum clock tolerance requirements. Notwithstanding this and other requirements, a considerable degree of freedom exists in the design and implementation of encoders, decoders, multiplexors, and demultiplexors.

Intro. 1 Transport stream

The transport stream is a stream definition which is tailored for communicating or storing one or more programs of coded data according to Rec. ITU-T H.262 | ISO/IEC 13818-2 and ISO/IEC 13818-3 and other data in environments in which significant errors may occur. Such errors may be manifested as bit value errors or loss of packets.

Transport streams may be either fixed or variable rate. In either case the constituent elementary streams may either be fixed or variable rate. The syntax and semantic constraints on the stream are identical in each of these cases. The transport stream rate is defined by the values and locations of program clock reference (PCR) fields, which in general are separate PCR fields for each program.

There are some difficulties with constructing and delivering a transport stream containing multiple programs with independent time bases such that the overall bit rate is variable. Refer to 2.4.2.3.

The transport stream may be constructed by any method that results in a valid stream. It is possible to construct transport streams containing one or more programs from elementary coded data streams, from program streams, or from other transport streams which may themselves contain one or more programs.

The transport stream is designed in such a way that several operations on a transport stream are possible with minimum effort. Among these are:

1) Retrieve the coded data from one program within the transport stream, decode it and present the decoded results as shown in Figure Intro. 2.

2) Extract the transport stream packets from one program within the transport stream and produce as output a different transport stream with only that one program as shown in Figure Intro. 3.

3) Extract the transport stream packets of one or more programs from one or more transport streams and produce as output a different transport stream (not illustrated).

4) Extract the contents of one program from the transport stream and produce as output a program stream containing that one program as shown in Figure Intro. 4.

5) Take a program stream, convert it into a transport stream to carry it over a lossy environment, and then recover a valid, and in certain cases, identical program stream.

Figure Intro. 2 and Figure Intro. 3 illustrate prototypical demultiplexing and decoding systems which take as input a transport stream. Figure Intro. 2 illustrates the first case, where a transport stream is directly demultiplexed and decoded. Transport streams are constructed in two layers:

– a system layer; and

– a compression layer.

The input stream to the transport stream decoder has a system layer wrapped about a compression layer. Input streams to the video and audio decoders have only the compression layer.

Operations performed by the prototypical decoder which accepts transport streams either apply to the entire transport stream ("multiplex-wide operations"), or to individual elementary streams ("stream-specific operations"). The transport stream system layer is divided into two sub-layers, one for multiplex-wide operations (the transport stream packet layer), and one for stream-specific operations (the PES packet layer).

A prototypical decoder for transport streams, including audio and video, is also depicted in Figure Intro. 2 to illustrate the function of a decoder. The architecture is not unique – some system decoder functions, such as decoder timing control, might equally well be distributed among elementary stream decoders and the channel-specific decoder – but this figure is useful for discussion. Likewise, indication of errors detected by the channel-specific decoder to the individual audio and video decoders may be performed in various ways and such communication paths are not shown in the diagram. The prototypical decoder design does not imply any normative requirement for the design of a transport stream decoder. Indeed non-audio/video data is also allowed, but not shown.



Figure Intro. 2 – Prototypical transport demultiplexing and decoding example

Figure Intro. 3 illustrates the second case, where a transport stream containing multiple programs is converted into a transport stream containing a single program. In this case the re-multiplexing operation may necessitate the correction of program clock reference (PCR) values to account for changes in the PCR locations in the bit stream.



Figure Intro. 3 – Prototypical transport multiplexing example

Figure Intro. 4 illustrates a case in which a multi-program transport stream is first demultiplexed and then converted into a program stream.

Figures Intro. 3 and Intro. 4 indicate that it is possible and reasonable to convert between different types and configurations of transport streams. There are specific fields defined in the transport stream and program stream syntax which facilitate the conversions illustrated. There is no requirement that specific implementations of demultiplexors or decoders include all of these functions.



Figure Intro. 4 – Prototypical transport stream to program stream conversion

Intro. 2 Program stream

The program stream is a stream definition which is tailored for communicating or storing one program of coded data and other data in environments where errors are very unlikely, and where processing of system coding, e.g., by software, is a major consideration.

Program streams may be either fixed or variable rate. In either case, the constituent elementary streams may be either fixed or variable rate. The syntax and semantics constraints on the stream are identical in each case. The program stream rate is defined by the values and locations of the system clock reference (SCR) and mux\_rate fields.

A prototypical audio/video program stream decoder system is depicted in Figure Intro. 5. The architecture is not unique – system decoder functions including decoder timing control might as equally well be distributed among elementary stream decoders and the channel-specific decoder – but this figure is useful for discussion. The prototypical decoder design does not imply any normative requirement for the design of a program stream decoder. Indeed non-audio/video data is also allowed, but not shown.



Figure Intro. 5 – Prototypical decoder for program streams

The prototypical decoder for program streams shown in Figure Intro. 5 is composed of system, video and audio decoders conforming to Parts 1, 2 and 3, respectively, of ISO/IEC 13818. In this decoder, the multiplexed coded representation of one or more audio and/or video streams is assumed to be stored or communicated on some channel in some channel-specific format. The channel-specific format is not governed by this Recommendation | International Standard, nor is the channel-specific decoding part of the prototypical decoder.

The prototypical decoder accepts as input a program stream and relies on a program stream decoder to extract timing information from the stream. The program stream decoder demultiplexes the stream, and the elementary streams so produced serve as inputs to video and audio decoders, whose outputs are decoded video and audio signals. Included in the design, but not shown in the figure, is the flow of timing information among the program stream decoder, the video and audio decoders, and the channel-specific decoder. The video and audio decoders are synchronized with each other and with the channel using this timing information.

Program streams are constructed in two layers: a system layer and a compression layer. The input stream to the program stream decoder has a system layer wrapped about a compression layer. Input streams to the video and audio decoders have only the compression layer.

Operations performed by the prototypical decoder either apply to the entire program stream ("multiplex-wide operations"), or to individual elementary streams ("stream-specific operations"). The program stream system layer is divided into two sub-layers, one for multiplex-wide operations (the pack layer), and one for stream-specific operations (the PES packet layer).

Intro. 3 Conversion between transport stream and program stream

It may be possible and reasonable to convert between transport streams and program streams by means of PES packets. This results from the specification of transport stream and program stream as embodied in 2.4.1 and 2.5.1 of the normative requirements of this Recommendation | International Standard. PES packets may, with some constraints, be mapped directly from the payload of one multiplexed bit stream into the payload of another multiplexed bit stream. It is possible to identify the correct order of PES packets in a program to assist with this if the program\_packet\_sequence\_counter is present in all PES packets.

Certain other information necessary for conversion, e.g., the relationship between elementary streams, is available in tables and headers in both streams. Such data, if available, shall be correct in any stream before and after conversion.

Intro. 4 Packetized elementary stream

Transport streams and program streams are each logically constructed from PES packets, as indicated in the syntax definitions in 2.4.3.6. PES packets shall be used to convert between transport streams and program streams; in some cases the PES packets need not be modified when performing such conversions. PES packets may be much larger than the size of a transport stream packet.

A continuous sequence of PES packets of one elementary stream with one stream ID may be used to construct a PES Stream. When PES packets are used to form a PES stream, they shall include elementary stream clock reference (ESCR) fields and elementary stream rate (ES\_Rate) fields, with constraints as defined in 2.4.3.8. The PES stream data shall be contiguous bytes from the elementary stream in their original order. PES streams do not contain some necessary system information which is contained in program streams and transport streams. Examples include the information in the pack header, system header, program stream map, program stream directory, program map table, and elements of the transport stream packet syntax.

The PES stream is a logical construct that may be useful within implementations of this Recommendation | International Standard; however, it is not defined as a stream for interchange and interoperability. Applications requiring streams containing only one elementary stream can use program streams or transport streams which each contain only one elementary stream. These streams contain all of the necessary system information. Multiple program streams or transport streams, each containing a single elementary stream, can be constructed with a common time base and therefore carry a complete program, i.e., with audio and video.

Intro. 5 Timing model

Systems, video and audio all have a timing model in which the end-to-end delay from the signal input to an encoder to the signal output from a decoder is a constant. This delay is the sum of encoding, encoder buffering, multiplexing, communication or storage, demultiplexing, decoder buffering, decoding, and presentation delays. As part of this timing model all video pictures and audio samples are presented exactly once, unless specifically coded to the contrary, and the inter-picture interval and audio sample rate are the same at the decoder as at the encoder. The system stream coding contains timing information which can be used to implement systems which embody constant end-to-end delay. It is possible to implement decoders which do not follow this model exactly; however, in such cases it is the decoder's responsibility to perform in an acceptable manner. The timing is embodied in the normative specifications of this Recommendation | International Standard, which must be adhered to by all valid bit streams, regardless of the means of creating them.

All timing is defined in terms of a common system clock, referred to as a system time clock (STC). In the program stream this clock may have an exactly specified ratio to the video or audio sample clocks, or it may have an operating frequency which differs slightly from the exact ratio while still providing precise end-to-end timing and clock recovery.

In the transport stream the system clock frequency is constrained to have the exactly specified ratio to the audio and video sample clocks at all times; the effect of this constraint is to simplify sample rate recovery in decoders.

Intro. 6 Conditional access

Encryption and scrambling for conditional access to programs encoded in the program and transport streams is supported by the system data stream definitions. Conditional access mechanisms are not specified here. The stream definitions are designed so that implementation of practical conditional access systems is reasonable, and there are some syntactical elements specified which provide specific support for such systems.

Intro. 7 Multiplex-wide operations

Multiplex-wide operations include the coordination of data retrieval of the channel, the adjustment of clocks, and the management of buffers. The tasks are intimately related. If the rate of data delivery of the channel is controllable, then data delivery may be adjusted so that decoder buffers neither overflow nor underflow; but if the data rate is not controllable, then elementary stream decoders must slave their timing to the data received from the channel to avoid overflow or underflow.

Program streams are composed of packs whose headers facilitate the above tasks. Pack headers specify intended times at which each byte is to enter the program stream Decoder from the channel, and this target arrival schedule serves as a reference for clock correction and buffer management. The schedule need not be followed exactly by decoders, but they must compensate for deviations about it.

Similarly, transport streams are composed of transport stream packets with headers containing information which specifies the times at which each byte is intended to enter a transport stream decoder from the channel. This schedule provides exactly the same function as that which is specified in the program stream.

An additional multiplex-wide operation is a decoder's ability to establish what resources are required to decode a transport stream or program stream. The first pack of each program stream conveys parameters to assist decoders in this task. Included, for example, are the stream's maximum data rate and the highest number of simultaneous video channels. The transport stream likewise contains globally useful information.

The transport stream and program stream each contain information which identifies the pertinent characteristics of, and relationships between, the elementary streams which constitute each program. Such information may include the language spoken in audio channels, as well as the relationship between video streams when multi-layer video coding is implemented.

Intro. 8 Individual stream operations (PES packet layer)

The principal stream-specific operations are:

1) demultiplexing; and

2) synchronizing playback of multiple elementary streams.

Intro. 8.1 Demultiplexing

On encoding, program streams are formed by multiplexing elementary streams, and transport streams are formed by multiplexing elementary streams, program streams, or the contents of other transport streams. Elementary streams may include private, reserved, and padding streams in addition to audio and video streams. The streams are temporally subdivided into packets, and the packets are serialized. A PES packet contains coded bytes from one and only one elementary stream.

In the program stream both fixed and variable packet lengths are allowed subject to constraints as specified in 2.5.1 and 2.5.2. For transport streams the packet length is 188 bytes. Both fixed and variable PES packet lengths are allowed, and will be relatively long in most applications.

On decoding, demultiplexing is required to reconstitute elementary streams from the multiplexed program stream or transport stream. Stream\_id codes in program stream packet headers, and packet ID codes in the transport stream make this possible.

Intro. 8.2 Synchronization

Synchronization among multiple elementary streams is accomplished with presentation time stamps (PTSs) in the program stream and transport streams. Time stamps are generally in units of 90 kHz, but the system clock reference (SCR), the program clock reference (PCR) and the optional elementary stream clock reference (ESCR) have extensions with a resolution of 27 MHz. Decoding of N-elementary streams is synchronized by adjusting the decoding of streams to a common master time base rather than by adjusting the decoding of one stream to match that of another. The master time base may be one of the N-decoders' clocks, the data source's clock, or it may be some external clock.

Each program in a transport stream, which may contain multiple programs, may have its own time base. The time bases of different programs within a transport stream may be different.

Because PTSs apply to the decoding of individual elementary streams, they reside in the PES packet layer of both the transport streams and program streams. End-to-end synchronization occurs when encoders save time stamps at capture time, when the time stamps propagate with associated coded data to decoders, and when decoders use those time stamps to schedule presentations.

Synchronization of a decoding system with a channel is achieved through the use of the SCR in the program stream and by its analogue, the PCR, in the transport stream. The SCR and PCR are time stamps encoding the timing of the bit stream itself, and are derived from the same time base used for the audio and video PTS values from the same program. Since each program may have its own time base, there are separate PCR fields for each program in a transport stream containing multiple programs. In some cases it may be possible for programs to share PCR fields. Refer to 2.4.4, program-specific information (PSI), for the method of identifying which PCR is associated with a program. A program shall have one and only one PCR time base associated with it.

Intro. 8.3 Relation to compression layer

The PES packet layer is independent of the compression layer in some senses, but not in all. It is independent in the sense that PES packet payloads need not start at compression layer start codes, as defined in Parts 2 and 3 of ISO/IEC 13818. For example, video start codes may occur anywhere within the payload of a PES packet, and start codes may be split by a PES packet header. However, time stamps encoded in PES packet headers apply to presentation times of compression layer constructs (namely, presentation units). In addition, when the elementary stream data conforms to Rec. ITU-T H.262 | ISO/IEC 13818-2 or ISO/IEC 13818-3, the PES\_packet\_data\_bytes shall be byte aligned to the bytes of this Recommendation | International Standard.

Intro. 9 System reference decoder

Part 1 of ISO/IEC 13818 employs a "system target decoder" (STD), one for transport streams (refer to 2.4.2) referred to as "transport system target decoder" (T-STD) and one for program streams (refer to 2.5.2) referred to as "program system target decoder" (P-STD), to provide a formalism for timing and buffering relationships. Because the STD is parameterized in terms of Rec. ITU-T H.222.0 | ISO/IEC 13818-1 fields (for example, buffer sizes) each elementary stream leads to its own parameterization of the STD. Encoders shall produce bit streams that meet the appropriate STD's constraints. Physical decoders may assume that a stream plays properly on its STD. The physical decoder must compensate for ways in which its design differs from that of the STD.

Intro. 10 Applications

The streams defined in this Recommendation | International Standard are intended to be as useful as possible to a wide variety of applications. Application developers should select the most appropriate stream.

Modern data communications networks may be capable of supporting Rec. ITU-T H.222.0 | ISO/IEC 13818-1 video and ISO/IEC 13818 audio. A real-time transport protocol is required. The program stream may be suitable for transmission on such networks.

The program stream is also suitable for multimedia applications on CD-ROM. Software processing of the program stream may be appropriate.

The transport stream may be more suitable for error-prone environments, such as those used for distributing compressed bit-streams over long-distance networks and in broadcast systems.

Many applications require storage and retrieval of Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstreams on various digital storage media (DSM). A digital storage media command and control (DSM-CC) protocol is specified in Annex B and Part 6 of ISO/IEC 13818 in order to facilitate the control of such media.

INTERNATIONAL STANDARD

ITU-T RECOMMENDATION

Information technology – Generic coding of moving pictures and   
associated audio information: Systems

SECTION 1 – GENERAL

## 1.1 Scope

This Recommendation | International Standard specifies the system layer of the coding. It was developed principally to support the combination of the video and audio coding methods defined in Parts 2 and 3 of ISO/IEC 13818. The system layer supports six basic functions:

1) the synchronization of multiple compressed streams on decoding;

2) the interleaving of multiple compressed streams into a single stream;

3) the initialization of buffering for decoding start up;

4) continuous buffer management;

5) time identification;

6) multiplexing and signalling of various components in a system stream.

A Rec. ITU-T H.222.0 | ISO/IEC 13818-1 multiplexed bit stream is either a transport stream or a program stream. Both streams are constructed from PES packets and packets containing other necessary information. Both stream types support multiplexing of video and audio compressed streams from one program with a common time base. The transport stream additionally supports the multiplexing of video and audio compressed streams from multiple programs with independent time bases. For almost error-free environments the program stream is generally more appropriate, supporting software processing of program information. The transport stream is more suitable for use in environments where errors are likely.

A Rec. ITU-T H.222.0 | ISO/IEC 13818-1 multiplexed bit stream, whether a transport stream or a program stream, is constructed in two layers: the outermost layer is the system layer, and the innermost is the compression layer. The system layer provides the functions necessary for using one or more compressed data streams in a system. The video and audio parts of this Specification define the compression coding layer for audio and video data. Coding of other types of data is not defined by this Specification, but is supported by the system layer provided that the other types of data adhere to the constraints defined in 2.7.

## 1.2 Normative references

The following Recommendations and International Standards contain provisions which, through reference in this text, constitute provisions of this Recommendation | International Standard. At the time of publication, the editions indicated were valid. All Recommendations and Standards are subject to revision, and parties to agreements based on this Recommendation | International Standard are encouraged to investigate the possibility of applying the most recent edition of the Recommendations and Standards listed below. Members of IEC and ISO maintain registers of currently valid International Standards. The Telecommunication Standardization Bureau of the ITU maintains a list of currently valid ITU-T Recommendations.

### 1.2.1 Identical Recommendations | International Standards

– Recommendation ITU-T H.262 (2012) | ISO/IEC 13818-2:2013, *Information technology – Generic coding of moving pictures and associated audio information: Video*.

– Recommendation ITU-T T.800 (2019) | ISO/IEC 15444-1:2019, *Information technology – JPEG 2000 image coding system: Core coding system.*

### 1.2.2 Paired Recommendations | International Standards equivalent in technical content

* Recommendation ITU-T H.264 (2019), *Advanced video coding for generic audiovisual services*.

ISO/IEC 14496-10:2020, *Information technology – Coding of audio-visual objects – Part 10: Advanced video coding*.

– Recommendation ITU-T H.265 (2019), *High efficiency video coding.*

ISO/IEC 23008‑2:2020, *Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 2: High efficiency video coding*.

– Recommendation. ITU-T H.273 (2021), *Coding-independent code points for video signal type identification*.

ISO/IEC 23091-2:2021, *Information technology — Coding-independent code points — Part 2: Video*.

– Recommendation ITU-T T.171 (1996), *Protocols for interactive audiovisual services: coded representation of multimedia and hypermedia objects*.

ISO/IEC 13522-1:1997, *Information technology – Coding of Multimedia and Hypermedia information – Part 1: MHEG object representation – Base notation (ASN.1)*.

* Recommendation ITU-T H.266 (2020), *Versatile video coding*.

ISO/IEC 23090-3:2021, *Information technology – Coded Representation of Immersive Media – Part 3: Versatile video coding*.

* Recommendation ITU-T H.274 (2020), *Versatile supplemental enhancement information messages for coded video bitstreams*.

ISO/IEC 23002-7:2021 – *Information Technology – MPEG Video technologies – Part 7: Versatile supplemental enhancement information messages for coded video bitstreams*.

### 1.2.3 Additional references

– Recommendation ITU-R BT.709-6 (2015), *Parameter values for the HDTV standards for production and international programme exchange*.

– Recommendation ITU-R BT.1886 (2011), *Reference electro-optical transfer function for flat panel displays used in HDTV studio production.*

– Recommendation ITU-R BT.2020 (2015), *Parameter values for ultra-high definition television systems for production and international programme exchange.*

– Recommendation ITU-R BT.2100-2 (2018), *Image parameter values for high dynamic range television for use in production and international programme exchange.*

– ISO 639-2:1998, *Codes for the representation of names of languages – Part 2: Alpha-3 code*.

* ISO 8859-1:1998, *Information technology – 8-bit single-byte coded graphic character sets – Part 1: Latin alphabet No. 1*.

– ISO 15706-1:2002, *Information and documentation – International Standard Audiovisual Number (ISAN) – Part 1: Audiovisual work identifier*.

– ISO 15706-2:2007, *Information and documentation – International Standard Audiovisual Number (ISAN) – Part 2: Version identifier*.

* ISO/IEC 11172-1:1993, *Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 1: Systems*.

– ISO/IEC 11172-2:1993, *Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 2: Video*.

– ISO/IEC 11172-3:1993, *Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 3: Audio*.

– ISO/IEC 13818-3:1998, *Information technology – Generic coding of moving pictures and associated audio information – Part 3: Audio*.

* ISO/IEC 13818-6:1998, *Information technology – Generic coding of moving pictures and associated audio information – Part 6: Extensions for DSM-CC*.

– ISO/IEC 13818-7:2006, *Information technology – Generic coding of moving pictures and associated audio information – Part 7: Advanced Audio Coding (AAC)*.

– ISO/IEC 13818-11:2004, *Information technology – Generic coding of moving pictures and associated audio information – Part 11: IPMP on MPEG-2 systems*.

* ISO/IEC 14496-1:2010, *Information technology – Coding of audio-visual objects – Part 1: Systems*.
* ISO/IEC 14496-2:2004, *Information technology – Coding of audio-visual objects – Part 2: Visual*.

– ISO/IEC 14496-3:2019, *Information technology – Coding of audio-visual objects – Part 3: Audio*.

– ISO/IEC 14496-17:2006, *Information technology, Coding of audio-visual objects – Part 17: Streaming text format*.

– ISO/IEC 21122-1:2019, *JPEG XS low-latency lightweight image coding system – Part 1: Core coding system*.

– ISO/IEC 21122-2:2019, *JPEG XS low-latency lightweight image coding system – Part 2: Profiles and buffer models*.

– ISO/IEC 21122-3:2019, *JPEG XS low-latency lightweight image coding system – Part 3: Transport and container formats*.

– ISO/IEC 23001-8:2016, *Information technology – MPEG systems technologies – Part 8: Coding-independent code-points.*

– ISO/IEC 23001-10:2020, *Information technology – MPEG systems technologies – Part 10:* *Carriage of timed metadata metrics of media in ISO base media file format.*

– ISO/IEC 23001-11:2019, *Information technology – MPEG systems technologies – Part 11: Energy‑efficient media consumption (Green Metadata)*.

– ISO/IEC 23001-12:2018, *Information technology – MPEG systems technologies – Part 12: Sample Variants*.

– ISO/IEC 23001-13:2019, *Information technology – MPEG systems technologies – Part 13: Media Orchestration*.

– ISO/IEC 23003-3:2020, *Information technology – MPEG audio technologies – Part 3: Unified speech and audio coding.*

– ISO/IEC 23003-4:2020, *Information technology – MPEG audio technologies – Part 4: Dynamic Range Control*.

– ISO/IEC 23008-3:2019, *Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 3: 3D audio.*

– ISO/IEC 23091-2:2019, *Coding-independent code points – Part 2: Video*.

* ISO/IEC 23094-1:2020*, Information technology — General video coding — Part 1: Essential video coding*
* ISO/IEC 23094-2:2021, *Information technology – General video coding – Part 2: Low complexity enhancement video coding.*
* IETF BCP 47: IETF RFC 4647 (2006) *Matching of Language Tags* combined with IETF RFC 5646 (2009), *Tags for Identifying Languages*.

– IETF RFC 3986 (2005), *Uniform Resource Identifier (URI): Generic Syntax*.

– IETF RFC 5484 (2009), *Associating Time-Codes with RTP Streams*.

– ANSI/SCTE 35 (2019), *Digital Program Insertion Cueing Message for Cable*.

SECTION 2 – TECHNICAL ELEMENTS

## 2.1 Definitions

For the purposes of this Recommendation | International Standard, the following definitions apply. If specific to a Part, this is parenthetically noted.

**2.1.1 access unit (system)**: A coded representation of a presentation unit. In the case of audio, an access unit is the coded representation of an audio frame, whereby each audio frame carries data from one or more audio channels; an audio frame may carry for example one mono channel, or two stereo channels or seven surround sound channels.

In the case of video, an access unit includes all the coded data for a picture, and any stuffing that follows it, up to but not including the start of the next access unit. If a picture is not preceded by a group\_start\_code or a sequence\_header\_code, the access unit begins with the picture start code. If a picture is preceded by a group\_start\_code and/or a sequence\_header\_code, the access unit begins with the first byte of the first of these start codes. If it is the last picture preceding a sequence\_end\_code in the bitstream, all bytes between the last byte of the coded picture and the sequence\_end\_code (including the sequence\_end\_code) belong to the access unit.

For the definition of an access unit for Rec. ITU-T H.264 | ISO/IEC 14496-10 video, see the AVC access unit definition in 2.1.3.

In the case of an ISO/IEC 14496-17 text stream, see ISO/IEC 14496-17 for the definition of an access unit.

**2.1.2 AVC 24-hour picture (system)**: An advanced video coding (AVC) access unit with a presentation time that is more than 24 hours in the future. For the purpose of this definition, AVC access unit n has a presentation time that is more than 24 hours in the future if the difference between the initial arrival time tai(n) and the DPB output time to,dpb(n) is more than 24 hours.

**2.1.3 AVC access unit (system)**: An access unit as defined for byte streams in Rec. ITU-T H.264 | ISO/IEC 14496‑10 with the constraints specified in 2.14.1.

**2.1.4 AVC slice (system)**: A byte\_stream\_nal\_unit as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10 with nal\_unit\_type values of 1 or 5, or a byte\_stream\_nal\_unit data structure with nal\_unit\_type value of 2 and any associated byte\_stream\_nal\_unit data structures with nal\_unit\_type equal to 3 and/or 4.

**2.1.5 AVC still picture (system)**: An AVC still picture consists of an AVC access unit containing an IDR picture, preceded by SPS and PPS NAL units that carry sufficient information to correctly decode the IDR picture. Preceding an AVC still picture, there shall be another AVC still picture or an end of sequence NAL unit terminating a preceding coded video sequence unless the AVC still picture is the very first access unit in the video stream.

**2.1.6 AVC video sequence (system)**: Coded video sequence as defined in 3.30 of Rec. ITU-T H.264 | ISO/IEC 14496-10.

**2.1.7 AVC video stream (system)**: A Rec. ITU-T H.264 | ISO/IEC 14496-10 stream. An AVC video stream consists of one or more AVC video sequences. An AVC video stream may also result from re-assembling video sub-bitstreams.

**2.1.8 AVC video sub-bitstream of MVC**: The video sub-bitstream that contains the base view as defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10, containing all VCL NAL units associated with the minimum value of view order index present in each AVC video sequence of the AVC video stream. The AVC video sub-bitstream of MVC may additionally contain the associated NAL units with nal\_unit\_type syntax element equal to 14 (prefix NAL units), as defined for MVC in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10.

**2.1.9 AVC video sub-bitstream of MVCD**: The video sub-bitstream that contains the base view as defined in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496-10, containing all VCL NAL units associated with the minimum value of view order index present in each AVC video sequence of the AVC video stream. The AVC video sub-bitstream of MVCD may additionally contain the associated NAL units with nal\_unit\_type syntax element equal to 14 (prefix NAL units), as defined for MVC in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10.

**2.1.10 AVC video sub-bitstream of SVC**: The video sub-bitstream that contains the base layer as defined in Annex G of Rec. ITU-T H.264 | ISO/IEC 14496-10 and that shall additionally contain NAL units with nal\_unit\_type equal to 14 (prefix NAL units) ) as defined for scalable video coding (SVC) in Annex G of Rec. ITU-T H.264 | ISO/IEC 14496-10. The AVC video sub-bitstream of SVC contains all VCL NAL units associated with dependency\_id equal to 0.

**2.1.11 bitrate**: The rate at which the compressed bit stream is delivered from the channel to the input of a decoder.

**2.1.12 byte aligned**: A bit in a coded bit stream is byte-aligned if its position is a multiple of 8-bits from the first bit in the stream.

**2.1.13 channel**: A digital medium that stores or transports a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream.

**2.1.14 coded B-frame**: A B-frame picture or a pair of B-field pictures.

**2.1.15 coded frame**: A coded frame is a coded I-frame, coded B-frame or a coded P-frame.

**2.1.16 coded I-frame**: An I-frame picture or a pair of field pictures where the first field picture is an I-picture and the second field picture is either an I-picture or a P-picture.

**2.1.17 coded P-frame**: A P-frame picture or a pair of P-field pictures.

**2.1.18 coded** **representation**: A data element as represented in its encoded form.

**2.1.19 compression**: Reduction in the number of bits used to represent an item of data.

**2.1.20 constant bitrate**: Operation where the bitrate is constant from start to finish of the compressed bit stream.

**2.1.21 constrained** **system parameter stream (CSPS) (system)**: A program stream for which the constraints defined in 2.7.9 apply.

**2.1.22 cyclic redundancy check (CRC)**: The CRC to verify the correctness of data.

**2.1.23 data** **element**: An item of data as represented before encoding and after decoding.

**2.1.24 decoded stream**: The decoded reconstruction of a compressed bit stream.

**2.1.25 decoder**: An embodiment of a decoding process.

**2.1.26 decoding (process)**: The process defined in this Recommendation | International Standard that reads an input-coded bit stream and outputs decoded pictures or audio samples.

**2.1.27 decoding time-stamp (DTS) (system)**: A field that may be present in a PES packet header that indicates the time that an access unit is decoded in the system target decoder.

**2.1.28 digital storage media (DSM)**: A digital storage or transmission device or system.

**2.1.29 DSM-CC**:Digital storage media command and control.

**2.1.30 entitlement control message (ECM)**:Entitlement control messages are private conditional access information which specify control words and possibly other, typically stream-specific, scrambling and/or control parameters.

**2.1.31 entitlement management message (EMM)**:Entitlement management messages are private conditional access information which specify the authorization levels or the services of specific decoders. They may be addressed to single decoders or groups of decoders.

**2.1.32 editing**: The process by which one or more compressed bit streams are manipulated to produce a new compressed bit stream. Edited bit streams meet the same requirements as streams which are not edited.

**2.1.33 elementary stream (ES) (system)**: A generic term for one of the coded video, coded audio or other coded bit streams in PES packets. One elementary stream is carried in a sequence of PES packets with one and only one stream\_id.

**2.1.34 elementary stream clock reference (ESCR) (system)**:A time stamp in the PES stream from which decoders of PES streams may derive timing.

**2.1.35 encoder**: An embodiment of an encoding process.

**2.1.36 encoding (process)**: A process, not specified in this Recommendation | International Standard, that reads a stream of input pictures or audio samples and produces a coded bit stream conforming to this Recommendation.

**2.1.37 entropy coding**: Variable length lossless coding of the digital representation of a signal to reduce redundancy.

**2.1.38 EVC video stream**: A raw bitstream as specified in ISO/IEC 23094-1 Annex B.

**2.1.39 EVC access unit**: An access unit as defined in ISO/IEC 23094-1 with the constraints specified in 2.24.1.

**2.1.40 EVC slice**: A slice as specified in ISO/IEC 23094-1.

**2.1.41 EVC video sequence (system)**: coded video sequence as defined in 23094-1.

**2.1.42 EVC still picture**: an EVC access unit containing an IDR picture, preceded by APS, SPS, PPS and other NAL units that carry sufficient information to correctly decode that picture, and also optionally preceded by another EVC still picture or an end of sequence NAL unit terminating a preceding coded video sequence when the current picture is not the first access unit in the video bitstream.

**2.1.43 EVC video sub-bitstream**: A subset of the NAL units of an EVC video stream in their original order.

**2.1.44 EVC temporal video sub-bitstream**: An EVC video sub-bitstream that contains all VCL NAL units and associated non-VCL NAL units of the temporal sub-layer, as specified in ISO/IEC 23094-1, associated to TemporalId equal to 0 and which may additionally contain all VCL NAL units and associated non-VCL NAL units of all temporal sub-layers associated to a contiguous range of TemporalId.

**2.1.45 event**: An event is defined as a collection of elementary streams with a common time base, an associated start time, and an associated end time.

**2.1.46 fast forward playback (video)**: The process of displaying a sequence, or parts of a sequence, of pictures in display-order faster than real-time.

**2.1.47 forbidden**: The term "forbidden", when used in the clauses of this Recommendation | International Standard defining the coded bit stream, indicates that the value specified shall never be used.

**2.1.48 green access unit** – An access unit that contains dynamic metadata in a message format as defined in 6.2.1 of ISO/IEC 23001-11.

**2.1.49 HEVC 24-hour picture (system)**: An HEVC access unit with a presentation time that is more than 24 hours in the future. For the purpose of this definition, HEVC access unit n has a presentation time that is more than 24 hours in the future if the difference between the initial arrival time tai(n) and the DPB output time to,dpb(n) is more than 24 hours.

**2.1.50 HEVC access unit**: An access unit as defined in Annex F of Rec. ITU-T H.265 | ISO/IEC 23008‑2 with the constraints specified in 2.17.1.

**2.1.51 HEVC base layer**: HEVC layer with nuh\_layer\_id equal to 0.

**2.1.52 HEVC base sub-partition**: HEVC video sub-bitstream that is also a conforming bitstream as specified in Rec. ITU-T H.265 | ISO/IEC 23008-2, which contains all VCL NAL units and the associated non-VCL NAL units of an HEVC base layer up to a target highest TemporalId identified by a target HEVC operation point.

**2.1.53 HEVC complete temporal representation**: A sub-layer representation as defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2 that contains all temporal sub-layers up to the temporal sub-layer with TemporalId equal to sps\_max\_sub\_layers\_minus1+1 as included in the active sequence parameter set, as specified in Rec. ITU-T H.265 | ISO/IEC 23008‑2.

**2.1.54 HEVC dependent slice segment**: An HEVC slice segment with the syntax element dependent\_slice\_segment\_flag in the slice header set to a value equal to 1, as defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2.

**2.1.55 HEVC enhancement sub-partition**: One HEVC layer with a particular value of nuh\_layer\_id greater than 0 in the NAL unit header syntax element or an HEVC temporal video sub-bitstream or HEVC temporal video subset thereof, of which the HEVC layer aggregation with an HEVC base sub-partition and zero or more other HEVC sub-partitions, according to HEVC layer list, results in a valid HEVC layered video stream.

**2.1.56 HEVC highest temporal sub-layer representation**: The sub-layer representation of the temporal sub-layer with the highest value of TemporalId, as defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2, in the associated HEVC temporal video sub-bitstream or HEVC temporal video subset.

**2.1.57 HEVC independent slice segment**: An HEVC slice segment with the syntax element dependent\_slice\_segment\_flag in the slice header set to a value 0 or inferred to be equal to 0, as defined in Rec. ITU‑T H.265 | ISO/IEC 23008‑2.

**2.1.58 HEVC layer**: HEVC video sub-bitstream that contains all VCL NAL units with a particular value of nuh\_layer\_id in the NAL unit header syntax element and associated non-VCL NAL units, as defined in Annex F of Rec. ITU-T H.265 | ISO/IEC 23008-2.

**2.1.59 HEVC layer aggregation**: Successive HEVC layer component aggregation of all HEVC layer components in an HEVC video sequence.

**2.1.60 HEVC layer component**: VCL NAL units and the associated non-VCL NAL units of an HEVC access unit which belong to an HEVC sub-partition.

**2.1.61 HEVC layer component aggregation**: Concatenation of all HEVC layer components with the same output time from all HEVC sub-partitions indicated in an HEVC layer list in the order indicated by the HEVC layer list, resulting in a valid HEVC access unit as defined in Annex F of Rec. ITU-T H.265 | ISO/IEC 23008-2.

**2.1.62 HEVC layer list**: Ordered list of HEVC sub-partitions for a target HEVC operation point of which the HEVC layer aggregation results in a valid HEVC layered video stream.

NOTE – An HEVC layer list is signalled for each target HEVC operation point using the HEVC operation point descriptor.

**2.1.63 HEVC layered video stream**: HEVC video stream that contains all VCL NAL units and associated non-VCL NAL units conforming to one or more profiles defined in Annex G or Annex H of Rec. ITU-T H.265 | ISO/IEC 23008‑2.

**2.1.64 HEVC operation point**: Operation point based on a target highest TemporalId, and a target layer identifier list as specified in Rec. ITU-T H.265 | ISO/IEC 23008‑2.

NOTE – Rec. ITU-T H.265 | ISO/IEC 23008‑2 specifies the sub-bitstream extraction process for an operation point according to which the operation point is a conforming bitstream. An operation point is associated with an HEVC layered video stream or HEVC base layer.

**2.1.65 HEVC slice**: An HEVC independent slice segment and zero or more subsequent HEVC dependent slice segments preceding the next HEVC independent slice segment (if any) within the same HEVC access unit.

**2.1.66 HEVC slice segment**: A byte\_stream\_nal\_unit with nal\_unit\_type in the range of 0 to 9 and 16 to 23, as defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2.

**2.1.67 HEVC still picture (system)**: An HEVC still picture consists of an HEVC access unit containing an IDR picture preceded by VPS, SPS and PPS NAL units, as defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2, that carry sufficient information to correctly decode this IDR picture. Preceding an HEVC still picture, there shall be another HEVC still picture or an end of sequence NAL unit terminating a preceding coded video sequence, as defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2.

**2.1.68 HEVC sub-partition**: Either an HEVC base sub-partition or an HEVC enhancement sub-partition.

NOTE – An HEVC sub-partition can either be an HEVC temporal video sub-bitstream if it includes VCL NAL units with the minimum value of TemporalId (i.e., including TemporalId equal to 0), or it can be an HEVC temporal video subset, if it complements an HEVC base sub-partition or HEVC enhancement sub-partition with the same target layer identifier.

**2.1.69 HEVC temporal enhancement sub-partition**: An HEVC temporal video subset of the same HEVC layer as another HEVC enhancement sub-partition of the same HEVC video stream which contains one or more complementary temporal sub-layers, as specified in Rec. ITU-T H.265 | ISO/IEC 23008-2.

**2.1.70 HEVC temporal video sub-bitstream**: An HEVC video sub-bitstream that contains all VCL NAL units and associated non-VCL NAL units of the temporal sub-layer of the same layer, as specified in Rec. ITU-T H.265 | ISO/IEC 23008‑2, associated with TemporalId equal to 0 and which may additionally contain all VCL NAL units and associated non-VCL NAL units of all temporal sub-layers of the same layer associated with a contiguous range of TemporalId from 1 to a value equal to or smaller than sps\_max\_sub\_layers\_minus1 included in the active sequence parameter set, as specified in Rec. ITU‑T H.265 | ISO/IEC 23008‑2.

**2.1.71 HEVC temporal video subset**: An HEVC video sub-bitstream that contains all VCL NAL units and the associated non-VCL NAL units of one or more temporal sub-layers of the same layer, as specified in Rec. ITU-T H.265 | ISO/IEC 23008‑2, with each temporal sub-layer not being present in the corresponding HEVC temporal video sub-bitstream and TemporalId associated with each temporal sub-layer forming a contiguous range of values that is equal to or smaller than sps\_max\_sub\_layers\_minus1 included in the active sequence parameter set, as specified in Rec. ITU‑T H.265 | ISO/IEC 23008‑2.

NOTE – According to the constraints for the transport of HEVC specified in 2.17.1, each temporal sub-layer of an *HEVC video stream* is present either in the HEVC temporal video sub-bitstream or in exactly one HEVC temporal video subset which is carried in a set of elementary streams that are associated by hierarchy descriptors or HEVC hierarchy extension descriptors. This prevents multiple inclusions of the same temporal sub-layer and allows aggregation of the HEVC temporal video sub-bitstreamwith associatedHEVC temporal video subsetsaccording to the hierarchy descriptors,as specified in 2.17.3 and according to the hierarchy descriptors or HEVC hierarchy extension descriptors, as specified in 2.17.4.

**2.1.72 HEVC tile of slices**: One or more consecutive HEVC slices which form the coded representation of a tile, as defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2.

**2.1.73 HEVC tile substream**: Substream of a Rec. ITU-T H.265 | ISO/IEC 23008 2 video stream that contains a Motion Constrained Tile Set, parameter sets, slice headers or a combination thereof.

**2.1.74 HEVC video sequence (system)**: A coded video sequence as defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2.

**2.1.75 HEVC video stream**: Byte stream as specified in Rec. ITU-T H. 265 | ISO/IEC 23008‑2 Annex B.

NOTE – This term represents either a byte stream as specified in Annex B of the first version of Rec. ITU-T H.265 | ISO/IEC 23008‑2 or an HEVC layered video sub-bitstream.

**2.1.76 HEVC video sub-bitstream**: A subset of the NAL units of an HEVC video stream in their original order.

**2.1.77 JPEG 2000 (J2K) video access unit**: The JPEG 2000 codestream or codestreams comprising a decodable and randomly accessible (portion of) image, preceded by all the parameters required to decode the access unit and display the decoded data.

**2.1.78 J2K block**: The JPEG 2000 codestream or codestreams corresponding to a rectangular portion of a video frame, as detailed in S.3.

NOTE – Usage of J2K blocks requires J2K block mode (defined in 2.1.79)

**2.1.79 J2K block mode**: Optional mode defined in S.3, dividing each frame of a J2K video stream in a certain amount of rectangular blocks, each encoded as an independent J2K block (defined in 2.1.78).

**2.1.80 J2K still picture (system)**: J2K video access unit as defined in 2.1.87 with constraints as specified in S.2.

**2.1.81 J2K stripe**: The JPEG 2000 codestream or codestreams comprising a decodable horizontally divided portion of an image, as detailed in S.4.

NOTE – Usage of J2K stripes requires J2K stripe mode (defined in 2.1.82) to be enabled in the J2K video descriptor. Such usage enables transport of a J2K video stream with a low end-to-end latency.

**2.1.82 J2K stripe mode**: Optional mode defined in S.4, dividing the (portion of) image transported in a J2K video access unit in a succession of horizontal stripes, each encoded as an independent J2K stripe (defined in 2.1.81).

**2.1.83 J2K video elementary stream**: Video elementary stream consisting of a succession of J2K video access units.

**2.1.84 J2K video sequence**: J2K video elementary stream where all the access units have the same profile/level, J2K video access unit coding parameters and video parameters.

**2.1.85 JPEG XS elementary stream header (jxes header)**: All parameters required to decode a JPEG XS video access unit and display the decoded data.

**2.1.86 JPEG XS still picture (system)**: JPEG XS video access unit as defined in 2.1.87 with constraints as specified in W.2.

**2.1.87 JPEG XS video access unit**: The JPEG XS codestream or multiple JPEG XS codestreams, as defined in ISO/IEC 21122-1, comprising a decodable and randomly accessible image, preceded by a JPEG XS elementary stream header.

**2.1.88 JPEG XS video elementary stream**: Video elementary stream consisting of a succession of JPEG XS video access units.

**2.1.89 JPEG XS video sequence**: 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 XSvideo access unit coding parameters, and video parameters.

**2.1.90 layer (video and systems)**: One of the levels in the data hierarchy of the video and system specifications defined in Parts 1 and 2 of this Recommendation | International Standard.

**2.1.90*bi*s LCEVC access unit (system)**: An access unit as defined for byte streams in ISO/IEC 23094-2 with the constraints specified in clause 2.25.1.

**2.1.90*ter* LCEVC video sequence (system)**: Coded video sequence as defined in ISO/IEC 23094-2.

**2.1.91 MCTS**: Motion Constrained Tile Set according to Rec. ITU-T H.265 | ISO/IEC 23008 2.

**2.1.92 metadata**: Information to describe audiovisual content and data essence in a format defined by ISO or any other authority.

**2.1.93 metadata access unit**: A global structure within metadata that defines the fraction of metadata that is intended to be decoded at a specific instant in time. The internal structure of a metadata Access Unit is defined by the format of the metadata.

**2.1.94 metadata application format**: Identifies the format of the application that uses the metadata; signals application specific information for transport of metadata.

**2.1.95 metadata decoder configuration information**: Data needed by a receiver to decode a specific metadata service. Depending on the format of the metadata, decoder configuration information may or may not be needed.

**2.1.96 metadata format**: Identifies the coding format of metadata.

**2.1.97 metadata service**: A coherent set of metadata of the same format delivered to a receiver for a specific purpose.

**2.1.98 metadata service id**: Identifier of a specific metadata service; used for some transport methods of the metadata.

**2.1.99 metadata stream**: The concatenation or collection of metadata Access Units from one or more metadata services.

**2.1.100 (multiplexed) stream (system)**: A bit stream composed of 0 or more elementary streams combined in a manner that conforms to this Recommendation | International Standard.

**2.1.101 MVC base view sub-bitstream**: The MVC base view sub-bitstream is defined to contain the AVC video sub‑bitstream of MVC conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496‑10 and one additional MVC video sub-bitstream associated with an MVC view\_id subset including the view order index that immediately follows the view order index associated with the base view.

NOTE – The MVC base view sub-bitstream is also an AVC video stream where no re-assembly is required before decoding.

**2.1.102 MVC operation point**: An MVC operation point is identified by a temporal\_id value representing a target temporal level and a set of view\_id values representing the target output views. One MVC operation point is associated with an AVC video stream which conforms to one or more profiles defined in Annex H of Rec. ITU‑T H.264 | ISO/IEC 14496-10. The AVC video stream associated with an MVC operation point is re‑assembled from a set consisting of one or more of the following items: AVC video sub-bitstream of MVC, MVC base view sub‑bitstream, MVC video sub-bitstreams.

**2.1.103 MVC slice (system)**:A byte\_stream\_nal\_unit with nal\_unit\_type syntax element equal to 20 of an AVC video stream which conforms to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10.

NOTE – As specified in Rec. ITU-T H.264 | ISO/IEC 14496-10, the value of svc\_extension\_flag is set equal to 0 for coded video sequences conforming to one or more profiles specified in Annex H. MVC slices should not include NAL units for which nal\_unit\_type is equal to 20 with svc\_extension\_flag equal to 1.

**2.1.104 MVC video sub-bitstream**: The MVC video sub-bitstream is defined to be all VCL NAL units with nal\_unit\_type equal to 20 associated with the same multiview video coding (MVC) view\_id subset of an advanced video coding (AVC) video stream and associated non-VCL NAL units which conform to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496‑10.

NOTE – In contrast to a sub-bitstream as specified in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10, an MVC video sub‑bitstream according to this Specification is not necessarily a decodable MVC video sub-bitstream. The one exception is when an MVC video sub-bitstream is also an MVC base view sub-bitstream. Re-assembling MVC video sub-bitstreams in an increasing order of view order index, starting from the lowest value of view order index up to any value of view order index, results in a decodable AVC video stream.

**2.1.105 MVC view\_id subset**: A set of one or more view\_id values, as defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10 in the NAL unit header syntax element, associated with one set of consecutive view order index values.

NOTE – An MVC video sub-bitstream or MVC base view sub-bitstream based on a specific MVC view\_id subset may not include view components for all view\_id values included in that MVC view\_id subset. One or more view order index values may be skipped if the view associated with a missing view order index value is not required for decoding the transmitted views.

**2.1.106 MVCD base view sub-bitstream**: The MVCD base view sub-bitstream is defined to contain the AVC video sub-bitstream of MVCD conforming to one or more profiles defined in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496‑10 and one additional MVCD video sub-bitstream associated with an MVCD view\_id subset including the view order index that immediately follows the view order index associated with the base view.

NOTE – The MVCD base view sub-bitstream is also an AVC video stream where no re-assembly is required before decoding.

**2.1.107 MVCD slice (system)**:A byte\_stream\_nal\_unit with nal\_unit\_type syntax element equal to 21 of an AVC video stream which conforms to one or more profiles defined in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496-10.

**2.1.108 MVCD video sub-bitstream**: The MVCD video sub-bitstream is defined to be all VCL NAL units with nal\_unit\_type equal to 21 associated with the same MVCD view\_id subset of an AVC video stream and associated non-VCL NAL units which conform to one or more profiles defined in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496‑10.

NOTE – In contrast to a sub-bitstream as specified in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496-10, an MVCD video sub‑bitstream according to this Specification is not necessarily a decodable MVCD video sub-bitstream. The one exception is when an MVCD video sub-bitstream is also an MVCD base view sub-bitstream. Re-assembling MVCD video sub-bitstreams in an increasing order of view order index, starting from the lowest value of view order index up to any value of view order index, results in a decodable AVC video stream.

**2.1.109 MVC view-component subset**: The VCL NAL units of an AVC access unit associated with the same MVC view\_id subset and associated non-VCL NAL units.

NOTE – Re-assembling MVC view-component subsets ordered according to the view order index, starting from the minimum view order index up to the highest view order index present in the access unit, while reordering the non-VCL NAL units conforming to the order of NAL units within an access unit, as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10, results in an AVC access unit.

**2.1.110 MVCD view-component subset**: The VCL NAL units of an AVC access unit associated with the same MVCD view\_id subset and associated non-VCL NAL units.

NOTE – Re-assembling MVCD view-component subsets ordered according to the view order index, starting from the minimum view order index up to the highest view order index present in the access unit, while reordering the non-VCL NAL units conforming to the order of NAL units within an access unit, as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10, results in an AVC access unit.

**2.1.111 MVCD view\_id subset**: A set of one or more view\_id values, as defined in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496-10 in the NAL unit header syntax element, associated with one set of consecutive view order index values.

NOTE – An MVCD video sub-bitstream or MVCD base view sub-bitstream based on a specific MVCD view\_id subset may not include view components for all view\_id values included in that MVCD view\_id subset. One or more view order index values may be skipped if the view associated with a missing view order index value is not required for decoding the transmitted views.

**2.1.112 pack (system)**: A pack consists of a pack header followed by zero or more packets. It is a layer in the system coding syntax described in 2.5.3.3.

**2.1.113 packet data (system)**: Contiguous bytes of data from an elementary stream present in a packet.

**2.1.114 packet identifier (PID) (system)**: A unique integer value used to identify elementary streams of a program in a single or multi-program transport stream as described in 2.4.3.

**2.1.115 padding (audio)**: A method to adjust the average length of an audio frame in time to the duration of the corresponding PCM samples, by conditionally adding a slot to the audio frame.

**2.1.116 payload**: Payload refers to the bytes which follow the header bytes in a packet. For example, the payload of some transport stream packets includes a PES\_packet\_header and its PES\_packet\_data\_bytes, or pointer\_field and PSI sections, or private data; but a PES\_packet\_payload consists of only PES\_packet\_data\_bytes. The transport stream packet header and adaptation fields are not payload.

**2.1.117 PES (system)**:An abbreviation for a Packetized Elementary Stream.

**2.1.118 PES packet (system)**: The data structure used to carry elementary stream data. A PES packet consists of a PES packet header followed by a number of contiguous bytes from an elementary data stream. It is a layer in the system coding syntax described in 2.4.3.6.

**2.1.119 PES packet header (system)**: The leading fields in a PES packet up to and not including the PES\_packet\_data\_byte fields, where the stream is not a padding stream. In the case of a padding stream the PES packet header is similarly defined as the leading fields in a PES packet up to and not including padding\_byte fields.

**2.1.120 packetized elementary stream (PES) (system)**:A PES system consists of PES packets, all of whose payloads consist of data from a single elementary stream, and all of which have the same stream\_id. Specific semantic constraints apply. Refer to Intro. 4.

**2.1.121 presentation time-stamp (PTS) (system)**: A field that may be present in a PES packet header that indicates the time that a presentation unit is presented in the system target decoder.

**2.1.122 presentation unit (PU) (system)**: A decoded audio access unit or a decoded picture.

**2.1.123 program (system)**: A program is a collection of program elements. Program elements may be elementary streams. Program elements need not have any defined time base; those that do, have a common time base and are intended for synchronized presentation.

**2.1.124 program clock reference (PCR) (system)**: A time stamp in the transport stream from which decoder timing is derived.

**2.1.125 program element (system)**: A generic term for one of the elementary streams or other data streams that may be included in a program.

**2.1.126 program-specific information (PSI) (system)**: PSI consists of normative data which is necessary for the demultiplexing of transport streams and the successful regeneration of programs and is described in 2.4.4.1. An example of privately defined PSI data is the non-mandatory network information table.

**2.1.127 quality access unit**: An access unit that contains dynamic quality metadata as defined in ISO/IEC 23001-10.

**2.1.128 random access**: The process of beginning to read and decode the coded bit stream at an arbitrary point.

**2.1.129 reserved**: The term "reserved", when used in the clauses defining the coded bit stream, indicates that the value may be used in the future for ISO defined extensions. Unless otherwise specified within this Recommendation | International Standard, all reserved bits shall be set to '1'.

**2.1.130 sample variant**: An assembled media sample replacing an original sample as defined in ISO/IEC 23001-12:2018.

NOTE – In ISO/IEC 23001-12:2018, the term “sample” is used commonly for samples as defined in ISO Base Media File Format (ISO/IEC 14496-12) and for access units as defined in 2.1.1.

**2.1.131 scrambling (system)**: The alteration of the characteristics of a video, audio or coded data stream in order to prevent unauthorized reception of the information in a clear form. This alteration is a specified process under the control of a conditional access system.

**2.1.132 service-compatible**: This is defined as 'simulcast' of two stereoscopic views which do not include scalable or temporal coding. The two views are independently compressed using MPEG-2 video or AVC or both and can be decoded independently.

**2.1.133 source stream**: A single non-multiplexed stream of samples before compression coding.

**2.1.134 splicing (system)**: The concatenation, performed on the system level, of two different elementary streams. The resulting system stream conforms totally to this Recommendation | International Standard. The splice may result in discontinuities in timebase, continuity counter, PSI, and decoding.

**2.1.135 start codes (system)**: 32-bit codes embedded in the coded bit stream. They are used for several purposes including identifying some of the layers in the coding syntax. Start codes consist of a 24-bit prefix (0x000001) and an 8‑bit stream\_id as shown in Table 2-22.

**2.1.136 STD input buffer (system)**: A first-in first-out buffer at the input of a system target decoder for storage of compressed data from elementary streams before decoding.

**2.1.137 still picture**: A still picture consists of a video sequence, coded as defined in Rec. ITU-T H.262 | ISO/IEC 13818-2, ISO/IEC 11172-2 or ISO/IEC 14496-2, that contains exactly one coded picture which is intra-coded. This picture has an associated PTS and in case of coding according to ISO/IEC 11172-2, Rec. ITU-T H.262 | ISO/IEC 13818-2 or ISO/IEC 14496-2, the presentation time of succeeding pictures, if any, is later than that of the still picture by at least two picture periods.

**2.1.138 SVC dependency representation**: The VCL NAL units of an AVC access unit associated with the same value of dependency\_id which is provided as part of the NAL unit header or the associated prefix NAL unit header, and the associated non-VCL NAL units. Re-assembling SVC dependency representations in a consecutive order of dependency\_id starting from the lowest value of dependency\_id present in the access unit up to any value of dependency\_id present in the access unit, while reordering the non-VCL NAL units conforming to the order of NAL units within an access unit as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10, results in an AVC access unit.

**2.1.139 SVC slice (system)**:A byte\_stream\_nal\_unit as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10 with nal\_unit\_type equal to 20 of an AVC video stream which conforms to one or more profiles defined in Annex G of Rec. ITU-T H.264 | ISO/IEC 14496-10.

NOTE – As specified in Rec. ITU-T H.264 | ISO/IEC 14496-10, the value of svc\_extension\_flag is set equal to 1 for coded video sequences conforming to one or more profiles specified in Annex G. SVC slices should not include NAL units for which nal\_unit\_type is equal to 20 with svc\_extension\_flag equal to 0.

**2.1.140 SVC video sub-bitstream**: The video sub-bitstream that contains VCL NAL units with nal\_unit\_type equal to 20 with the same NAL unit header syntax element dependency\_id not equal to 0.

**2.1.141 system header (system)**: The system header is a data structure defined in 2.5.3.5 that carries information summarizing the system characteristics of Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program stream.

**2.1.142 system clock reference (SCR) (system)**: A time stamp in the program stream from which decoder timing is derived.

**2.1.143 system target decoder (STD) (system)**: A hypothetical reference model of a decoding process used to define the semantics of a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 multiplexed bit stream.

**2.1.144 time-stamp (system)**: A term that indicates the time of a specific action such as the arrival of a byte or the presentation of a Presentation Unit.

**2.1.145 transport stream packet header (system)**: The leading fields in a transport stream packet, up to and including the continuity\_counter field.

**2.1.146 variable bitrate**: An attribute of transport streams or program streams wherein the rate of arrival of bytes at the input to a decoder varies with time.

**2.1.147 video sub-bitstream**: A video sub-bitstream is defined to be all VCL NAL units associated with the same value of dependency\_id of an AVC video stream which conforms to one or more profiles defined in Annex G of Rec. ITU-T H.264 | ISO/IEC 14496-10 and all associated non-VCL NAL units in decoding order as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10. Re-assembling video sub-bitstreams in a consecutive order of dependency\_id, starting from the dependency\_id equal to 0 up to any value of dependency\_id, results in an AVC video stream. A video sub-bitstream shall have the AVC byte stream format as defined in Annex B of Rec. ITU-T H.264 | ISO/IEC 14496-10.

**2.1.148 view order index**: An index that indicates the decoding order of MVC view components in an AVC access unit as defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10 or MVCD view components in an AVC access unit as defined in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496-10. The association of view order index values to the NAL unit header syntax element view\_id is indicated for an AVC video sequence in the sequence parameter set MVC extension as defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10 or in the sequence parameter set MVCD extension as defined in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496-10.

**2.1.149 VVC video stream**: A byte stream as specified in Rec. ITU-T H.266 | ISO/IEC 23090-3 Annex B.

**2.1.150 VVC access unit**: An access unit as defined in Rec. ITU-T H.266 | ISO/IEC 23090-3 with the constraints specified in 2.23.1.

**2.1.151 VVC 24-hour picture (system)**: A VVC access unit with a presentation time that is more than 24 hours in the future. For the purpose of this definition, VVC access unit n has a presentation time that is more than 24 hours in the future if the difference between the initial arrival time tai(n) and the DPB output time to,dpb(n) is more than 24 hours.

**2.1.152 VVC slice**: A slice as specified in Rec. ITU-T H.266 | ISO/IEC 23090-3.

**2.1.153 VVC subpicture**: A subpicture as specified in Rec. ITU-T H.266 | ISO/IEC 23090-3.

**2.1.154 VVC tile of slices**: One or more consecutive VVC slices which form the coded representation of a tile as defined in Rec. ITU-T H.266 | ISO/IEC 23090-3.

**2.1.155 VVC still picture (system)**: A VVC still picture consists of a VVC access unit containing an IDR picture preceded by VPS, SPS and PPS NAL units, as defined in Rec. ITU-T H.266 | ISO/IEC 23090-3, that carry sufficient information to correctly decode this IDR picture. Preceding a VVC still picture, there shall be another VVC still picture or an End of Sequence NAL unit terminating a preceding coded video sequence as defined in Rec. ITU-T H.266 | ISO/IEC 23090-3.

**2.1.156 VVC video sequence (system)**: coded video sequence as defined in Rec. ITU-T H.266 | ISO/IEC 23090-3.

**2.1.157 VVC video sub-bitstream**: A subset of the NAL units of a VVC video stream in their original order.

**2.1.158 VVC temporal video sub-bitstream**: A VVC video sub-bitstream that contains all VCL NAL units and associated non-VCL NAL units of the temporal sublayer, as specified in Rec. ITU-T H.266 | ISO/IEC 23090-3, associated to TemporalId equal to 0 and which may additionally contain all VCL NAL units and associated non-VCL NAL units of all temporal sublayers associated to a contiguous range of TemporalId from 1 to a value equal to or smaller than sps\_max\_sublayers\_minus1 included in the referred sequence parameter set, as specified in Rec. ITU-T H.266 | ISO/IEC 23090-3.

**2.1.159 VVC temporal video subset**: A VVC video sub-bitstream that contains all VCL NAL units and the associated non-VCL NAL units of one or more temporal sublayers, as specified in Rec. ITU-T H.266 | ISO/IEC 23090-3, with each temporal sublayer not being present in the corresponding VVC temporal video sub-bitstream and TemporalId associated with each temporal sublayer forming a contiguous range of values.

NOTE – According to the constraints for the transport of VVC specified in 2.23.1, each temporal sublayer of a VVC video stream is present either in the VVC temporal video sub-bitstream or in exactly one VVC temporal video subset which are carried in a set of elementary streams that are associated by hierarchy descriptors. This prevents multiple inclusion of the same temporal sublayer and allows aggregation of the VVC temporal video sub-bitstream with associated VVC temporal video subsets according to the hierarchy descriptors as specified in 2.23.3.

**2.1.160 VVC highest temporal sublayer representation**: The sublayer representation of the temporal sublayer with the highest value of TemporalId, as defined in Rec. ITU-T H.266 | ISO/IEC 23090-3, in the associated VVC temporal video sub-bitstream or VVC temporal video subset.

**2.1.161 VVC complete temporal representation**: A sublayer representation as defined in Rec. ITU-T H.266 | ISO/IEC 23090-3 that contains all temporal sublayers up to the temporal sublayer with TemporalId equal to sps\_max\_sublayers\_minus1 as included in the referred sequence parameter set, as specified in Rec. ITU-T H.266 | ISO/IEC 23090-3.

## 2.2 Symbols and abbreviations

The mathematical operators used to describe this Recommendation | International Standard are similar to those used in the C-programming language. However, integer division with truncation and rounding are specifically defined. The bitwise operators are defined assuming two's-complement representation of integers. Numbering and counting loops generally begin from 0.

### 2.2.1 Arithmetic operators

+ Addition

– Subtraction (as a binary operator) or negation (as a unary operator)

++ Increment

– – Decrement

\* or × Multiplication

^ Power

/ Integer division with truncation of the result toward 0. For example, 7/4 and –7/–4 are truncated to 1 and –7/4 and 7/–4 are truncated to –1.

// Integer division with rounding to the nearest integer. Half-integer values are rounded away from 0 unless otherwise specified. For example 3//2 is rounded to 2, and –3//2 is rounded to –2.

DIV Integer division with truncation of the result towards – ∞.

% Modulus operator. Defined only for positive numbers.

Sign( ) Sign(x) = 1 x > 0

0 x = = 0

–1 x < 0

NINT( ) Nearest integer operator. Returns the nearest integer value to the real-valued argument. Half-integer values are rounded away from 0.

sin Sine

cos Cosine

exp Exponential

√ Square root

log10 Logarithm to base ten

loge Logarithm to base e

### 2.2.2 Assignment

= Assignment operator

### 2.2.3 Bitwise operators

& AND

| OR

>> Shift right with sign extension

<< Shift left with 0 fill

### 2.2.4 Constants

π 3.14159265359

e 2.71828182845

### 2.2.5 Logical operators

|| Logical OR

&& Logical AND

! Logical NOT

### 2.2.6 Mnemonics

The following mnemonics are defined to describe the different data types used in the coded bit-stream.

bslbf Bit string, left bit first, where "left" is the order in which bit strings are written in this Recommendation | International Standard. Bit strings are written as a string of 1s and 0s within single quote marks, e.g., '1000 0001'. Blanks within a bit string are for ease of reading and have no significance.

NOTE – In some tables of this specification, two bit strings are used to define a range of values. In other tables, bit strings are used to specify integer values. In these cases, each bit string is equivalent to a binary number of which the least significant bit (unity position) equals the rightmost bit of the bit string and the significance of bits increases from right to left.

ch Channel

gr Granule of 3 \* 32 sub-band samples in audio Layer II, 18 \* 32 sub-band samples in audio Layer III.

main\_data The main\_data portion of the bit stream contains the scale factors, Huffman encoded data, and ancillary information.

main\_data\_beg This gives the location in the bit stream of the beginning of the main\_data for the frame. The location is equal to the ending location of the previous frame's main\_data plus 1 bit. It is calculated from the main\_data\_end value of the previous frame.

part2\_length This value contains the number of main\_data bits used for scale factors.

rpchof Remainder polynomial coefficients, highest order first

sb Sub-band

scfsi Scalefactor selector information

switch\_point\_l Number of scalefactor band (long block scalefactor band) from which point on window switching is used

switch\_point\_s Number of scalefactor band (short block scalefactor band) from which point on window switching is used

tcimsbf Two's complement integer, msb (sign) bit first

uimsbf Unsigned integer, most significant bit first

vlclbf Variable length code, left bit first, where "left" refers to the order in which the variable length codes are written

window Number of actual time slot in case of block\_type = = 2, 0 ≤ window ≤ 2.

The byte order of multi-byte words is most significant byte first.

### 2.2.7 Range operator

.. Range operator. n .. m defines the inclusive range of numbers from n up to m

### 2.2.8 Relational operators

> Greater than

≥ Greater than or equal to

< Less than

≤ Less than or equal to

= = Equal to

!= Not equal to

max [,...,] The maximum value in the argument list

min [,...,] The minimum value in the argument list

## 2.3 Method of describing bit stream syntax

The bit streams retrieved by the decoder are described in 2.4.1 and 2.5.1. Each data item in the bit stream is in bold type. It is described by its name, its length in bits, and a mnemonic for its type and order of transmission.

The action caused by a decoded data element in a bit stream depends on the value of that data element and on data elements previously decoded. The decoding of the data elements and definition of the state variables used in their decoding are described in the clauses containing the semantic description of the syntax. The following constructs are used to express the conditions when data elements are present, and are in normal type.

Note this syntax uses the "C"-code convention that a variable or expression evaluating to a non-zero value is equivalent to a condition that is true:

|  |  |
| --- | --- |
| while ( condition ) {  **data\_element  . . .** } | If the condition is true, then the group of data elements occurs next in the data stream. This repeats until the condition is not true. |
| do {  **data\_element  . . .**  } while ( condition ) | The data element always occurs at least once. The data element is repeated until the condition is not true. |
| if ( condition ) {  **data\_element  . . .** } | If the condition is true, then the first group of data elements occurs next in the data stream. |
| else {  **data\_element**  **. . .** } | If the condition is not true, then the second group of data elements occurs next in the data stream. |
| for (i = 0; i < n; i++) {  **data\_element**  **. . .** } | The group of data elements occurs n times. Conditional constructs within the group of data elements may depend on the value of the loop control variable i, which is set to zero for the first occurrence, incremented to 1 for the second occurrence, and so forth. |

As noted, the group of data elements may contain nested conditional constructs. For compactness, the {} are omitted when only one data element follows:

|  |  |
| --- | --- |
| **data\_element []** | data\_element [] is an array of data. The number of data elements is indicated by the context. |
| **data\_element [n]** | data\_element [n] is the n+1th element of an array of data. |
| **data\_element [m][n]** | data\_element [m][n] is the m+1,n+1th element of a two-dimensional array of data. |
| **data\_element [l][m][n]** | data\_element [l][m][n] is the l+1,m+1,n+1th element of a three-dimensional array of data. |
| **data\_element [m..n]** | is the inclusive range of bits between bit m and bit n in the data\_element. |

While the syntax is expressed in procedural terms, it should not be assumed that either Figure 2-1 or Figure 2-2 implements a satisfactory decoding procedure. In particular, they define a correct and error-free input bitstream. Actual decoders must include a means to look for start codes and sync bytes (transport stream) in order to begin decoding correctly, and to identify errors, erasures or insertions while decoding. The methods to identify these situations, and the actions to be taken, are not standardized.

## 2.4 Transport stream bitstream requirements

### 2.4.1 Transport stream coding structure and parameters

The Rec. ITU-T H.222.0 | ISO/IEC 13818-1 transport stream coding layer allows one or more programs to be combined into a single stream. Data from each elementary stream are multiplexed together with information that allows synchronized presentation of the elementary streams within a program.

A transport stream consists of one or more programs. Audio and video elementary streams consist of access units.

Elementary stream data is carried in PES packets. A PES packet consists of a PES packet header followed by packet data. PES packets are inserted into transport stream packets. The first byte of each PES packet header is located at the first available payload location of a transport stream packet.

The PES packet header begins with a 32-bit start-code that also identifies the stream or stream type to which the packet data belongs. The PES packet header may contain decoding and presentation time stamps (DTS and PTS). The PES packet header also contains other optional fields. The PES packet data field contains a variable number of contiguous bytes from one elementary stream.

Transport stream packets begin with a 4-byte prefix, which contains a 13-bit packet ID (PID), defined in Table 2-2. The PID identifies, via the program-specific information (PSI) tables, the contents of the data contained in the transport stream packet. Transport stream packets of one PID value carry data of one and only one elementary stream.

The PSI tables are carried in the transport stream. There are six PSI tables:

• Program association table (PAT);

• Program map table (MPT);

• Conditional access table (CAT);

• Network information table (NIT);

• Transport stream description table (TSDT);

• IPMP control information table.

These tables contain the necessary and sufficient information to demultiplex and present programs. The program map table in Table 2-33 specifies, among other information, which PIDs, and therefore which elementary streams, are associated to form each program. This table also indicates the PID of the transport stream packets which carry the PCR for each program. The conditional access table shall be present if scrambling is employed. The network information table is optional and its contents are not specified by this Recommendation | International Standard. The IPMP control information table shall be present if IPMP as described in ISO/IEC 13818-11 is used by any of the components in the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream.

Transport stream packets may be null packets. Null packets are intended for padding of transport streams. They may be inserted or deleted by re-multiplexing processes and, therefore, the delivery of the payload of null packets to the decoder cannot be assumed.

This Recommendation | International Standard does not specify the coded data which may be used as part of conditional access systems. This Specification does, however, provide mechanisms for program service providers to transport and identify this data for decoder processing, and to reference correctly data which are specified by this Specification. This type of support is provided both through transport stream packet structures and in the conditional access table (refer to Table 2-32).

### 2.4.2 Transport stream system target decoder

#### 2.4.2.1 General

The semantics of the transport stream specified in 2.4.3 and the constraints on these semantics specified in 2.7 require exact definitions of byte arrival and decoding events and the times at which these occur. The definitions needed are set out in this Recommendation | International Standard using a hypothetical decoder known as the transport stream system target decoder (T-STD). Informative Annex D contains further explanation of the T-STD.

The T-STD is a conceptual model used to define these terms precisely and to model the decoding process during the construction or verification of transport streams. The T-STD is defined only for this purpose. There are three types of decoders in the T-STD: video, audio, and systems. Figure 2-1 illustrates an example. Neither the architecture of the T‑STD nor the timing described precludes uninterrupted, synchronized play-back of transport streams from a variety of decoders with different architectures or timing schedules.



Figure 2-1 – Transport stream system target decoder notation

The following notation is used to describe the transport stream system target decoder and is partially illustrated in Figure 2-1 above.

i, i′, i″ are indices to bytes in the transport stream. The first byte has index 0.

j is an index to access units in the elementary streams.

k, k′, k″ are indices to presentation units in the elementary streams.

n is an index to the elementary streams.

p is an index to transport stream packets in the transport stream.

t(i) indicates the time in seconds at which the i-th byte of the transport stream enters the system target decoder. The value t(0) is an arbitrary constant.

PCR(i) is the time encoded in the PCR field measured in units of the period of the 27-MHz system clock where i is the byte index of the final byte of the program\_clock\_reference\_base field.

An(j) is the j-th access unit in elementary stream n. An(j) is indexed in decoding order.

tdn(j) is the decoding time, measured in seconds, in the system target decoder of the j-th access unit in elementary stream n.

Pn(k) is the k-th presentation unit in elementary stream n. Pn(k) results from decoding An(j). Pn(k) is indexed in presentation order.

tpn(k) is the presentation time, measured in seconds, in the system target decoder of the k-th presentation unit in elementary stream n.

t is time measured in seconds.

Fn(t) is the fullness, measured in bytes, of the system target decoder input buffer for elementary stream n at time t.

Bn is the main buffer for elementary stream n. It is present only for audio elementary streams.

BSn is the size of buffer, Bn, measured in bytes.

Bsys is the main buffer in the system target decoder for system information for the program that is in the process of being decoded.

BSsys is the size of Bsys, measured in bytes.

MBn is the multiplexing buffer, for elementary stream n. It is present only for video elementary streams.

MBSn is the size of MBn, measured in bytes.

EBn is the elementary stream buffer for elementary stream n. It is present only for video elementary streams.

EBSn is the size of the elementary stream buffer EBn, measured in bytes.

TBsys is the transport buffer for system information for the program that is in the process of being decoded.

TBSsys is the size of TBsys, measured in bytes.

TBn is the transport buffer for elementary stream n.

TBSn is the size of TBn, measured in bytes.

Dsys is the decoder for system information in program stream n.

Dn is the decoder for elementary stream n.

On is the re-order buffer for video elementary stream n.

Rsys is the rate at which data are removed from Bsys.

Rxn is the rate at which data are removed from TBn.

Rbxn is the rate at which PES packet payload data are removed from MBn when the leak method is used. Defined only for video elementary streams.

Rbxn(j) is the rate at which PES packet payload data are removed from MBn when the vbv\_delay method is used. Defined only for video elementary streams.

Rxsys is the rate at which data are removed from TBsys.

Res is the video elementary stream rate coded in a sequence header.

#### 2.4.2.2 System clock frequency

Timing information referenced in the T-STD is carried by several data fields defined in this Specification. Refer to 2.4.3.4 and 2.4.3.6. In PCR fields this information is coded as the sampled value of a program's system clock. The PCR fields are carried in the adaptation field of the transport stream packets with a PID value equal to the PCR\_PID defined in the TS\_program\_map\_section of the program being decoded.

Practical decoders may reconstruct this clock from these values and their respective arrival times. The following are minimum constraints which apply to the program's system clock frequency as represented by the values of the PCR fields when they are received by a decoder.

The value of the system clock frequency is measured in Hz and shall meet the following constraints:

27 000 000 – 810 ≤ system\_clock\_frequency ≤ 27 000 000 + 810

rate of change of system\_clock\_frequency with time ≤ 75 × 10–3 Hz/s

NOTE – Sources of coded data should follow a tighter tolerance in order to facilitate compliant operation of consumer recorders and playback equipment.

A program's system\_clock\_frequency may be more accurate than required. Such improved accuracy may be transmitted to the decoder via the system clock descriptor described in 2.6.20.

Bit rates defined in this Specification are measured in terms of system\_clock\_frequency. For example, a bit rate of 27 000 000 bits per second in the T-STD would indicate that one byte of data is transferred every eight (8) cycles of the system clock.

The notation "system\_clock\_frequency" is used in several places in this Specification to refer to the frequency of a clock meeting these requirements. For notational convenience, equations in which PCR, PTS, or DTS appear, lead to values of time which are accurate to some integral multiple of (300 × 233/system\_clock\_frequency) seconds. This is due to the encoding of PCR timing information as 33 bits of 1/300 of the system clock frequency plus 9 bits for the remainder, and encoding as 33 bits of the system clock frequency divided by 300 for PTS and DTS.

#### 2.4.2.3 Input to the transport stream system target decoder

Input to the transport stream system target decoder (T-STD) is a transport stream. A transport stream may contain multiple programs with independent time bases. However, the T-STD decodes only one program at a time. In the T‑STD model all timing indications refer to the time base of that program.

Data from the transport stream enters the T-STD at a piecewise constant rate. The time t(i) at which the i-th byte enters the T-STD is defined by decoding the program clock reference (PCR) fields in the input stream, encoded in the transport stream packet adaptation field of the program to be decoded and by counting the bytes in the complete transport stream between successive PCRs of that program. The PCR field (see equation 2-1) is encoded in two parts: one, in units of the period of 1/300 times the system clock frequency, called program\_clock\_reference\_base (see equation 2-2), and one in units of the system clock frequency called program\_clock\_reference\_extension (see equation 2-3). The values encoded in these are computed by PCR\_base(i) (see equation 2-2) and PCR\_ext(i) (see equation 2-3) respectively. The value encoded in the PCR field indicates the time t(i), where i is the index of the byte containing the last bit of the program\_clock\_reference\_base field.

Specifically:

 (2-1)

where:

(2-2)

(2-3)

For all other bytes the input arrival time, t(i) shown in equation 2-4 below, is computed from PCR(i″) and the transport rate at which data arrive, where the transport rate is determined as the number of bytes in the transport stream between the bytes containing the last bit of two successive program\_clock\_reference\_base fields of the same program divided by the difference between the time values encoded in these same two PCR fields.

 (2-4)

where:

i is the index of any byte in the transport stream for i″ < i < i′.

i″ is the index of the byte containing the last bit of the most recent program\_clock\_reference\_base field applicable to the program being decoded.

PCR(i″) is the time encoded in the program clock reference base and extension fields in units of the system clock.

The transport rate for any byte i between byte i″ and byte i′ is given by:

 (2-5)

where:

i′ is the index of the byte containing the last bit of the immediately following program\_clock\_reference\_base field applicable to the program being decoded.

In the case of a timebase discontinuity, indicated by the discontinuity\_indicator in the transport packet adaptation field, the definition given in equation 2-4 and equation 2-5 for the time of arrival of bytes at the input to the T-STD is not applicable between the last PCR of the old timebase and the first PCR of the new timebase. In this case the time of arrival of these bytes is determined according to equation 2-4 with the modification that the transport rate used is that applicable between the last and next to last PCR of the old timebase.

A tolerance is specified for the PCR values. The PCR tolerance is defined as the maximum inaccuracy allowed in received PCRs. This inaccuracy may be due to imprecision in the PCR values or to PCR modification during re‑multiplexing. It does not include errors in packet arrival time due to network jitter or other causes. The PCR tolerance is ± 500 ns.

In the T-STD model, the inaccuracy will be reflected as an inaccuracy in the calculated transport rate using equation 2‑5.

Transport streams with multiple programs and variable rate

Transport streams may contain multiple programs which have independent time bases. Separate sets of PCRs, as indicated by the respective PCR\_PID values, are required for each such independent program, and therefore the PCRs cannot be co-located. The transport stream rate is piecewise constant for the program entering the T-STD. Therefore, if the transport stream rate is variable it can only vary at the PCRs of the program under consideration. Since the PCRs, and therefore the points in the transport Stream where the rate varies, are not co-located, the rate at which the transport stream enters the T-STD would have to differ depending on which program is entering the T-STD. Therefore, it is not possible to construct a consistent T-STD delivery schedule for an entire transport stream when that transport stream contains multiple programs with independent time bases and the rate of the transport stream is variable. It is straightforward, however, to construct constant bit rate transport streams with multiple variable rate programs.

#### 2.4.2.4 Buffering

Complete transport stream packets containing system information, for the program selected for decoding, enter the system transport buffer, TBsys, at the transport stream rate. These include transport stream packets whose PID values are 0, 1, 2 or 3, and all transport stream packets identified via the program association table (see Table 2-30) as having the program\_map\_PID value for the selected program. Network information table (NIT) data as specified by the NIT PID is not transferred to TBsys.

NOTE 1 – Size of IPMP control information table could be large, and the repetition rate of this table should be adjusted to meet the buffer requirement.

All bytes that enter the buffer TBn are removed at the rate Rxn specified below. Bytes which are part of the PES packet header or its contents are delivered to the main buffer Bn for audio elementary streams and system data, and to the multiplexing buffer MBn for video elementary streams. Other bytes are not, and may be used to control the system. Duplicate transport stream packets are not delivered to Bn, MBn, or Bsys.

The buffer TBn is emptied as follows:

– When there is no data in TBn, Rxn is equal to zero.

– Otherwise for video:



where:

Rmax[profile, level] is specified according to the profile and level which can be found in Table 8-13 of Rec. ITU‑T H.262 | ISO/IEC 13818-2. This table specifies the upper bound of the rate of each elementary video stream within a specific profile and level.

Rxn is equal to 1, 2 × Rmax for ISO/IEC 11172-2 constrained parameter video streams, where Rmax refers to the maximum bitrate for a constrained parameters bitstream in ISO/IEC 11172-2.

For ISO/IEC 13818-7 ADTS audio:

|  |  |
| --- | --- |
| Number of Channels | Rxn [bit/s] |
| 1 .. 2 | 2 000 000 |
| 3 .. 8 | 5 529 600 |
| 9 .. 12 | 8 294 400 |
| 13 .. 48 | 33 177 600 |

Channels: The number of full-bandwidth audio output channels plus the number of independently switched coupling channel elements within the same elementary audio stream. For example, in the typical case that there are no independently switched coupling channel elements, mono is 1 channel, stereo is 2 channels and 5.1 channel surround is five channels (the low-frequency effects (LFE) channel is not counted).

For other audio,



For systems data:



Rxn is measured with respect to the system clock frequency.

Complete transport stream packets containing system information, for the program selected for decoding, enter the system transport buffer, TBsys, at the transport stream rate. These include transport stream packets whose PID values are 0, 1, 2 and 3 (if present), and all transport stream packets identified via the program association table (PAT) (see Table 2-30) as having the program\_map\_PID value for the selected program. Network information table (NIT) data as specified by the NIT PID is not transferred to TBsys.

Bytes are removed from TBsys at the rate Rxsys and delivered to Bsys. Each byte is transferred instantaneously.

Duplicate transport stream packets are not delivered to Bsys.

Transport packets which do not enter any TBn or TBsys are discarded.

The transport buffer size is fixed at 512 bytes.

The elementary stream buffer sizes EBS1 through EBSn are defined for video as equal to the vbv\_buffer\_size as it is carried in the sequence header. Refer to the summary of constrained parameters in ISO/IEC 11172-2 and Table 8-14 of Rec. ITU-T H.262 | ISO/IEC 13818-2.

NOTE 2 – In the following equations, unit conversion should be implicitly performed as appropriate. Values expressed in bits are implicitly converted into values expressed in bytes by: *number\_of\_bytes* = (*number\_of\_bits* +7) / 8.

The multiplexing buffer size MBS1 through MBSn are defined for video as follows:

For Low and Main level:



where BSoh, PES packet overhead buffering is defined as:



and BSmux, additional multiplex buffering is defined as:



and where VBVmax[profile, level] is defined in Table 8-14 of Rec. ITU-T H.262 | ISO/IEC 13818-2 and Rmax[profile, level] is defined in Table 8-13 of Rec. ITU-T H.262 | ISO/IEC 13818-2, and vbv buffer size is carried in the sequence header described in 6.2.2 of Rec. ITU-T H.262 | ISO/IEC 13818-2.

For High 1440 and High level:



where BSoh is defined as:



and BSmux is defined as:



and where Rmax[profile, level] is defined in Table 8-13 of Rec. ITU-T H.262 | ISO/IEC 13818-2.

For Constrained Parameters ISO/IEC 11172-2 bitstreams:



where BSoh is defined as:



and BSmux is defined as:



and where Rmax and vbv\_max refer to the maximum bitrate and the maximum vbv\_buffer\_size for a Constrained Parameters bitstream in ISO/IEC 11172-2 respectively.

A portion BSmux = 4 ms × Rmax[profile, level] of the MBSn is allocated for buffering to allow multiplexing. The remainder is available for BSoh and may also be available for initial multiplexing.

NOTE 3 – Buffer occupancy by PES packet overhead is directly bounded in PES streams by the PES-STD which is defined in 2.5.2.4. It is possible, but not necessary, to utilize PES streams to construct transport streams.

Buffer BSn

The main buffer sizes BS1 through BSn are defined as follows.

Audio

For ISO/IEC 13818-7 ADTS audio:

|  |  |
| --- | --- |
| Number of Channels | BSn [bytes] |
| 1 .. 2 | 3 584 |
| 3 .. 8 | 8 976 |
| 9 .. 12 | 12 804 |
| 13 .. 48 | 51 216 |

Channels: The number of full-bandwidth audio output channels plus the number of independently switched coupling channel elements within the same elementary audio stream. For example, in the typical case that there are no independently switched coupling channel elements, mono is 1 channel, stereo is 2 channels and 5.1 channel surround is 5 channels (the LFE channel is not counted).

For other audio:



The size of the access unit decoding buffer BSdec, and the PES packet overhead buffer BSoh are constrained by:



A portion (736 bytes) of the 3584 byte buffer is allocated for buffering to allow multiplexing. The rest, 2848 bytes, are shared for access unit buffering BSdec, BSoh and additional multiplexing.

Systems

The main buffer Bsys for system data is of size BSsys = 1536 bytes.

Video

For video elementary streams, data is transferred from MBn to EBn using one of two methods: the leak method or the video buffering verifier (VBV) delay method.

Leak method

The leak method transfers data from MBn to EBn using a leak rate Rbx. The leak method is used whenever any of the following is true:

• the STD descriptor (refer to 2.6.32) for the elementary stream is not present in the transport stream;

• the STD descriptor is present and the leak\_valid flag has a value of '1';

• the STD descriptor is present, the leak\_valid has a value of '0', and the vbv\_delay fields coded in the video stream have the value 0xFFFF; or

• trick mode status is true (refer to 2.4.3.7).

For Low and Main level:



For High-1440 and High level:



For Constrained Parameters bitstream in ISO/IEC 11172-2:



where Rmax is the maximum bit rate for a Constrained Parameters bitstream in ISO/IEC 11172-2.

If there is PES packet payload data in MBn, and buffer EBn is not full, the PES packet payload is transferred from MBn to EBn at a rate equal to Rbxn. If EBn is full, data are not removed from MBn. When a byte of data is transferred from MBn to EBn, all PES packet header bytes that are in MBn and immediately precede that byte, are instantaneously removed and discarded. When there is no PES packet payload data present in MBn, no data is removed from MBn. All data that enters MBn leaves it. All PES packet payload data bytes enter EBn instantaneously upon leaving MBn.

Vbv\_delay method

The vbv\_delay method specifies precisely the time at which each byte of coded video data is transferred from MBn to EBn, using the vbv\_delay values coded in the video elementary stream. The vbv\_delay method is used whenever the STD descriptor (refer to 2.6.32) for this elementary stream is present in the transport stream, the leak\_valid flag in the descriptor has the value '0', and vbv\_delay fields coded in the video stream are not equal to 0xFFFF. If any vbv\_delay values in a video sequence are not equal to 0xFFFF, none of the vbv\_delay fields in that sequence shall be equal to 0xFFFF (refer to ISO/IEC 11172-2 and Rec. ITU-T H.262 | ISO/IEC 13818-2).

When the vbv\_delay method is used, the final byte of the video picture start code for picture j is transferred from MBn to the EBn at the time tdn(j) – vbv\_delay(j), where tdn(j) is the decoding time of picture j, as defined above, and vbv\_delay(j) is the delay time, in seconds, indicated by the vbv\_delay field of picture j. The transfer of bytes between the final bytes of successive picture start codes (including the final byte of the second start code), into the buffer EBn, is at a piecewise constant rate, Rbx(j), which is specified for each picture j. Specifically, the rate, Rbx(j), of transfer into this buffer is given by:

 (2-6)

where NB(j) is the number of bytes between the final bytes of the picture start codes (including the final byte of the second start code) of pictures j and j + 1, excluding PES packet header bytes.

NOTE 4 – vbv\_delay(j + 1) and tdn(j + 1) may have values that differ from those normally expected for periodic video display if the low\_delay flag in the video sequence extension is set to '1'. It may not be possible to determine the correct values by examination of the bit stream.

The Rbx(j) derived from equation 2-6 shall be less than or equal to Rmax[profile, level] for elementary streams of stream type 0x02 (refer to Table 2-34), where Rmax[profile, level] is defined in Rec. ITU-T H.262 | ISO/IEC 13818-2, and shall be less than or equal to the maximum bit rate allowed for constrained parameter video elementary streams of stream type 0x01, refer to ISO/IEC 11172-2.

When a byte of data is transferred from MBn to EBn, all PES packet header bytes that are in MBn and immediately precede that byte are instantaneously removed and discarded. All data that enters MBn leaves it. All PES packet payload data bytes enter EBn instantaneously upon leaving MBn.

Removal of access units

For each elementary stream buffer EBn and main buffer Bn all data for the access unit that has been in the buffer longest, An(j), and any stuffing bytes that immediately precede it that are present in the buffer at the time tdn(j) are removed instantaneously at time tdn(j). The decoding time tdn(j) is specified in the DTS or PTS fields (refer to 2.4.3.6). Decoding times tdn(j + 1), tdn(j + 2), ... of access units without encoded DTS or PTS fields which directly follow access unit j may be derived from information in the elementary stream. Refer to Annex C of Rec. ITU-T H.262 | ISO/IEC 13818-2, ISO/IEC 13818-3, or ISO/IEC 11172. Also refer to 2.7.5. In the case of audio, all PES packet headers that are stored immediately before the access unit or that are embedded within the data of the access unit are removed simultaneously with the removal of the access unit. As the access unit is removed it is instantaneously decoded to a presentation unit.

System data

In the case of system data, data is removed from the main buffer Bsys at a rate of Rsys whenever there is at least 1 byte available in buffer Bsys.

*Rsys* = max (80 000 bits/s, (2-7)

NOTE 5 – The intention of increasing Rsys in the case of high transport rates is to allow an increased data rate for the Program‑specific information.

Low delay

When the low\_delay flag in the video sequence extension is set to '1' (see 6.2.2.3 of Rec. ITU-T H.262 | ISO/IEC 13818‑2) the EBn buffer may underflow. In this case, when the T-STD elementary stream buffer EBn is examined at the time specified by tdn(j), the complete data for the access unit may not be present in the buffer EBn. When this case arises, the buffer shall be re-examined at intervals of two field-periods until the data for the complete access unit is present in the buffer. At this time the entire access unit shall be removed from buffer EBn instantaneously. Overflow of buffer EBn shall not occur.

When the low\_delay\_mode flag is set to '1', EBn underflow is allowed to occur continuously without limit. The T-STD decoder shall remove access unit data from buffer EBn at the earliest time consistent with the paragraph above and any DTS or PTS values encoded in the bit stream. Note that the decoder may be unable to re-establish correct decoding and display times as indicated by DTS and PTS until the EBn buffer underflow situation ceases and a PTS or DTS is found in the bit stream.

Trick mode

When the DSM\_trick\_mode flag (2.4.3.6) is set to '1' in the PES Packet header of a packet containing the start of a B‑type video access unit and the trick\_mode\_control field is set to '001' (slow motion) or '010' (freeze frame), or '100' (slow reverse) the B-picture access unit is not removed from the video data buffer EBn until the last time of possibly multiple times that any field of the picture is decoded and presented. Repetition of the presentation of fields and pictures is defined in 2.4.3.8 under slow motion, slow reverse, and field\_id\_cntrl. The access unit is removed instantaneously from EBn at the indicated time, which is dependent on the value of rep\_cntrl.

When the DSM\_trick\_mode flag is set to '1' in the PES packet header of a packet containing the first byte of a picture start code, trick\_mode status becomes true when that picture start code in the PES packet is removed from buffer EBn Trick mode status remains true until a PES packet header is received by the T-STD in which the DSM\_trick\_mode flag is set to '0' and the first byte of the picture start code after that PES packet header is removed from buffer EBn. When trick mode status is true, the buffer EBn may underflow. All other constraints from normal streams are retained when trick mode status is true.

#### 2.4.2.5 Decoding

Elementary streams buffered in B1 through Bn and EB1 through EBn are decoded instantaneously by decoders D1 through Dn and may be delayed in re-order buffers O1 through On before being presented at the output of the T-STD. Re-order buffers are used only in the case of a video elementary stream when some access units are not carried in presentation order. These access units will need to be re-ordered before presentation. In particular, if Pn(k) is an I‑picture or a P-picture carried before one or more B-pictures, then it must be delayed in the re-order buffer, On, of the T-STD before being presented. Any picture previously stored in On is presented before the current picture can be stored. Pn(k) should be delayed until the next I-picture or P-picture is decoded. While it is stored in the re-order buffer, the subsequent B-pictures are decoded and presented.

The time at which a presentation unit Pn(k) is presented is tpn(k). For presentation units that do not require re-ordering delay, tpn(k) is equal to tdn(j) since the access units are decoded instantaneously; this is the case, for example, for B‑frames. For presentation units that are delayed, tpn(k) and tdn(j) differ by the time that Pn(k) is delayed in the re-order buffer, which is a multiple of the nominal picture period. Care should be taken to use adequate re-ordering delay from the beginning of video elementary streams to meet the requirements of the entire stream. For example, a stream which initially has only I- and P-pictures but later includes B-pictures should include re-ordering delay starting at the beginning of the stream.

Rec. ITU-T H.262 | ISO/IEC 13818-2 explains re-ordering of video pictures in greater detail.

#### 2.4.2.6 Presentation

The function of a decoding system is to reconstruct presentation units from compressed data and to present them in a synchronized sequence at the correct presentation times. Although real audio and visual presentation devices generally have finite and different delays and may have additional delays imposed by post-processing or output functions, the system target decoder models these delays as zero.

In the T-STD in Figure 2-1 the display of a video presentation unit (a picture) occurs instantaneously at its presentation time, tpn(k).

In the T-STD the output of an audio presentation unit starts at its presentation time, tpn(k), when the decoder instantaneously presents the first sample. Subsequent samples in the presentation unit are presented in sequence at the audio sampling rate.

#### 2.4.2.7 Buffer management

Transport streams shall be constructed so that conditions defined in this subclause are satisfied. This subclause makes use of the notation defined for the system target decoder.

TBn and TBsys shall not overflow. TBn and TBsys shall empty at least once every second. Bn shall not overflow nor underflow. Bsys shall not overflow.

EBn shall not underflow except when the low delay flag in the video sequence extension is set to '1' (refer to 6.2.2.3 in Rec. ITU-T H.262 | ISO/IEC 13818-2) or trick\_mode status is true.

When the leak method for specifying transfers is in effect, MBn shall not overflow, and shall empty at least once every second. EBn shall not overflow.

When the vbv\_delay method for specifying transfers is in effect, MBn shall not overflow nor underflow, and EBn shall not overflow.

The delay of any data through the system target decoder buffers shall be less than or equal to one second except for still picture video data, ISO/IEC 14496, ISO/IEC 23008-2, ISO/IEC 23090-3, ISO/IEC 23094-1 and ISO/IEC 23094-2 streams. Specifically: tdn(j) – t(i) ≤ 1 second for all j, and all bytes i in access unit An(j).

For still picture video data, the delay is constrained by tdn(j) – t(i) ≤ 60 seconds for all j, and all bytes i in access unit An(j).

For ISO/IEC 14496, ISO/IEC 23008-2, ISO/IEC 23090-3, ISO/IEC 23094-1 and ISO/IEC 23094-2 streams, the delay is constrained by tdn(j) – t(i) ≤ 10 seconds for all j, and all bytes i in access unit An(j).

Definition of overflow and underflow

Let Fn(t) be the instantaneous fullness of T-STD buffer Bn.

Fn(t) = 0 instantaneously before t = t(0)

Overflow does not occur if:



for all t and n.

Underflow does not occur if:



for all t and n.

#### 2.4.2.8 T-STD extensions for carriage of ISO/IEC 14496 data

For decoding of ISO/IEC 14496 data carried in a transport stream the T-STD model is extended. T-STD parameters for decoding of individual ISO/IEC 14496 elementary streams are defined in 2.11.2, while 2.11.3 defines T‑STD extensions and parameters for decoding of ISO/IEC 14496 scenes and associated streams.

#### 2.4.2.9 T-STD extensions for carriage of Rec. ITU-T H.264 | ISO/IEC 14496-10 video

To define the decoding in the T-STD of Rec. ITU-T H.264 | ISO/IEC 14496-10 video streams carried in a transport stream, the T-STD model needs to be extended. The T-STD extension and T-STD parameters for decoding of AVC video streams conforming to one or more profiles defined in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496-10 video streams are defined in 2.14.3.1, T-STD extension and T-STD parameters for decoding of AVC video streams conforming to one or more profiles defined in Annex G of Rec. ITU-T H.264 | ISO/IEC 14496-10 are defined in 2.14.3.5, and T-STD extension and T-STD parameters for decoding of AVC video streams conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10 are defined in 2.14.3.7.

#### 2.4.2.10 T-STD extensions for carriage of ISO/IEC 14496-17 text streams

To define the decoding in the T-STD of ISO/IEC 14496-17 text streams carried in a transport stream, the T‑STD model needs to be extended. The T-STD extension and T-STD parameters for decoding of ISO/IEC 14496-17 text streams are defined in 2.15.3.1.

#### 2.4.2.11 T-STD extensions for carriage of J2K video elementary streams

The interpretation, extensions, use and constraints for syntax elements in the adaptation header (2.4.3.4 and 2.4.3.5) for JPEG 2000 part 1 video are defined in S.5.

The interpretation, extensions, use and constraints for syntax elements in the PES header (2.4.3.6 and 2.4.3.7) for JPEG 2000 Part 1 video are defined in S.5.

To define the decoding of J2K video elementary streams carried in a transport stream, the T-STD model needs to be extended. The T-STD extensions and T-STD parameters for decoding of J2K video elementary streams conforming to one or more profiles defined in Rec. ITU-T T.800 (2002) | ISO/IEC 15444-1:2004 are defined in S.6.

NOTE – No extensions are specified for P-STD model, as carriage of J2K video elementary streams in program streams is not supported.

#### 2.4.2.12 T-STD extensions for carriage of HEVC

T-STD extensions and T-STD parameters for the decoding of HEVC video streams are defined in 2.17.2 and 2.17.3. Program stream support including P-STD extensions and P-STD parameters are not specified for HEVC video streams.

#### 2.4.2.13 T-STD extensions for carriage of MVCD video sub-bitstream

T-STD extensions and T-STD parameters for decoding of MVCD video sub-bitstreams are defined in 2.14.1 and 2.14.3.7.

NOTE – Program stream extensions are not specified for MVCD video sub-bitstreams.

#### 2.4.2.14 T-STD extensions for carriage of MV HEVC and SHVC

T-STD extensions and T-STD parameters for decoding of HEVC layered video streams are defined in 2.17.4. Program stream support including P-STD extensions and P-STD parameters are not specified for HEVC extension video streams.

#### 2.4.2.15 T-STD extensions for carriage of JPEG XS video elementary streams

The interpretation, extensions, use and constraints for syntax elements in the adaptation header (2.4.3.4 and 2.4.3.5) for JPEG XS Part-1 video are defined in W.5.

The interpretation, extensions, use and constraints for syntax elements in the PES header (2.4.3.6 and 2.4.3.7) for JPEG XS Part-1 video are defined in W.5.

To define the decoding of JPEG XS video elementary streams carried in a Transport Stream, the T-STD model needs to be extended. The T-STD extensions and T-STD parameters for decoding of JPEG XS video elementary streams conforming to one or more profiles defined in ISO/IEC 21122-2 are defined in W.6.

NOTE – No extensions are specified for P-STD model, as carriage of JPEG XS video elementary streams in program streams is not supported.

#### 2.4.2.16 T-STD extensions for carriage of VVC

T-STD extensions and T-STD parameters for decoding of VVC video streams are defined in 2.23.2 and 2.23.3. Program stream support including P-STD extensions and P-STD parameters are not specified for VVC video streams.

#### 2.4.2.17 T-STD extensions for carriage of EVC

T-STD extensions and T-STD parameters for decoding of EVC video streams are defined in 2.24.2. Program stream support including P-STD extensions and P-STD parameters are not specified for EVC video streams.

### 2.4.3 Specification of the transport stream syntax and semantics

The following syntax describes a stream of bytes. Transport stream packets shall be 188 bytes long.

#### 2.4.3.1 Transport stream

See Table 2-1.

Table 2-1 – Transport stream

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| MPEG\_transport\_stream() { |  |  |
| do { |  |  |
| transport\_packet() |  |  |
| } while (nextbits() = = sync\_byte) |  |  |
| } |  |  |

#### 2.4.3.2 Transport stream packet layer

See Table 2-2.

Table 2-2 – Transport packet of this Recommendation | International Standard

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| transport\_packet(){ |  |  |
| **sync\_byte** | **8** | **bslbf** |
| **transport\_error\_indicator** | **1** | **bslbf** |
| **payload\_unit\_start\_indicator** | **1** | **bslbf** |
| **transport\_priority** | **1** | **bslbf** |
| **PID** | **13** | **uimsbf** |
| **transport\_scrambling\_control** | **2** | **bslbf** |
| **adaptation\_field\_control** | **2** | **bslbf** |
| **continuity\_counter** | **4** | **uimsbf** |
| if(adaptation\_field\_control = = '10' || adaptation\_field\_control = = '11'){ |  |  |
| adaptation\_field() |  |  |
| } |  |  |
| if(adaptation\_field\_control = = '01' || adaptation\_field\_control = = '11') { |  |  |
| for (i = 0; i < N; i++){ |  |  |
| **data\_byte** | **8** | **bslbf** |
| } |  |  |
| } |  |  |
| } |  |  |

#### 2.4.3.3 Semantic definition of fields in transport stream packet layer

**sync\_byte** –The sync\_byte is a fixed 8-bit field whose value is '0100 0111' (0x47). Sync\_byte emulation in the choice of values for other regularly occurring fields, such as PID, should be avoided.

**transport\_error\_indicator** – The transport\_error\_indicator is a 1-bit flag. When set to '1' it indicates that at least 1 uncorrectable bit error exists in the associated transport stream packet. This bit may be set to '1' by entities external to the transport layer. When set to '1' this bit shall not be reset to '0' unless the bit value(s) in error have been corrected.

**payload\_unit\_start\_indicator** –The payload\_unit\_start\_indicator is a 1-bit flag which has normative meaning for transport stream packets that carry PES packets (refer to 2.4.3.6) or transport stream section data (refer to Table 2-31 in 2.4.4.5).

When the payload of the transport stream packet contains PES packet data, the payload\_unit\_start\_indicator has the following significance: a '1' indicates that the payload of this transport stream packet will commence with the first byte of a PES packet and a '0' indicates no PES packet shall start in this transport stream packet. If the payload\_unit\_start\_indicator is set to '1', then one and only one PES packet starts in this transport stream packet. This also applies to private streams of stream\_type 6 (refer to Table 2-34).

When the payload of the transport stream packet contains transport stream section data, the payload\_unit\_start\_indicator has the following significance: if the transport stream packet carries the first byte of a section, the payload\_unit\_start\_indicator value shall be '1', indicating that the first byte of the payload of this transport stream packet carries the pointer\_field. If the transport stream packet does not carry the first byte of a section, the payload\_unit\_start\_indicator value shall be '0', indicating that there is no pointer\_field in the payload. Refer to 2.4.4.2 and 2.4.4.3. This also applies to private streams of stream\_type 5 (refer to Table 2-34).

For null packets the payload\_unit\_start\_indicator shall be set to '0'.

The meaning of this bit for transport stream packets carrying only private data is not defined in this Specification.

**transport\_priority** – The transport\_priority is a 1-bit indicator. When set to '1' it indicates that the associated packet is of greater priority than other packets having the same PID which do not have the bit set to '1'. The transport mechanism can use this to prioritize its data within an elementary stream. Depending on the application the transport\_priority field may be coded regardless of the PID or within one PID only. This field may be changed by channel-specific encoders or decoders.

**PID** –The PID is a 13-bit field, indicating the type of the data stored in the packet payload. PID value 0x0000 is reserved for the program association table (see Table 2-30). PID value 0x0001 is reserved for the conditional access table (see Table 2-32). PID value 0x0002 is reserved for the transport stream description table (see Table 2-36), PID value 0x0003 is reserved for IPMP control information table (see ISO/IEC 13818-11) and PID values 0x0004‑0x000F are reserved. PID value 0x1FFF is reserved for null packets (see Table 2-3).

Table 2-3 – PID table

|  |  |
| --- | --- |
| Value | Description |
| 0x0000 | Program association table |
| 0x0001 | Conditional access table |
| 0x0002 | Transport stream description table |
| 0x0003 | IPMP control information table |
| 0x0004 | Adaptive streaming information (see Note 2) |
| 0x0005 .. 0x000F | Reserved |
| 0x0010 .. 0x1FFE | May be assigned as network\_PID, Program\_map\_PID, elementary\_PID, or for other purposes |
| 0x1FFF | Null packet |
| NOTE 1 – The transport packets with PID values 0x0000, 0x0001, and 0x0010-0x1FFE are allowed to carry a PCR.  NOTE 2 – Payload syntax is defined in 5.10.3.3.5 of ISO/IEC 23009-1. | |

**transport\_scrambling\_control** –This 2-bit field indicates the scrambling mode of the transport stream packet payload. The transport stream packet header, and the adaptation field when present, shall not be scrambled. In the case of a null packet the value of the transport\_scrambling\_control field shall be set to '00' (see Table 2-4).

Table 2-4 – Scrambling control values

|  |  |
| --- | --- |
| Value | Description |
| '00' | Not scrambled |
| '01' | User-defined |
| '10' | User-defined |
| '11' | User-defined |

**adaptation\_field\_control** –This 2-bit field indicates whether this transport stream packet header is followed by an adaptation field and/or payload (see Table 2-5).

Table 2-5 – Adaptation field control values

|  |  |
| --- | --- |
| Value | Description |
| '00' | Reserved for future use by ISO/IEC |
| '01' | No adaptation\_field, payload only |
| '10' | Adaptation\_field only, no payload |
| '11' | Adaptation\_field followed by payload |

Rec. ITU-T H.222.0 | ISO/IEC 13818-1 decoders shall discard transport stream packets with the adaptation\_field\_control field set to a value of '00'. In the case of a null packet the value of the adaptation\_field\_control shall be set to '01'.

**continuity\_counter** –The continuity\_counter is a 4-bit field incrementing with each transport stream packet with the same PID. The continuity\_counter wraps around to 0 after its maximum value. The continuity\_counter shall not be incremented when the adaptation\_field\_control of the packet equals '00' or '10'.

In transport streams, duplicate packets may be sent as two, and only two, consecutive transport stream packets of the same PID. The duplicate packets shall have the same continuity\_counter value as the original packet and the adaptation\_field\_control field shall be equal to '01' or '11'. In duplicate packets each byte of the original packet shall be duplicated, with the exception that in the program clock reference fields, if present, a valid value shall be encoded.

The continuity\_counter in a particular transport stream packet is continuous when it differs by a positive value of one from the continuity\_counter value in the previous transport stream packet of the same PID, or when either of the non‑incrementing conditions (adaptation\_field\_control set to '00' or '10', or duplicate packets as described above) are met. The continuity counter may be discontinuous when the discontinuity\_indicator is set to '1' (refer to 2.4.3.4). In the case of a null packet the value of the continuity\_counter is undefined.

**data\_byte** – Data bytes shall be contiguous bytes of data from the PES packets (refer to 2.4.3.6), transport stream sections (refer to 2.4.4), packet stuffing bytes after transport stream sections, or private data not in these structures as indicated by the PID. In the case of null packets with PID value 0x1FFF, data\_bytes may be assigned any value. The number of data\_bytes, N, is specified by 184 minus the number of bytes in the adaptation\_field(), as described in 2.4.3.4.

#### 2.4.3.4 Adaptation field

See Table 2-6.

| Table 2-6 – Transport stream adaptation field | | | |
| --- | --- | --- | --- |
| Syntax | No. of bits | Mnemonic | |
| adaptation\_field() { |  |  | |
| **adaptation\_field\_length** | **8** | **uimsbf** | |
| if (adaptation\_field\_length > 0) { |  |  | |
| **discontinuity\_indicator** | **1** | **bslbf** | |
| **random\_access\_indicator** | **1** | **bslbf** | |
| **elementary\_stream\_priority\_indicator** | **1** | **bslbf** | |
| **PCR\_flag** | **1** | **bslbf** | |
| **OPCR\_flag** | **1** | **bslbf** | |
| **splicing\_point\_flag** | **1** | **bslbf** | |
| **transport\_private\_data\_flag** | **1** | **bslbf** | |
| **adaptation\_field\_extension\_flag** | **1** | **bslbf** | |
| if (PCR\_flag = = '1') { |  |  | |
| **program\_clock\_reference\_base** | **33** | **uimsbf** | |
| **reserved** | **6** | **bslbf** | |
| **program\_clock\_reference\_extension** | **9** | **uimsbf** | |
| } |  |  | |
| if (OPCR\_flag = = '1') { |  |  | |
| **original\_program\_clock\_reference\_base** | **33** | **uimsbf** | |
| **reserved** | **6** | **bslbf** | |
| **original\_program\_clock\_reference\_extension** | **9** | **uimsbf** | |
| } |  |  | |
| if (splicing\_point\_flag = = '1') { |  |  | |
| **splice\_countdown** | **8** | **tcimsbf** | |
| } |  |  | |
| if (transport\_private\_data\_flag = = '1') { |  |  | |
| **transport\_private\_data\_length** | **8** | **uimsbf** | |
| for (i = 0; i < transport\_private\_data\_length; i++) { |  |  | |
| **private\_data\_byte** | **8** | **bslbf** | |
| } |  |  | |
| } |  |  | |
| if (adaptation\_field\_extension\_flag = = '1') { |  |  | |
| **adaptation\_field\_extension\_length** | **8** | **uimsbf** | |
| **ltw\_flag** | **1** | **bslbf** | |
| **piecewise\_rate\_flag** | **1** | **bslbf** | |
| **seamless\_splice\_flag** | **1** | **bslbf** | |
| **af\_descriptor\_not\_present\_flag** | **1** | **bslbf** | |
| **reserved** | **4** | **bslbf** | |
| if (ltw\_flag = = '1') { |  |  | |
| **ltw\_valid\_flag** | **1** | **bslbf** | |
| **ltw\_offset** | **15** | **uimsbf** | |
| } |  |  | |
| if (piecewise\_rate\_flag = = '1') { |  |  | |
| **reserved** | **2** | **bslbf** | |
| **piecewise\_rate** | **22** | **uimsbf** | |
| } |  |  | |
| if (seamless\_splice\_flag = = '1') { |  |  | |
| **Splice\_type** | **4** | **bslbf** | |
| **DTS\_next\_AU[32..30]** | **3** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| **DTS\_next\_AU[29..15]** | **15** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| **DTS\_next\_AU[14..0]** | **15** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| } |  |  | |
| if (af\_descriptor\_not\_present\_flag   '0') { |  |  | |
| for (i  0; i  N1; i) { |  |  | |
| af\_descriptor() |  |  | |
| } |  |  | |
| } |  |  | |
| else { |  |  | |
| for (i = 0; i < N2; i++) { |  |  | |
| **reserved** | **8** | **bslbf** | |
| **}** |  |  | |
| } |  |  | |
| } |  |  | |
| for (i = 0; i < N3; i++) { |  |  | |
| **stuffing\_byte** | **8** | **bslbf** | |
| } |  |  | |
| } |  |  | |
| } |  |  | |

#### 2.4.3.5 Semantic definition of fields in adaptation field

**adaptation\_field\_length** –The adaptation\_field\_length is an 8-bit field specifying the number of bytes in the adaptation\_field immediately following the adaptation\_field\_length. The value '0' is for inserting a single stuffing byte in the adaptation field of a transport stream packet. When the adaptation\_field\_control value is '11', the value of the adaptation\_field\_length shall be in the range 0 to 182. When the adaptation\_field\_control value is '10', the value of the adaptation\_field\_length shall be 183. For transport stream packets carrying PES packets, stuffing is needed when there is insufficient PES packet data to completely fill the transport stream packet payload bytes. Stuffing is accomplished by defining an adaptation field longer than the sum of the lengths of the data elements in it, so that the payload bytes remaining after the adaptation field exactly accommodates the available PES packet data. The extra space in the adaptation field is filled with stuffing bytes.

This is the only method of stuffing allowed for transport stream packets carrying PES packets. For transport stream packets carrying sections, an alternative stuffing method is described in 2.4.4.1.

**discontinuity\_indicator** – This is a 1-bit field which when set to '1' indicates that the discontinuity state is true for the current transport stream packet. When the discontinuity\_indicator is set to '0' or is not present, the discontinuity state is false. The discontinuity indicator is used to indicate two types of discontinuities, system time-base discontinuities and continuity\_counter discontinuities.

A system time-base discontinuity is indicated by the use of the discontinuity\_indicator in transport stream packets of a PID designated as a PCR\_PID (refer to 2.4.4.10). When the discontinuity state is true for a transport stream packet of a PID designated as a PCR\_PID, the next PCR in a transport stream packet with that same PID represents a sample of a new system time clock for the associated program. The system time-base discontinuity point is defined to be the instant in time when the first byte of a packet containing a PCR of a new system time-base arrives at the input of the T-STD. The discontinuity\_indicator shall be set to '1' in the packet in which the system time-base discontinuity occurs. The discontinuity\_indicator bit may also be set to '1' in transport stream packets of the same PCR\_PID prior to the packet which contains the new system time-base PCR. In this case, once the discontinuity\_indicator has been set to '1', it shall continue to be set to '1' in all transport stream packets of the same PCR\_PID up to and including the transport stream packet which contains the first PCR of the new system time-base. After the occurrence of a system time-base discontinuity, no fewer than two PCRs for the new system time-base shall be received before another system time-base discontinuity can occur. Further, except when trick mode status is true, data from no more than two system time-bases shall be present in the set of T-STD buffers for one program at any time.

Prior to the occurrence of a system time-base discontinuity, the first byte of a transport stream packet which contains a PTS or DTS which refers to the new system time-base shall not arrive at the input of the T-STD. After the occurrence of a system time-base discontinuity, the first byte of a transport stream packet which contains a PTS or DTS which refers to the previous system time-base shall not arrive at the input of the T-STD.

A continuity\_counter discontinuity is indicated by the use of the discontinuity\_indicator in any transport stream packet. When the discontinuity state is true in any transport stream packet of a PID not designated as a PCR\_PID, the continuity\_counter in that packet may be discontinuous with respect to the previous transport stream packet of the same PID. When the discontinuity state is true in a transport stream packet of a PID that is designated as a PCR\_PID, the continuity\_counter may only be discontinuous in the packet in which a system time-base discontinuity occurs. A continuity counter discontinuity point occurs when the discontinuity state is true in a transport stream packet and the continuity\_counter in the same packet is discontinuous with respect to the previous transport stream packet of the same PID. A continuity counter discontinuity point shall occur at most one time from the initiation of the discontinuity state until the conclusion of the discontinuity state. Furthermore, for all PIDs that are not designated as PCR\_PIDs, when the discontinuity\_indicator is set to '1' in a packet of a specific PID, the discontinuity\_indicator may be set to '1' in the next transport stream packet of that same PID, but shall not be set to '1' in three consecutive transport stream packet of that same PID.

For the purpose of this clause, an elementary stream access point is defined as follows:

• ISO/IEC 11172-2 video and Rec. ITU-T H.262 | ISO/IEC 13818-2 video – The first byte of a video sequence header.

• ISO/IEC 14496-2 visual – The first byte of the visual object sequence header.

• AVC video streams conforming to one or more profiles defined in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496-10 – The first byte of an AVC access unit. The SPS and PPS parameter sets referenced in this and all subsequent AVC access units in the coded video stream shall be provided after this access point in the byte stream and prior to their activation.

• Video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex G of Rec. ITU-T H.264 | ISO/IEC 14496-10 – The first byte of an SVC dependency representation is an elementary stream access point if the following conditions are met:

• The subset sequence parameter sets and picture parameter sets referenced in this and all subsequent SVC dependency representation in the video sub-bitstream shall be provided after this access point in the byte stream and prior to their activation.

• If this SVC video sub-bitstream access point requires the elementary stream access point of the same AVC access unit, if any, contained in the corresponding elementary stream that needs to be present in decoding order before decoding the elementary stream associated with this elementary stream access point, then the corresponding elementary stream shall also include an elementary stream access point.

NOTE 1 – If the hierarchy descriptor is present for this SVC video sub-bitstream then the video sub-bitstream of which the hierarchy\_layer\_index equals the hierarchy\_embedded\_layer\_index of this SVC sub-bitstream should have an elementary stream access point in the same access unit.

• MVC video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10 – The first byte of an MVC view-component subset is an elementary stream access point if the following two conditions are met:

– The subset sequence parameter sets and picture parameter sets referenced in this and all subsequent MVC view-component subsets in the MVC video sub-bitstream shall be provided after this access point in the byte stream and prior to their activation.

– If this MVC video sub-bitstream access point requires the elementary stream access point of the same AVC access unit, if any, contained in the corresponding elementary stream that needs to be present in decoding order before decoding the elementary stream associated with this elementary stream access point, then the corresponding elementary stream shall also include an elementary stream access point.

NOTE 2 – If the hierarchy descriptor is present for this MVC video sub-bitstream, then the MVC video sub-bitstream of which the hierarchy\_layer\_index equals the hierarchy\_embedded\_layer\_index of this MVC sub-bitstream should have an elementary stream access point in this same access unit.

• Audio – The first byte of an audio frame.

• ISO/IEC 14496-17 text stream – The first byte of a text access unit. In case in-band sample descriptions are used, each in-band sample description shall be provided in the ISO/IEC 14496-17 stream after this access point and prior to its use by an access unit.

• HEVC video streams or HEVC temporal video sub-bitstreams – The first byte of an HEVC access unit. The VPS, SPS and PPS parameter sets, as defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2, referenced in this and all subsequent HEVC access units in the HEVC video sequence shall be provided after this access point in the byte stream and prior to their activation.

• MVCD video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496-10 – The first byte of an MVCD view-component subset is an elementary stream access point if the following two conditions are met:

– The subset sequence parameter sets and picture parameter sets referenced in this and all subsequent MVCD view-component subsets in the MVCD video sub-bitstream shall be provided after this access point in the byte stream and prior to their activation.

– If this MVCD video sub-bitstream access point requires the elementary stream access point of the same AVC access unit, if any, contained in the corresponding elementary stream that needs to be present in decoding order before decoding the elementary stream associated with this elementary stream access point, then the corresponding elementary stream shall also include an elementary stream access point.

NOTE 3 – If the hierarchy descriptor is present for this MVCD video sub-bitstream, then the MVCD video sub-bitstream of which the hierarchy\_layer\_index equals the hierarchy\_embedded\_layer\_index of this MVCD sub-bitstream should have an elementary stream access point in this same access unit.

• VVC video streams or VVC temporal video sub-bitstreams – The first byte of a VVC access unit. The DCI NAL unit (if present) and VPS (if present), SPS and PPS parameter sets, as defined in Rec. ITU-T H.266 | ISO/IEC 23090-3, referenced in this and all subsequent VVC access units in the VVC video sequence shall be provided after this access point in the byte stream and prior to being referred to.

• EVC video streams or EVC temporal video sub-bitstreams – The first byte of an EVC access unit. The APS, SPS and PPS, as defined in ISO/IEC 23094-1, referenced in this and all subsequent EVC access units in the EVC video sequence shall be provided after this access point in the byte stream and prior to their activation.

After a continuity counter discontinuity in a transport packet which is designated as containing elementary stream data, the first byte of elementary stream data in a transport stream packet of the same PID shall be the first byte of an elementary stream access point. In the case of ISO/IEC 11172-2, or Rec. ITU-T H.262 | ISO/IEC 13818-2 or ISO/IEC 14496-2 video, the first byte of an elementary stream access point may also be the first byte of a sequence\_end\_code followed by an elementary stream access point.

Each transport stream packet which contains elementary stream data with a PID not designated as a PCR\_PID, and in which a continuity counter discontinuity point occurs, and in which a PTS or DTS occurs, shall arrive at the input of the T-STD after the system time-base discontinuity for the associated program occurs. In the case where the discontinuity state is true, if two consecutive transport stream packets of the same PID occur which have the same continuity\_counter value and have adaptation\_field\_control values set to '01' or '11', the second packet may be discarded. A transport stream shall not be constructed in such a way that discarding such a packet will cause the loss of PES packet payload data or transport stream section data.

After the occurrence of a discontinuity\_indicator set to '1' in a transport stream packet which contains PSI information, a single discontinuity in the version\_number of PSI sections may occur. At the occurrence of such a discontinuity, a version of the TS\_program\_map\_sections of the appropriate program shall be sent with section\_length = = 13 and the current\_next\_indicator = = 1, such that there are no program\_descriptors and no elementary streams described. This shall then be followed by a version of the TS\_program\_map\_section for each affected program with the version\_number incremented by one and the current\_next\_indicator = = 1, containing a complete program definition. This indicates a version change in PSI data.

**random\_access\_indicator** – The random\_access\_indicator is a 1-bit field that indicates that the current transport stream packet, and possibly subsequent transport stream packets with the same PID, contain some information to aid random access at this point.

Specifically, when the bit is set to '1', the next PES packet to start in the payload of transport stream packets with the current PID shall contain an elementary stream access point as defined in the semantics for the discontinuity\_indicator field. In addition, in the case of video, a presentation timestamp shall be present for the first picture following the elementary stream access point.

In the case of audio, the presentation timestamp shall be present in the PES packet containing the first byte of the audio frame. In the PCR\_PID the random\_access\_indicator may only be set to '1' in a transport stream packet containing the PCR fields.

**elementary\_stream\_priority\_indicator** – The elementary\_stream\_priority\_indicator is a 1-bit field. It indicates, among packets with the same PID, the priority of the elementary stream data carried within the payload of this transport stream packet. A '1' indicates that the payload has a higher priority than the payloads of other transport stream packets.

In the case of ISO/IEC 11172-2 or Rec. ITU-T H.262 | ISO/IEC 13818-2 or ISO/IEC 14496-2 video, this field may be set to '1' only if the payload contains one or more bytes from an intra-coded slice.

In the case of Rec. ITU-T H.264 | ISO/IEC 14496-10 video, this field may be set to '1' only if the payload contains one or more bytes from a slice with slice\_type set to 2, 4, 7, or 9.

A value of '0' indicates that the payload has the same priority as all other packets which do not have this bit set to '1'.

For MVC video sub-bitstreams or MVC base view sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10, this field may be set to '1' only if the payload contains one or more bytes from an anchor picture, indicated by the slice\_type equal to 2, 4, 7, or 9 and the anchor\_pic\_flag syntax element equal to 1 for all prefix NAL units and slice extension NAL units.

For MVCD video sub-bitstreams or MVCD base view sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496-10, this field may be set to '1' only if the payload contains one or more bytes from an anchor picture, indicated by the slice\_type equal to 2, 4, 7, or 9 and the anchor\_pic\_flag syntax element equal to 1 for all prefix NAL units and slice extension NAL units. In the case of HEVC video streams or HEVC temporal video sub-bitstreams or HEVC temporal video subsets, this field may be set to '1' only if the payload contains one or more bytes from a slice with slice\_type set to 2. A value of '0' indicates that the payload has the same priority as all other packets which do not have this bit set to '1'.

In the case of VVC video streams or VVC temporal video sub-bitstreams or VVC temporal video subsets, this field may be set to '1' only if the payload contains one or more bytes from a slice with slice\_type set to 2. A value of '0' indicates that the payload has the same priority as all other packets which do not have this bit set to '1'.

In the case of EVC video streams or EVC temporal video sub-bitstreams, this field may be set to '1' only if the payload contains one or more bytes from a slice with slice\_type set to 2. A value of '0' indicates that the payload has the same priority as all other packets which do not have this bit set to '1'.

**PCR\_flag** – The PCR\_flag is a 1-bit flag. A value of '1' indicates that the adaptation\_field contains a PCR field coded in two parts. A value of '0' indicates that the adaptation field does not contain any PCR field.

**OPCR\_flag** – The OPCR\_flag is a 1-bit flag. A value of '1' indicates that the adaptation\_field contains an OPCR field coded in two parts. A value of '0' indicates that the adaptation field does not contain any OPCR field.

**splicing\_point\_flag** – The splicing\_point\_flag is a 1-bit flag. When set to '1', a splice\_countdown field shall be present in this adaptation field, specifying the occurrence of a splicing point. A value of '0' indicates that a splice\_countdown field is not present in the adaptation field.

**transport\_private\_data\_flag** – The transport\_private\_data\_flag is a 1-bit flag. A value of '1' indicates that the adaptation field contains one or more private\_data bytes. A value of '0' indicates the adaptation field does not contain any private\_data bytes.

**adaptation\_field\_extension\_flag** –The adaptation\_field\_extension\_flag is a 1-bit field which when set to '1' indicates the presence of an adaptation field extension. A value of '0' indicates that an adaptation field extension is not present in the adaptation field.

**program\_clock\_reference\_base; program\_clock\_reference\_extension** –The program\_clock\_reference (PCR) is a 42‑bit field coded in two parts. The first part, program\_clock\_reference\_base, is a 33-bit field whose value is given by PCR\_base(i), as given in equation 2-2. The second part, program\_clock\_reference\_extension, is a 9-bit field whose value is given by PCR\_ext(i), as given in equation 2-3. The PCR indicates the intended time of arrival of the byte containing the last bit of the program\_clock\_reference\_base at the input of the system target decoder.

**original\_program\_clock\_reference\_base; original\_program\_clock\_reference\_extension** – The optional original program reference (OPCR) is a 42-bit field coded in two parts. These two parts, the base and the extension, are coded identically to the two corresponding parts of the PCR field. The presence of the OPCR is indicated by the OPCR\_flag. The OPCR field shall be coded only in transport stream packets in which the PCR field is present. OPCRs are permitted in both single program and multiple program transport streams.

OPCR assists in the reconstruction of a single program transport stream from another transport stream. When reconstructing the original single program transport stream, the OPCR may be copied to the PCR field. The resulting PCR value is valid only if the original single program transport stream is reconstructed exactly in its entirety. This would include at least any PSI and private data packets which were present in the original transport stream and would possibly require other private arrangements. It also means that the OPCR must be an identical copy of its associated PCR in the original single program transport stream.

The OPCR is expressed as follows:

 (2-8)

where:

 (2-9)

 (2-10)

The OPCR field is ignored by the decoder. The OPCR field shall not be modified by any multiplexor or decoder.

**splice\_countdown** –The splice\_countdown is an 8-bit field, representing a value which may be positive or negative. A positive value specifies the remaining number of transport stream packets, of the same PID, following the associated transport stream packet until a splicing point is reached. Duplicate transport stream packets and transport stream packets which only contain adaptation fields are excluded. The splicing point is located immediately after the last byte of the transport stream packet in which the associated splice\_countdown field reaches zero. In the transport stream packet where the splice\_countdown reaches zero, the last data byte of the transport stream packet payload shall be the last byte of a coded audio frame or a coded picture. In the case of video, the corresponding access unit may or may not be terminated by a sequence\_end\_code. Transport stream packets with the same PID, which follow, may contain data from a different elementary stream of the same type.

The payload of the next transport stream packet of the same PID (duplicate packets and packets without payload being excluded) shall commence with the first byte of a PES packet. In the case of audio, the PES packet payload shall commence with an access point. In the case of video, the PES packet payload shall commence with an access point, or with a sequence\_end\_code, followed by an access point. Thus, the previous coded audio frame or coded picture aligns with the packet boundary, or is padded to make this so. Subsequent to the splicing point, the countdown field may also be present. When the splice\_countdown is a negative number whose value is minus n (–n), it indicates that the associated transport stream packet is the n-th packet following the splicing point (duplicate packets and packets without payload being excluded).

For the definition of an elementary stream access point, see the semantics of discontinuity\_indicator.

**transport\_private\_data\_length** –The transport\_private\_data\_length is an 8-bit field specifying the number of private\_data bytes immediately following the transport private\_data\_length field. The number of private\_data bytes shall be such that private data does not extend beyond the adaptation field.

**private\_data\_byte** – The private\_data\_byte is an 8-bit field that shall not be specified by ITU-T | ISO/IEC.

**adaptation\_field\_extension\_length** – The adaptation\_field\_extension\_length is an 8-bit field. It indicates the number of bytes of the extended adaptation field data immediately following this field, including reserved bytes if present.

**ltw\_flag** (legal time window\_flag)–This is a 1-bit field which when set to '1' indicates the presence of the ltw\_offset field.

**piecewise\_rate\_flag** – This is a 1-bit field which when set to '1' indicates the presence of the piecewise\_rate field.

**seamless\_splice\_flag** – This is a 1-bit flag which when set to '1' indicates that the splice\_type and DTS\_next\_AU fields are present. A value of '0' indicates that neither splice\_type nor DTS\_next\_AU fields are present. This field shall be set to '0' in transport stream packets in which the splicing\_point\_flag is set to '0'. Once it is set to '1' in a transport stream packet in which the splice\_countdown is positive, it shall be set to '1' in all the subsequent transport stream packets of the same PID that have the splicing\_point\_flag set to '1', until the packet in which the splice\_countdown reaches zero (including this packet).

When this flag is set, and if the elementary stream carried in this PID is not a Rec. ITU-T H.262 | ISO/IEC 13818-2 video stream, then the splice\_type field shall be set to '0000'. If the elementary stream carried in this PID is an Rec. ITU-T H.262 | ISO/IEC 13818-2 video stream, it shall fulfil the constraints indicated by the splice\_type value.

**af\_descriptor\_not\_present\_flag** – This 1-bit field when set to '0' signals the presence of one or several af\_descriptor() constructs in the adaptation header. When this flag is set to '1' it indicates that the af\_descriptor() is not present in the adaptation header.

**ltw\_valid\_flag** (legal time window\_valid\_flag) – This is a 1-bit field which when set to '1' indicates that the value of the ltw\_offset shall be valid. A value of '0' indicates that the value in the ltw\_offset field is undefined.

**ltw\_offset** (legal time window offset) – This is a 15-bit field, the value of which is defined only if the ltw\_valid flag has a value of '1'. When defined, the legal time window offset is in units of (300/fs) seconds, where fs is the system clock frequency of the program that this PID belongs to, and fulfils:





where i is the index of the first byte of this transport stream packet, offset is the value encoded in this field, t(i) is the arrival time of byte i in the T-STD, and t1(i) is the upper bound in time of a time interval called the legal time window which is associated with this transport stream packet.

The legal time window has the property that if this transport stream is delivered to a T-STD starting at time t1(i), i.e., at the end of its legal time window, and all other transport stream packets of the same program are delivered at the end of their legal time windows, then:

• For video – The MBn buffer for this PID in the T-STD shall contain less than 184 bytes of elementary stream data at the time the first byte of the payload of this transport stream packet enters it, and no buffer violations in the T-STD shall occur.

• For audio – The Bn buffer for this PID in the T-STD shall contain less thanBSdec + 1 bytes of elementary stream data at the time the first byte of this transport stream packet enters it, and no buffer violations in the T-STD shall occur.

Depending on factors including the size of the buffer MBn and the rate of data transfer between MBn and EBn, it is possible to determine another time t0(i), such that if this packet is delivered anywhere in the interval [t0(i), t1(i)], no T‑STD buffer violations will occur. This time interval is called the legal time window. The value of t0 is not defined in this Recommendation | International Standard.

The information in this field is intended for devices such as remultiplexers which may need this information in order to reconstruct the state of the buffers MBn.

**piecewise\_rate** – The meaning of this 22-bit field is only defined when both the ltw\_flag and the ltw\_valid\_flag are set to '1'. When defined, it is a positive integer specifying a hypothetical bitrate R which is used to define the end times of the Legal Time Windows of transport stream packets of the same PID that follow this packet but do not include the legal\_time\_window\_offset field.

Assume that the first byte of this transport stream packet and the N following transport stream packets of the same PID have indices Ai, Ai+1, ..., Ai+N, respectively, and that the N latter packets do not have a value encoded in the field legal\_time\_window\_offset. Then the values t1(Ai+j) shall be determined by:



where j goes from 1 to N.

All packets between this packet and the next packet of the same PID to include a legal\_time\_window\_offset field shall be treated as if they had the value:



corresponding to the value t1(.) as computed by the formula above encoded in the legal\_time\_window\_offset field. t(j) is the arrival time of byte j in the T-STD.

The meaning of this field is not defined when it is present in a transport stream packet with no legal\_time\_window\_offset field.

**splice\_type** – This is a 4-bit field. From the first occurrence of this field onwards, it shall have the same value in all the subsequent transport stream packets of the same PID in which it is present, until the packet in which the splice\_countdown reaches zero (including this packet). If the elementary stream carried in that PID is not a Rec. ITU‑T H.262 | ISO/IEC 13818-2 video stream, then this field shall have the value '1111' (unspecified). If the elementary stream carried in that PID is a Rec. ITU-T H.262 | ISO/IEC 13818-2 video stream, then this field indicates the conditions that shall be respected by this elementary stream for splicing purposes. These conditions are defined as a function of profile, level and splice\_type in Table 2-7 through Table 2-20.

In these tables, a value for 'splice\_decoding\_delay' and 'max\_splice\_rate' means that the following conditions shall be satisfied by the video elementary stream:

1) The last byte of the coded picture ending in the transport stream packet in which the splice\_countdown reaches zero shall remain in the VBV buffer of the VBV model for an amount of time equal to (splice\_decoding\_delay tn+1 – tn), where for the purpose of this subclause:

• n is the index of the coded picture ending in the transport stream packet in which the splice\_countdown reaches zero, i.e., the coded picture referred to above.

• tn is defined in C.3.1 of Rec. ITU-T H.262 | ISO/IEC 13818-2.

• (tn+1 – tn) is defined in C.9 through C.12 of Rec. ITU-T H.262 | ISO/IEC 13818-2.

NOTE 4 – tn is the time when coded picture n is removed from the VBV buffer, and (tn+1 – tn) is the duration for which picture n is presented.

2) The VBV buffer of the VBV model shall not overflow if its input is switched at the splicing point to a stream of a constant rate equal to 'max\_splice\_rate' for an amount of time equal to 'splice\_decoding\_delay'.

Table 2-7 – Splice parameters Table 1  
Simple Profile Main Level, Main Profile Main Level,   
SNR Profile Main Level (both layers),  
Spatial Profile High-1440 Level (base layer),  
High Profile Main Level (middle + base layers),  
Multi-view Profile Main Level (base layer) Video

|  |  |
| --- | --- |
| splice\_type | Conditions |
| '0000' | splice\_decoding\_delay = 120 ms; max\_splice\_rate = 15.0 × 106 bit/s |
| '0001' | splice\_decoding\_delay = 150 ms; max\_splice\_rate = 12.0 × 106 bit/s |
| '0010' | splice\_decoding\_delay = 225 ms; max\_splice\_rate = 8.0 × 106 bit/s |
| '0011' | splice\_decoding\_delay = 250 ms; max\_splice\_rate = 7.2 × 106 bit/s |
| '0100' .. '1011' | Reserved |
| '1100' .. '1110' | User-defined |
| '1111' | unspecified |

Table 2-8 – Splice parameters Table 2  
Main Profile Low Level, SNR Profile Low Level (both layers),  
High Profile Main Level (base layer),   
Multi-view Profile Low Level (base layer) Video

|  |  |
| --- | --- |
| splice\_type | Conditions |
| '0000' | splice\_decoding\_delay = 115 ms; max\_splice\_rate = 4.0 × 106 bit/s |
| '0001' | splice\_decoding\_delay = 155 ms; max\_splice\_rate = 3.0 × 106 bit/s |
| '0010' | splice\_decoding\_delay = 230 ms; max\_splice\_rate = 2.0 × 106 bit/s |
| '0011' | splice\_decoding\_delay = 250 ms; max\_splice\_rate = 1.8 ×106 bit/s |
| '0100' .. '1011' | Reserved |
| '1100' .. '1110' | User-defined |
| '1111' | unspecified |

Table 2-9 – Splice parameters Table 3  
Main Profile High-1440 Level, Spatial Profile High-1440 Level (all layers),  
High Profile High-1440 Level (middle + base layers),   
Multi-view Profile High-1440 Level (base layer) Video

|  |  |
| --- | --- |
| splice\_type | Conditions |
| '0000' | splice\_decoding\_delay = 120 ms; max\_splice\_rate = 60.0 × 106 bit/s |
| '0001' | splice\_decoding\_delay = 160 ms; max\_splice\_rate = 45.0 × 106bit/s |
| '0010' | splice\_decoding\_delay = 240 ms; max\_splice\_rate = 30.0 × 106 bit/s |
| '0011' | splice\_decoding\_delay = 250 ms; max\_splice\_rate = 28.5 × 106 bit/s |
| '0100' .. '1011' | Reserved |
| '1100' .. '1110' | User-defined |
| '1111' | unspecified |

Table 2-10 – Splice parameters Table 4  
Main Profile High Level, High Profile High-1440 Level (all layers),  
High Profile High Level (middle + base layers),   
Multi-view Profile High Level (base layer) Video

|  |  |
| --- | --- |
| splice\_type | Conditions |
| '0000' | splice\_decoding\_delay = 120 ms; max\_splice\_rate = 80.0 × 106 bit/s |
| '0001' | splice\_decoding\_delay = 160 ms; max\_splice\_rate = 60.0 × 106 bit/s |
| '0010' | splice\_decoding\_delay = 240 ms; max\_splice\_rate = 40.0 × 106 bit/s |
| '0011' | splice\_decoding\_delay = 250 ms; max\_splice\_rate = 38.0 × 106 bit/s |
| '0100' .. '1011' | Reserved |
| '1100' .. '1110' | User-defined |
| '1111' | unspecified |

Table 2-11 – Splice parameters Table 5  
SNR Profile Low Level (base layer) Video

|  |  |
| --- | --- |
| splice\_type | Conditions |
| '0000' | splice\_decoding\_delay = 115 ms; max\_splice\_rate = 3.0 × 106 bit/s |
| '0001' | splice\_decoding\_delay = 175 ms; max\_splice\_rate = 2.0 × 106 bit/s |
| '0010' | splice\_decoding\_delay = 250 ms; max\_splice\_rate = 1.4 × 106 bit/s |
| '0011' .. '1011' | Reserved |
| '1100' .. '1110' | User-defined |
| '1111' | unspecified |

Table 2-12 – Splice parameters Table 6  
SNR Profile Main Level (base layer) Video

|  |  |
| --- | --- |
| splice\_type | Conditions |
| '0000' | splice\_decoding\_delay = 115 ms; max\_splice\_rate = 10.0 × 106 bit/s |
| '0001' | splice\_decoding\_delay = 145 ms; max\_splice\_rate = 8.0 × 106 bit/s |
| '0010' | splice\_decoding\_delay = 235 ms; max\_splice\_rate = 5.0 × 106 bit/s |
| '0011' | splice\_decoding\_delay = 250 ms; max\_splice\_rate = 4.7 × 106 bit/s |
| '0100' .. '1011' | Reserved |
| '1100' .. '1110' | User-defined |
| '1111' | unspecified |

Table 2-13 – Splice parameters Table 7  
Spatial Profile High-1440 Level (middle + base layers) Video

|  |  |
| --- | --- |
| splice\_type | Conditions |
| '0000' | splice\_decoding\_delay = 120 ms; max\_splice\_rate = 40.0 × 106 bit/s |
| '0001' | splice\_decoding\_delay = 160 ms; max\_splice\_rate = 30.0 × 106 bit/s |
| '0010' | splice\_decoding\_delay = 240 ms; max\_splice\_rate = 20.0 × 106 bit/s |
| '0011' | splice\_decoding\_delay = 250 ms; max\_splice\_rate = 19.0 × 106 bit/s |
| '0100' .. '1011' | Reserved |
| '1100' .. '1110' | User-defined |
| '1111' | unspecified |

Table 2-14 – Splice parameters Table 8  
High Profile Main Level (all layers), High Profile High-1440 Level (base layer) Video

|  |  |
| --- | --- |
| splice\_type | Conditions |
| '0000' | splice\_decoding\_delay = 120 ms; max\_splice\_rate = 20.0 × 106 bit/s |
| '0001' | splice\_decoding\_delay = 160 ms; max\_splice\_rate = 15.0 × 106 bit/s |
| '0010' | splice\_decoding\_delay = 240 ms; max\_splice\_rate = 10.0 × 106 bit/s |
| '0011' | splice\_decoding\_delay = 250 ms; max\_splice\_rate = 9.5 × 106 bit/s |
| '0100' .. '1011' | Reserved |
| '1100' .. '1110' | User-defined |
| '1111' | unspecified |

Table 2-15 – Splice parameters Table 9  
High Profile High Level (base layer),   
Multi-view Profile Main Level (both layers) Video

|  |  |
| --- | --- |
| splice\_type | Conditions |
| '0000' | splice\_decoding\_delay = 120 ms; max\_splice\_rate = 25.0 × 106 bit/s |
| '0001' | splice\_decoding\_delay = 165 ms; max\_splice\_rate = 18.0 × 106 bit/s |
| '0010' | splice\_decoding\_delay = 250 ms; max\_splice\_rate = 12.0 × 106 bit/s |
| '0011' .. '1011' | Reserved |
| '1100' .. '1110' | User-defined |
| '1111' | unspecified |

Table 2-16 – Splice parameters Table 10  
High Profile High Level (all layers),  
Multi-view Profile High-1440 Level (both layers) Video

|  |  |
| --- | --- |
| splice\_type | Conditions |
| '0000' | splice\_decoding\_delay = 120 ms; max\_splice\_rate = 100.0 × 106 bit/s |
| '0001' | splice\_decoding\_delay = 160 ms; max\_splice\_rate = 75.0 × 106 bit/s |
| '0010' | splice\_decoding\_delay = 240 ms; max\_splice\_rate = 50.0 × 106 bit/s |
| '0011' | splice\_decoding\_delay = 250 ms; max\_splice\_rate = 48.0 × 106 bit/s |
| '0100' .. '1011' | Reserved |
| '1100' .. '1110' | User-defined |
| '1111' | unspecified |

Table 2-17 – Splice parameters Table 11  
4:2:2 Profile Main Level Video

|  |  |
| --- | --- |
| splice\_type | Conditions |
| '0000' | splice\_decoding\_delay = 45 ms; max\_splice\_rate = 50.0 × 106 bit/s |
| '0001' | splice\_decoding\_delay = 90 ms; max\_splice\_rate = 50.0 × 106 bit/s |
| '0010' | splice\_decoding\_delay = 180 ms; max\_splice\_rate = 50.0 × 106 bit/s |
| '0011' | splice\_decoding\_delay = 225 ms; max\_splice\_rate = 40.0 × 106 bit/s |
| '0100' | splice\_decoding\_delay = 250 ms; max\_splice\_rate = 36.0 × 106 bit/s |
| '0101' .. '1011' | Reserved |
| '1100' .. '1110' | User-defined |
| '1111' | unspecified |

Table 2-18 – Splice parameters Table 12  
Multi-view Profile Low Level (both layers) Video

|  |  |
| --- | --- |
| splice\_type | Conditions |
| '0000' | splice\_decoding\_delay = 115 ms; max\_splice\_rate = 8.0 × 106 bit/s |
| '0001' | splice\_decoding\_delay = 155 ms; max\_splice\_rate = 6.0 × 106 bit/s |
| '0010' | splice\_decoding\_delay = 230 ms; max\_splice\_rate = 4.0 × 106 bit/s |
| '0011' | splice\_decoding\_delay = 250 ms; max\_splice\_rate = 3.7 × 106 bit/s |
| '0100' .. '1011' | Reserved |
| '1100' .. '1110' | User-defined |
| '1111' | unspecified |

Table 2-19 – Splice parameters Table 13  
Multi-view Profile High Level (both layers) Video

|  |  |
| --- | --- |
| splice\_type | Conditions |
| '0000' | splice\_decoding\_delay = 120 ms; max\_splice\_rate = 130.0 × 106 bit/s |
| '0001' | splice\_decoding\_delay = 150 ms; max\_splice\_rate = 104.0 × 106 bit/s |
| '0010' | splice\_decoding\_delay = 240 ms; max\_splice\_rate = 65.0 × 106 bit/s |
| '0011' | splice\_decoding\_delay = 250 ms; max\_splice\_rate = 62.4 × 106 bit/s |
| '0100' .. '1011' | Reserved |
| '1100' .. '1110' | User-defined |
| '1111' | unspecified |

Table 2-20 – Splice parameters Table 14  
4:2:2 Profile High Level Video

|  |  |
| --- | --- |
| splice\_type | Conditions |
| '0000' | splice\_decoding\_delay = 45 ms; max\_splice\_rate = 300.0 × 106 bit/s |
| '0001' | splice\_decoding\_delay = 90 ms; max\_splice\_rate = 300.0 × 106 bit/s |
| '0010' .. '0011' | Reserved |
| '0100' | splice\_decoding\_delay = 250 ms; max\_splice\_rate = 180.0 × 106 bit/s |
| '0101' .. '1011' | Reserved |
| '1100' .. '1110' | User-defined |
| '1111' | unspecified |

**DTS\_next\_AU** (decoding time stamp next access unit) – This is a 33-bit field, coded in three parts. In the case of continuous and periodic decoding through this splicing point it indicates the decoding time of the first access unit following the splicing point. This decoding time is expressed in the time base which is valid in the transport stream packet in which the splice\_countdown reaches zero. From the first occurrence of this field onwards, it shall have the same value in all the subsequent transport stream packets of the same PID in which it is present, until the packet in which the splice\_countdown reaches zero (including this packet).

The af\_descriptor() field may carry one or more descriptors as defined in Annex U. For descriptors carrying information associated with specific access units of an elementary stream, the descriptor applies to the first access unit that starts in the PES packet immediately following this adaptation field. There may be several TS packets carrying no payload before the start of the PES, in which case these descriptors apply to the next TS packet with payload on the same PID.

The adaptation field shall contain only complete af\_descriptor() descriptors, i.e., a single descriptor is always contained in a single transport stream packet.

NOTE 5 – The adaptation field should remain relatively small; it is therefore recommended for large descriptors to use PES carriage as defined in Annex U.

**stuffing\_byte** – This is a fixed 8-bit value equal to '1111 1111' that can be inserted by the encoder. It is discarded by the decoder.

#### 2.4.3.6 PES packet

See Table 2-21.

| Table 2-21 – PES packet | | | | |
| --- | --- | --- | --- | --- |
| Syntax | No. of bits | Mnemonic | |
| PES\_packet() { |  |  | |
| **packet\_start\_code\_prefix** | **24** | **bslbf** | |
| **stream\_id** | **8** | **uimsbf** | |
| **PES\_packet\_length** | **16** | **uimsbf** | |
| if (stream\_id != program\_stream\_map  && stream\_id != padding\_stream  && stream\_id != private\_stream\_2  && stream\_id != ECM  && stream\_id != EMM  && stream\_id != program\_stream\_directory  && stream\_id != DSMCC\_stream  && stream\_id != ITU-T Rec. H.222.1 type E stream) { |  |  | |
| '10' | **2** | **bslbf** | |
| **PES\_scrambling\_control** | **2** | **bslbf** | |
| **PES\_priority** | **1** | **bslbf** | |
| **data\_alignment\_indicator** | **1** | **bslbf** | |
| **copyright** | **1** | **bslbf** | |
| **original\_or\_copy** | **1** | **bslbf** | |
| **PTS\_DTS\_flags** | **2** | **bslbf** | |
| **ESCR\_flag** | **1** | **bslbf** | |
| **ES\_rate\_flag** | **1** | **bslbf** | |
| **DSM\_trick\_mode\_flag** | **1** | **bslbf** | |
| **additional\_copy\_info\_flag** | **1** | **bslbf** | |
| **PES\_CRC\_flag** | **1** | **bslbf** | |
| **PES\_extension\_flag** | **1** | **bslbf** | |
| **PES\_header\_data\_length** | **8** | **uimsbf** | |
| if (PTS\_DTS\_flags == '10') { |  |  | |
| **'0010'** | **4** | **bslbf** | |
| **PTS [32..30]** | | **3** | **bslbf** |
| **marker\_bit** | **1** | **bslbf** | |
| **PTS [29..15]** | **15** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| **PTS [14..0]** | **15** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| } |  |  | |
| if (PTS\_DTS\_flags == '11') { |  |  | |
| **'0011'** | **4** | **bslbf** | |
| **PTS [32..30]** | **3** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| **PTS [29..15]** | **15** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| **PTS [14..0]** | **15** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| '0001' | **4** | **bslbf** | |
| **DTS [32..30]** | **3** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| **DTS [29..15]** | **15** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| **DTS [14..0]** | **15** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| } |  |  | |
| if (ESCR\_flag == '1') { |  |  | |
| **reserved** | **2** | **bslbf** | |
| **ESCR\_base[32..30]** | **3** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| **ESCR\_base[29..15]** | **15** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| **ESCR\_base[14..0]** | **15** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| **ESCR\_extension** | **9** | **uimsbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| } |  |  | |
| if (ES\_rate\_flag == '1') { |  |  | |
| **marker\_bit** | **1** | **bslbf** | |
| **ES\_rate** | **22** | **uimsbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| } |  |  | |
| if (DSM\_trick\_mode\_flag == '1') { |  |  | |
| **trick\_mode\_control** | **3** | **uimsbf** | |
| if ( trick\_mode\_control == fast\_forward ) { |  |  | |
| **field\_id** | **2** | **bslbf** | |
| **intra\_slice\_refresh** | **1** | **bslbf** | |
| **frequency\_truncation** | **2** | **bslbf** | |
| } |  |  | |
| else if ( trick\_mode\_control == slow\_motion ) { |  |  | |
| **rep\_cntrl** | **5** | **uimsbf** | |
| } |  |  | |
| else if ( trick\_mode\_control == freeze\_frame ) { |  |  | |
| **field\_id** | **2** | **uimsbf** | |
| **reserved** | **3** | **bslbf** | |
| } |  |  | |
| else if ( trick\_mode\_control == fast\_reverse ) { |  |  | |
| **field\_id** | **2** | **bslbf** | |
| **intra\_slice\_refresh** | **1** | **bslbf** | |
| **frequency\_truncation** | **2** | **bslbf** | |
| else if ( trick\_mode\_control == slow\_reverse ) { |  |  | |
| **rep\_cntrl** | **5** | **uimsbf** | |
| } |  |  | |
| else |  |  | |
| **reserved** | **5** | **bslbf** | |
| } |  |  | |
| if ( additional\_copy\_info\_flag == '1') { |  |  | |
| **marker\_bit** | **1** | **bslbf** | |
| **additional\_copy\_info** | **7** | **bslbf** | |
| } |  |  | |
| if ( PES\_CRC\_flag == '1') { |  |  | |
| **previous\_PES\_packet\_CRC** | **16** | **bslbf** | |
| } |  |  | |
| if ( PES\_extension\_flag == '1') { |  |  | |
| **PES\_private\_data\_flag** | **1** | **bslbf** | |
| **pack\_header\_field\_flag** | **1** | **bslbf** | |
| **program\_packet\_sequence\_counter\_flag** | **1** | **bslbf** | |
| **P-STD\_buffer\_flag** | **1** | **bslbf** | |
| **reserved** | **3** | **bslbf** | |
| **PES\_extension\_flag\_2** | **1** | **bslbf** | |
| if ( PES\_private\_data\_flag == '1') { |  |  | |
| **PES\_private\_data** | **128** | **bslbf** | |
| } |  |  | |
| if (pack\_header\_field\_flag == '1') { |  |  | |
| **pack\_field\_length** | **8** | **uimsbf** | |
| pack\_header() |  |  | |
| } |  |  | |
| if (program\_packet\_sequence\_counter\_flag == '1') { |  |  | |
| **marker\_bit** | **1** | **bslbf** | |
| **program\_packet\_sequence\_counter** | **7** | **uimsbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| **MPEG1\_MPEG2\_identifier** | **1** | **bslbf** | |
| **original\_stuff\_length** | **6** | **uimsbf** | |
| } |  |  | |
| if ( P-STD\_buffer\_flag == '1') { |  |  | |
| **'01'** | **2** | **bslbf** | |
| **P-STD\_buffer\_scale** | **1** | **bslbf** | |
| **P-STD\_buffer\_size** | **13** | **uimsbf** | |
| } |  |  | |
| if ( PES\_extension\_flag\_2 == '1') { |  |  | |
| **marker\_bit** | **1** | **bslbf** | |
| **PES\_extension\_field\_length** | **7** | **uimsbf** | |
| **stream\_id\_extension\_flag** | **1** | **bslbf** | |
| if ( stream\_id\_extension\_flag == '0') { |  |  | |
| **stream\_id\_extension** | **7** | **uimsbf** | |
| **}** |  |  | |
| else { |  |  | |
| **reserved** | **6** | **bslbf** | |
| **tref\_extension\_flag** | **1** | **bslbf** | |
| if ( tref\_extension\_flag   '0' ) { |  |  | |
| **reserved** | **4** | **bslbf** | |
| **TREF[32..30]** | **3** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| **TREF[29..15]** | **15** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| **TREF[14..0]** | **15** | **bslbf** | |
| **marker\_bit** | **1** | **bslbf** | |
| **}** |  |  | |
| **}** |  |  | |
| for ( i  0; i  N3; i) { |  |  | |
| **reserved** | **8** | **bslbf** | |
| **}** |  |  | |
| } |  |  | |
| } |  |  | |
| for (i < 0; i < N1; i++) { |  |  | |
| **stuffing\_byte** | **8** | **bslbf** | |
| } |  |  | |
| for (i < 0; i < N2; i++) { |  |  | |
| **PES\_packet\_data\_byte** | **8** | **bslbf** | |
| } |  |  | |
| } |  |  | |
| else if ( stream\_id == program\_stream\_map  || stream\_id == private\_stream\_2  || stream\_id == ECM  || stream\_id == EMM  || stream\_id == program\_stream\_directory  || stream\_id == DSMCC\_stream  || stream\_id == ITU-T Rec. H.222.1 type E stream ) { |  |  | |
| for (i = 0; i < PES\_packet\_length; i++) { |  |  | |
| **PES\_packet\_data\_byte** | **8** | **bslbf** | |
| } |  |  | |
| } |  |  | |
| else if ( stream\_id == padding\_stream) { |  |  | |
| for (i < 0; i < PES\_packet\_length; i++) { |  |  | |
| **padding\_byte** | **8** | **bslbf** | |
| } |  |  | |
| } |  |  | |
| } |  |  | |
| NOTE – The value N3 equals the byte count given by PES\_extension\_field\_length minus the bytes which contain the stream\_id\_extension\_flag and the occurring data elements in the if-else-construct after the stream\_id\_extension\_flag. | | | |

#### 2.4.3.7 Semantic definition of fields in PES packet

**packet\_start\_code\_prefix** – The packet\_start\_code\_prefix is a 24-bit code. Together with the stream\_id that follows it constitutes a packet start code that identifies the beginning of a packet. The packet\_start\_code\_prefixis the bit string '0000 0000 0000 0000 0000 0001' (0x000001).

**stream\_id** – In program streams, the stream\_id specifies the type and number of the elementary stream as defined by the stream\_id Table 2-22. In transport streams, the stream\_id may be set to any valid value which correctly describes the elementary stream type as defined in Table 2-22. In transport streams, the elementary stream type is specified in the program-specific information as specified in 2.4.4. For AVC video streams conforming to one or more profiles defined in Annex G of Rec. ITU-T H.264 | ISO/IEC 14496-10, all video sub-bitstreams of the same AVC video stream shall have the same stream\_id value. For AVC video streams conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496‑10, all MVC video sub-bitstreams of the same AVC video stream shall have the same stream\_id value. For AVC video streams conforming to one or more profiles defined in Annex I of Rec. ITU‑T H.264 | ISO/IEC 14496‑10, all MVCD video sub-bitstreams of the same AVC video stream shall have the same stream\_id value.

**PES\_packet\_length** –A 16-bit field specifying the number of bytes in the PES packet following the last byte of the field. A value of 0 indicates that the PES packet length is neither specified nor bounded and is allowed only in PES packets whose payload consists of bytes from a video elementary stream contained in transport stream packets.

**PES\_scrambling\_control** –The 2-bit PES\_scrambling\_control field indicates the scrambling mode of the PES packet payload. When scrambling is performed at the PES level, the PES packet header, including the optional fields when present, shall not be scrambled (see Table 2-23).

| Table 2-22 – Stream\_id assignments | | |
| --- | --- | --- |
| Stream\_id | Note | stream coding |
| '1011 1100' | 1 | program\_stream\_map |
| '1011 1101' | 2,9,10,11 | private\_stream\_1 |
| '1011 1110' |  | padding\_stream |
| '1011 1111' | 3 | private\_stream\_2 |
| '110x xxxx' |  | ISO/IEC 13818-3 or ISO/IEC 11172-3 or ISO/IEC 13818-7 or ISO/IEC 14496-3 or ISO/IEC 23008-3 audio stream number 'x xxxx' |
| '1110 xxxx' |  | Rec. ITU-T H.262 | ISO/IEC 13818-2, ISO/IEC 11172-2, ISO/IEC 14496-2, Rec. ITU-T H.264 | ISO/IEC 14496-10, Rec. ITU-T H.265 | ISO/IEC 23008-2, Rec. ITU-T H.266 | ISO/IEC 23090-3, ISO/IEC 23094-1 or ISO/IEC 23094-2 video stream number xxxx |
| '1111 0000' | 3 | ECM\_stream |
| '1111 0001' | 3 | EMM\_stream |
| '1111 0010' | 5 | Rec. ITU-T H.222.0 | ISO/IEC 13818-1 Annex A or ISO/IEC 13818‑6\_DSMCC\_stream |
| '1111 0011' | 2 | ISO/IEC\_13522\_stream |
| '1111 0100' | 6 | Rec. ITU-T H.222.1 type A |
| '1111 0101' | 6 | Rec. ITU-T H.222.1 type B |
| '1111 0110' | 6 | Rec. ITU-T H.222.1 type C |
| '1111 0111' | 6 | Rec. ITU-T H.222.1 type D |
| '1111 1000' | 6 | Rec. ITU-T H.222.1 type E |
| '1111 1001' | 7 | ancillary\_stream |
| '1111 1010' |  | ISO/IEC 14496-1\_SL-packetized\_stream |
| '1111 1011' |  | ISO/IEC 14496-1\_M4Mux\_stream |
| '1111 1100' |  | metadata stream |
| '1111 1101' | 8 | extended\_stream\_id |
| '1111 1110' |  | reserved data stream |
| '1111 1111' | 4 | program\_stream\_directory |
| The notation x means that the values '0' or '1' are both permitted and results in the same stream type. The stream number is given by the values taken by the x's.  NOTE 1 – PES packets of type program\_stream\_map have unique syntax specified in 2.5.4.1.  NOTE 2 – PES packets of type private\_stream\_1 and ISO/IEC\_13552\_stream follow the same PES packet syntax as those for Rec. ITU-T H.262 | ISO/IEC 13818-2 video and ISO/IEC 13818-3 audio streams.  NOTE 3 – PES packets of type private\_stream\_2, ECM\_stream and EMM\_stream are similar to private\_stream\_1 except that no syntax is specified after PES\_packet\_length field.  NOTE 4 – PES packets of type program\_stream\_directory have a unique syntax specified in 2.5.5.  NOTE 5 – PES packets of type DSM-CC\_stream have a unique syntax specified in ISO/IEC 13818-6.  NOTE 6 – This stream\_id is associated with stream\_type 0x09 in Table 2-34.  NOTE 7 – This stream\_id is only used in PES packets, which carry data from a program stream or an ISO/IEC 11172-1 System Stream, in a transport stream (refer to 2.4.3.8).  NOTE 8 – The use of stream\_id 0xFD (extended\_stream\_id) identifies that this PES packet employs an extended syntax to permit additional stream types to be identified.  NOTE 9 – JPEG 2000 video streams (stream\_type = 0x21) are carried using the same PES packet syntax as private\_stream\_1.  NOTE 10 – Timeline and External Media Information streams (stream\_type = 0x27) are carried using the same PES packet syntax as private\_stream\_1.  NOTE 11 – JPEG XS video streams (stream\_type = 0x32) are carried using the same PES packet syntax as private\_stream\_1. | | |

Table 2-23 – PES scrambling control values

|  |  |
| --- | --- |
| Value | Description |
| '00' | Not scrambled |
| '01' | User-defined |
| '10' | User-defined |
| '11' | User-defined |

**PES\_priority** –This is a 1-bit field indicating the priority of the payload in this PES packet. A '1' indicates a higher priority of the payload of the PES packet payload than a PES packet payload with this field set to '0'. A multiplexor can use the PES\_priority bit to prioritize its data within an elementary stream. This field shall not be changed by the transport mechanism.

**data\_alignment\_indicator** – This is a 1-bit flag. When set to a value of '1', it indicates that the PES packet header is immediately followed by the video syntax element or audio sync word indicated in the data\_stream\_alignment\_descriptor in 2.6.10 if this descriptor is present. If set to a value of '1' and the descriptor is not present, alignment as indicated in alignment\_type '01' in Table 2-53, Table 2-54 or Table 2-55 is required. When set to a value of '0', it is not defined whether any such alignment occurs or not.

**copyright** –This is a 1-bit field. When set to '1' it indicates that the material of the associated PES packet payload is protected by copyright. When set to '0' it is not defined whether the material is protected by copyright. A copyright descriptor described in 2.6.24 is associated with the elementary stream which contains this PES packet and the copyright flag is set to '1' if the descriptor applies to the material contained in this PES packet.

**original\_or\_copy** – This is a 1-bit field. When set to '1' the contents of the associated PES packet payload is an original. When set to '0' it indicates that the contents of the associated PES packet payload are a copy.

**PTS\_DTS\_flags** –This is a 2-bit field. When the PTS\_DTS\_flags field is set to '10', the PTS fields shall be present in the PES packet header. When the PTS\_DTS\_flags field is set to '11', both the PTS fields and DTS fields shall be present in the PES packet header. When the PTS\_DTS\_flags field is set to '00' no PTS or DTS fields shall be present in the PES packet header. The value '01' is forbidden.

**ESCR\_flag** – A 1-bit flag, which when set to '1' indicates that ESCR base and extension fields are present in the PES packet header. When set to '0' it indicates that no ESCR fields are present.

**ES\_rate\_flag** –A 1-bit flag, which when set to '1' indicates that the ES\_rate field is present in the PES packet header. When set to '0' it indicates that no ES\_rate field is present.

**DSM\_trick\_mode\_flag** – A 1-bit flag, which when set to '1' it indicates the presence of an 8-bit trick mode field. When set to '0' it indicates that this field is not present.

**additional\_copy\_info\_flag** – A 1-bit flag, which when set to '1' indicates the presence of the additional\_copy\_info field. When set to '0' it indicates that this field is not present.

**PES\_CRC\_flag** –A 1-bit flag, which when set to '1' indicates that a CRC field is present in the PES packet. When set to '0' it indicates that this field is not present.

**PES\_extension\_flag** – A 1-bit flag, which when set to '1' indicates that an extension field exists in this PES packet header. When set to '0' it indicates that this field is not present.

**PES\_header\_data\_length** –An 8-bit field specifying the total number of bytes occupied by the optional fields and any stuffing bytes contained in this PES packet header. The presence of optional fields is indicated in the byte that precedes the PES\_header\_data\_length field.

**marker\_bit** – A marker\_bit is a 1-bit field that has the value '1'.

**PTS (presentation time stamp)** – Presentation times shall be related to decoding times as follows: The PTS is a 33-bit number coded in three separate fields. It indicates the time of presentation, tpn(k), in the system target decoder of a presentation unit k of elementary stream n. The value of PTS is specified in units of the period of the system clock frequency divided by 300 (yielding 90 kHz). The presentation time is derived from the PTS according to equation 2-11 below. Refer to 2.7.4 for constraints on the frequency of coding presentation timestamps.

 (2-11)

where tpn(k) is the presentation time of presentation unit Pn(k).

In the case of audio, if a PTS is present in PES packet header it shall refer to the first access unit commencing in the PES packet. An audio access unit commences in a PES packet if the first byte of the audio access unit is present in the PES packet.

In the case of ISO/IEC 11172-2 video or ISO/IEC 14496-2 video, if a PTS is present in a PES packet header, it shall refer to the access unit containing the first picture start code that commences in this PES packet. A picture start code commences in a PES packet if the first byte of the picture start code is present in the PES packet. For I- and P-pictures in non-low\_delay sequences and in the case when there is no decoding discontinuity between access units (AUs) k and k', the presentation time tpn(k) shall be equal to the decoding time tdn(k') of the next transmitted I- or P-picture (refer to 2.7.5). If there is a decoding discontinuity, or the stream ends, the difference between tpn(k) and tdn(k) shall be the same as if the original stream had continued without a discontinuity and without ending.

NOTE 1 – A low\_delay sequence is an ISO/IEC 14496-2 video sequence in which the low\_delay flag is set to '1' (refer to 6.2.3 of ISO/IEC 14496-2).

For Rec. ITU-T H.262 | ISO/IEC 13818-2 video, if a PTS is present in a PES packet header, it shall refer to the access unit containing the first picture start code that commences in this PES packet. A picture start code commences in a PES packet if the first byte of the picture start code is present in the PES packet. For I- and P-coded frames in non-low\_delay sequences and in the case when there is no decoding discontinuity between access units (AUs) k and k', the presentation time tpn(k) shall be equal to the decoding time tdn(k') of the next transmitted I- or P-coded frame (refer to 2.7.5). If there is a decoding discontinuity, or the stream ends, the difference between tpn(k) and tdn(k) shall be the same as if the original stream had continued without a discontinuity and without ending.

NOTE 2 – A low\_delay sequence is an Rec. ITU-T H.262 | ISO/IEC 13818-2 video sequence in which the low\_delay flag is set to '1' (refer to 6.2.2.3 of Rec. ITU-T H.262 | ISO/IEC 13818-2). Also note that for field pictures the presentation time refers to the first field picture of the coded frame.

For AVC video streams conforming to one or more profiles defined in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496‑10 video, if a PTS is present in the PES packet header, it shall refer to the first AVC access unit that commences in this PES packet. An AVC access unit commences in a PES packet if the first byte of the AVC access unit is present in the PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496-10, for each decoded AVC access unit, the PTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal DPB output time in the HRD, defined herein as to,n,dpb(n) = tr,n(n) + tc \* dpb\_output\_delay(n), where tr,n(n), tc, and dpb\_output\_delay(n) are defined as in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496-10.

For video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex G of Rec. ITU‑T H.264 | ISO/IEC 14496-10, if a PTS is present in the PES packet header, it shall refer to the first SVC dependency representation that commences in this PES packet. An SVC dependency representation commences in a PES packet if the first byte of the SVC dependency representation is present in the PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496-10, for each re‑assembled and decoded AVC access unit, the PTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal DPB output time in the HRD, defined herein as to,n,dpb(n) = tr,n(n) + tc \* dpb\_output\_delay(n), where tr,n(n), tc, and dpb\_output\_delay(n) are defined as in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496-10.

NOTE 3 – Different clocks may be used for derivation of PTS and to,n,dpb(n).

For MVC video sub-bitstreams, MVC base view sub-bitstream or AVC video sub-bitstream of MVC of AVC video streams conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10, if a PTS is present in the PES packet header, it shall refer to the first MVC view-component subset that commences in this PES packet. An MVC view-component subset commences in a PES packet if the first byte of the MVC view-component subset is present in the PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496-10, for each re-assembled and decoded AVC access unit, the PTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal DPB output time in the HRD, defined herein as to,n,dpb(n) = tr,n(n) + tc \* dpb\_output\_delay(n), where tr,n(n), tc, and dpb\_output\_delay(n) are defined as in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496-10.

In the case of an ISO/IEC 14496-17 text stream, if a PTS is present in PES packet header, it shall refer to the first text access unit commencing in the PES packet. A text access unit commences in a PES packet if the first byte of the text access unit is present in the PES packet.

For MVCD video sub-bitstreams, MVCD base view sub-bitstream or AVC video sub-bitstream of MVCD of AVC video streams conforming to one or more profiles defined in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496-10, if a PTS is present in the PES packet header, it shall refer to the first MVCD view-component subset that commences in this PES packet. An MVCD view-component subset commences in a PES packet if the first byte of the MVCD view-component subset is present in the PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496-10, for each re-assembled and decoded AVC access unit, the PTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal DPB output time in the HRD, defined herein as to,n,dpb(n) = tr,n(n) + tc \* dpb\_output\_delay(n), where tr,n(n), tc, and dpb\_output\_delay(n) are defined as in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496-10.

The presentation time tpn(k) shall be equal to the decoding time tdn(k) for:

• audio access units;

• access units in Rec. ITU-T H.262 | ISO/IEC 13818-2 or ISO/IEC 14496-2 low delay video sequences;

• B-pictures in ISO/IEC 11172-2, Rec. ITU-T H.262 | ISO/IEC 13818-2 or ISO/IEC 14496-2 video streams.

• text access units in ISO/IEC 14496-17.

If there is filtering in audio, it is assumed by the system model that filtering introduces no delay, hence the sample referred to by PTS at encoding is the same sample referred to by PTS at decoding. In the case of scalable coding refer to 2.7.6.

For HEVC video streams, HEVC temporal video sub-bitstreams and HEVC temporal video subsets, if a PTS is present in the PES packet header, it shall refer to the first HEVC access unit that commences in this PES packet. For HEVC video sub-partitions, if a PTS is present in the PES packet header, it shall refer to the first HEVC layer component that commences in this PES packet. An HEVC layer component commences in a PES packet if the first byte of the HEVC layer component is present in the PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of Rec. ITU-T H.265 | ISO/IEC 23008‑2, for each HEVC access unit the PTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal DPB output time in the HRD, as defined in Annex C of Rec. ITU-T H.265 | ISO/IEC 23008‑2.

For VVC video streams, VVC temporal video sub-bitstreams and VVC temporal video subsets, if a PTS is present in the PES packet header, it shall refer to the first VVC access unit that commences in this PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of Rec. ITU-T H.266 | ISO/IEC 23090-3, for each VVC access unit the PTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal DPB output time in the HRD, as defined in Annex C of Rec. ITU-T H.266 | ISO/IEC 23090-3.

For EVC video streams and EVC temporal video sub-bitstreams, if a PTS is present in the PES packet header, it shall refer to the first and only EVC access unit that commences in this PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of ISO/IEC 23094-1, for each EVC access unit the PTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal DPB output time in the HRD, as defined in Annex C of ISO/IEC 23094-1.

For LCEVC video streams, the PTS shall be present in the PES packet header, and it shall refer to only one LCEVC access unit that commences in this PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of ISO/IEC 23094-2, for each LCEVC access unit the PTS value in the STD shall, within its accuracy, indicate the same instant in time as the nominal DPB output time in the HRD, as defined in Annex C of ISO/IEC 23094‑2.

**DTS (decoding time stamp)** – The DTS is a 33-bit number coded in three separate fields. It indicates the decoding time, tdn(j), in the system target decoder of an access unit j of elementary stream n. The value of DTS is specified in units of the period of the system clock frequency divided by 300 (yielding 90 kHz). The decoding time derived from the DTS according to equation 2-12 below:

(2-12)

where tdn(j) is the decoding time of access unit An(j).

In the case of ISO/IEC 11172-2 video, Rec. ITU-T H.262 | ISO/IEC 13818-2 video, or ISO/IEC 14496-2 video, if a DTS is present in a PES packet header, it shall refer to the access unit containing the first picture start code that commences in this PES packet. A picture start code commences in a PES packet if the first byte of the picture start code is present in the PES packet.

For AVC video streams conforming to one or more profiles defined in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496‑10 video, if a DTS is present in the PES packet header, it shall refer to the first AVC access unit that commences in this PES packet. An AVC access unit commences in a PES packet if the first byte of the AVC access unit is present in the PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of Rec. ITU‑T H.264 | ISO/IEC 14496-10, for each AVC access unit the DTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal CPB removal time tr,n( n ) in the HRD, as defined in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496-10.

For video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex G of Rec. ITU‑T H.264 | ISO/IEC 14496-10, if a DTS is present in the PES packet header, it shall refer to the first SVC dependency representation that commences in this PES packet. An SVC dependency representation commences in a PES packet if the first byte of the SVC dependency representation is present in the PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496-10, for each re‑assembled AVC access unit the DTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal CPB removal time tr,n( n ) in the HRD, as defined in Annex C of Rec. ITU‑T H.264 | ISO/IEC 14496-10.

For MVC video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10, if a DTS is present in the PES packet header, it shall refer to the first MVC view‑component subset that commences in this PES packet. An MVC view-component subset commences in a PES packet if the first byte of the MVC view-component subset is present in the PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496-10, for each re‑assembled AVC access unit the DTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal CPB removal time tr,n( n ) in the HRD, as defined in Annex C of Rec. ITU‑T H.264 | ISO/IEC 14496-10.

NOTE 4 – Different clocks may be used for derivation of DTS and tr,n( n ).

In the case of scalable coding refer to 2.7.6.

For HEVC video streams, HEVC temporal video sub-bitstreams and HEVC temporal video subsets, if a DTS is present in the PES packet header, it shall refer to the first HEVC access unit that commences in this PES packet. For HEVC video sub-partitions, if a DTS is present in the PES packet header, it shall refer to the first HEVC layer component that commences in this PES packet. An HEVC layer component commences in a PES packet if the first byte of the HEVC layer component is present in the PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of Rec. ITU-T H.265 | ISO/IEC 23008‑2, for each HEVC access unit the DTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal CPB removal time in the HRD, as defined in Annex C of Rec. ITU-T H.265 | ISO/IEC 23008‑2.

For MVCD video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496-10, if a DTS is present in the PES packet header, it shall refer to the first MVCD view‑component subset that commences in this PES packet. An MVCD view-component subset commences in a PES packet if the first byte of the MVCD view-component subset is present in the PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496-10, for each re-assembled AVC access unit the DTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal CPB removal time tr,n( n ) in the HRD, as defined in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496-10.

For VVC video streams, VVC temporal video sub-bitstreams and VVC temporal video subsets, if a DTS is present in the PES packet header, it shall refer to the first VVC access unit that commences in this PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of Rec. ITU-T H.266 | ISO/IEC 23090-3, for each VVC access unit the DTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal CPB removal time in the HRD, as specified in Annex C of Rec. ITU-T H.266 | ISO/IEC 23090-3.

For EVC video streams and EVC temporal video sub-bitstreams, if a DTS is present in the PES packet header, it shall refer to the first and only EVC access unit that commences in this PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of ISO/IEC 23094-1, for the n-th EVC access unit the DTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal CPB removal time tr,n(n) in the HRD, as defined in Annex C of Rec. ISO/IEC 23094-1.

For LCEVC video streams, the DTS shall not be present in the PES packet header because the LCEVC decoding process is strictly in presentation order. This restriction on the strict coincidence of decoding order and presentation order of the access units refers exclusively to the LCEVC video stream. On the other hand, the base video stream that the LCEVC video stream enhances has no such restriction, which means that in this case the decoding order and the presentation order can be different.

**ESCR\_base; ESCR\_extension** – The elementary stream clock reference is a 42-bit field coded in two parts. The first part, ESCR\_base, is a 33-bit field whose value is given by ESCR\_base(i), as given in equation 2-14. The second part, ESCR\_ext, is a 9-bit field whose value is given by ESCR\_ext(i), as given in equation 2-15. The ESCR field indicates the intended time of arrival of the byte containing the last bit of the ESCR\_base at the input of the PES-STD for PES streams (refer to 2.5.2.4).

Specifically:

(2-13)

where:

(2-14)

(2-15)

The ESCR and ES\_rate field (refer to semantics immediately following) contain timing information relating to the sequence of PES streams. These fields shall satisfy the constraints defined in 2.7.3.

**ES\_rate (elementary stream rate)** –The ES\_rate field is a 22-bit unsigned integer specifying the rate at which the system target decoder receives bytes of the PES packet in the case of a PES stream. The ES\_rate is valid in the PES packet in which it is included and in subsequent PES packets of the same PES stream until a new ES\_rate field is encountered. The value of the ES\_rate is measured in units of 50 bytes/second. The value '0' is forbidden. The value of the ES\_rate is used to define the time of arrival of bytes at the input of a P-STD for PES streams defined in 2.5.2.4. The value encoded in the ES\_rate field may vary from PES\_packet to PES\_packet.

**trick\_mode\_control** – A 3-bit field that indicates which trick mode is applied to the associated video stream (see Table 2-24). In cases of other types of elementary streams, the meanings of this field and those defined by the followingfive bits are undefined. For the definition of trick\_mode status, refer to the trick mode section of 2.4.2.4.

When trick\_mode status is false, the number of times N, a picture is output by the decoding process for progressive sequences, is specified for each picture by the repeat\_first\_field and top\_field\_first fields in the case of Rec. ITU‑T H.262 | ISO/IEC 13818-2 video, and is specified through the sequence header in the case of ISO/IEC 11172‑2 Video.

For interlaced sequences, when trick\_mode status is false, the number of times N, a picture is output by the decoding process for progressive sequences, is specified for each picture by the repeat\_first\_field and progressive\_frame fields in the case of Rec. ITU-T H.262 | ISO/IEC 13818-2 video.

When trick mode status is true, the number of times that a picture shall be displayed depends on the value of N.

When the value of this field changes or trick mode operations cease, any combination of the following may occur:

• discontinuity in the time base;

• decoding discontinuity;

• continuity counter discontinuity.

Table 2-24 – Trick mode control values

|  |  |
| --- | --- |
| Value | Description |
| '000' | Fast forward |
| '001' | Slow motion |
| '010' | Freeze frame |
| '011' | Fast reverse |
| '100' | Slow reverse |
| '101' .. '111' | Reserved |

In the context of trick mode, the non-normal speed of decoding and presentation may cause the values of certain fields defined in video elementary stream data to be incorrect. Likewise, the semantic constraint on the slice structure may be invalid. The video syntax elements to which this exception applies are:

• bit\_rate;

• vbv\_delay;

• repeat\_first\_field;

• v\_axis\_positive;

• field\_sequence;

• subcarrier;

• burst\_amplitude;

• subcarrier\_phase.

A decoder cannot rely on the values encoded in these fields when in trick mode.

Decoders are not normatively required to decode the trick\_mode\_control field. However, the following normative requirements shall apply to decoders that do decode the trick\_mode\_control field.

**fast forward** – The value '000', in the trick\_mode\_control field. When this value is present it indicates a fast forward video stream and defines the meaning of the following five bits in the PES packet header. The intra\_slice\_refresh bit may be set to '1' indicating that there may be missing macroblocks which the decoder may replace with co-sited macroblocks of previously decoded pictures. The field\_id field, defined in Table 2-25, indicates which field or fields should be displayed. The frequency\_truncation field indicates that a restricted set of coefficients may be included. The meaning of the values of this field are shown in Table 2-26.

**slow motion** – The value '001', in the trick\_mode\_control field. When this value is present it indicates a slow motion video stream and defines the meaning of the following five bits in the PES packet header. In the case of progressive sequences, the picture should be displayed N × rep\_cntrl times, where N is defined above.

In the case of ISO/IEC 11172-2 video and Rec. ITU-T H.262 | ISO/IEC 13818-2 video progressive sequences, the picture should be displayed for N × rep\_cntrl picture duration.

In the case of Rec. ITU-T H.262 | ISO/IEC 13818-2 interlaced sequences, the picture should be displayed for N × rep\_cntrl field duration. If the picture is a frame picture, the first field to be displayed is the top field if top\_field\_first is 1, and the bottom field if top\_field\_first is '0' (refer to Rec. ITU-T H.262 | ISO/IEC 13818-2). This field is displayed for N × rep\_cntrl / 2 field duration. The other field of the picture is then displayed for N − N × rep\_cntrl / 2 field duration.

**freeze frame** – The value '010', in the trick\_mode\_control field. When this value is present it indicates a freeze frame video stream and defines the meaning of the following five bits in the PES packet header. The field\_id field, defined in Table 2-25, identifies which field(s) should be displayed. The field\_id field refers to the first video access unit that commences in the PES packet which contains the field\_id field, unless the PES packet contains zero payload bytes. In the latter case the field\_id field refers to the most recent previous video access unit.

**fast reverse** – The value '011', in the trick\_mode\_control field. When this value is present it indicates a fast reverse video stream and defines the meaning of the following five bits in the PES packet header. The intra\_slice\_refresh bit may be set to '1' indicating that there may be missing macroblocks which the decoder may replace with co-sited macroblocks of previously decoded pictures. The field\_id field, defined in Table 2-25, indicates which field or fields should be displayed. The frequency\_truncation field indicates that a restricted set of coefficients may be included. The meaning of the values of this field are shown in Table 2-26.

**slow reverse** – The value '100', in the trick\_mode\_control field. When this value is present it indicates a slow reverse video stream and defines the meaning of the following five bits in the PES packet header. In the case of ISO/IEC 11172‑2 video and Rec. ITU-T H.262 | ISO/IEC 13818-2 video progressive sequences, the picture should be displayed for N × rep\_cntrl picture duration, where N is defined above.

In the case of Rec. ITU-T H.262 | ISO/IEC 13818-2 interlaced sequences, the picture should be displayed for N × rep\_cntrl field duration. If the picture is a frame picture, the first field to be displayed is the bottom field if top\_field\_first is 1, and the top field if top\_field\_first is '0' (refer to Rec. ITU-T H.262 | ISO/IEC 13818-2). This field is displayed for N × rep\_cntrl / 2 field duration. The other field of the picture is then displayed for N – N × rep\_cntrl / 2 field duration.

**field\_id** – A 2-bit field that indicates which field(s) should be displayed. It is coded according to Table 2-25.

Table 2-25 – Field\_id field control values

|  |  |
| --- | --- |
| Value | Description |
| '00' | Display from top field only |
| '01' | Display from bottom field only |
| '10' | Display complete frame |
| '11' | Reserved |

**intra\_slice\_refresh** – A 1-bit flag, which when set to '1', indicates that there may be missing macroblocks between coded slices of video data in this PES packet. When set to '0' this may not occur. For more information, see Rec. ITU‑T H.262 | ISO/IEC 13818-2. The decoder may replace missing macroblocks with co-sited macroblocks of previously decoded pictures.

**frequency\_truncation** – A 2-bit field which indicates that a restricted set of coefficients may have been used in coding the video data in this PES packet. The values are defined in Table 2-26.

Table 2-26 – Coefficient selection values

|  |  |
| --- | --- |
| Value | Description |
| '00' | Only DC coefficients are non-zero |
| '01' | Only the first three coefficients are non-zero |
| '10' | Only the first six coefficients are non-zero |
| '11' | All coefficients may be non-zero |

**rep\_cntrl** – A 5-bit field that indicates the number of times each field in an interlaced picture should be displayed, or the number of times that a progressive picture should be displayed. It is a function of the trick\_mode\_control field and the top\_field\_first bit in the video sequence header whether the top field or the bottom field should be displayed first in the case of interlaced pictures. The value '0' is forbidden.

**additional\_copy\_info** – This 7-bit field contains private data relating to copyright information.

**previous\_PES\_packet\_CRC** –The previous\_PES\_packet\_CRC is a 16-bit field that contains the CRC value that yields a zero output of the 16 registers in the decoder similar to the one defined in Annex A, but with the polynomial:

x16 + x12 + x5 + 1

after processing the data bytes of the previous PES packet, exclusive of the PES packet header.

NOTE 5 – This CRC is intended for use in network maintenance such as isolating the source of intermittent errors. It is not intended for use by elementary stream decoders. It is calculated only over the data bytes because PES packet header data can be modified during transport.

**PES\_private\_data\_flag** –A 1-bit flag which when set to '1' indicates that the PES packet header contains private data. When set to a value of '0' it indicates that private data is not present in the PES header.

**pack\_header\_field\_flag** – A 1-bit flag which when set to '1' indicates that an ISO/IEC 11172-1 pack header or a program stream pack header is stored in this PES packet header. If this field is in a PES packet that is contained in a program stream, then this field shall be set to '0'. In a transport stream, when set to the value '0' it indicates that no pack header is present in the PES header.

**program\_packet\_sequence\_counter\_flag** – A 1-bit flag which when set to '1' indicates that the program\_packet\_sequence\_counter, MPEG1\_MPEG2\_identifier, and original\_stuff\_length fields are present in this PES packet. When set to a value of '0' it indicates that these fields are not present in the PES header.

**P-STD\_buffer\_flag** –A 1-bit flag which when set to'1' indicates that the P-STD\_buffer\_scale and P-STD\_buffer\_size are present in the PES packet header. When set to a value of '0' it indicates that these fields are not present in the PES header.

**PES\_extension\_flag\_2** –A 1-bit field which when set to '1' indicates the presence of the PES\_extension\_field\_length field and associated fields. When set to a value of '0' this indicates that the PES\_extension\_field\_length field and any associated fields are not present.

**PES\_private\_data** –This is a 16-byte field which contains private data. This data, combined with the fields before and after, shall not emulate the packet\_start\_code\_prefix (0x000001).

**pack\_field\_length** – This is an 8-bit field which indicates the length, in bytes, of the pack\_header\_field().

**program\_packet\_sequence\_counter** –The program\_packet\_sequence\_counter field is a 7-bit field. It is an optional counter that increments with each successive PES packet from a program stream or from an ISO/IEC 11172-1 stream or the PES packets associated with a single program definition in a transport stream, providing functionality similar to a continuity counter (refer to 2.4.3.2). This allows an application to retrieve the original PES packet sequence of a program stream or the original packet sequence of the original ISO/IEC 11172-1 stream. The counter will wrap around to 0 after its maximum value. Repetition of PES packets shall not occur. Consequently, no two consecutive PES packets in the program multiplex shall have identical program\_packet\_sequence\_counter values.

**MPEG1\_MPEG2\_identifier** – A 1-bit flag which when set to '1' indicates that this PES packet carries information from an ISO/IEC 11172-1 stream. When set to '0' it indicates that this PES packet carries information from a program stream.

**original\_stuff\_length** –This 6-bit field specifies the number of stuffing bytes used in the original Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 PES packet header or in the original ISO/IEC 11172-1 packet header.

**P-STD\_buffer\_scale** – The P-STD\_buffer\_scale is a 1-bit field, the meaning of which is only defined if this PES packet is contained in a program stream. It indicates the scaling factor used to interpret the subsequent P‑STD\_buffer\_size field. If the preceding stream\_id indicates an audio stream, P-STD\_buffer\_scale shall have the value '0'. If the preceding stream\_id indicates a video stream, P-STD\_buffer\_scale shall have the value '1'. For all other stream types, the value may be either '1' or '0'.

**P-STD\_buffer\_size** – The P-STD\_buffer\_size is a 13-bit unsigned integer, the meaning of which is only defined if this PES packet is contained in a program stream. It defines the size of the input buffer, BSn, in the P-STD. If P‑STD\_buffer\_scale has the value '0', then the P-STD\_buffer\_size measures the buffer size in units of 128 bytes. IfP‑STD\_buffer\_scale has the value '1', then the P-STD\_buffer\_size measures the buffer size in units of 1024 bytes. Thus:

 (2-16)

else:

(2-17)

The encoded value of the P-STD buffer size takes effect immediately when the P-STD\_buffer\_size field is received by the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 System Target Decoder (refer to 2.7.7).

For AVC video streams conforming to one or more profiles defined in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496‑10, the size BSn shall be larger than or equal to the size of the CPB signalled by the CpbSize[ cpb\_cnt\_minus1 ] specified by the NAL hrd\_parameters() in the AVC video stream. If the NAL hrd\_parameters() are not present in the AVC video stream, then BSn shall be larger than or equal to the size of the NAL CPB for the byte stream format defined in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496-10 as 1200 × MaxCPB for the applied level.

For video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex G of Rec. ITU‑T H.264 | ISO/IEC 14496-10, the size BSn shall be larger than or equal to the size of the CPB signalled by the CpbSize[ cpb\_cnt\_minus1 ] specified by the NAL hrd\_parameters() for the video sub-bitstream carried in elementary stream ESn as defined in 2.14.3.6. If the NAL hrd\_parameters() are not present in the video sub-bitstream, the size BSn shall be larger than or equal to the size of the NAL CPB for the byte stream format defined in Rec. ITU-T H.264 | ISO/IEC 14496-10 as 1200 × MaxCPB for the applied level for the elementary stream ESn.

For MVC video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10, the size BSn shall be larger than or equal to the size of the CPB signalled by the CpbSize[ cpb\_cnt\_minus1 ] specified by the NAL hrd\_parameters() for the MVC video sub-bitstreams carried in elementary stream ESn, as defined in 2.14.3.6. If the NAL hrd\_parameters() are not present in the MVC video sub‑bitstreams, the size BSn shall be larger than or equal to the size of the NAL CPB for the byte stream format defined in Rec. ITU-T H.264 | ISO/IEC 14496-10 as 1200 × MaxCPB for the applied level for the elementary stream ESn.

**PES\_extension\_field\_length** –This is a 7-bit field which specifies the length, in bytes, of the data following this field in the PES extension field up to and including any reserved bytes.

**stream\_id\_extension\_flag** – A 1-bit flag, which when set to '0' indicates that a stream\_id\_extension field is present in the PES packet header. The value of '1' for this flag is reserved.

**stream\_id\_extension** – In program streams, the stream\_id\_extension specifies the type and number of the elementary stream as defined by the stream\_id\_extension in Table 2-27. In transport streams, the stream\_id\_extension may be set to any valid value which correctly describes the elementary stream type as defined in Table 2-27. In transport streams, the elementary stream type is specified in the program-specific information as specified in 2.4.4. Note that this field is used as an extension of the stream\_id defined above. This field shall not be used unless the value of stream\_id is '1111 1101'.

| Table 2-27 – Stream\_id\_extension assignments | | |
| --- | --- | --- |
| stream\_id\_extension | Note | stream coding |
| '000 0000' | 1 | IPMP control information stream |
| '000 0001' | 2 | IPMP stream |
| '000 0010' .. '000 1111' |  | ISO/IEC 14496-17 text stream |
| '001 0000' .. '001 1111' |  | ISO/IEC 23002-3 auxiliary video stream |
| '010 0000' .. '011 1111' |  | reserved\_data\_stream |
| '100 0000' .. '111 1111' |  | private\_stream |
| NOTE 1 – PES packets of stream\_id\_extension 0b000 0000 (IPMP control information stream) have a unique syntax specified in ISO/IEC 13818-11 (MPEG-2 IPMP).  NOTE 2 – PES packets of stream\_id\_extension 0b000 0001 (IPMP stream) have a unique syntax specified in ISO/IEC 13818-11 (MPEG-2 IPMP). | | |

**stuffing\_byte** – This is a fixed 8-bit value equal to '1111 1111' that can be inserted by the encoder, for example to meet the requirements of the channel. It is discarded by the decoder. No more than 32 stuffing bytes shall be present in one PES packet header.

**PES\_packet\_data\_byte** –PES\_packet\_data\_bytes shall be contiguous bytes of data from the elementary stream indicated by the packet's stream\_id or PID. When the elementary stream data conforms to Rec. ITU‑T H.262 | ISO/IEC 13818-2 or ISO/IEC 13818-3, the PES\_packet\_data\_bytes shall be byte aligned to the bytes of this Recommendation | International Standard. The byte-order of the elementary stream shall be preserved. The number of PES\_packet\_data\_bytes, N, is specified by the PES\_packet\_length field. N shall be equal to the value indicated in the PES\_packet\_length minus the number of bytes between the last byte of the PES\_packet\_length field and the first PES\_packet\_data\_byte.

In the case of a private\_stream\_1, private\_stream\_2, ECM\_stream, or EMM\_stream, the contents of the PES\_packet\_data\_byte field are user definable and will not be specified by ITU-T | ISO/IEC in the future.

**padding\_byte** –This is a fixed 8-bit value equal to '1111 1111'. It is discarded by the decoder.

**tref\_extension\_flag** – A 1-bit flag, which when set to '0' indicates that a TREF field is present in the PES packet header. The value of '1' for this flag is reserved.

**TREF (timestamp reference)** – The TREF is a 33-bit number coded in three separate fields. It indicates the decoding time value, tdn(j), in the system target decoder as indicated by the DTS, or in absence of the DTS, by the PTS of the PES header of the same j-th access unit in a corresponding elementary stream n.

#### 2.4.3.8 Carriage of program streams and ISO/IEC 11172-1 Systems streams in the transport stream

The transport stream contains optional fields to support the carriage of program streams and ISO/IEC 11172-1 Systems streams, in a way that allows simple reconstruction of the respective stream at the decoder.

When placing a program stream into a transport stream, program stream PES packets with stream\_id values of private\_stream\_1, Rec. ITU-T H.262 | ISO/IEC 13818-2 or ISO/IEC 11172-2 video, and ISO/IEC 13818-3 or ISO/IEC 11172-3 audio, are carried in transport stream packets.

For these PES packets, when reconstructing the program stream at the transport stream decoder, the PES packet data is copied to the program stream being reconstructed.

For program streams PES packets with stream\_id values of program\_stream\_map, padding\_stream, private\_stream\_2, ECM, EMM, DSM\_CC\_stream, or program\_stream\_directory, all the bytes of the program stream PES packet, except for the packet\_start\_code\_prefix, are placed into the data\_bytes fields of a new PES packet. The stream\_id of this new PES packet has the value of ancillary\_stream (refer to Table 2-22). This new PES packet is then carried in transport stream packets.

When reconstructing the program stream at the transport stream decoder, for PES packets with a stream\_id value of ancillary\_stream\_id, packet\_start\_code\_prefix is written to the program stream being reconstructed, followed by the data\_byte fields from these transport stream PES packets.

ISO/IEC 11172-1 streams are carried within transport streams by first replacing ISO/IEC 11172-1 packet headers with Rec. ITU-T H.262 | ISO/IEC 13818-2 PES packet headers. ISO/IEC 11172-1 packet header field values are copied to the equivalent Rec. ITU-T H.262 | ISO/IEC 13818-2 PES packet header fields.

The program\_packet\_sequence\_counter field is included within the header of each PES packet carrying data from a program stream, or an ISO/IEC 11172-1 System stream. This allows the order of PES packets in the original program stream, or packets in the original ISO/IEC 11172-1 system stream, to be reproduced at the decoder.

The pack\_header() field of a program stream, or an ISO/IEC 11172-1 system stream, is carried in the transport stream in the header of the immediately following PES packet.

### 2.4.4 Program-specific information

#### 2.4.4.1 General

Program-specific information (PSI) includes both Rec. ITU-T H.222.0 | ISO/IEC 13818-1 normative data and private data that enable demultiplexing of programs by decoders. Programs are composed of one or more elementary streams, each labelled with a PID. Programs, elementary streams or parts thereof may be scrambled for conditional access. However, program-specific information shall not be scrambled.

In transport streams, program-specific information is classified into six table structures as shown in Table 2-28. While these structures may be thought of as simple tables, they shall be segmented into sections and inserted in transport stream packets, some with predetermined PIDs and others with user selectable PIDs.

Table 2-28 – Program-specific information

|  |  |  |  |
| --- | --- | --- | --- |
| Structure Name | Stream Type | Reserved PID # | Description |
| Program association table | Rec. ITU-T H.222.0 |  ISO/IEC 13818-1 | 0x00 | Associates program number and program map table PID |
| Program map table | Rec. ITU-T H.222.0 |  ISO/IEC 13818-1 | Assigned in the PAT | Specifies PID values for components of one or more programs |
| Network information table | Private | Assigned in the PAT | Physical network parameters such as FDM frequencies, transponder numbers, etc. |
| Conditional access table | Rec. ITU-T H.222.0 |  ISO/IEC 13818-1 | 0x01 | Associates one or more (private) EMM streams each with a unique PID value |
| Transport stream description table | Rec. ITU-T H.222.0 |  ISO/IEC 13818-1 | 0x02 | Associates one or more descriptors from Table 2-45 to an entire transport stream |
| IPMP control information table | Rec. ITU-T H.222.0 |  ISO/IEC 13818-1 | 0x03 | Contains IPMP tool list, rights container, tool container defined in ISO/IEC 13818-11 |

Rec. ITU-T H.222.0 | ISO/IEC 13818-1 defined PSI tables shall be segmented into one or more sections that are carried within transports packets. A section is a syntactic structure that shall be used for mapping each Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 defined PSI table into transport stream packets.

Along with Rec. ITU-T H.222.0 | ISO/IEC 13818-1 defined PSI tables, it is possible to carry private data tables. The means by which private information is carried within transport stream packets is not defined by this Specification. It may be structured in the same manner used for carrying of Rec. ITU-T H.222.0 | ISO/IEC 13818-1 defined PSI tables, such that the syntax for mapping this private data is identical to that used for the mapping of Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 defined PSI tables. For this purpose, a private section is defined. If the private data is carried in transport stream packets with the same PID value as transport stream packets carrying program map tables (as identified in the program association table), then the private\_section syntax and semantics shall be used. The data carried in the private\_data\_bytes may be scrambled. However, no other fields of the private\_section shall be scrambled. This private\_section allows data to be transmitted with a minimum of structure. When this structure is not used, the mapping of private data within transport stream packets is not defined by this Recommendation | International Standard.

Sections may be variable in length. The beginning of a section is indicated by a pointer\_field in the transport stream packet payload. The syntax of this field is specified in Table 2-29.

Adaptation fields may occur in transport stream packets carrying transport stream sections.

Within a transport stream, packet stuffing bytes of value 0xFF may be found in the payload of transport stream packets carrying sections only after the last byte of a section. In this case all bytes until the end of the transport stream packet shall also be stuffing bytes of value 0xFF. These bytes may be discarded by a decoder. In such a case, the payload of the next transport stream packet with the same PID value shall begin with a pointer\_field of value 0x00 indicating that the next section starts immediately thereafter.

Each transport stream shall contain one or more transport stream packets with PID value 0x0000. These transport stream packets together shall contain a complete program association table, providing a complete list of all programs within the transport stream. The most recently transmitted version of the table with the current\_next\_indicator set to a value of'1' shall always apply to the current data in the transport stream. Any changes in the programs carried within the transport stream shall be described in an updated version of the program association table carried in transport stream packets with PID value 0x0000. These sections shall all use table\_id value 0x00. Only sections with this value of table\_id are permitted within transport stream packets with PID value of 0x0000. For a new version of the PAT to become valid, all sections (as indicated in the last\_section\_number) with a new version\_number and with the current\_next\_indicator set to '1' must exit Bsys defined in the T-STD (refer to 2.4.2.1). The PAT becomes valid when the last byte of the section needed to complete the table exits Bsys.

Whenever one or more elementary streams within a transport stream are scrambled, transport stream packets with a PID value 0x0001 shall be transmitted containing a complete conditional access table including CA\_descriptors associated with the scrambled streams. The transmitted transport stream packets will together form one complete version of the conditional access table. The most recently transmitted version of the table with the current\_next\_indicator set to a value of '1' shall always apply to the current data in the transport stream. Any changes in scrambling making the existing table invalid or incomplete shall be described in an updated version of the conditional access table. These sections will all use table\_id value 0x01. Only sections with this table\_id value are permitted within transport stream packets with a PID value of 0x0001. For a new version of the CAT to become valid, all sections (as indicated in the last\_section\_number) with a new version\_number and with the current\_next\_indicator set to '1' must exit Bsys. The CAT becomes valid when the last byte of the section needed to complete the table exits Bsys.

Each transport stream shall contain one or more transport stream packets with PID values which are labelled under the program association table as transport stream packets containing TS program map sections. Each program listed in the program association table shall be described in a unique TS\_program\_map\_section. Every program shall be fully defined within the transport stream itself. Private data which has an associated elementary\_PID field in the appropriate program map table section is part of the program. Other private data may exist in the transport stream without being listed in the program map table section. The most recently transmitted version of the TS\_program\_map\_section with the current\_next\_indicator set to a value of '1' shall always apply to the current data within the transport stream. Any changes in the definition of any of the programs carried within the transport stream shall be described in an updated version of the corresponding section of the program map table carried in transport stream packets with the PID value identified as the program\_map\_PID for that specific program. All transport stream packets which carry a given TS\_program\_map\_section shall have the same PID value. During the continuous existence of a program, including all of its associated events, the program\_map\_PID shall not change. A program definition shall not span more than one TS\_program\_map\_section. A new version of a TS\_program\_map\_section becomes valid when the last byte of that section with a new version\_number and with the current\_next\_indicator set to '1' exits Bsys.

Sections with a table\_id value of 0x02 shall contain program map table information. Such sections may be carried in transport stream packets with different PID values.

The network information table is optional and its contents are private. If present it is carried within transport stream packets that will have the same PID value, called the network\_PID. The network\_PID value is defined by the user and, when present, shall be found in the program association table under the reserved program\_number 0x0000. If the network information table exists, it shall take the form of one or more private\_sections.

The maximum number of bytes in a section of a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 defined PSI table is 1024 bytes. The maximum number of bytes in a private\_section is 4096 bytes.

The transport stream description table is optional. When present, the transport stream description is carried within transport stream packets that have a PID value 0x0002 as specified in Table 2-28 and shall apply to the entire transport stream. Sections of the transport stream description shall use a table\_id value of 0x03 as specified in Table 2-31 and its contents are restricted to descriptors specified in Table 2-45. The TS\_description\_section becomes valid when the last byte of the section required to complete the table exits Bsys.

There are no restrictions on the occurrence of start codes, sync bytes or other bit patterns in transport stream section data.

#### 2.4.4.2 Pointer

The pointer\_field syntax is defined in Table 2-29.

Table 2-29 – Program-specific information pointer

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| **pointer\_field** | **8** | **uimsbf** |

#### 2.4.4.3 Semantics definition of fields in pointer syntax

**pointer\_field** – This is an 8-bit field whose value shall be the number of bytes, immediately following the pointer\_field until the first byte of the first section that is present in the payload of the transport stream packet (so a value of 0x00 in the pointer\_field indicates that the section starts immediately after the pointer\_field). When at least one section begins in a given transport stream packet, then the payload\_unit\_start\_indicator (refer to 2.4.3.2) shall be set to '1' and the first byte of the payload of that transport stream packet shall contain the pointer. When no section begins in a given transport stream packet, then the payload\_unit\_start\_indicator shall be set to '0' and no pointer shall be sent in the payload of that packet.

#### 2.4.4.4 Program association table

The program association table (PAT) provides the correspondence between a program\_number and the PID value of the transport stream packets which carry the program definition. The program\_number is the numeric label associated with a program.

The overall table is contained in one or more sections with the following syntax. It may be segmented to occupy multiple sections (see Table 2-30).

| Table 2-30 – Program association section | | |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| program\_association\_section() { |  |  |
| **table\_id** | **8** | **uimsbf** |
| **section\_syntax\_indicator** | **1** | **bslbf** |
| **'0'** | **1** | **bslbf** |
| **reserved** | **2** | **bslbf** |
| **section\_length** | **12** | **uimsbf** |
| **transport\_stream\_id** | **16** | **uimsbf** |
| **reserved** | **2** | **bslbf** |
| **version\_number** | **5** | **uimsbf** |
| **current\_next\_indicator** | **1** | **bslbf** |
| **section\_number** | **8** | **uimsbf** |
| **last\_section\_number** | **8** | **uimsbf** |
| for (i = 0; i < N; i++) { |  |  |
| **program\_number** | **16** | **uimsbf** |
| **reserved** | **3** | **bslbf** |
| if (program\_number = = '0') { |  |  |
| **network\_PID** | **13** | **uimsbf** |
| } |  |  |
| else { |  |  |
| **program\_map\_PID** | **13** | **uimsbf** |
| } |  |  |
| } |  |  |
| **CRC\_32** | **32** | **rpchof** |
| } |  |  |

#### 2.4.4.5 Table\_id assignments

The table\_id field identifies the contents of a transport stream section as shown in Table 2-31.

| Table 2-31 – table\_id assignment values | | |
| --- | --- | --- |
| Value | Description |
| 0x00 | program\_association\_section |
| 0x01 | conditional\_access\_section (CA\_section) |
| 0x02 | TS\_program\_map\_section |
| 0x03 | TS\_description\_section |
| 0x04 | ISO\_IEC\_14496\_scene\_description\_section |
| 0x05 | ISO\_IEC\_14496\_object\_descriptor\_section |
| 0x06 | Metadata\_section |
| 0x07 | IPMP Control Information Section (defined in ISO/IEC 13818-11) |
| 0x08 | ISO\_IEC\_14496\_section |
| 0x09 | ISO/IEC 23001-11 (Green access unit) section |
| 0x0A | ISO/IEC 23001-10 (Quality access unit) section |
| 0x0B | ISO/IEC 23001-13 (Media Orchestration access unit) section |
| 0x0C .. 0x37 | Rec. ITU-T H.222.0 | ISO/IEC 13818-1 reserved |
| 0x38 .. 0x3F | Defined in ISO/IEC 13818-6 |
| 0x40 .. 0xFE | User private |
| 0xFF | Forbidden |

#### 2.4.4.6 Semantic definition of fields in program association section

**table\_id** – This is an 8-bit field, which shall be set to 0x00 as shown in Table 2-31.

**section\_syntax\_indicator** – The section\_syntax\_indicator is a 1-bit field which shall be set to '1'.

**section\_length** – This is a 12-bit field, the first two bits of which shall be '00'. The remaining 10 bits specify the number of bytes of the section, starting immediately following the section\_length field, and including the CRC. The value in this field shall not exceed 1021 (0x3FD).

**transport\_stream\_id** – This is a 16-bit field which serves as a label to identify this transport stream from any other multiplex within a network. Its value is defined by the user.

**version\_number** – This 5-bit field is the version number of the whole program association table. The version number shall be incremented by 1 modulo 32 whenever the definition of the program association table changes. When the current\_next\_indicator is set to '1', then the version\_number shall be that of the currently applicable program association table. When the current\_next\_indicator is set to '0', then the version\_number shall be that of the next applicable program association table.

**current\_next\_indicator** – A 1-bit indicator, which when set to '1' indicates that the program association table sent is currently applicable. When the bit is set to '0', it indicates that the table sent is not yet applicable and shall be the next table to become valid.

**section\_number** – This 8-bit field gives the number of this section. The section\_number of the first section in the Program Association Table shall be 0x00. It shall be incremented by 1 with each additional section in the program association table.

**last\_section\_number** – This 8-bit field specifies the number of the last section (that is, the section with the highest section\_number) of the complete program association table.

**program\_number** – Program\_number is a 16-bit field. It specifies the program to which the program\_map\_PID is applicable. When set to 0x0000, then the following PID reference shall be the network PID. For all other cases the value of this field is user defined. This field shall not take any single value more than once within one version of the program association table.

NOTE – The program\_number may be used as a designation for a broadcast channel, for example.

**network\_PID** – The network\_PID is a 13-bit field, which is used only in conjunction with the value of the program\_number set to 0x0000, specifies the PID of the transport stream packets which shall contain the network information table. The value of the network\_PID field is defined by the user, but shall only take values as specified in Table 2-3. The presence of the network\_PID is optional.

**program\_map\_PID** – The program\_map\_PID is a 13-bit field specifying the PID of the transport stream packets which shall contain the program\_map\_section applicable for the program as specified by the program\_number. No program\_number shall have more than one program\_map\_PID assignment. The value of the program\_map\_PID is defined by the user, but shall only take values as specified in Table 2-3.

**CRC\_32** – This is a 32-bit field that contains the CRC value that gives a zero output of the registers in the decoder defined in Annex A after processing the entire program association section.

#### 2.4.4.7 Conditional access table

The conditional access table (CAT) provides the association between one or more conditional access (CA) systems, their EMM streams and any special parameters associated with them. Refer to 2.6.16 for a definition of the descriptor() field in Table 2-32.

The table is contained in one or more sections with the following syntax. It may be segmented to occupy multiple sections.

Table 2-32 – Conditional access section

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| CA\_section() { |  |  |
| **table\_id** | **8** | **uimsbf** |
| **section\_syntax\_indicator** | **1** | **bslbf** |
| **'0'** | **1** | **bslbf** |
| **reserved** | **2** | **bslbf** |
| **section\_length** | **12** | **uimsbf** |
| **reserved** | **18** | **bslbf** |
| **version\_number** | **5** | **uimsbf** |
| **current\_next\_indicator** | **1** | **bslbf** |
| **section\_number** | **8** | **uimsbf** |
| **last\_section\_number** | **8** | **uimsbf** |
| for (i = 0; i < N; i++) { |  |  |
| descriptor() |  |  |
| } |  |  |
| **CRC\_32** | **32** | **rpchof** |
| } |  |  |

#### 2.4.4.8 Semantic definition of fields in conditional access section

**table\_id** – This is an 8-bit field, which shall be set to 0x01 as specified in Table 2-31.

**section\_syntax\_indicator** – The section\_syntax\_indicator is a 1-bit field which shall be set to '1'.

**section\_length** – This is a 12-bit field, the first two bits of which shall be '00'. The remaining 10-bits specify the number of bytes of the section starting immediately following the section\_length field, and including the CRC. The value in this field shall not exceed 1021 (0x3FD).

**version\_number** – This 5-bit field is the version number of the entire conditional access table. The version number shall be incremented by 1 modulo 32 when a change in the information carried within the CA table occurs. When the current\_next\_indicator is set to '1', then the version\_number shall be that of the currently applicable conditional access table. When the current\_next\_indicator is set to '0', then the version\_number shall be that of the next applicable conditional access table.

**current\_next\_indicator** – A 1-bit indicator, which when set to '1' indicates that the conditional access table sent is currently applicable. When the bit is set to '0', it indicates that the conditional access table sent is not yet applicable and shall be the next conditional access table to become valid.

**section\_number** – This 8-bit field gives the number of this section. The section\_number of the first section in the conditional access table shall be 0x00. It shall be incremented by 1 with each additional section in the conditional access table.

**last\_section\_number** – This 8-bit field specifies the number of the last section (that is, the section with the highest section\_number) of the conditional access table.

**CRC\_32** – This is a 32-bit field that contains the CRC value that gives a zero output of the registers in the decoder defined in Annex A after processing the entire conditional access section.

#### 2.4.4.9 Program map table

The program map table provides the mappings between program numbers and the program elements that comprise them. A single instance of such a mapping is referred to as a "program definition". The program map table is the complete collection of all program definitions for a transport stream. This table shall be transmitted in packets, the PID values of which are selected by the encoder. More than one PID value may be used, if desired. The table is contained in one or more sections with the following syntax. It may be segmented to occupy multiple sections. In each section, the section number field shall be set to zero. Sections are identified by the program\_number field.

Definition for the descriptor() fields may be found in 2.6 (see Table 2-33).

Table 2-33 – Transport stream program map section

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| TS\_program\_map\_section() { |  |  |
| **table\_id** | **8** | **uimsbf** |
| **section\_syntax\_indicator** | **1** | **bslbf** |
| **'0'** | **1** | **bslbf** |
| **reserved** | **2** | **bslbf** |
| **section\_length** | **12** | **uimsbf** |
| **program\_number** | **16** | **uimsbf** |
| **reserved** | **2** | **bslbf** |
| **version\_number** | **5** | **uimsbf** |
| **current\_next\_indicator** | **1** | **bslbf** |
| **section\_number** | **8** | **uimsbf** |
| **last\_section\_number** | **8** | **uimsbf** |
| **reserved** | **3** | **bslbf** |
| **PCR\_PID** | **13** | **uimsbf** |
| **reserved** | **4** | **bslbf** |
| **program\_info\_length** | **12** | **uimsbf** |
| for (i = 0; i < N; i++) { |  |  |
| descriptor() |  |  |
| } |  |  |
| for (i = 0; i < N1; i++) { |  |  |
| **stream\_type** | **8** | **uimsbf** |
| **reserved** | **3** | **bslbf** |
| **elementary\_PID** | **13** | **uimsbf** |
| **reserved** | **4** | **bslbf** |
| **ES\_info\_length** | **12** | **uimsbf** |
| for (i = 0; i < N2; i++) { |  |  |
| descriptor() |  |  |
| } |  |  |
| } |  |  |
| **CRC\_32** | **32** | **rpchof** |
| } |  |  |

#### 2.4.4.10 Semantic definition of fields in transport stream program map section

**table\_id** – This is an 8-bit field, which in the case of a TS\_program\_map\_section shall be always set to 0x02 as shown in Table 2-31.

**section\_syntax\_indicator** – The section\_syntax\_indicator is a 1-bit field which shall be set to '1'.

**section\_length** – This is a 12-bit field, the first two bits of which shall be '00'. The remaining 10 bits specify the number of bytes of the section starting immediately following the section\_length field, and including the CRC. The value in this field shall not exceed 1021 (0x3FD).

**program\_number** – program\_number is a 16-bit field. It specifies the program to which the program\_map\_PID is applicable. One program definition shall be carried within only one TS\_program\_map\_section. This implies that a program definition is never longer than 1016 (0x3F8). See Informative Annex C for ways to deal with the cases when that length is not sufficient. The program\_number may be used as a designation for a broadcast channel, for example. By describing the different program elements belonging to a program, data from different sources (e.g., sequential events) can be concatenated together to form a continuous set of streams using a program\_number. For examples of applications refer to Annex C.

**version\_number** – This 5-bit field is the version number of the TS\_program\_map\_section. The version number shall be incremented by 1 modulo 32 when a change in the information carried within the section occurs. Version number refers to the definition of a single program, and therefore to a single section. When the current\_next\_indicator is set to '1', then the version\_number shall be that of the currently applicable TS\_program\_map\_section. When the current\_next\_indicator is set to '0', then the version\_number shall be that of the next applicable TS\_program\_map\_section.

**current\_next\_indicator** – A 1-bit field, which when set to '1' indicates that the TS\_program\_map\_section sent is currently applicable. When the bit is set to '0', it indicates that the TS\_program\_map\_section sent is not yet applicable and shall be the next TS\_program\_map\_section to become valid.

**section\_number** – The value of this 8-bit field shall be 0x00.

**last\_section\_number** – The value of this 8-bit field shall be 0x00.

**PCR\_PID** – This is a 13-bit field indicating the PID of the transport stream packets which shall contain the PCR fields valid for the program specified by program\_number. If no PCR is associated with a program definition for private streams, then this field shall take the value of 0x1FFF. Refer to the semantic definition of PCR in 2.4.3.5 and Table 2-3 for restrictions on the choice of PCR\_PID value.

**program\_info\_length** – This is a 12-bit field, the first two bits of which shall be '00'. The remaining 10 bits specify the number of bytes of the descriptors immediately following the program\_info\_length field.

**stream\_type** – This is an 8-bit field specifying the type of program element carried within the packets with the PID whose value is specified by the elementary\_PID. The values of stream\_type are specified in Table 2-34.

NOTE – A Rec. ITU-T H.222.0 | ISO/IEC 13818-1 auxiliary stream is available for data types defined by this Specification, other than audio, video, and DSM-CC, such as the program stream directory and the program stream map.

| Table 2-34 – Stream type assignments | | |
| --- | --- | --- |
| Value | Description | |
| 0x00 | ITU-T | ISO/IEC Reserved | |
| 0x01 | ISO/IEC 11172-2 Video | |
| 0x02 | Rec. ITU-T H.262 | ISO/IEC 13818-2 Video or ISO/IEC 11172-2 constrained parameter video stream (see Note 2) | |
| 0x03 | ISO/IEC 11172-3 Audio | |
| 0x04 | ISO/IEC 13818-3 Audio | |
| 0x05 | Rec. ITU-T H.222.0 | ISO/IEC 13818-1 private\_sections | |
| 0x06 | Rec. ITU-T H.222.0 | ISO/IEC 13818-1 PES packets containing private data | |
| 0x07 | ISO/IEC 13522 MHEG | |
| 0x08 | Rec. ITU-T H.222.0 | ISO/IEC 13818-1 Annex A DSM-CC | |
| 0x09 | Rec. ITU-T H.222.1 | |
| 0x0A | ISO/IEC 13818-6 type A | |
| 0x0B | ISO/IEC 13818-6 type B | |
| 0x0C | ISO/IEC 13818-6 type C | |
| 0x0D | ISO/IEC 13818-6 type D | |
| 0x0E | Rec. ITU-T H.222.0 | ISO/IEC 13818-1 auxiliary | |
| 0x0F | ISO/IEC 13818-7 Audio with ADTS transport syntax | |
| 0x10 | ISO/IEC 14496-2 Visual | |
| 0x11 | ISO/IEC 14496-3 Audio with the LATM transport syntax as defined in ISO/IEC 14496-3 | |
| 0x12 | ISO/IEC 14496-1 SL-packetized stream or M4Mux stream carried in PES packets | |
| 0x13 | ISO/IEC 14496-1 SL-packetized stream or M4Mux stream carried in ISO/IEC 14496\_sections | |
| 0x14 | ISO/IEC 13818-6 Synchronized Download Protocol | |
| 0x15 | Metadata carried in PES packets | |
| 0x16 | Metadata carried in metadata\_sections | |
| 0x17 | Metadata carried in ISO/IEC 13818-6 Data Carousel | |
| 0x18 | Metadata carried in ISO/IEC 13818-6 Object Carousel | |
| 0x19 | Metadata carried in ISO/IEC 13818-6 Synchronized Download Protocol | |
| 0x1A | IPMP stream (defined in ISO/IEC 13818-11, MPEG-2 IPMP) | |
| 0x1B | AVC video stream conforming to one or more profiles defined in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496-10 or AVC video sub-bitstream of SVC as defined in 2.1.10 or MVC base view sub‑bitstream, as defined in 2.1.101, or AVC video sub-bitstream of MVC, as defined in 2.1.8 or MVCD base view sub‑bitstream, as defined in 2.1.106, or AVC video sub-bitstream of MVCD, as defined in 2.1.9, or AVC base layer of an HEVC video stream conforming to one or more profiles defined in Annex G or Annex H of Rec. ITU-T H.265 | ISO/IEC 23008-2 | |
| 0x1C | ISO/IEC 14496-3 Audio, without using any additional transport syntax, such as DST, ALS and SLS | |
| 0x1D | ISO/IEC 14496-17 Text | |
| 0x1E | Auxiliary video stream as defined in ISO/IEC 23002-3 | |
| 0x1F | SVC video sub-bitstream of an AVC video stream conforming to one or more profiles defined in Annex G of Rec. ITU-T H.264 | ISO/IEC 14496-10 | |
| 0x20 | MVC video sub-bitstream of an AVC video stream conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10 | |
| 0x21 | Video stream conforming to one or more profiles as defined in Rec. ITU-T T.800 | ISO/IEC 15444-1 | |
| 0x22 | Additional view Rec. ITU-T H.262 | ISO/IEC 13818-2 video stream for service-compatible stereoscopic 3D services (see Notes 3 and 4) | |
| 0x23 | Additional view Rec. ITU-T H.264 | ISO/IEC 14496-10 video stream conforming to one or more profiles defined in Annex A for service-compatible stereoscopic 3D services (see Notes 3 and 4) | |
| 0x24 | Rec. ITU-T H.265 | ISO/IEC 23008-2 video stream or an HEVC temporal video sub-bitstream (see Note 5) | |
| 0x25 | HEVC temporal video subset of an HEVC video stream conforming to one or more profiles defined in Annex A of Rec. ITU-T H.265 | ISO/IEC 23008-2 | |
| 0x26 | MVCD video sub-bitstream of an AVC video stream conforming to one or more profiles defined in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496-10 | |
| 0x27 | Timeline and External Media Information Stream (see Annex U) | |
| 0x28 | HEVC enhancement sub-partition which includes TemporalId 0 of an HEVC video stream where all NALs units contained in the stream conform to one or more profiles defined in Annex G of Rec. ITU‑T H.265 | ISO/IEC 23008-2 | |
| 0x29 | HEVC temporal enhancement sub-partition of an HEVC video stream where all NAL units contained in the stream conform to one or more profiles defined in Annex G of Rec. ITU-T H.265 | ISO/IEC 23008-2 | |
| 0x2A | HEVC enhancement sub-partition which includes TemporalId 0 of an HEVC video stream where all NAL units contained in the stream conform to one or more profiles defined in Annex H of Rec. ITU-T H.265 | ISO/IEC 23008-2 | |
| 0x2B | HEVC temporal enhancement sub-partition of an HEVC video stream where all NAL units contained in the stream conform to one or more profiles defined in Annex H of Rec. ITU-T H.265 | ISO/IEC 23008-2 | |
| 0x2C | Green access units carried in MPEG-2 sections | |
| 0x2D | ISO/IEC 23008-3 Audio with MHAS transport syntax – main stream | |
| 0x2E | ISO/IEC 23008-3 Audio with MHAS transport syntax – auxiliary stream | |
| 0x2F | Quality access units carried in sections | |
| 0x30 | Media Orchestration Access Units carried in sections | |
| 0x31 | Substream of a Rec. ITU-T H.265 | ISO/IEC 23008 2 video stream that contains a Motion Constrained Tile Set, parameter sets, slice headers or a combination thereof. See 2.17.5.1. | |
| 0x32 | JPEG XS video stream conforming to one or more profiles as defined in ISO/IEC 21122-2 | |
| 0x33 | VVC video stream or a VVC temporal video sub-bitstream conforming to one or more profiles defined in Annex A of Rec. ITU-T H.266 | ISO/IEC 23090-3 | |
| 0x34 | VVC temporal video subset of a VVC video stream conforming to one or more profiles defined in Annex A of Rec. ITU-T H.266 | ISO/IEC 23090-3 | |
| 0x35 | EVC video stream or an EVC temporal video sub-bitstream conforming to one or more profiles defined in ISO/IEC 23094-1 | |
| 0x36 | LCEVC video stream conforming to one or more profiles defined in ISO/IEC 23094-2 | |
| 0x37 .. 0x7E | Rec. ITU-T H.222.0 | ISO/IEC 13818-1 reserved | | |
| 0x7F | IPMP stream | |
| 0x80 .. 0xFF | User Private | |

NOTE 1 – In the above table various stream types are assigned for carriage of audio signals, with or without a transport syntax. Typically, the transport syntax is used for providing sync words. The use of a specific transport syntax, if at all, is specified in the clauses in this Specification specifying the transport of the various audio signals.

NOTE 2 – Rec. ITU-T H.262 | ISO/IEC 13818-2 video with frame packing arrangement information is signalled using stream\_type value 0x02.

NOTE 3 – The base view of service-compatible stereoscopic 3D services is signalled using stream\_type value 0x02 for Rec. ITU‑T H.262 | ISO/IEC 13818-2 video or stream\_type 0x1B for Rec. ITU-T H.264 | ISO/IEC 14496-10 video.

NOTE 4 – The additional view for service-compatible stereoscopic 3D services is signalled using stream\_type value 0x22 for Rec. ITU-T H.262 | ISO/IEC 13818-2 video or stream\_type value 0x23 for Rec. ITU-T H.264 | ISO/IEC 14496-10 video conforming to one or more profiles defined in Annex A. For service-compatible stereoscopic 3D services, the additional view is not signalled using stream\_type values 0x02 or 0x1B.

NOTE 5 – In the case of a single ES carrying multiple HEVC tile substreams that are identified through adaptation field descriptors, stream type 0x24 can be used if the ES is deemed to be a backward compatible HEVC stream and is decodable on a legacy device.

**elementary\_PID** – This is a 13-bit field specifying the PID of the transport stream packets which carry the associated program element.

**ES\_info\_length** – This is a 12-bit field, the first two bits of which shall be '00'. The remaining 10 bits specify the number of bytes of the descriptors of the associated program element immediately following the ES\_info\_length field.

**CRC\_32** – This is a 32-bit field that contains the CRC value that gives a zero output of the registers in the decoder defined in Annex B after processing the entire transport stream program map section.

#### 2.4.4.11 Syntax of the private section

When private data is sent in transport stream packets with a PID value designated as a program map PID in the program association table the private\_section shall be used. The private\_section allows data to be transmitted with a minimum of structure while enabling a decoder to parse the stream. The sections may be used in two ways: if the section\_syntax\_indicator is set to '1', then the whole structure common to all tables shall be used; if the indicator is set to '0', then only the fields 'table\_id' through 'private\_section\_length' shall follow the common structure syntax and semantics and the rest of the private\_section may take any form the user determines. Examples of extended use of this syntax are found in informative Annex C.

A private table may be made of several private\_sections, all with the same table\_id (see Table 2-35).

Table 2-35 – Private section

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| private\_section() { |  |  |
| **table\_id** | **8** | **uimsbf** |
| **section\_syntax\_indicator** | **1** | **bslbf** |
| **private\_indicator** | **1** | **bslbf** |
| **reserved** | **2** | **bslbf** |
| **private\_section\_length** | **12** | **uimsbf** |
| if (section\_syntax\_indicator = = '0') { |  |  |
| for (i = 0; i < N; i++) { |  |  |
| **private\_data\_byte** | **8** | **bslbf** |
| } |  |  |
| } |  |  |
| else { |  |  |
| **table\_id\_extension** | **16** | **uimsbf** |
| **reserved** | **2** | **bslbf** |
| **version\_number** | **5** | **uimsbf** |
| **current\_next\_indicator** | **1** | **bslbf** |
| **section\_number** | **8** | **uimsbf** |
| **last\_section\_number** | **8** | **uimsbf** |
| for (i = 0; i < private\_section\_length-9; i++) { |  |  |
| **private\_data\_byte** | **8** | **bslbf** |
| } |  |  |
| **CRC\_32** | **32** | **rpchof** |
| } |  |  |
| } |  |  |

#### 2.4.4.12 Semantic definition of fields in private section

**table\_id** – This 8-bit field, the value of which identifies the Private Table this section belongs to. Only values defined in Table 2-31 as "user private" may be used.

**section\_syntax\_indicator** – This is a 1-bit indicator. When set to '1', it indicates that the private section follows the generic section syntax beyond the private\_section\_length field. When set to '0', it indicates that the private\_data\_bytes immediately follow the private\_section\_length field.

**private\_indicator** – This is a 1-bit user-definable flag that shall not be specified by ITU-T | ISO/IEC in the future.

**private\_section\_length** – A 12-bit field. It specifies the number of remaining bytes in the private section immediately following the private\_section\_length field up to the end of the private\_section. The value in this field shall not exceed 4093 (0xFFD).

**private\_data\_byte** – The private\_data\_byte field is user definable and shall not be specified by ITU-T | ISO/IEC in the future.

**table\_id\_extension** – This is a 16-bit field. Its use and value are defined by the user.

**version\_number** – This 5-bit field is the version number of the private\_section. The version\_number shall be incremented by 1 modulo 32 when a change in the information carried within the private\_section occurs. When the current\_next\_indicator is set to '0', then the version\_number shall be that of the next applicable private\_section with the same table\_id and section\_number.

**current\_next\_indicator** – A 1-bit field, which when set to '1' indicates that the private\_section sent is currently applicable. When the current\_next\_indicator is set to '1', then the version\_number shall be that of the currently applicable private\_section. When the bit is set to '0', it indicates that the private\_section sent is not yet applicable and shall be the next private\_section with the same section\_number and table\_id to become valid.

**section\_number** – This 8-bit field gives the number of the private\_section. The section\_number of the first section in a private table shall be 0x00. The section\_number shall be incremented by 1 with each additional section in this private table.

**last\_section\_number** – This 8-bit field specifies the number of the last section (that is, the section with the highest section\_number) of the private table of which this section is a part.

**CRC\_32** – This is a 32-bit field that contains the CRC value that gives a zero output of the registers in the decoder defined in Annex A after processing the entire private section.

#### 2.4.4.13 Syntax of the transport stream section

Rec. ITU-T H.222.0 | ISO/IEC 13818-1 compliant bitstreams may carry the information defined in Table 2-36. Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 compliant decoders may decode the information defined in this table.

The transport stream description table is defined to support the carriage of descriptors as found in 2.6 for an entire transport stream. The descriptors shall apply to the entire transport stream. This table uses a table\_id value of 0x03 as specified in Table 2-31 and is carried in transport stream packets whose PID value is 0x0002 as specified in Table 2-3.

Table 2-36 – The transport stream description table

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| TS\_description\_section() { |  |  |
| **table\_id** | **8** | **uimsbf** |
| **section\_syntax\_indicator** | **1** | **bslbf** |
| **'0'** | **1** | **bslbf** |
| **reserved** | **2** | **bslbf** |
| **section\_length** | **12** | **uimsbf** |
| **reserved** | **18** | **bslbf** |
| **version\_number** | **5** | **uimsbf** |
| **current\_next\_indicator** | **1** | **bslbf** |
| **section\_number** | **8** | **uimsbf** |
| **last\_section\_number** | **8** | **uimsbf** |
| for (i = 0; i < N; i++) { |  |  |
| descriptor() |  |  |
| } |  |  |
| **CRC\_32** | **32** | **rpchof** |
| } |  |  |

#### 2.4.4.14 Semantic definition of fields in the transport stream section

**table\_id** –This is an 8-bit field, which shall be set to '0x03' as specified in Table 2-31.

**section\_length** – This is a 12-bit field, the first two bits of which shall be '00'. The remaining 10 bits specify the number of bytes of the section, starting immediately following the section\_length field, and including the CRC. The value in this field shall not exceed 1021 (0x3FD).

**version\_number** – This 5-bit field is the version number of the whole transport stream description table. The version number shall be incremented by 1 modulo 32 whenever the definition of the transport stream description table changes. When the current\_next\_indicator is set to '1', then the version\_number shall be that of the currently applicable transport stream description table. When the current\_next\_indicator is set to '0', then the version\_number shall be that of the next applicable transport stream description table.

**current\_next\_indicator** – A 1-bit indicator, which, when set to '1', indicates that the transport stream description table sent is currently applicable. When the bit is set to '0', it indicates that the table sent is not yet applicable and shall be the next table to become valid.

**section\_number** – This 8-bit field gives the number of this section. The section\_number of the first section in the transport stream description table shall be 0x00. It shall be incremented by 1 with each additional section in the transport stream description table.

**last\_section\_number** – This 8-bit field specifies the number of the last section (that is, the section with the highest section\_number) of the complete transport stream description table.

**CRC\_32** – This is a 32-bit field that contains the CRC value that gives a zero output of the registers in the decoder defined in Annex A after processing the entire transport stream description section.

## 2.5 Program stream bitstream requirements

### 2.5.1 Program stream coding structure and parameters

The Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program stream coding layer allows one program of one or more elementary streams to be combined into a single stream. Data from each elementary stream are multiplexed together with information that allows synchronized presentation of the elementary streams within the program.

A program stream consists of one or more elementary streams from one program multiplexed together. Audio and video elementary streams consist of access units.

Elementary stream data is carried in PES packets. A PES packet consists of a PES packet header followed by packet data. PES packets are inserted into program stream packs.

The PES packet header begins with a 32-bit start-code that also identifies the stream (refer to Table 2-22) to which the packet data belongs. The PES packet header may contain just a presentation time stamp (PTS) or both a presentation timestamp and a decoding time stamp (DTS). The PES packet header also contains other optional fields. The packet data contains a variable number of contiguous bytes from one elementary stream.

In a program stream, PES packets are organized in packs. A pack commences with a pack header and is followed by zero or more PES packets. The pack header begins with a 32-bit start-code. The pack header is used to store timing and bitrate information.

The program stream begins with a system header that optionally may be repeated. The system header carries a summary of the system parameters defined in the stream.

This Recommendation | International Standard does not specify the coded data which may be used as part of conditional access systems. This Recommendation | International Standard does, however, provide mechanisms for program service providers to transport and identify this data for decoder processing, and to correctly reference data which are here specified.

### 2.5.2 Program stream system target decoder

The semantics of the program stream and the constraints on these semantics require exact definitions of decoding events and the times at which these events occur. The definitions needed are set out in this Specification using a hypothetical decoder known as the program stream system target decoder (P-STD).

The P-STD is a conceptual model used to define these terms precisely and to model the decoding process during the construction of program streams. The P-STD is defined only for this purpose. Neither the architecture of the P-STD nor the timing described precludes uninterrupted, synchronized playback of program streams from a variety of decoders with different architectures or timing schedules.



Figure 2-2 – Program stream system target decoder notation

The following notation is used to describe the program stream system target decoder and is partially illustrated in Figure 2‑2.

i, i′ are indices to bytes in the program stream. The first byte has index 0.

j is an index to access units in the elementary streams.

k, k′, k″ are indices to presentation units in the elementary streams.

n is an index to the elementary streams.

t(i) indicates the time in seconds at which the i-th byte of the program stream enters the system target decoder. The value t(0) is an arbitrary constant.

SCR(i) is the time encoded in the SCR field measured in units of the 27 MHz system clock where i is the byte index of the final byte of the system\_clock\_reference\_base field.

An(j) is the j-th access unit in elementary stream n. An(j) is indexed in decoding order.

tdn(j) is the decoding time, measured in seconds, in the system target decoder of the j-th access unit in elementary stream n.

Pn(k) is the k-th presentation unit in elementary stream n. Pn(k) is indexed in presentation order.

tpn(k) is the presentation time, measured in seconds, in the system target decoder of the k-th presentation unit in elementary stream n.

t is time measured in seconds.

Fn(t) is the fullness, measured in bytes, of the system target decoder input buffer for elementary stream n at time t.

Bn the input buffer in the system target decoder for elementary stream n.

BSn is the size of the system target decoder input buffer, measured in bytes, for elementary stream n.

Dn is the decoder for elementary stream n.

On is the reorder buffer for video elementary stream n.

#### 2.5.2.1 System clock frequency

Timing information referenced in P-STD is carried by several data fields defined in this Specification. The fields are defined in 2.5.3.3 and 2.4.3.6. This information is coded as the sampled value of a system clock.

The value of the system clock frequency is measured in Hz and shall meet the following constraints:

– 27 000 000 – 810 <= system\_clock\_frequency <= 27 000 000 + 810;

– rate of change of system\_clock\_frequency with time <= 75 × 10–3 Hz/s.

The notation "system\_clock\_frequency" is used in several places in this Recommendation | International Standard to refer to the frequency of a clock meeting these requirements. For notational convenience, equations in which SCR, PTS, or DTS appear, lead to values of time which are accurate to some integral multiple of (300 × 233/system\_clock\_frequency) seconds. This is due to the encoding of SCR timing information as 33 bits of 1/300 of the system clock frequency plus 9 bits for the remainder, and encoding as 33 bits of the system clock frequency divided by 300 for PTS and DTS.

#### 2.5.2.2 Input to the program stream system target decoder

Data from the program stream enters the system target decoder. The i-th byte enters at time t(i). The time at which this byte enters the system target decoder can be recovered from the input stream by decoding the input system clock reference (SCR) fields and the program\_mux\_rate field encoded in the pack header. The SCR, as defined in equation 2‑18, is coded in two parts: one, in units the period of 1/300 × the system clock frequency, called system\_clock\_reference\_base (see equation 2-19), and one, called system\_clock\_reference\_ext equation (see equation 2‑20), in units of the period of the system clock frequency. In the following the values encoded in these fields are denoted by SCR\_base(i) and SCR\_ext(i). The value encoded in the SCR field indicates time t(i), where i refers to the byte containing the last bit of the system\_clock\_reference\_base field.

Specifically:

(2-18)

where:

(2-19)

(2-20)

The input arrival time, t(i), as given in equation 2-21, for all other bytes shall be constructed from SCR(i) and the rate at which data arrives, where the arrival rate within each pack is the value represented in the program\_mux\_rate field in that pack's header.

 (2-21)

where:

i′ is the index of the byte containing the last bit of the system\_clock\_reference\_base field in the pack header

i is the index of any byte in the pack, including the pack header

SCR(i′) is the time encoded in the system clock reference base and extension fields in units of the system clock

program\_mux\_rate is a field defined in 2.5.3.3.

After delivery of the last byte of a pack there may be a time interval during which no bytes are delivered to the input of the P-STD.

#### 2.5.2.3 Buffering

The PES packet data from elementary stream n is passed to the input buffer for stream n, Bn. Transfer of byte i from the system target decoder input to Bn is instantaneous, so that byte i enters the buffer for stream n, of size BSn, at time t(i).

Bytes present in the pack header, system headers, program stream maps, program stream directories, or PES packet headers of the program stream such as SCR, DTS, PTS, and packet\_length fields, are not delivered to any of the buffers, but may be used to control the system.

The input buffer sizes BS1 through BSn are given by the P-STD buffer size parameter in the syntax in equations 2-16 and 2-17.

At the decoding time, tdn(j), all data for the access unit that has been in the buffer longest, An(j), and any stuffing bytes that immediately precede it that are present in the buffer at the time tdn(j), are removed instantaneously at time tdn(j). The decoding time tdn(j) is specified in the DTS or PTS fields. Decoding times tdn(j + 1), tdn(j + 2), ... of access units without encoded DTS or PTS fields which directly follow access unit j may be derived from information in the elementary stream. Refer to Annex C of Rec. ITU-T H.262 | ISO/IEC 13818-2, ISO/IEC 13818-3, ISO/IEC 11172-2 or ISO/IEC 11172-3. Also refer to 2.7.5. As the access unit is removed from the buffer, it is instantaneously decoded to a presentation unit.

The program stream shall be constructed and t(i) shall be chosen so that the input buffers of size BS1 through BSn neither overflow nor underflow in the program system target decoder. That is:



for all t and n,

and:



instantaneously before t = t(0).

Fn(t) is the instantaneous fullness of P-STD buffer Bn.

An exception to this condition is that the P-STD buffer Bn may underflow when the low\_delay flag in the video sequence header is set to '1' (refer to 2.4.2.7) or when trick\_mode status is true (refer to 2.4.3.8).

For all program streams, the delay caused by system target decoder input buffering shall be less than or equal to one second except for still picture video data and ISO/IEC 14496 streams. The input buffering delay is the difference in time between a byte entering the input buffer and when it is decoded.

Specifically: in the case of no still picture video data and no ISO/IEC 14496 stream the delay is constrained by:



in the case of still picture video data the delay is constrained by:



in the case of ISO/IEC 14496 streams the delay is constrained by:



for all bytes contained in access unit j.

For program streams, all bytes of each pack shall enter the P-STD before any byte of a subsequent pack.

When the **low\_delay** flag in the video sequence extension is set to '1' (refer to 6.2.2.3 of Rec. ITU-T H.262 | ISO/IEC 13818-2), the VBV buffer may underflow. In this case when the P-STD elementary stream buffer Bn is examined at the time specified by tdn(j), the complete data for the access unit may not be present in the buffer Bn. When this case arises, the buffer shall be re-examined at intervals of two field-periods until the data for the complete access unit is present in the buffer. At this time the entire access unit shall be removed from buffer Bn instantaneously.

VBV buffer underflow is allowed to occur continuously without limit. The P-STD decoder shall remove access unit data from buffer Bn at the earliest time consistent with the paragraph above and any DTS or PTS values encoded in the bitstream. The decoder may be unable to re-establish correct decoding and display times as indicated by DTS and PTS until the VBV buffer underflow situation ceases and a PTS or DTS is found in the bitstream.

#### 2.5.2.4 PES streams

It is possible to construct a stream of data as a contiguous stream of PES packets each containing data of the same elementary stream and with the same stream\_id. Such a stream is called a PES stream. The PES-STD model for a PES stream is identical to that for the program stream, with the exception that the Elementary Stream Clock Reference (ESCR) is used in place of the SCR, and ES\_rate in place of program\_mux\_rate. The demultiplexor sends data to only one elementary stream buffer.

NOTE – In the following equations, unit conversion should be implicitly performed as appropriate.

As a PES stream only carries a single elementary stream, the buffer sizes in the PES-STD do not account for multiplexing with other elementary streams, but only for multiplexing of the elementary stream carried in the PES stream with PES headers, pack headers and system headers.

Buffer sizes BSn in the PES-STD model are defined as follows:

– *For Rec. ITU-T H.262 | ISO/IEC 13818-2 video:*

BSn = VBVmax[profile, level] + BSoh

BSoh = (1/750) seconds × Rmax[profile, level], where VBVmax[profile, level] and Rmax[profile, level] are the maximum VBV size and bit rate per profile, level, and layer as defined in Tables 8-14 and 8-13, respectively, of Rec. ITU-T H.262 | ISO/IEC 13818-2. BSoh is allocated for PES packet header overhead.

– *For ISO/IEC 11172-2 video:*

BSn = VBVmax + BSoh

BSoh = (1/750) seconds × Rmax, where Rmax and vbv\_max refer to the maximum bitrate and maximum vbv\_buffer\_size for a constrained parameter bitstream in ISO/IEC 11172-2 respectively.

– *For ISO/IEC 11172-3 or ISO/IEC 13818-3 audio:*

BSn = 2848 bytes

– *For ISO/IEC 13818-7 ADTS audio*:

BSn = 2848 bytes if 1-2 channels  
 BSn = 7200 bytes if 3-8 channels  
 BSn = 10800 bytes if 9-12 channels  
 BSn = 43200 bytes if 13-48 channels

Note that the above numbers differ from the BSn numbers specified in 2.4.3.2 due to the fact that a PES stream carries a single elementary stream only.

– *For ISO/IEC 14496-3 audio, except for ISO/IEC 14496-3 DST, ALS and SLS*:

BSn = 2848 bytes if 1-2 channels  
 BSn = 7200 bytes if 3-8 channels  
 BSn = 10800 bytes if 9-12 channels  
 BSn = 43200 bytes if 13-48 channels

Note that the above numbers differ from the BSn numbers specified in 2.11.2.2 due to the fact that a PES stream carries a single elementary stream only.

– *For ISO/IEC 14496-3 DST-64 audio*:

BSn = 5000 × (number of channels) bytes; hence for stereo BSn = 10 000 bytes  
 and for 5.1 surround sound audio BSn = 30 000 bytes

– *For ISO/IEC 14496-3 DST-128 audio*:

BSn = 10 000 × (number of channels) bytes; hence for stereo BSn = 20 000 bytes  
 and for 5.1 surround sound audio BSn = 60 000 bytes

– *For ISO/IEC 14496-3 DST-256 audio*:

BSn = 20 000 × (number of channels) bytes; hence for stereo BSn = 40 000 bytes  
 and for 5.1 surround sound audio BSn = 120 000 bytes

– *For ISO/IEC 14496-3 ALS and SLS audio*:

BSn = 33 000 × (number of channels) bytes; hence for stereo BSn = 66 000 bytes  
 and for 5.1 surround sound audio BSn = 198 000 bytes

– *For Rec. ITU-T H.264 | ISO/IEC 14496-10 video:*

BSn = 1200 × MaxCPB[level] + BSoh

where MaxCPB[level] is defined in Table A.1 (Level Limits) in Rec. ITU-T H.264 | ISO/IEC 14496-10 for each level.

#### 2.5.2.5 Decoding and presentation

Decoding and presentation in the program stream system target decoder are the same as defined for the transport stream system target decoder in 2.4.2.5 and 2.4.2.6 respectively.

#### 2.5.2.6 P-STD extensions for carriage of ISO/IEC 14496 data

For decoding of ISO/IEC 14496 data carried in a program stream the P-STD model is extended. For decoding of individual ISO/IEC 14496 elementary streams in the P-STD see 2.11.2. Clause 2.11.3 defines P-STD extensions and parameters for decoding of ISO/IEC 14496 scenes and associated streams.

#### 2.5.2.7 P-STD extensions for carriage of Rec. ITU-T H.264 | ISO/IEC 14496-10 video

For decoding of AVC video streams conforming to one or more profiles defined in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496-10 video streams carried in a program stream in the P-STD model, see 2.14.3.2, for decoding of AVC video streams conforming to one or more profiles defined in Annex G of Rec. ITU-T H.264 | ISO/IEC 14496-10 carried in a program stream in the P-STD model, see 2.14.3.6 and for decoding of AVC video streams conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10 carried in a program stream in the P-STD model, see 2.14.3.8.

#### 2.5.2.8 P-STD extensions for carriage of ISO/IEC 14496-17 text streams

For decoding of ISO/IEC 14496-17 text streams carried in a program stream in the P-STD model, see 2.15.3.2.

### 2.5.3 Specification of the program stream syntax and semantics

The following syntax describes a stream of bytes.

#### 2.5.3.1 Program stream

See Table 2-37.

Table 2-37 – Program stream

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| MPEG2\_program\_stream() { |  |  |
| do { |  |  |
| pack() |  |  |
| } while (nextbits() = = pack\_start\_code) |  |  |
| **MPEG\_program\_end\_code** | **32** | **bslbf** |
| } |  |  |

#### 2.5.3.2 Semantic definition of fields in program stream

**MPEG\_program\_end\_code** *–* The MPEG\_program\_end\_code is the bit string '0000 0000 0000 0000 0000 0001 1011 1001' (0x000001B9). It terminates the program stream.

#### 2.5.3.3 Pack layer of program stream

See Tables 2-38 and 2-39.

Table 2-38 – Program stream pack

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| pack() { |  |  |
| pack\_header() |  |  |
| while (nextbits() = -= packet\_start\_code\_prefix) { |  |  |
| PES\_packet() |  |  |
| } |  |  |
| } |  |  |

Table 2-39 – Program stream pack header

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| pack\_header() { |  |  |
| **pack\_start\_code** | **32** | **bslbf** |
| '01' | **2** | **bslbf** |
| **system\_clock\_reference\_base [32..30]** | **3** | **bslbf** |
| **marker\_bit** | **1** | **bslbf** |
| **system\_clock\_reference\_base [29..15]** | **15** | **bslbf** |
| **marker\_bit** | **1** | **bslbf** |
| **system\_clock\_reference\_base [14..0]** | **15** | **bslbf** |
| **marker\_bit** | **1** | **bslbf** |
| **system\_clock\_reference\_extension** | **9** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **program\_mux\_rate** | **22** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **marker\_bit** | **1** | **bslbf** |
| **reserved** | **5** | **bslbf** |
| **pack\_stuffing\_length** | **3** | **uimsbf** |
| for (i = 0; i < pack\_stuffing\_length; i++) { |  |  |
| stuffing\_byte | **8** | **bslbf** |
| } |  |  |
| if (nextbits() = = system\_header\_start\_code) { |  |  |
| system\_header () |  |  |
| } |  |  |
| } |  |  |

#### 2.5.3.4 Semantic definition of fields in program stream pack

**pack\_start\_code** – The pack\_start\_code is the bit string '0000 0000 0000 0000 0000 0001 1011 1010' (0x000001BA). It identifies the beginning of a pack.

**system\_clock\_reference\_base; system\_clock\_reference\_extension** –The system clock reference (SCR) is a 42-bit field coded in two parts. The first part, system\_clock\_reference\_base, is a 33-bit field whose value is given by SCR\_base(i) as given in equation 2-19. The second part, system\_clock\_reference\_extension, is a 9-bit field whose value is given by SCR\_ext(i), as given in equation 2-20. The SCR indicates the intended time of arrival of the byte containing the last bit of the system\_clock\_reference\_base at the input of the program target decoder.

The frequency of coding requirements for the SCR field are given in 2.7.1.

**marker\_bit** – A marker\_bit is a 1-bit field that has the value '1'.

**program\_mux\_rate** – This is a 22-bit integer specifying the rate at which the P-STD receives the program stream during the pack in which it is included. The value of program\_mux\_rate is measured in units of 50 bytes/second. The value '0' is forbidden. The value represented in program\_mux\_rate is used to define the time of arrival of bytes at the input to the P-STD in 2.5.2. The value encoded in the program\_mux\_rate field may vary from pack to pack in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program multiplexed stream.

**pack\_stuffing\_length** – A 3-bit integer specifying the number of stuffing bytes which follow this field.

**stuffing\_byte** – This is a fixed 8-bit value equal to '1111 1111' that can be inserted by the encoder, for example to meet the requirements of the channel. It is discarded by the decoder. In each pack header no more than 7 stuffing bytes shall be present.

#### 2.5.3.5 System header

See Table 2-40.

Table 2-40 – Program stream system header

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| system\_header () { |  |  |
| **system\_header\_start\_code** | **32** | **bslbf** |
| **header\_length** | **16** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **rate\_bound** | **22** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **audio\_bound** | **6** | **uimsbf** |
| **fixed\_flag** | **1** | **bslbf** |
| **CSPS\_flag** | **1** | **bslbf** |
| **system\_audio\_lock\_flag** | **1** | **bslbf** |
| **system\_video\_lock\_flag** | **1** | **bslbf** |
| **marker\_bit** | **1** | **bslbf** |
| **video\_bound** | **5** | **uimsbf** |
| **packet\_rate\_restriction\_flag** | **1** | **bslbf** |
| **reserved\_bits** | **7** | **bslbf** |
| while (nextbits ()   '1') { |  |  |
| **stream\_id** | **8** | **uimsbf** |
| if (stream\_id   '1011 0111') { |  |  |
| **'11'** | **2** | **bslbf** |
| **'000 0000'** | **7** | **bslbf** |
| **stream\_id\_extension** | **7** | **uimsbf** |
| **'1011 0110'** | **8** | **bslbf** |
| **'11'** | **2** | **bslbf** |
| **P-STD\_buffer\_bound\_scale** | **1** | **bslbf** |
| **P-STD\_buffer\_size\_bound** | **13** | **uimsbf** |
| } |  |  |
| else { |  |  |
| **'11'** | **2** | **bslbf** |
| **P-STD\_buffer\_bound\_scale** | **1** | **bslbf** |
| **P-STD\_buffer\_size\_bound** | **13** | **uimsbf** |
| } |  |  |
| } |  |  |
| } |  |  |

#### 2.5.3.6 Semantic definition of fields in system header

**system\_header\_start\_code** – The system\_header\_start\_code is the bit string '0000 0000 0000 0000 0000 0001 1011 1011' (0x000001BB). It identifies the beginning of a system header.

**header\_length** –This 16-bit field indicates the length in bytes of the system header following the header\_length field. Future extensions of this Specification may extend the system header.

**rate\_bound** – A 22-bit field. The rate\_bound is an integer value greater than or equal to the maximum value of the program\_mux\_rate field coded in any pack of the program stream. It may be used by a decoder to assess whether it is capable of decoding the entire stream.

**audio\_bound** –A 6-bit field. The audio\_bound is an integer in the inclusive range from 0 to 32 and is set to a value greater than or equal to the maximum number of ISO/IEC 13818-3 and ISO/IEC 11172-3 audio streams in the program stream for which the decoding processes are simultaneously active. For the purpose of this subclause, the decoding process of an ISO/IEC 13818-3 or ISO/IEC 11172-3 audio stream is active if the STD buffer is not empty or if a Presentation Unit is being presented in the P-STD model.

**fixed\_flag** –The fixed\_flag is a 1-bit flag. When set to '1' fixed bitrate operation is indicated. When set to '0' variable bitrate operation is indicated. During fixed bitrate operation, the value encoded in all system\_clock\_reference fields in the multiplexed Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream shall adhere to the following linear equation:

*SCR\_base*(*i*) = ((c1 × i + c2) *DIV* 300) *%* 233(2-22)

*SCR\_ext*(*i*) = ((c1 × i + c2) *DIV* 300) *%* 300(2-23)

where:

c1 is a real-valued constant valid for all i.

c2 is a real-valued constant valid for all i.

i is the index in the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 multiplexed stream of the byte containing the final bit of any system\_clock\_reference field in the stream.

**CSPS\_flag** – The CSPS\_flag is a 1-bit field. If its value is set to '1' the program stream meets the constraints defined in 2.7.9.

**system\_audio\_lock\_flag** – The system\_audio\_lock\_flag is a 1-bit field indicating that there is a specified, constant rational relationship between the audio sampling rate and the system\_clock\_frequency in the system target decoder. The system\_clock\_frequency is defined in 2.5.2.1 and the audio sampling rate is specified in ISO/IEC 13818-3. The system\_audio\_lock\_flag may only be set to '1' if, for all presentation units in all audio elementary streams in the program stream, the ratio of system\_clock\_frequency to the actual audio sampling rate, SCASR, is constant and equal to the value indicated in the following table at the nominal sampling rate indicated in the audio stream.

 (2-24)

The notation  denotes real division.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Nominal audio sampling frequency (kHz) | 16 | 32 | 22.05 | 44.1 | 24 | 48 |
| SCASR | 27 000 000 ------------- 16 000 | 27 000 000 ------------- 32 000 | 27 000 000 ------------- 22 050 | 27 000 000 ------------- 44 100 | 27 000 000 ------------- 24 000 | 27 000 000 ------------- 48 000 |

**system\_video\_lock\_flag** – The system\_video\_lock\_flag is a 1-bit field indicating that there is a specified, constant rational relationship between the video time base and the system clock frequency in the system target decoder. The system\_video\_lock\_flag may only be set to '1' if, for all presentation units in all video elementary streams in the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, the ratio of system\_clock\_frequency to the frequency of the actual video time base is constant.

For ISO/IEC 11172-2 and Rec. ITU-T H.262 | ISO/IEC 13818-2 video streams, if the system\_video\_lock\_flag is set to '1', then the ratio of system\_clock\_frequency to the actual video frame rate, SCFR, shall be constant and equal to the value indicated in the following table at the nominal frame rate indicated in the video stream.

For ISO/IEC 14496-2 video streams, if the system\_video\_lock\_flag is set to '1', then the time base of the ISO/IEC 14496‑2 video stream, as defined by vop\_time\_increment\_resolution, shall be locked to the STC and shall be exactly equal to N times system\_clock\_frequency divided by K, with N and K integers that have a fixed value within each visual object sequence, with K greater than or equal to N.

For Rec. ITU-T H.264 | ISO/IEC 14496-10 video streams, the frequency of the AVC time base is defined by the AVC parameter time\_scale. If the system\_video\_lock\_flag is set to '1' for an AVC video stream or for a video sub-bitstream, then the frequency of the AVC time base shall be locked to the STC and shall be exactly equal to N times system\_clock\_frequency divided by K, with N and K integers that have a fixed value within each AVC video sequence, with K greater than or equal to N.

 (2-25)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nominal frame rate (Hz) | 23.976 | 24 | 25 | 29.97 | 30 | 50 | 59.94 | 60 |
| SCFR | 1 126 125 | 1 125 000 | 1 080 000 | 900 900 | 900 000 | 540 000 | 450 450 | 450 000 |

The values of the ratio SCFR are exact. The actual frame rate differs slightly from the nominal rate in cases where the nominal rate is 23.976, 29.97, or 59.94 frames per second.

**video\_bound** – The video\_bound is a 5-bit integer in the inclusive range from 0 to 16 and is set to a value greater than or equal to the maximum number of video streams in the program stream of which the decoding processes are simultaneously active. For the purpose of this subclause, the decoding process of a video stream is active if one of the buffers in the P‑STD model is not empty, or if a Presentation Unit is being presented in the P-STD model.

**packet\_rate\_restriction\_flag** – The packet\_rate\_restriction\_flag is a 1-bit flag. If the CSPS flag is set to '1', the packet\_rate\_restriction\_flag indicates which constraint is applicable to the packet rate, as specified in 2.7.9. If the CSPS flag is set to value of '0', then the meaning of the packet\_rate\_restriction\_flag is undefined.

**reserved\_bits** – This 7-bit field is reserved for future use by ISO/IEC. Until otherwise specified by ITU-T | ISO/IEC it shall have the value '111 1111'.

**stream\_id** –The stream\_id is an 8-bit field that indicates the coding and elementary stream number of the stream to which the following P-STD\_buffer\_bound\_scale and P-STD\_buffer\_size\_bound fields refer.

If stream\_id equals '1011 1000' the P-STD\_buffer\_bound\_scale and P-STD\_buffer\_size\_bound fields following the stream\_id refer to all audio streams in the program stream.

If stream\_id equals '1011 1001' the P-STD\_buffer\_bound\_scale and P-STD\_buffer\_size\_bound fields following the stream\_id refer to all video streams in the program stream.

If stream\_id equals '1111 1101', the P-STD\_buffer\_bound\_scale and P-STD\_buffer\_size\_bound fields following the stream\_id refer to all elementary streams with an extended\_stream\_id in the program stream, independent of the coded value of the stream\_id\_extension in the PES header of those streams.

If stream\_id equals '1011 0111', the following stream\_id\_extension field shall be interpreted as referring to the stream coding and elementary stream number according to Table 2-27.

If the stream\_id takes on any other value it shall be a byte value greater than or equal to '1011 1100' and shall be interpreted as referring to the stream coding and elementary stream number according to Table 2-22.

Each elementary stream present in the program stream shall have its P-STD\_buffer\_bound\_scale and P‑STD\_buffer\_size\_bound specified exactly once by this mechanism in each system header.

**stream\_id\_extension** –The stream\_id\_extension is a 7-bit field. In case the stream\_id field is coded with the value '1011 0111', then the stream\_id\_extension indicates the coding and elementary stream number of the stream with an extended\_stream\_id to which the P-STD\_buffer\_bound\_scale and P-STD\_buffer\_size\_bound fields following the stream\_id\_extension field refer.

**P-STD\_buffer\_bound\_scale** – The P-STD\_buffer\_bound\_scale is a 1-bit field that indicates the scaling factor used to interpret the subsequent P-STD\_buffer\_size\_bound field. If the preceding stream\_id indicates an audio stream, P‑STD\_buffer\_bound\_scale shall have the value '0'. If the preceding stream\_id indicates a video stream, P‑STD\_buffer\_bound\_scale shall have the value '1'. For all other stream types, the value of the P‑STD\_buffer\_bound\_scale may be either '1' or '0'.

**P-STD\_buffer\_size\_bound** – The P-STD\_buffer\_size\_bound is a 13-bit unsigned integer defining a value greater than or equal to the maximum P-STD input buffer size, BSn, over all packets for stream n in the program stream. If P‑STD\_buffer\_bound\_scale has the value '0', then P-STD\_buffer\_size\_bound measures the buffer size bound in units of 128 bytes. If P-STD\_buffer\_bound\_scale has the value '1', then P-STD\_buffer\_size\_bound measures the buffer size bound in units of 1024 bytes. Thus:



else:

*BSn* ≤ *P* – *STD\_buffer\_size\_bound* × 1024

#### 2.5.3.7 Packet layer of program stream

The packet layer of the program stream is defined by the PES packet layer in 2.4.3.6.

### 2.5.4 Program stream map

The Program Stream Map (PSM) provides a description of the elementary streams in the program stream and their relationship to one another. When carried in a transport stream this structure shall not be modified. The PSM is present as a PES packet when the stream\_id value is 0xBC (refer to Table 2-22).

NOTE – This syntax differs from the PES packet syntax described in 2.4.3.6.

Definition for the descriptor() fields may be found in 2.6.

#### 2.5.4.1 Syntax of program stream map

See Table 2-41.

Table 2-41 – Program stream map

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| program\_stream\_map() { |  |  |
| **packet\_start\_code\_prefix** | **24** | **bslbf** |
| **map\_stream\_id** | **8** | **uimsbf** |
| **program\_stream\_map\_length** | **16** | **uimsbf** |
| **current\_next\_indicator** | **1** | **bslbf** |
| **single\_extension\_stream\_flag** | **1** | **bslbf** |
| **reserved** | **1** | **bslbf** |
| **program\_stream\_map\_version** | **5** | **uimsbf** |
| **reserved** | **7** | **bslbf** |
| **marker\_bit** | **1** | **bslbf** |
| **program\_stream\_info\_length** | **16** | **uimsbf** |
| for (i  0; i  N; i++) { |  |  |
| descriptor() |  |  |
| } |  |  |
| **elementary\_stream\_map\_length** | **16** | **uimsbf** |
| for (i  0; i  N1; i++) { |  |  |
| **stream\_type** | **8** | **uimsbf** |
| **elementary\_stream\_id** | **8** | **uimsbf** |
| **elementary\_stream\_info\_length** | **16** | **Uimsbf** |
| if ( elementary\_stream\_id = = 0xFD &&  single\_extension\_stream\_flag = = 0) { |  |  |
| **pseudo\_descriptor\_tag** | **8** | **Uimsbf** |
| **pseudo\_descriptor\_length** | **8** | **Uimsbf** |
| **marker\_bit** | **1** | **Bslbf** |
| **elementary\_stream\_id\_extension** | **7** | **Uimsbf** |
| for (i  3; i  N2; i++) { |  |  |
| descriptor() |  |  |
| } |  |  |
| } |  |  |
| else { |  |  |
| for (i  0; i  N2; i++) { |  |  |
| descriptor() |  |  |
| } |  |  |
| } |  |  |
| } |  |  |
| **CRC**\_**32** | **32** | **rpchof** |
| } |  |  |

#### 2.5.4.2 Semantic definition of fields in program stream map

**packet\_start\_code\_prefix** –The packet\_start\_code\_prefix is a 24-bit code. Together with the map\_stream\_id that follows it constitutes a packet start code that identifies the beginning of a packet. The packet\_start\_code\_prefix is the bit string '0000 0000 0000 0000 0000 0001' (0x000001 in hexadecimal).

**map\_stream\_id** – This is an 8-bit field whose value shall be 0xBC.

**program\_stream\_map\_length** –The program\_stream\_map\_length is a 16-bit field indicating the total number of bytes in the program\_stream\_map immediately following this field. The maximum value of this field is 1018 (0x3FA).

**single\_extension\_stream\_flag** – This is a 1-bit field indicating, when set to '1', that the program stream contains at most one elementary stream with stream\_id equal to 0xFD.

**current\_next\_indicator** – This is a 1-bit field, when set to '1' indicates that the program stream map sent is currently applicable. When the bit is set to '0', it indicates that the program stream map sent is not yet applicable and shall be the next table to become valid.

**program\_stream\_map\_version** – This 5-bit field is the version number of the whole program stream map. The version number shall be incremented by 1 modulo 32 whenever the definition of the program stream map changes. When the current\_next\_indicator is set to '1', then the program\_stream\_map\_version shall be that of the currently applicable program stream map. When the current\_next\_indicator is set to '0', then the program\_stream\_map\_version shall be that of the next applicable program stream map.

**program\_stream\_info\_length** –The program\_stream\_info\_length is a 16-bit field indicating the total length of the descriptors immediately following this field.

**marker\_bit** – A marker\_bit is a 1-bit field that has the value '1'.

**elementary\_stream\_map\_length** –This is a 16-bit field specifying the total length, in bytes, of all elementary stream information in this program stream map. It includes the stream\_type, elementary\_stream\_id, and elementary\_stream\_info\_length fields.

**stream\_type** –This 8-bit field specifies the type of the stream according to Table 2-34. The stream\_type field shall only identify elementary streams contained in PES packets. A value of 0x05 is prohibited.

**elementary\_stream\_id** – The elementary\_stream\_id is an 8-bit field indicating the value of the stream\_id field in the PES packet headers of PES packets in which this elementary stream is stored. When elementary\_stream\_id is equal to 0xFD, the following applies:

– If single\_extension\_stream\_flag is equal to 1, this indicates that the program stream contains only one elementary stream with stream\_id equal to 0xFD. Note that the type of this elementary stream is signalled by the encoded value of the stream\_id\_extension field in the PES headers of PES packets carrying this elementary stream.

– Otherwise (single\_extension\_stream\_flag is equal to 0), the elementary\_stream\_id\_extension field is present to identify the elementay stream.

**elementary\_stream\_info\_length** –The elementary\_stream\_info\_length is a 16-bit field indicating the length in bytes of the descriptors and, when present, the pseudo\_descriptor\_tag, the pseudo\_descriptor\_length, and the elementary\_stream\_id\_extension (and associated marker\_bit) data immediately following this field.

**pseudo\_descriptor\_tag** – This is an 8-bit unsigned integer that shall be coded with the value 0x01; note that the use of value 0x01 for descriptor tags is forbidden in Table 2-45.

**pseudo\_descriptor\_length** – The pseudo\_descriptor\_length is an 8-bit unsigned integer that shall be coded with the value 1.

**elementary\_stream\_id\_extension** –This 7-bit field, when present, indicates the encoded value of the elementary\_stream\_id\_extension field in the PES packet headers of PES packets in which this elementary stream is stored.

**CRC\_32** – This is a 32-bit field that contains the CRC value that gives a zero output of the registers in the decoder defined in Annex A after processing the entire program stream map.

### 2.5.5 Program stream directory

The directory for an entire stream is made up of all the directory data carried by program stream directory packets identified with the directory\_stream\_id. The syntax for program\_stream\_directory packets is defined in Table 2-42.

NOTE 1 – This syntax differs from the PES packet syntax described in 2.4.3.6.

Directory entries may be required to reference I-pictures in a video stream as defined in Rec. ITU‑T H.262 | ISO/IEC 13818-2 and ISO/IEC 11172-2. If an I-picture that is referenced in a directory entry is preceded by a sequence header with no intervening picture headers, the directory entry shall reference the first byte of the sequence header. If an I‑picture that is referenced in a directory entry is preceded by a group of pictures header with no intervening picture headers and no immediately preceding sequence header, the directory entry shall reference the first byte of the group of pictures header. Any other picture that a directory entry references shall be referenced by the first byte of the picture header.

NOTE 2 – It is recommended that I-pictures immediately following a sequence header should be referenced in directory structures so that the directory contains an entry at every point where the decoder may be reset completely.

For AVC video streams conforming to one or more profiles defined in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496‑10, directory entries may be required to reference IDR picture or pictures associated with a recovery point SEI message in an AVC video stream. Each such directory entry shall refer to the first byte of an AVC access unit.

For video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex G of Rec. ITU‑T H.264 | ISO/IEC 14496-10, directory entries may be required to reference IDR picture or pictures to be re‑assembled from video sub-bitstreams and associated with a recovery point SEI message present in a video sub‑bitstream. Each such directory entry shall refer to the first byte of an SVC dependency representation.

For MVC video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10, directory entries may be required to reference IDR picture or pictures to be re‑assembled from MVC video sub-bitstreams and associated with a recovery point SEI message present in an MVC video sub-bitstream. Each such directory entry shall refer to the first byte of an MVC view-component subset.

Directory references to audio streams as defined in ISO/IEC 13818-3 and ISO/IEC 11172-3 shall be the syncword of the audio frame.

NOTE 3 – It is recommended that the distance between referenced access units not exceed half a second.

Access units shall be referenced in a program\_stream\_directory packet in the same order that they appear in the bitstream.

#### 2.5.5.1 Syntax of program stream directory packet

See Table 2-42.

Table 2-42 – Program stream directory packet

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| directory\_PES\_packet(){ |  |  |
| **packet\_start\_code\_prefix** | **24** | **bslbf** |
| **directory\_stream\_id** | **8** | **uimsbf** |
| **PES\_packet\_length** | **16** | **uimsbf** |
| **number\_of\_access\_units** | **15** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **prev\_directory\_offset[44..30]** | **15** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **prev\_directory\_offset[29..15]** | **15** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **prev\_directory\_offset[14..0]** | **15** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **next\_directory\_offset[44..30]** | **15** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **next\_directory\_offset[29..15]** | **15** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **next\_directory\_offset[14..0]** | **15** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| for (i  0; i  number\_of\_access\_units; i) { |  |  |
| **packet\_stream\_id** | **8** | **uimsbf** |
| **PES\_header\_position\_offset\_sign** | **1** | **tcimsbf** |
| **PES\_header\_position\_offset[43..30]** | **14** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **PES\_header\_position\_offset[29..15]** | **15** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **PES\_header\_position\_offset[14..0]** | **15** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **reference\_offset** | **16** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| if (packet\_stream\_id = = 0xFD) { |  |  |
| **packet\_stream\_id\_extension\_msbs** | **3** | **uimsbf** |
| } |  |  |
| else { |  |  |
| **reserved** | **3** | **bslbf** |
| } |  |  |
| **PTS[32..30]** | **3** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **PTS[29..15]** | **15** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **PTS[14..0]** | **15** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **bytes\_to\_read[22..8]** | **15** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **bytes\_to\_read[7..0]** | **8** | **uimsbf** |
| **marker\_bit** | **1** | **bslbf** |
| **intra\_coded\_indicator** | **1** | **bslbf** |
| **coding\_parameters\_indicator** | **2** | **bslbf** |
| if (packet\_stream\_id = = 0xFD) { |  |  |
| **packet\_stream\_id\_extension\_lsbs** | **4** | **uimsbf** |
| } |  |  |
| else { |  |  |
| **reserved** | **4** | **bslbf** |
| } |  |  |
| } |  |  |
| } |  |  |

#### 2.5.5.2 Semantic definition of fields in program stream directory

**packet\_start\_code\_prefix** – The packet\_start\_code\_prefix is a 24-bit code. Together with the stream\_id that follows, it constitutes a packet startcode that identifies the beginning of a packet. The packet\_start\_code\_prefix is the bit string '0000 0000 0000 0000 0000 0001' (0x000001 in hexadecimal).

**directory\_stream\_id** – This 8-bit field shall have a value '1111 1111' (0xFF).

**PES\_packet\_length** – The PES\_packet\_length is a 16-bit field indicating the total number of bytes in the program\_stream\_directory immediately following this field (refer to Table 2-21).

**number\_of\_access\_units** – This 15-bit field is the number of access\_units that are referenced in this Directory PES packet.

**prev\_directory\_offset** – This 45-bit unsigned integer gives the byte address offset of the first byte of the packet start code of the previous program stream directory packet. This address offset is relative to the first byte of the start code of the packet which contains this previous\_directory\_offset field. The value '0' indicates that there is no previous program stream directory packet.

**next\_directory\_offset** – This 45-bit unsigned integer gives the byte address offset of the first byte of the packet start code of the next program stream directory packet. This address offset is relative to the first byte of the start code of the packet which contains this next\_directory\_offset field. The value '0' indicates that there is no next program stream directory packet.

**packet\_stream\_id** – This 8-bit field is the stream\_id of the elementary stream that contains the access unit referenced by this directory entry.

**PES\_header\_position\_offset\_sign** – This 1-bit field is the arithmetic sign for the PES\_header\_position\_offset described immediately following. A value of '0' indicates that the PES\_header\_position\_offset is a positive offset. A value of '1' indicates that the PES\_header\_position\_offset is a negative offset.

**PES\_header\_position\_offset** – This 44-bit unsigned integer gives the byte offset address of the first byte of the PES packet containing the access unit referenced. The offset address is relative to the first byte of the start-code of the packet containing this PES\_header\_position\_offset field. The value '0' indicates that no access unit is referenced.

**reference\_offset** – This 16-bit field is an unsigned integer indicating the position of the first byte of the referenced access unit, measured in bytes relative to the first byte of the PES packet containing the first byte of the referenced access unit.

**PTS (presentation\_time\_stamp)** – This 33-bit field is the PTS of the access unit that is referenced. The semantics of the coding of the PTS field are as described in 2.4.3.6.

**bytes\_to\_read** –This 23-bit unsigned integer is the number of bytes in the program stream after the byte indicated by reference\_offset that are needed to decode the access unit completely. This value includes any bytes multiplexed at the systems layer including those containing information from other streams.

**intra\_coded\_indicator** – This is a 1-bit flag. When set to '1' it indicates that the referenced access unit is not predictively coded. This is independent of other coding parameters that might be needed to decode the access unit. For example, this field shall be coded as '1' for video Intra frames, whereas for 'P' and 'B' frames this bit shall be coded as '0'. For all PES packets containing data which is not from a Rec. ITU-T H.262 | ISO/IEC 13818-2 video stream, this field is undefined (see Table 2-43).

Table 2-43 – Intra\_coded indicator

|  |  |
| --- | --- |
| Value | Meaning |
| '0' | Not Intra |
| '1' | Intra |

**coding\_parameters\_indicator** – This 2-bit field is used to indicate the location of coding parameters that are needed to decode the access units referenced (see Table 2-44). For example, this field can be used to determine the location of quantization matrices for video frames.

Table 2-44 – Coding\_parameters indicator

|  |  |
| --- | --- |
| Value | Meaning |
| '00' | All coding parameters are set to their default values |
| '01' | All coding parameters are set in this access unit, at least one of them is not set to a default |
| '10' | Some coding parameters are set in this access unit |
| '11' | No coding parameters are coded in this access unit |

**packet\_stream\_id\_extension\_msbs** –This 3-bit field is present if packet\_stream\_id equals 0xFD; its coding is specified below.

**packet\_stream\_id\_extension\_lsbs** –This 4-bit field is present if packet\_stream\_id equals 0xFD; its coding is specified below.

If packet\_stream\_id is equal to 0xFD, the packet\_stream\_id\_extension indicates the encoded value of the stream\_id\_extension in the PES header of the PES packet(s) containing the access unit referenced by this directory entry. The value of the packet\_stream\_id\_extension is specified by:

*packet\_stream\_id\_extension = packet\_stream\_id\_extension\_msbs \* 16 + packet\_stream\_id\_extension\_lsbs*

## 2.6 Program and program element descriptors

Program and program element descriptors are structures which may be used to extend the definitions of programs and program elements. All descriptors have a format which begins with an 8-bit tag value. The tag value is followed by an 8‑bit descriptor length and data fields.

### 2.6.1 Semantic definition of fields in program and program element descriptors

The following semantics apply to the descriptors defined in 2.6.2 through the end of 2.6.

**descriptor\_tag** – The descriptor\_tag is an 8-bit field which identifies each descriptor.

Table 2-45 provides the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 defined, Rec. ITU-T H.222.0 | ISO/IEC 13818-1 reserved, and user available descriptor tag values. An 'X' in the TS or PS columns indicates the applicability of the descriptor to either the transport stream or program stream respectively. Note that the meaning of fields in a descriptor may depend on which stream it is used in. Each case is specified in the descriptor semantics below.

**descriptor\_length** – The descriptor\_lengthis an 8-bit field specifying the number of bytes of the descriptor immediately following descriptor\_length field.

| Table 2-45 – Program and program element descriptors | | | | |
| --- | --- | --- | --- | --- |
| descriptor\_tag | TS | PS | Identification | |
| 0 | n/a | n/a | Reserved | |
| 1 | n/a | X | Forbidden | |
| 2 | X | X | video\_stream\_descriptor | |
| 3 | X | X | audio\_stream\_descriptor | |
| 4 | X | X | hierarchy\_descriptor | |
| 5 | X | X | registration\_descriptor | |
| 6 | X | X | data\_stream\_alignment\_descriptor | |
| 7 | X | X | target\_background\_grid\_descriptor | |
| 8 | X | X | video\_window\_descriptor | |
| 9 | X | X | CA\_descriptor | |
| 10 | X | X | ISO\_639\_language\_descriptor | |
| 11 | X | X | system\_clock\_descriptor | |
| 12 | X | X | multiplex\_buffer\_utilization\_descriptor | |
| 13 | X | X | copyright\_descriptor | |
| 14 | X |  | maximum\_bitrate\_descriptor | |
| 15 | X | X | private\_data\_indicator\_descriptor | |
| 16 | X | X | smoothing\_buffer\_descriptor | |
| 17 | X |  | STD\_descriptor | |
| 18 | X | X | IBP\_descriptor | |
| 19 .. 26 | X |  | Defined in ISO/IEC 13818-6 | |
| 27 | X | X | MPEG-4\_video\_descriptor | |
| 28 | X | X | MPEG-4\_audio\_descriptor | |
| 29 | X | X | IOD\_descriptor | |
| 30 | X |  | SL\_descriptor | |
| 31 | X | X | FMC\_descriptor | |
| 32 | X | X | external\_ES\_ID\_descriptor | |
| 33 | X | X | MuxCode\_descriptor | |
| 34 | X | X | M4MuxBufferSize\_descriptor | |
| 35 | X |  | multiplexBuffer\_descriptor | |
| 36 | X | X | content\_labeling\_descriptor | |
| 37 | X | X | metadata\_pointer\_descriptor | |
| 38 | X | X | metadata\_descriptor | |
| 39 | X | X | metadata\_STD\_descriptor | |
| 40 | X | X | AVC video descriptor | |
| 41 | X | X | IPMP\_descriptor (defined in ISO/IEC 13818-11, MPEG-2 IPMP) | |
| 42 | X | X | AVC timing and HRD descriptor | |
| 43 | X | X | MPEG-2\_AAC\_audio\_descriptor | |
| 44 | X | X | M4MuxTiming\_descriptor | |
| 45 | X | X | MPEG-4\_text\_descriptor | |
| 46 | X | X | MPEG-4\_audio\_extension\_descriptor | |
| 47 | X | X | Auxiliary\_video\_stream\_descriptor | |
| 48 | X | X | SVC extension descriptor | |
| 49 | X | X | MVC extension descriptor | |
| 50 | X | n/a | J2K video descriptor | |
| 51 | X | X | MVC operation point descriptor | |
| 52 | X | X | MPEG2\_stereoscopic\_video\_format\_descriptor | |
| 53 | X | X | Stereoscopic\_program\_info\_descriptor | |
| 54 | X | X | Stereoscopic\_video\_info\_descriptor | |
| 55 | X | n/a | Transport\_profile\_descriptor | |
| 56 | X | n/a | HEVC video descriptor | |
| 57 | X | n/a | VVC video descriptor | |
| 58 | X | n/a | EVC video descriptor | |
| 59 .. 62 | n/a | n/a | Rec. ITU-T H.222.0 | ISO/IEC 13818-1 Reserved | |
| 63 | X | X | Extension\_descriptor | |
| 64 .. 255 | n/a | n/a | User Private | |

### 2.6.2 Video stream descriptor

The video stream descriptorprovides basic information which identifies the coding parameters of a video elementary stream as described in Rec. ITU-T H.262 | ISO/IEC 13818-2 or ISO/IEC 11172-2 (see Table 2-46).

Table 2-46 – Video stream descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| video\_stream\_descriptor(){ |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **multiple\_frame\_rate\_flag** | **1** | **bslbf** |
| **frame\_rate\_code** | **4** | **uimsbf** |
| **MPEG\_1\_only\_flag** | **1** | **bslbf** |
| **constrained\_parameter\_flag** | **1** | **bslbf** |
| **still\_picture\_flag** | **1** | **bslbf** |
| if (MPEG\_1\_only\_flag = = '0'){ |  |  |
| **profile\_and\_level\_indication** | **8** | **uimsbf** |
| **chroma\_format** | **2** | **uimsbf** |
| **frame\_rate\_extension\_flag** | **1** | **bslbf** |
| **reserved** | **5** | **bslbf** |
| } |  |  |
| } |  |  |

### 2.6.3 Semantic definitions of fields in video stream descriptor

**multiple\_frame\_rate\_flag** – This 1-bit field when set to '1' indicates that multiple frame rates may be present in the video stream. When set to a value of '0' only a single frame rate is present.

**frame\_rate\_code –** This is a 4-bit field as defined in 6.3.3 of Rec. ITU-T H.262 | ISO/IEC 13818-2, except that when the multiple\_frame\_rate\_flag is set to a value of '1' the indication of a particular frame rate also permits certain other frame rates to be present in the video stream, as specified in Table 2-47:

Table 2-47 – Frame rate code

|  |  |
| --- | --- |
| Coded as | Also includes |
| 23.976 |  |
| 24.0 | 23.976 |
| 25.0 |  |
| 29.97 | 23.976 |
| 30.0 | 23.976 24.0 29.97 |
| 50.0 | 25.0 |
| 59.94 | 23.976 29.97 |
| 60.0 | 23.976 24.0 29.97 30.0 59.94 |

**MPEG\_1\_only\_flag –** This is a 1-bit field which when set to '1' indicates that the video stream contains only ISO/IEC 11172-2 data. If set to '0' the video stream may contain both Rec. ITU-T H.262 | ISO/IEC 13818-2 video data and constrained parameter ISO/IEC 11172-2 video data.

**constrained\_parameter\_flag –** This is a 1-bit field which when set to '1' indicates that the video stream shall not contain unconstrained ISO/IEC 11172-2 video data. If this field is set to '0' the video stream may contain both constrained parameters and unconstrained ISO/IEC 11172-2 video streams. If the MPEG\_1\_only\_flag is set to '0', the constrained\_parameter\_flag shall be set to '1'.

**still\_picture\_flag –** This is a 1-bit field, which when set to '1' indicates that the video stream contains only still pictures. If the bit is set to '0' then the video stream may contain either moving or still picture data.

**profile\_and\_level\_indication –** This 8-bit field is coded in the same manner as the profile\_and\_level\_indication fields in the Rec. ITU-T H.262 | ISO/IEC 13818-2 video stream. The value of this field indicates a profile and level that is equal to or higher than any profile and level in any sequence in the associated video stream. For the purposes of this subclause, an ISO/IEC 11172-2 constrained parameter stream is considered to be a Main Profile at Low Level stream (MP @ LL).

**chroma\_format –** This 2-bit field is coded in the same manner as the chroma\_format fields in the Rec. ITU-T H.262 | ISO/IEC 13818-2 video stream. The value of this field shall be at least equal to or higher than the value of the chroma\_format field in any video sequence of the associated video stream. For the purposes of this subclause, an ISO/IEC 11172-2 video stream is considered to have chroma\_format field with the value '01', indicating 4:2:0.

**frame\_rate\_extension\_flag –** This is a 1-bit flag which when set to '1' indicates that either or both the frame\_rate\_extension\_n and the frame\_rate\_extension\_d fields are non-zero in any video sequences of the Rec. ITU‑T H.262 | ISO/IEC 13818-2 video stream. For the purposes of this subclause, an ISO/IEC 11172-2 video stream is constrained to have both fields set to zero.

### 2.6.4 Audio stream descriptor

The audio stream descriptorprovides basic information which identifies the coding version of an audio elementary stream as described in ISO/IEC 13818-3 or ISO/IEC 11172-3 (see Table 2-48).

| Table 2-48 – Audio stream descriptor | | |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| audio\_stream\_descriptor(){ |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **free\_format\_flag** | **1** | **bslbf** |
| **ID** | **1** | **bslbf** |
| **layer** | **2** | **bslbf** |
| **variable\_rate\_audio\_indicator** | **1** | **bslbf** |
| **reserved** | **3** | **bslbf** |
| } |  |  |

### 2.6.5 Semantic definition of fields in audio stream descriptor

**free\_format\_flag –** This 1-bit field when set to '1' indicates that the audio stream may contain one or more audio frames with the bitrate\_index set to '0000'. If set to '0', then the bitrate\_index is not '0000' (refer to 2.4.2.3 of ISO/IEC 13818-3) in any audio frame of the audio stream.

**ID** – This 1-bit field when set to '1' indicates that the ID field is set to '1' in each audio frame in the audio stream (refer to 2.4.2.3 of ISO/IEC 13818-3).

**layer** – This 2-bit field is coded in the same manner as the layer field in the ISO/IEC 13818-3 or ISO/IEC 11172-3 audio streams (refer to 2.4.2.3 of ISO/IEC 13818-3). The layer indicated in this field shall be equal to or higher than the highest layer specified in any audio frame of the audio stream.

**variable\_rate\_audio\_indicator –** This 1-bit flag, when set to '0' indicates that the encoded value of the bit rate field shall not change in consecutive audio frames which are intended to be presented without discontinuity.

### 2.6.6 Hierarchy descriptor

The hierarchy descriptor provides information to identify the program elements containing components of hierarchically-coded video, audio, and private streams. (See Table 2-49.)

| Table 2-49 – Hierarchy descriptor | | | |
| --- | --- | --- | --- |
| Syntax | No. of bits | Mnemonic | |
| hierarchy\_descriptor() { |  |  | |
| **descriptor\_tag** | **8** | **uimsbf** | |
| **descriptor\_length** | **8** | **uimsbf** | |
| **no\_view\_scalability\_flag** | **1** | **bslbf** | |
| **no\_temporal\_scalability\_flag** | **1** | **bslbf** | |
| **no\_spatial\_scalability\_flag** | **1** | **bslbf** | |
| **no\_quality\_scalability\_flag** | **1** | **bslbf** | |
| **hierarchy\_type** | **4** | **uimsbf** | |
| **reserved** | **2** | **bslbf** | |
| **hierarchy\_layer\_index** | **6** | **uimsbf** | |
| **tref\_present\_flag** | **1** | **bslbf** | |
| **reserved** | **1** | **bslbf** | |
| **hierarchy\_embedded\_layer\_index** | **6** | **uimsbf** | |
| **reserved** | **2** | **bslbf** | |
| **hierarchy\_channel** | **6** | **uimsbf** | |
| } |  |  | |

### 2.6.7 Semantic definition of fields in hierarchy descriptor

**no\_view\_scalability\_flag –**A 1-bit flag, which when set to '0' indicates that the associated program element enhances the number of views of the bit-stream resulting from the program element referenced by the hierarchy\_embedded\_layer\_index. The value of '1' for this flag is reserved.

**no\_temporal\_scalability\_flag** – A 1-bit flag, which when set to '0' indicates that the associated program element enhances the frame rate of the bit-stream resulting from the program element referenced by the hierarchy\_embedded\_layer\_index. The value of '1' for this flag is reserved.

**no\_spatial\_scalability\_flag** – A 1-bit flag, which when set to '0' indicates that the associated program element enhances the spatial resolution of the bit-stream resulting from the program element referenced by the hierarchy\_embedded\_layer\_index. The value of '1' for this flag is reserved.

**no\_quality\_scalability\_flag** – A 1-bit flag, which when set to '0' indicates that the associated program element enhances the SNR quality or fidelity of the bit-stream resulting from the program element referenced by the hierarchy\_embedded\_layer\_index. The value of '1' for this flag is reserved.

**hierarchy\_type** – The hierarchical relation between the associated hierarchy layer and its hierarchy embedded layer is defined in Table 2-50. If scalability applies in more than one dimension, this field shall be set to the value of '8' ("Combined Scalability"), and the flags no\_view\_scalability\_flag, no\_temporal\_scalability\_flag, no\_spatial\_scalability\_flag and no\_quality\_scalability\_flag shall be set accordingly. For MVC video sub-bitstreams, this field shall be set to the value of '9' ("MVC video sub-bitstream") and the flags no\_view\_scalability\_flag, no\_temporal\_scalability\_flag, no\_spatial\_scalability\_flag and no\_quality\_scalability\_flag shall be set to '1'. For MVC base view sub-bitstreams, this field shall be set to the value of '15' and the flags no\_view\_scalability\_flag, no\_temporal\_scalability\_flag, no\_spatial\_scalability\_flag and no\_quality\_scalability\_flag shall be set to '1'. For MVCD video sub-bitstreams, this field shall be set to the value of '9' ("MVCD video sub-bitstream") and the flags no\_view\_scalability\_flag, no\_temporal\_scalability\_flag, no\_spatial\_scalability\_flag and no\_quality\_scalability\_flag shall be set to '1'. For MVCD base view sub-bitstreams, this field shall be set to the value of '15' and the flags no\_view\_scalability\_flag, no\_temporal\_scalability\_flag, no\_spatial\_scalability\_flag and no\_quality\_scalability\_flag shall be set to '1'.

**hierarchy\_layer\_index** – The hierarchy\_layer\_index is a 6-bit field that defines a unique index of the associated program element in a table of coding layer hierarchies. Indices shall be unique within a single program definition. For video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex G of Rec. ITU‑T H.264 | ISO/IEC 14496-10, this is the program element index, which is assigned in a way that the bitstream order will be correct if associated SVC dependency representations of the video sub-bitstreams of the same access unit are re-assembled in increasing order of hierarchy\_layer\_index. For MVC video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10, this is the program element index, which is assigned in a way that the bitstream order will be correct if associated MVC view-component subsets of the MVC video sub-bitstreams of the same access unit are re-assembled in increasing order of hierarchy\_layer\_index. For MVCD video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496-10, this is the program element index, which is assigned in a way that the bitstream order will be correct if associated MVCD view-component subsets of the MVCD video sub-bitstreams of the same access unit are re-assembled in increasing order of hierarchy\_layer\_index.

**tref\_present\_flag** – A 1-bit flag, which when set to '0' indicates that the TREF field may be present in the PES packet headers in the associated elementary stream. The value of '1' for this flag is reserved.

**hierarchy\_embedded\_layer\_index** –The hierarchy\_embedded\_layer\_index is a 6-bit field that defines the hierarchy\_layer\_index of the program element that needs to be accessed and be present in decoding order before decoding of the elementary stream associated with this hierarchy\_descriptor. This field is undefined if the hierarchy\_type value is 15.

**hierarchy\_channel –** The hierarchy\_channel is a 6-bit field that indicates the intended channel number for the associated program element in an ordered set of transmission channels. The most robust transmission channel is defined by the lowest value of this field with respect to the overall transmission hierarchy definition.

NOTE – A given hierarchy\_channel may at the same time be assigned to several program elements.

| Table 2-50 – Hierarchy\_type field values | |
| --- | --- |
| Value | Description |
| 0 | Reserved |
| 1 | Spatial Scalability |
| 2 | SNR Scalability |
| 3 | Temporal Scalability |
| 4 | Data partitioning |
| 5 | Extension bitstream |
| 6 | Private Stream |
| 7 | Multi-view Profile |
| 8 | Combined Scalability or MV-HEVC sub-partition. |
| 9 | MVC video sub-bitstream or MVCD video sub-bitstream |
| 10 | Auxiliary picture layer as defined in Annex F of Rec. ITU-T H.265 | ISO/IEC 23008-2. |
| 11 .. 14 | Reserved |
| 15 | Base layer or MVC base view sub-bitstream or AVC video sub‑bitstream of MVC or HEVC temporal video sub‑bitstream or HEVC base sub-partition or Base layer of MVCD base view sub-bitstream or AVC video sub‑bitstream of MVCD or VVC temporal video sub-bitstream or EVC temporal video sub-bitstream. |

### 2.6.8 Registration descriptor

The registration\_descriptor provides a method to uniquely and unambiguously identify formats of private data (see Table 2-51).

Table 2-51 – Registration descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Identifier |
| registration\_descriptor() { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **format\_identifier** | **32** | **uimsbf** |
| for (i = 0; i < N; i++){ |  |  |
| **additional\_identification\_info** | **8** | **bslbf** |
| **}** |  |  |
| **}** |  |  |

### 2.6.9 Semantic definition of fields in registration descriptor

**format\_identifier** – The format\_identifier is a 32-bit value obtained from a Registration Authority as designated by ISO/IEC JTC 1/SC 29.

**additional\_identification\_info** – The meaning of additional\_identification\_info bytes, if any, are defined by the assignee of that format\_identifier, and once defined they shall not change.

### 2.6.10 Data stream alignment descriptor

The data stream alignment descriptor describes which type of alignment is present in the associated elementary stream. If the data\_alignment\_indicator in the PES packet header is set to '1' and the descriptor is present, alignment – as specified in this descriptor – is required (see Table 2-52).

Table 2-52 – Data stream alignment descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| data\_stream\_alignment\_descriptor() { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **alignment\_type** | **8** | **uimsbf** |
| } |  |  |

### 2.6.11 Semantics of fields in data stream alignment descriptor

**alignment\_type** – Table 2-53 describes the alignment type for ISO/IEC 11172-2 video, Rec. ITU-T H.262 | ISO/IEC 13818-2 video, or ISO/IEC 14496-2 visual streams when the data\_alignment\_indicator in the PES packet header has a value of '1'. For these video streams, the first PES\_packet\_data\_byte following the PES header shall be the first byte of a start code of the type indicated in Table 2-53. At the beginning of a video sequence, the alignment shall occur at the start code of the first sequence header.

NOTE – Specifying alignment type '1' from Table 2-53 does not preclude the alignment from beginning at a GOP or SEQ header.

The definition of an access unit is given in 2.1.1.

Table 2-53 – Video stream alignment values

|  |  |
| --- | --- |
| **Alignment type** | **Description** |
| 0 | Reserved |
| 1 | Slice, or video access unit |
| 2 | Video access unit |
| 3 | GOP, or SEQ |
| 4 | SEQ |
| 5 .. 255 | Reserved |

Table 2-54 describes the alignment type for Rec. ITU-T H.264 | ISO/IEC 14496-10 video when the data\_alignment\_indicator in the PES packet header has a value of '1'.

In this case:

• For AVC video streams conforming to one or more profiles defined in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496-10, the first PES\_packet\_data\_byte following the PES header shall be the first byte of an AVC access unit or the first byte of an AVC slice, as signalled by the alignment\_type value.

• For video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex G of Rec. ITU-T H.264 | ISO/IEC 14496-10, the first PES\_packet\_data\_byte following the PES header shall be the first byte of an SVC dependency representation or the first byte of an SVC slice, as signalled by the alignment\_type value.

• For MVC video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10, the first PES\_packet\_data\_byte following the PES header shall be the first byte of an MVC view-component subset or the first byte of an MVC slice, as signalled by the alignment\_type value.

• For MVCD video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496-10, the first PES\_packet\_data\_byte following the PES header shall be the first byte of an MVCD view-component subset, the first byte of an MVC slice or the first byte of MVCD slice, as signalled by the alignment\_type value.

Table 2-54 – AVC video stream alignment values

|  |  |
| --- | --- |
| Alignment type | Description |
| 0 | Reserved |
| 1 | AVC slice or AVC access unit |
| 2 | AVC access unit |
| 3 | SVC slice or SVC dependency representation |
| 4 | SVC dependency representation |
| 5 | MVC slice or MVC view-component subset |
| 6 | MVC view-component subset |
| 7 | MVCD slice or MVCD view-component subset |
| 8 | MVCD view-component subset |
| 9 .. 255 | Reserved |

Table 2-55 describes the alignment type for HEVC when the data\_alignment\_indicator in the PES packet header has a value of '1'.

Table 2-55 – HEVC video stream alignment values

|  |  |
| --- | --- |
| Alignment type | Description |
| 0 | Reserved |
| 1 | HEVC access unit |
| 2 | HEVC slice |
| 3 | HEVC access unit or slice |
| 4 | HEVC tile of slices |
| 5 | HEVC access unit or tile of slices |
| 6 | HEVC slice or tile of slices |
| 7 | HEVC access unit or slice or tile of slices |
| 8 | HEVC slice segment |
| 9 | HEVC slice segment or access unit |
| 10 | HEVC slice segment or slice |
| 11 | HEVC slice segment or access unit or slice |
| 12 | HEVC slice segment or tile of slices |
| 13 | HEVC slice segment or access unit or tile of slices |
| 14 | HEVC slice segment or slice or tile of slices |
| 15 | HEVC slice segment or access unit or slice or tile of slices |
| 16 .. 255 | Reserved |

Table 2-56 describes the audio alignment type when the data\_alignment\_indicator in the PES packet header has a value of '1'. In this case the first PES\_packet\_data\_byte following the PES header is the first byte of an audio sync word.

Table 2-56 – Audio stream alignment values

|  |  |
| --- | --- |
| Alignment type | Description |
| 0 | Reserved |
| 1 | Sync word |
| 2 .. 255 | Reserved |

Table 2-57 describes the alignment type for VVC when the data\_alignment\_indicator in the PES packet header has a value of '1'.

Table 2-57 – VVC video stream alignment values

|  |  |
| --- | --- |
| Alignment type | Description |
| 00 | Reserved |
| 01 | VVC access unit |
| 02 | VVC slice |
| 03 | VVC access unit or slice |
| 04 | VVC tile of slices |
| 05 | VVC access unit or tile of slices |
| 06 | VVC slice or tile of slices |
| 07 | VVC access unit or slice or tile of slices |
| 08 | VVC subpicture |
| 09 | VVC access unit or subpicture |
| 10 | VVC slice or subpicture |
| 11 | VVC subpicture or tile of slices |
| 12 | VVC access unit or slice or subpicture |
| 13 | VVC access unit or subpicture or tile of slices |
| 14 | VVC slice or subpicture or tile of slices |
| 15 | VVC access unit or slice or subpicture or tile of slices |
| 16-255 | Reserved |

Table 2-58 describes the alignment type for EVC when the data\_alignment\_indicator in the PES packet header has a value of '1'.

Table 2-58 – EVC video stream alignment values

|  |  |
| --- | --- |
| **Alignment type** | **Description** |
| 00 | Reserved |
| 01 | EVC access unit |
| 02 | EVC slice |
| 03 | EVC access unit or slice |
| 04 - 255 | Reserved |

### 2.6.12 Target background grid descriptor

It is possible to have one or more video streams which, when decoded, are not intended to occupy the full display area (e.g., a monitor). The combination of target\_background\_grid\_descriptor and video\_window\_descriptors allows the display of these video windows in their desired locations. The target\_background\_grid\_descriptor is used to describe a grid of unit pixels projected on to the display area. The video\_window\_descriptor is then used to describe, for the associated stream, the location on the grid at which the top left pixel of the display window or display rectangle of the video presentation unit should be displayed. This is represented in Figure 2-3.



Figure 2-3 – Target background grid descriptor display area

### 2.6.13 Semantics of fields in target background grid descriptor

**horizontal\_size** – The horizontal size of the target background grid in pixels.

**vertical\_size** – The vertical size of the target background grid in pixels.

**aspect\_ratio\_information** –Specifies the sample aspect ratio or display aspect ratio of the target background grid. Aspect\_ratio\_information is defined in Rec. ITU-T H.262 | ISO/IEC 13818-2 (see Table 2-59).

Table 2-59 – Target background grid descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| target\_background\_grid\_descriptor() { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **horizontal\_size** | **14** | **uimsbf** |
| **vertical\_size** | **14** | **uimsbf** |
| **aspect\_ratio\_information** | **4** | **uimsbf** |
| } |  |  |

### 2.6.14 Video window descriptor

The video window descriptor (see Table 2-60) is used to describe the window characteristics of the associated video elementary stream. Its values reference the target background grid descriptor for the same stream. Also see target\_background\_grid\_descriptor in 2.6.12.

Table 2-60 – Video window descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| video\_window\_descriptor() { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **horizontal\_offset** | **14** | **uimsbf** |
| **vertical\_offset** | **14** | **uimsbf** |
| **window\_priority** | **4** | **uimsbf** |
| } |  |  |

### 2.6.15 Semantic definition of fields in video window descriptor

**horizontal\_offset** – The value indicates the horizontal position of the top left pixel of the current video display window or display rectangle if indicated in the picture display extension on the target background grid for display as defined in the target\_background\_grid\_descriptor. The top left pixel of the video window shall be one of the pixels of the target background grid (refer to Figure 2-3).

**vertical\_offset** –The value indicates the vertical position of the top left pixel of the current video display window or display rectangle if indicated in the picture display extension on the target background grid for display as defined in the target\_background\_grid\_descriptor. The top left pixel of the video window shall be one of the pixels of the target background grid (refer to Figure 2-3).

**window\_priority** – The value indicates how windows overlap. A value of 0 being lowest priority and a value of 15 is the highest priority, i.e., windows with priority 15 are always visible.

### 2.6.16 Conditional access descriptor

The conditional access descriptor is used to specify both system-wide conditional access management information such as EMMs and elementary stream-specific information such as ECMs. It may be used in both the TS\_program\_map\_section (refer to 2.4.4.9) and the program\_stream\_map (refer to 2.5.3). If any elementary stream is scrambled, a CA descriptor shall be present for the program containing that elementary stream. If any system-wide conditional access management information exists within a transport stream, a CA descriptor shall be present in the conditional access table.

When the CA descriptor is found in the TS\_program\_map\_section (table\_id = 0x02), the CA\_PID points to packets containing program related access control information, such as ECMs. Its presence as program information indicates applicability to the entire program. In the same case, its presence as extended ES information indicates applicability to the associated program element. Provision is also made for private data.

When the CA descriptor is found in the CA\_section (table\_id = 0x01), the CA\_PID points to packets containing system‑wide and/or access control management information, such as EMMs.

The contents of the transport stream packets containing conditional access information are privately defined (see Table 2‑61).

Table 2-61 – Conditional access descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| CA\_descriptor() { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **CA\_system\_ID** | **16** | **uimsbf** |
| **reserved** | **3** | **bslbf** |
| **CA\_PID** | **13** | **uimsbf** |
| for (i = 0; i < N; i++) { |  |  |
| **private\_data\_byte** | **8** | **uimsbf** |
| } |  |  |
| } |  |  |

### 2.6.17 Semantic definition of fields in conditional access descriptor

**CA\_system\_ID** –This is a 16-bit field indicating the type of CA system applicable for either the associated ECM and/or EMM streams. The coding of this is privately defined and is not specified by ITU-T | ISO/IEC.

**CA\_PID** –This is a 13-bit field indicating the PID of the transport stream packets which shall contain either ECM or EMM information for the CA systems as specified with the associated CA\_system\_ID. The contents (ECM or EMM) of the packets indicated by the CA\_PID is determined from the context in which the CA\_PID is found, i.e., a TS\_program\_map\_section or the CA table in the transport stream, or the stream\_id field in the program stream.

In transport streams, the presence of PID 0x03 indicates that there is IPMP as described in ISO/IEC 13818-11 used by components in the transport stream. In program streams, the presence of stream\_ID\_extension value 0x00 indicates that IPMP as described in ISO/IEC 13818-11 is used by components in the program stream. Within a given Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 stream, components could use both IPMP as described in ISO/IEC 13818-11 as well as CA as defined in ISO/IEC 13818-1:2006. Compatibility between the two schemes is described in ISO/IEC 13818‑11.

### 2.6.18 ISO 639 language descriptor

The language descriptor is used to specify the language of the associated program element (see Table 2-62).

Table 2-62 – ISO 639 language descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| ISO\_639\_language\_descriptor() { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| for (i = 0; i < N; i++) { |  |  |
| **ISO\_639\_language\_code** | **24** | **bslbf** |
| **audio\_type** | **8** | **bslbf** |
| } |  |  |
| } |  |  |

### 2.6.19 Semantic definition of fields in ISO 639 language descriptor

**ISO\_639\_language\_code** – Identifies the language or languages used by the associated program element. The ISO\_639\_language\_code contains a 3-character code as specified by ISO 639, Part 2. Each character is coded into 8 bits according to ISO 8859-1 and inserted in order into this 24-bit field. In the case of multilingual audio streams the sequence of ISO\_639\_language\_code fields shall reflect the content of the audio stream.

**audio\_type** – The audio\_type is an 8-bit field which specifies the type of stream defined in Table 2-63. This field is used to address complete main elementary streams. If multiple substreams exists, then audio\_type shall define the group of subcomponents. Any further detail of specific subcomponents is left up to in-stream information.

Multiple ISO\_639\_language\_code / audio\_type pairs can be described within a single ISO 639 language descriptor.

NOTE – An example of this would be an english track that is suitable for both Hearing-impaired in English with Visual impaired commentary in Spanish. i=0{“ENG”/0x02} and i=1{“ESP”/0x03}.

Table 2-63 – Audio type values

| Value | Description |
| --- | --- |
| 0x00 | Undefineda |
| 0x01 | Clean effectsb |
| 0x02 | Hearing impairedc |
| 0x03 | Visual impaired commentaryd |
| 0x04 .. 0x7F | User Private |
| 0x80 | Primarye |
| 0x81 | Nativef |
| 0x82 | Emergencyg |
| 0x83 | Primary commentaryh |
| 0x84 | Alternate commentaryi |
| 0x85 .. 0xFF | Reserved |
| a Undefined: This is a default value which indicates the purpose of the stream is unknown. It is different from the languages code “und” which would indicate the language of the audio is unknown.  b Clean effects: This value indicates that the referenced program element has no language.  c Hearing impaired: This value indicates that the referenced program element is prepared for the hearing impaired. An alternate label for this is enhanced audio intelligibility.  d Visual\_impaired\_commentary: This value indicates that the referenced program element is prepared for the visually impaired viewer. Alternate labels for this are audio descriptive services (ADS) or descriptive video services (DVS).  e Primary: This value indicates the default recommended language of the program. This is a separate value from the language code itself. If the language code is Welsh with and Undefined audio\_type, it means “this is Welsh”. When the language code is welsh with audio\_type 0x80, it means “this is Welsh”, and it is the primary language of this program.”  f Native: This value indicates the language in which the program was produced. An absence of this value indicates that it is either undefined or a dubbed language.  g Emergency: This value indicates the purpose for the stream is used for emergency event information.  h Primary commentary: This value indicates the referenced program element is prepared with narration in addition to the natural sound of the content. For example, this narration may be an announcer to a sporting event or director’s comments of a film or program and is the default commentary for the content.  i Alternate commentary: This value indicates the referenced program element is prepared with alternate naration in addition to the natural sound of the content. This additional commentary is of the same language as the primary commentary and may be used for example in cases such as an additional away announcer to a sporting event or producer’s comments of the film or program. | |

### 2.6.20 System clock descriptor

This descriptor conveys information about the system clock that was used to generate the timestamps.

If an external clock reference was used, the external\_clock\_reference\_indicator may be set to '1'. The decoder optionally may use the same external reference if it is available.

If the system clock is more accurate than the 30-ppm accuracy required, then the accuracy of the clock can be communicated by encoding it in the clock\_accuracy fields. The clock frequency accuracy is:

*clock\_accuracy\_integer* × 10–*clock\_accuracy\_exponent ppm* (2-26)

If clock\_accuracy\_integer is set to '0', then the system clock accuracy is 30 ppm. When the external\_clock\_reference\_indicator is set to '1', the clock accuracy pertains to the external reference clock (see Table 2‑64).

Table 2-64 – System clock descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| system\_clock\_descriptor() { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **external\_clock\_reference\_indicator** | **1** | **bslbf** |
| **reserved** | **1** | **bslbf** |
| **clock\_accuracy\_integer** | **6** | **uimsbf** |
| **clock\_accuracy\_exponent** | **3** | **uimsbf** |
| **reserved** | **5** | **bslbf** |
| } |  |  |

### 2.6.21 Semantic definition of fields in system clock descriptor

**external\_clock\_reference\_indicator** – This is a 1-bit indicator. When set to '1', it indicates that the system clock has been derived from an external frequency reference that may be available at the decoder.

**clock\_accuracy\_integer** – This is a 6-bit integer. Together with the clock\_accuracy\_exponent, it gives the fractional frequency accuracy of the system clock in parts per million.

**clock\_accuracy\_exponent** – This is a 3-bit integer. Together with the clock\_accuracy\_integer, it gives the fractional frequency accuracy of the system clock in parts per million.

### 2.6.22 Multiplex buffer utilization descriptor

The multiplex buffer utilization descriptor provides bounds on the occupancy of the STD multiplex buffer. This information is intended for devices such as remultiplexers, which may use this information to support a desired re‑multiplexing strategy (see Table 2-65).

Table 2-65 – Multiplex buffer utilization descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| Multiplex\_buffer\_utilization\_descriptor() { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **bound\_valid\_flag** | **1** | **bslbf** |
| **LTW\_offset\_lower\_bound** | **15** | **uimsbf** |
| **reserved** | **1** | **bslbf** |
| **LTW\_offset\_upper\_bound** | **15** | **uimsbf** |
| } |  |  |

### 2.6.23 Semantic definition of fields in multiplex buffer utilization descriptor

**bound\_valid\_flag** – A value of '1' indicates that the LTW\_offset\_lower\_bound and the LTW\_offset\_upper\_bound fields are valid.

**LTW\_offset\_lower\_bound** – This 15-bit field is defined only if the bound\_valid flag has a value of '1'. When defined, this field has the units of (27 MHz/300) clock periods, as defined for the LTW\_offset (refer to 2.4.3.4). The LTW\_offset\_lower\_bound represents the lowest value that any LTW\_offset field would have, if that field were coded in every packet of the stream or streams referenced by this descriptor. Actual LTW\_offset fields may or may not be coded in the bitstream when the multiplex buffer utilization descriptor is present. This bound is valid until the next occurrence of this descriptor.

**LTW\_offset\_upper\_bound** – This 15-bit field is defined only if the bound\_valid has a value of '1'. When defined, this field has the units of (27 MHz/300) clock periods, as defined for the LTW\_offset (refer to 2.4.3.4). The LTW\_offset\_upper\_bound represents the largest value that any LTW\_offset field would have, if that field were coded in every packet of the stream or streams referenced by this descriptor. Actual LTW\_offset fields may or may not be coded in the bitstream when the multiplex buffer utilization descriptor is present. This bound is valid until the next occurrence of this descriptor.

### 2.6.24 Copyright descriptor

The copyright\_descriptor provides a method to enable audiovisual works identification. This copyright\_descriptor applies to programs or program elements within programs (see Table 2-66).

Table 2-66 – Copyright descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Identifier |
| copyright\_descriptor() { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **copyright\_identifier** | **32** | **uimsbf** |
| for (i = 0; i < N; i++){ |  |  |
| **additional\_copyright\_info** | **8** | **bslbf** |
| } |  |  |
| } |  |  |

### 2.6.25 Semantic definition of fields in copyright descriptor

**copyright\_identifier** – This field is a 32-bit value obtained from the Registration Authority.

**additional\_copyright\_info** – The meaning of additional\_copyright\_info bytes, if any, are defined by the assignee of that copyright\_identifier, and once defined, they shall not change.

### 2.6.26 Maximum bitrate descriptor

See Table 2-67.

Table 2-67 – Maximum bitrate descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Identifier |
| maximum\_bitrate\_descriptor() { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **reserved** | **2** | **bslbf** |
| **maximum\_bitrate** | **22** | **uimsbf** |
| } |  |  |

### 2.6.27 Semantic definition of fields in maximum bitrate descriptor

**maximum\_bitrate** – The maximum bitrate is coded as a 22-bit positive integer in this field. The value indicates an upper bound of the bitrate, including transport overhead, that will be encountered in this program element or program. The value of maximum\_bitrate is expressed in units of 50 bytes/second. The maximum\_bitrate\_descriptor is included in the Program Map Table (PMT). Its presence as extended program information indicates applicability to the entire program. Its presence as ES information indicates applicability to the associated program element.

### 2.6.28 Private data indicator descriptor

See Table 2-68.

Table 2-68 – Private data indicator descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Identifier |
| private\_data\_indicator\_descriptor() { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **private\_data\_indicator** | **32** | **uimsbf** |
| } |  |  |

### 2.6.29 Semantic definition of fields in Private data indicator descriptor

**private\_data\_indicator** – The value of the private\_data\_indicator is private and shall not be defined by ITU‑T | ISO/IEC.

### 2.6.30 Smoothing buffer descriptor

This descriptor is optional and conveys information about the size of a smoothing buffer, SBn, associated with this descriptor, and the associated leak rate out of that buffer, for the program element(s) that it refers to.

In the case of transport streams, bytes of transport stream packets of the associated program element(s) present in the transport stream are input to a buffer SBn of size given by sb\_size, at the time defined by equation 2-4.

In the case of program streams, bytes of all PES packets of the associated elementary streams, are input to a buffer SBn of size given by sb\_size, at the time defined by equation 2-21.

When there is data present in this buffer, bytes are removed from this buffer at a rate defined by sb\_leak\_rate. The buffer, SBn shall never overflow.During the continuous existence of a program, the value of the elements of the Smoothing Buffer descriptor of the different program element(s) in the program, shall not change.

The meaning of the smoothing buffer\_descriptor (see Table 2-69) is only defined when it is included in the PMT or the program stream Map*.*

If, in the case of a transport stream, it is present in the ES info in the Program Map Table, all transport stream packets of the PID of that program element enter the smoothing buffer.

If, in the case of a transport stream, it is present in the program information, the following transport stream packets enter the smoothing buffer:

• all transport stream packets of all PIDs listed as elementary\_PIDs in the extended program information as well as;

• all transport stream packets of the PID which is equal to the PMT\_PID of this section;

• all transport stream packets of the PCR\_PID of the program.

All bytes that enter the associated buffer also exit it.

At any given time there shall be at most one descriptor referring to any individual program element and at most one descriptor referring to the program in its entirety.

Table 2-69 – Smoothing buffer descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| smoothing\_buffer\_descriptor () { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **reserved** | **2** | **bslbf** |
| **sb\_leak\_rate** | **22** | **uimsbf** |
| **reserved** | **2** | **bslbf** |
| **sb\_size** | **22** | **uimsbf** |
| } |  |  |

### 2.6.31 Semantic definition of fields in smoothing buffer descriptor

**sb\_leak\_rate** – This 22-bit field is coded as a positive integer. Its contents indicate the value of the leak rate out of the SBn buffer for the associated elementary stream or other data in units of 400 bits/s.

**sb\_size** – This 22-bit field is coded as a positive integer. Its contents indicate the value of the size of the multiplexing buffer smoothing buffer SBn for the associated elementary stream or other data in units of 1 byte (see Table 2-68).

### 2.6.32 STD descriptor

This descriptor is optional and applies only to the T-STD model and to Rec. ITU-T H.262 | ISO/IEC 13818-2 video elementary streams, and is used as specified in 2.4.2. This descriptor does not apply to program streams (see Table 2‑70).

Table 2-70 – STD descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| STD\_descriptor () { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **reserved** | **7** | **bslbf** |
| **leak\_valid\_flag** | **1** | **bslbf** |
| } |  |  |

### 2.6.33 Semantic definition of fields in STD descriptor

**leak\_valid\_flag** – The leak\_valid\_flag is a 1-bit flag. When set to '1', the transfer of data from the buffer MBn to the buffer EBn in the T-STD uses the leak method as defined in 2.4.2.4. If this flag has a value equal to '0', and the vbv\_delay fields present in the associated video stream do not have the value 0xFFFF, the transfer of data from the buffer MBn to the buffer EBn uses the vbv\_delay method as defined in 2.4.2.4.

### 2.6.34 IBP descriptor

This optional descriptor provides information about some characteristics of the sequence of frame types in an ISO/IEC 11172-2, Rec. ITU-T H.262 | ISO/IEC 13818-2, or ISO/IEC 14496-2 video stream (see Table 2-71).

Table 2-71 – IBP descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| ibp\_descriptor() { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **closed\_gop\_flag** | **1** | **uimsbf** |
| **identical\_gop\_flag** | **1** | **uimsbf** |
| **max\_gop-length** | **14** | **uimsbf** |
| } |  |  |

### 2.6.35 Semantic definition of fields in IBP descriptor

**closed\_gop\_flag** – This 1-bit flag when set to '1' indicates that a group of pictures header is encoded before every I‑frame and that the closed\_gop flag is set to '1' in all group of pictures headers in the video sequence.

**identical\_gop\_flag** – This 1-bit flag when set to '1' indicates that the number of P-frames and B-frames between I‑frames, and the picture coding types and sequence of picture types between I-pictures is the same throughout the sequence, except possibly for the pictures up to the second I-picture.

**max\_gop\_length** – This 14-bit unsigned integer indicates the maximum number of the coded pictures between any two consecutive I-pictures in the sequence. The value of '0' is forbidden.

### 2.6.36 MPEG-4 video descriptor

For individual ISO/IEC 14496-2 streams directly carried in PES packets, as defined in 2.11.2, the MPEG-4 video descriptor (see Table 2-72) provides basic information for identifying the coding parameters of such visual elementary streams. The MPEG-4 video descriptor does not apply to ISO/IEC 14496-2 streams encapsulated in SL-packets and in M4Mux packets, as defined in 2.11.3.

Table 2-72 – MPEG-4 video descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| MPEG-4\_video\_descriptor () { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **MPEG-4\_visual\_profile\_and\_level** | **8** | **uimsbf** |
| } |  |  |

### 2.6.37 Semantic definition of fields in MPEG-4 video descriptor

**MPEG-4\_video\_profile\_and\_level** – This 8-bit field shall identify the profile and level of the ISO/IEC 14496-2 video stream. This field shall be coded with the same value as the profile\_and\_level\_indication field in the Visual Object Sequence Header in the associated ISO/IEC 14496-2 stream.

### 2.6.38 MPEG-4 audio descriptor

For individual ISO/IEC 14496-3 streams directly carried in PES packets, as defined in 2.11.2, the MPEG-4 audio descriptor (see Table 2-73) provides basic information for identifying the coding parameters of such audio elementary streams. The MPEG-4 audio descriptor does not apply to ISO/IEC 14496-3 streams encapsulated in SL-packets and in M4Mux packets, as defined in 2.11.3.

Table 2-73 – MPEG-4 audio descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| MPEG-4\_audio\_descriptor () { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **MPEG-4\_audio\_profile\_and\_level** | **8** | **uimsbf** |
| } |  |  |

### 2.6.39 Semantic definition of fields in MPEG-4 audio descriptor

**MPEG-4\_audio\_profile\_and\_level** – This 8-bit field identifies the profile and level of the ISO/IEC 14496-3 audio stream corresponding to Table 2-74. If encoded with the value 0x0F, then it is signalled that no profile and level is defined for the associated MPEG-4 audio stream. The encoded value 0xFF indicates that the audio profile and level is not specified by the MPEG-4\_audio\_profile\_and\_level field; in that case, in addition to the MPEG-4 audio descriptor, an MPEG-4 audio extension descriptor shall be associated with the same MPEG-4 audio stream. In all other cases, next to an MPEG-4 audio descriptor, also an MPEG-4 audio extension descriptor may be associated with the same MPEG-4 audio stream.

| Table 2-74 – MPEG-4\_audio\_profile\_and\_level assignment values | | |
| --- | --- | --- |
| Value | Description | |
| 0x00 .. 0x0E | Reserved | |
| 0x0F | No audio profile and level defined for the associated MPEG-4 audio stream | |
| 0x10 | Main profile, level 1 | |
| 0x11 | Main profile, level 2 | |
| 0x12 | Main profile, level 3 | |
| 0x13 | Main profile, level 4 | |
| 0x14 .. 0x17 | Reserved | |
| 0x18 | Scalable Profile, level 1 | |
| 0x19 | Scalable Profile, level 2 | |
| 0x1A | Scalable Profile, level 3 | |
| 0x1B | Scalable Profile, level 4 | |
| 0x1C .. 0x1F | Reserved | |
| 0x20 | Speech profile, level 1 | |
| 0x21 | Speech profile, level 2 | |
| 0x22 .. 0x27 | Reserved | |
| 0x28 | Synthesis profile, level 1 | |
| 0x29 | Synthesis profile, level 2 | |
| 0x2A | Synthesis profile, level 3 | |
| 0x2B .. 0x2F | Reserved | |
| 0x30 | High quality audio profile, level 1 | |
| 0x31 | High quality audio profile, level 2 | |
| 0x32 | High quality audio profile, level 3 | |
| 0x33 | High quality audio profile, level 4 | |
| 0x34 | High quality audio profile, level 5 | |
| 0x35 | High quality audio profile, level 6 | |
| 0x36 | High quality audio profile, level 7 | |
| 0x37 | High quality audio profile, level 8 | |
| 0x38 | Low delay audio profile, level 1 | |
| 0x39 | Low delay audio profile, level 2 | |
| 0x3A | Low delay audio profile, level 3 | |
| 0x3B | Low delay audio profile, level 4 | |
| 0x3C | Low delay audio profile, level 5 | |
| 0x3D | Low delay audio profile, level 6 | |
| 0x3E | Low delay audio profile, level 7 | |
| 0x3F | Low delay audio profile, level 8 | |
| 0x40 | Natural audio profile, level 1 | |
| 0x41 | Natural audio profile, level 2 | |
| 0x42 | Natural audio profile, level 3 | |
| 0x43 | Natural audio profile, level 4 | |
| 0x44 .. 0x47 | Reserved | |
| 0x48 | Mobile audio internetworking profile, level 1 | |
| 0x49 | Mobile audio internetworking profile, level 2 | |
| 0x4A | Mobile audio internetworking profile, level 3 | |
| 0x4B | Mobile audio internetworking profile, level 4 | |
| 0x4C | Mobile audio internetworking profile, level 5 | |
| 0x4D | Mobile audio internetworking profile, level 6 | |
| 0x4E .. 0x4F | Reserved | |
| 0x50 | AAC profile, level 1 | |
| 0x51 | AAC profile, level 2 | |
| 0x52 | AAC profile, level 4 | |
| 0x53 | AAC profile, level 5 | |
| 0x54 | AAC profile, level 6 | |
| 0x55 | AAC profile, level 7 | |
| 0x56 .. 0x57 | Reserved | |
| 0x58 | High efficiency AAC profile, level 2 | |
| 0x59 | High efficiency AAC profile, level 3 | |
| 0x5A | High efficiency AAC profile, level 4 | |
| 0x5B | High efficiency AAC profile, level 5 | |
| 0x5C | High efficiency AAC profile, level 6 | |
| 0x5D | High efficiency AAC profile, level 7 | |
| 0x5E .. 0x5F | Reserved | |
| 0x60 | High efficiency AAC v2 profile, level 2 | |
| 0x61 | High efficiency AAC v2 profile, level 3 | |
| 0x62 | High efficiency AAC v2 profile, level 4 | |
| 0x63 | High efficiency AAC v2 profile, level 5 | |
| 0x64 | High efficiency AAC v2 profile, level 6 | |
| 0x65 | High efficiency AAC v2 profile, level 7 | |
| 0x66 .. 0x67 | Reserved | |
| 0x68 | Extended HE AAC Profile, level 1 | |
| 0x69 | Extended HE AAC Profile, level 2 | |
| 0x6A | Extended HE AAC Profile, level 3 | |
| 0x6B | Extended HE AAC Profile, level 4 | |
| 0x6C | Extended HE AAC Profile, level 6 | |
| 0x6D | Extended HE AAC Profile, level 7 | |
| 0x6E .. 0x6F | Reserved | |
| 0x70 | Baseline USAC Profile, level 1 | |
| 0x71 | Baseline USAC Profile, level 2 | |
| 0x72 | Baseline USAC Profile, level 3 | |
| 0x73 | Baseline USAC Profile, level 4 | |
| 0x74 .. 0x7F | Reserved | |
| 0x80 | Low Delay AAC Profile, level 1 | |
| 0x81 .. 0x87 | Reserved | |
| 0x88 | Low Delay AAC v2 Profile, level 1 | |
| 0x89 | Low Delay AAC v2 Profile, level 2 | |
| 0x8A | Low Delay AAC v2 Profile, level 3 | |
| 0x8B | Low Delay AAC v2 Profile, level 4 | |
| 0x8C .. 0x8F | Reserved | |
| 0x90 | High Definition AAC Profile, level 1 | |
| 0x91 .. 0x97 | Reserved | |
| 0x98 | ALS Simple Profile, level 1 | |
| 0x99 .. 0xFE | Reserved | |
| 0xFF | Audio profile and level not specified by the MPEG-4\_audio\_profile\_and\_level field in this descriptor | |

### 2.6.40 IOD descriptor

The IOD descriptor (see Table 2-75) encapsulates the InitialObjectDescriptor structure. An initial object descriptor allows access to a set of ISO/IEC 14496 streams by identifying the ES\_ID values of the ISO/IEC 14496-1 scene description and object descriptor streams. Both the scene description stream and the object descriptor stream contain further information about the ISO/IEC 14496 streams that are part of the scene. See Annex R for a description of the content access procedure. The InitialObjectDescriptor is specified in 7.2.6.4 of ISO/IEC 14496-1.

Within a transport stream, the IOD descriptor shall be conveyed in the descriptor loop immediately following the program\_info\_length field in the Program Map Table. If a program stream map is present in a program stream, the IOD descriptor shall be conveyed in the descriptor loop immediately following the program\_stream\_info\_length field in the program stream map. More than one IOD descriptor may be associated with a program.

NOTE – This Specification does not specify how the IOD\_label may be used by higher level service information to uniquely select one of the ISO/IEC 14496 presentations identified by multiple IOD descriptors.

Table 2-75 – IOD descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| IOD\_descriptor () { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **Scope\_of\_IOD\_label** | **8** | **uimsbf** |
| **IOD\_label** | **8** | **uimsbf** |
| InitialObjectDescriptor () |  |  |
| } |  |  |

### 2.6.41 Semantic definition of fields in IOD descriptor

**Scope\_of\_IOD\_label** – This 8-bit field specifies the scope of the IOD\_label field. A value of 0x10 indicates that the IOD\_label is unique within the program stream or within the specific program in a transport stream in which the IOD descriptor is carried. A value of 0x11 indicates that the IOD\_label is unique within the transport stream in which the IOD descriptor is carried. All other values of the Scope\_of\_IOD\_label field are reserved.

**IOD\_label** – This 8-bit field specifies the label of the IOD descriptor.

**InitialObjectDescriptor ()** – This structure is defined in 7.2.6.4 of ISO/IEC 14496-1.

### 2.6.42 SL descriptor

The SL descriptor (see Table 2-76) shall be used when a single ISO/IEC 14496-1 SL-packetized stream is encapsulated in PES packets. The SL descriptor associates the ES\_ID of this SL-packetized stream to an elementary\_PID in case of a transport stream or to an elementary\_stream\_id in case of a program stream. Within a transport stream, the SL descriptor shall be conveyed for the corresponding elementary stream in the descriptor loop immediately following the ES\_info\_length field in the Program Map Table. If a program stream map is present in a program stream, the SL descriptor shall be conveyed in the descriptor loop immediately following the elementary\_stream\_info\_length field within the Program Stream Map.

NOTE – SL packetized streams may be used in a program stream. However, only one stream\_id exists for ISO/IEC 14496‑1 SL‑packetized streams. In order to associate multiple such streams within a program stream to an ISO/IEC 14496-1 scene, M4Mux has to be used and signalled appropriately by an FMC descriptor. This limitation does not exist in a transport stream where the SL descriptor provides unambiguous mapping between an ISO/IEC 14496-1 ES\_ID value and a Rec. ITU-T H.222.0 | ISO/IEC 13818‑1 elementary\_PID value.

Table 2-76 – SL descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| SL\_descriptor () { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **ES\_ID** | **16** | **uimsbf** |
| } |  |  |

### 2.6.43 Semantic definition of fields in SL descriptor

**ES\_ID** – This 16-bit field shall specify the identifier of an ISO/IEC 14496-1 SL-packetized stream.

### 2.6.44 FMC descriptor

The FMC descriptor (see Table 2-77) indicates that the ISO/IEC 14496-1 M4Mux tool has been used to multiplex ISO/IEC 14496-1 SL‑packetized streams into a M4Mux stream before encapsulation in PES packets or ISO//IEC14496\_sections. The FMC descriptor associates M4Mux channels to the ES\_ID values of the SL-packetized streams in the M4Mux stream.

An FMC descriptor is required for each program element referenced by an elementary\_PID value in a transport stream and for each elementary\_stream\_id in a program stream that conveys a M4Mux stream. Within a transport stream, the FMC descriptor shall be conveyed for the corresponding elementary stream in the descriptor loop immediately following the ES\_info\_length field in the Program Map Table. If a Program Stream Map is present in a program stream, the FMC descriptor shall be conveyed in the descriptor loop immediately following the elementary\_stream\_info\_length field in the program stream Map.

For each SL\_packetized stream in a M4Mux stream, the M4Mux channel shall be identified by a single entry in the FMC descriptor.

Table 2-77 – FMC descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| FMC\_descriptor () { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| for (i = 0; i < descriptor\_length; i + = 3) { |  |  |
| **ES\_ID** | **16** | **uimsbf** |
| **M4MuxChannel** | **8** | **uimsbf** |
| } |  |  |
| } |  |  |

### 2.6.45 Semantic definition of fields in FMC descriptor

**ES\_ID** – This 16-bit field specifies the identifier of an ISO/IEC 14496-1 SL-packetized stream.

**M4MuxChannel** – This 8-bit field specifies the number of the M4Mux channel used for this SL-packetized stream.

### 2.6.46 External\_ES\_ID descriptor

The External\_ES\_ID descriptor (see Table 2-78) assigns an ES\_ID, as defined in ISO/IEC 14496-1, to a program element to which no ES\_ID value has been assigned by other means. This ES\_ID allows reference to a non-ISO/IEC 14496 component in the scene description or, for example, to associate a non-ISO/IEC 14496 component with an IPMP stream.

Within a transport stream, the assignment of an ES\_ID shall be made by conveying an External\_ES\_ID descriptor for the corresponding elementary stream in the descriptor loop immediately following the ES\_info\_length field in the Program Map Table. If a program stream map is present in a program stream, the External\_ES\_ID descriptor shall be conveyed in the descriptor loop immediately following the elementary\_stream\_info\_length field in the Program Stream Map.

Table 2-78 – External\_ES\_ID descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| External\_ES\_ID\_descriptor () { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **External\_ES\_ID** | **16** | **uimsbf** |
| } |  |  |

### 2.6.47 Semantic definition of fields in External\_ES\_ID descriptor

**External\_ES\_ID** – This 16-bit field assigns an ES\_ID identifier, as defined in ISO/IEC 14496-1, to a component of a program.

### 2.6.48 Muxcode descriptor

The Muxcode descriptor (see Table 2-79) conveys MuxCodeTableEntry structures as defined in 7.4.2.5 of ISO/IEC 14496-1. MuxCodeTableEntries configure the MuxCode mode of M4Mux.

One or more Muxcode descriptors may be associated with each elementary\_PID or elementary\_stream\_id, respectively, conveying an ISO/IEC 14496-1 M4Mux stream that utilizes the MuxCode mode. Within a transport stream, the Muxcode descriptor shall be conveyed for the corresponding elementary stream in the descriptor loop immediately following the ES\_info\_length field in the Program Map Table. If a Program Stream Map is present in a program stream, the Muxcode descriptor shall be conveyed in the descriptor loop immediately following the elementary\_stream\_info\_length field in the Program Stream Map.

MuxCodeTableEntries may be updated with new versions. In case of such updates, the version\_number of each Program Map Table or the program\_stream\_map\_version of each Program Stream Map, respectively, carrying the MuxCode descriptor in their descriptor loop shall be incremented by 1 modulo 32.

Table 2-79 – Muxcode descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| Muxcode\_descriptor () { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| for (i = 0; i < N; i++) { |  |  |
| MuxCodeTableEntry () |  |  |
| } |  |  |
| } |  |  |

### 2.6.49 Semantic definition of fields in Muxcode descriptor

**MuxCodeTableEntry ()** – This structure is defined in 7.4.2.5 of ISO/IEC 14496-1.

### 2.6.50 M4MuxBufferSize descriptor

The M4MuxBufferSize descriptor (see Table 2-80) conveys the size of the M4Mux buffer (FB) for each SL packetized stream multiplexed in a M4Mux stream.

One M4MuxBufferSize descriptor shall be associated with each elementary\_PID or elementary\_stream\_id, respectively, conveying an ISO/IEC 14496-1 M4Mux stream. Within a transport stream, the M4MuxBufferSize descriptor shall be conveyed for the corresponding elementary stream in the descriptor loop immediately following the ES\_info\_length field in the Program Map Table. If a Program Stream Map is present in a program stream, the M4MuxBufferSize descriptor shall be conveyed in the descriptor loop immediately following the elementary\_stream\_info\_length field within the Program Stream Map.

Table 2-80 – M4MuxBufferSize descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| M4MuxBufferSize\_descriptor () { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **DefaultM4MuxBufferDescriptor()** |  |  |
| for (i=0; i<descriptor\_length; i += 4) { |  |  |
| **M4MuxBufferDescriptor()** |  |  |
| } |  |  |
| } |  |  |

### 2.6.51 Semantic definition of fields in M4MuxBufferSize descriptor

**DefaultM4MuxBufferDescriptor()** – This descriptor specifies the default M4Mux buffer size for this M4Mux stream. It is defined in 7.4.2.8 of ISO/IEC 14496-1:2010.

**M4MuxBufferDescriptor()** – This descriptor specifies the M4Mux buffer size for one SL-packetized stream carried within the M4Mux stream. It is defined in 7.4.2.7 of ISO/IEC 14496-1:2010.

### 2.6.52 MultiplexBuffer descriptor

The MultiplexBuffer descriptor (see Table 2-81) conveys the size of the multiplex buffer MBn, as well as the leak rate Rxn at which data is transferred from transport buffer TBn into buffer MBn for a specific Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program element referenced by an elementary\_PID value in the Program Map Table.

One MultiplexBuffer descriptor shall be associated with each elementary\_PID that contains an ISO/IEC 14496 M4Mux stream or SL-packetized stream, including those containing ISO\_IEC\_14496\_sections. See 2.11.3.9 for the definition of buffers and rates in the T-STD model for decoding of ISO/IEC 14496 content.

The MultiplexBuffer descriptor shall be conveyed in the descriptor loop immediately following the ES\_info\_length field in the Program Map Table.

Table 2-81 – MultiplexBuffer descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| MultiplexBuffer\_descriptor () { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **MB\_buffer\_size** | **24** | **uimsbf** |
| **TB\_leak\_rate** | **24** | **uimsbf** |
| } |  |  |

### 2.6.53 Semantic definition of fields in MultiplexBuffer descriptor

**MB\_buffer\_size** – This 24-bit field shall specify the size in byte of buffer MBn of the elementary stream n that is associated with this descriptor.

**TB\_leak\_rate** – This 24-bit field shall specify in units of 400 bits per second the rate at which data is transferred from transport buffer TBn to multiplex buffer MBn for the elementary stream n that is associated with this descriptor.

### 2.6.54 M4MuxTiming descriptor

See Table 2-82.

Table 2-82 – M4MuxTiming descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| M4MuxTiming\_descriptor () {  **descriptor\_tag**  **descriptor\_length**  **FCR\_ES\_ID**  **FCRResolution**  **FCRLength**  **FmxRateLength**  } | **8**  **8**  **16**  **32**  **8**  **8** | **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf** |

### 2.6.55 Semantic definition of fields in M4MuxTiming descriptor

**FCR\_ES\_ID** – is the ES\_ID associated with this clock reference stream

**FCRResolution** – is the resolution of the object time base in cycles per second

**FCRLength** – is the length of the fmxClockReference field in M4Mux packets with index = 238 A length of zero shall indicate that no M4Mux packets with index = 238 are present in this M4Mux stream. FCRlength shall take values between zero and 64.

**FmxRateLength** – Is the length of the fmxRate field in M4Mux packets with index = 238? FmxRateLength shall take values between 1 and 32.

### 2.6.56 Content labelling descriptor

The content labelling descriptor assigns a label to content; the label can be used by metadata to reference the associated content. This label, the content\_reference\_id\_record, is metadata application format specific. The content labelling descriptor is associated with a content segment. For the purpose of this clause, a content segment is defined as a portion in time of a program, an elementary stream (such as audio or video) or any combination of programs or elementary streams. The descriptor may be included in the PMT in the descriptor loop for either the program or an elementary stream, but may also be contained in tables not defined in this Specification, for example tables to describe segments of programs or elementary streams. The content labelling descriptor also provides information on which content time base is used and on the offset between the content time base and the metadata time base. When the Normal Play Time (NPT) concept of DSM-CC, as specified in ISO/IEC 13818-6, is used as the content time base, the ID of the NPT time base is provided. The descriptor allows for carriage of private data. See Table 2-83.

Table 2-83 – Content labelling descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| Content\_labeling\_descriptor () {  **descriptor\_tag**  **descriptor\_length**  **metadata\_application\_format**  if (metadata\_application\_format== 0xFFFF){  **metadata\_application\_format\_identifier** }  **content\_reference\_id\_record\_flag  content\_time\_base\_indicator  reserved**  if (content\_reference\_id\_record\_flag == '1'){  **content\_reference\_id\_record\_length**  for (i=0; i<content\_reference\_id\_record\_length;i++){  **content\_reference\_id\_byte**  }  }  if (content\_time\_base\_indicator == 1 || content\_time\_base\_indicator == 2){  **reserved  content\_time\_base\_value  reserved  metadata\_time\_base\_value** }  if (content\_time\_base\_indicator== 2){  **reserved  contentId** }  if (content\_time\_base\_indicator ≥ 3 && content\_time\_base\_indicator ≤ 7){  **time\_base\_association\_data\_length**  for (i=0; i< time\_base\_association\_data\_length;i++){  **reserved**  }  } for (i=0; i<N;i++){  **private\_data\_byte**  } } | **8 8 16  32  1 4 3  8  8    7 33 7 33   1 7   8  8    8** | **uimsbf uimsbf uimsbf  uimsbf  bslbf uimsbf bslbf  uimsbf  bslbf    bslbf uimsbf bslbf uimsbf   bslbf uimsbf   uimsbf  bslbf**   **bslbf** |

### 2.6.57 Semantic definition of fields in content labelling descriptor

**metadata\_application\_format**: The metadata\_application\_format is a 16-bit field, coded as defined in Table 2-84, that specifies the application responsible for defining usage, syntax and semantics of the content\_reference\_id record and of any other privately defined fields in this descriptor. See also 2.12.1. The value 0xFFFF indicates that the format is signalled by the value carried in the metadata\_application\_format\_identifier field.

| Table 2-84 – Metadata\_application\_format | |
| --- | --- |
| Value | Description |
| 0x0000 .. 0x000F | Reserved |
| 0x0010 | ISO 15706-1 (ISAN) encoded in its binary form (see Notes 1 and 3) |
| 0x0011 | ISO 15706-2 (V-ISAN) encoded in its binary form (see Notes 2 and 3) |
| 0x0012 .. 0x00FF | Reserved |
| 0x0100 .. 0xFFFE | User defined |
| 0xFFFF | Defined by the metadata\_application\_format\_identifier field |
| NOTE 1 – For ISAN, the content\_reference\_id\_byte is set to binary encoding and the content\_reference\_id\_record\_length is set to 0x08.  NOTE 2 – For V-ISAN, the content\_reference\_id\_byte is set to binary encoding and the content\_reference\_id\_record\_length is set to 0x0C.  NOTE 3 – For interoperability amongst metadata applications that use the metadata\_application\_format values of 0x0010 and 0x0011, it is recommended that the content\_reference\_id\_flag be set to '1' and the content\_time\_base\_indicator be set to '00'. | |

**metadata\_application\_format\_identifier**: The coding of this 32-bit field is fully equivalent to the coding of the format\_identifier field in the registration\_descriptor, as defined in 2.6.8.

NOTE – The assigned Registration Authority for the format\_identifier field is SMPTE.

**content\_reference\_id\_record\_flag**: The content\_reference\_id\_record\_flag is a 1-bit flag that signals the presence of a content\_reference\_id\_record in this descriptor.

**content\_time\_base\_indicator**: The content\_time\_base\_indicator (see Table 2-85) is a 4-bit field which specifies the used content time base. If the descriptor is associated with a program, then the content time base applies to all streams that are part of that program. A value of 1 indicates usage of the STC, while a value of 2 indicates usage of NPT, the Normal Play Time as defined in ISO/IEC 13818-6. The values between 8 and 15 indicate usage of a privately defined content time base. If coded with a value of 0, no content time base is defined in this descriptor. If no content time base is specified for a program or stream, then the mapping of time references in the metadata to the content is not defined in this Specification.

Table 2-85 – Content\_time\_base\_indicator values

|  |  |
| --- | --- |
| Value | Description |
| 0 | No content time base defined in this descriptor |
| 1 | Use of STC |
| 2 | Use of NPT |
| 3 .. 7 | Reserved |
| 8 .. 15 | Use of privately defined content time base |

**content\_reference\_id\_record\_length**: The content\_reference\_id\_record\_length is an 8-bit field that specifies the number of content\_reference\_id\_bytes immediately following this field. This field shall not be coded with the value '0'.

**content\_reference\_id\_byte**: The content\_reference\_id\_byte is part of a string of one or more contiguous bytes that assigns one or more reference identifications (labels) to the content to which this descriptor is associated. The format of this byte string is defined by the body indicated by the coded value in the metadata\_application\_format field.

**content\_time\_base\_value**: The content\_time\_base\_value is a 33-bit field that specifies a value in units of 90 kHz of the content time base indicated by the content\_time\_base\_indicator field.

**metadata\_time\_base\_value**: The metadata\_time\_base\_value is a 33-bit field that is coded in units of 90 kHz. The field is coded with the value of the metadata time base at the instant in time in which the time base indicated by content\_time\_base\_indicator reaches the value encoded in the content\_time\_base\_value field. Note that the metadata time base may use any time-scale, but that its value is to be coded in units of 90 kHz. For example, if a SMPTE type of time code is used, then the number of hours, minutes, seconds and frames is expressed in the corresponding number of 90-kHz units.

**contentId**: The contentId is a 7-bit field that specifies the value of the content\_Id field in the NPT Reference Descriptor for the applied NPT time base.

**time\_base\_association\_data\_length**: The time\_base\_association\_data\_length is an 8-bit field that specifies the number of reserved bytes immediately following this field. The reserved bytes can be used to carry time base association data for time bases defined in future.

**private\_data\_byte**: The private\_data\_byte is an 8-bit field. The private\_data\_bytes represent data, the format of which is defined privately. These bytes can be used to provide additional information as deemed appropriate. The use of these bytes is defined by the metadata application format.

### 2.6.58 Metadata pointer descriptor

The metadata pointer descriptor (see Table 2-86) points to a single metadata service and associates this metadata service with audiovisual content in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream. The metadata is associated with the content within the context of the descriptor. The context is defined by the location of the descriptor. In a transport stream, the descriptor may be located in the PMT in the descriptor loop for either the program or an elementary stream, but may also be located in tables not defined in this Specification, such as tables describing bouquets of broadcast services. The metadata may be located in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, but the same metadata may also be provided on alternative locations, such as the Internet.

The descriptor may contain location information of metadata that is not carried in a Rec. ITU-T H.222.0 | ISO/IEC 13818‑1 stream; the coding of the location information is metadata application format specific. The descriptor allows for carriage of private data.

For metadata carried in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, the descriptor specifies the tools used for such carriage. If the metadata is carried in PES packets, metadata sections, or ISO/IEC 13818-6 synchronized download sections, the metadata\_service\_id field identifies the metadata service in the referenced metadata stream. If an ISO/IEC 13818-6 carousel is used to carry the metadata, then the private data may provide information to signal the metadata service, such as the applied value of the module\_id for carriage of the metadata in a data carousel, and the file name of the metadata when the object carousel is used.

Receivers should be aware that multiple metadata services may be pointed to from the same program or audiovisual stream (as defined by the context of the descriptor). A unique metadata pointer descriptor shall be used to point to each metadata service used by the program or audiovisual stream. Similarly, the same metadata service can be pointed to from several programs or audiovisual streams by using a separate metadata pointer descriptors for each association.

Table 2-86 – Metadata pointer descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| Metadata\_pointer\_descriptor () {  **descriptor\_tag**  **descriptor\_length**  **metadata\_application\_format**  if (metadata\_application\_format== 0xFFFF){  **metadata\_application\_format\_identifier** }  **metadata\_format**  if (metadata\_format== 0xFF){  **metadata\_format\_identifier** }  **metadata\_service\_id  metadata\_locator\_record\_flag  MPEG\_carriage\_flags  reserved** if (metadata\_locator\_record\_flag == '1'){  **metadata\_locator\_record\_length**  for ( i = 0; i < metadata\_locator\_record\_length; i++){  **metadata\_locator\_record\_byte**  }  }  if (MPEG\_carriage\_flags ≤ 2){   **program\_number**  }  if (MPEG\_carriage\_flags == 1){ **transport\_stream\_location**  **transport\_stream\_id**  }  for (i=0; i<N;i++){  **private\_data\_byte**  } } | **8 8 16  32  8  32  8 1 2 5  8  8    16   16 16   8** | **uimsbf uimsbf uimsbf  uimsbf  uimsbf  uimsbf  uimsbf bslbf** **uimsbf bslbf**  **uimsbf   bslbf    uimsbf    uimsbf uimsbf    bslbf** |

### 2.6.59 Semantic definition of fields in metadata pointer descriptor

**metadata\_application\_format**: The metadata\_application\_format is a 16-bit field that specifies the application responsible for defining usage, syntax and semantics of the metadata\_locator\_record record and any other privately defined fields in this descriptor. The coding of this field is defined in Table 2-84 in 2.6.57.

**metadata\_application\_format\_identifier**: The coding of this field is defined in 2.6.57.

**metadata\_format**: The metadata\_format is an 8-bit field that indicates the format and coding of the metadata. The coding of this field is specified in Table 2-87.

Table 2-87 – Metadata format values

|  |  |
| --- | --- |
| Value | Description |
| 0x00 .. 0x0F | Reserved |
| 0x10 | ISO/IEC 15938-1 TeM |
| 0x11 | ISO/IEC 15938-1 BiM |
| 0x12 .. 0x3E | Reserved |
| 0x3F | Defined by metadata application format |
| 0x40 .. 0xFE | Private use |
| 0xFF | Defined by metadata\_format\_identifier field |

The values 0x10 and 0x11 identify ISO/IEC 15938-1 defined data. The value 0x3F indicates that the format is defined by the body indicated by the metadata\_application\_format field. The values in the inclusive range of 0x40 up to 0xFE are available to signal use of private formats. The value 0xFF indicates that the format is signalled by the metadata\_format\_identifier field.

**metadata\_format\_identifier**: The coding of this 32-bit field is fully equivalent to the coding of the format\_identifier field in the registration\_descriptor, as defined in 2.6.8.

NOTE – SMPTE is assigned as Registration Authority for the format\_identifier field.

**metadata\_service\_id**: This 8-bit field references the metadata service. It is used for retrieving a metadata service from within a metadata stream.

**metadata\_locator\_record\_flag**: The metadata\_locator\_record\_flag is a 1-bit field which, when set to '1' indicates that associated metadata is available on a location outside of a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, specified in a metadata\_locator\_record.

**MPEG\_carriage\_flags**: The MPEG\_carriage\_flags is a 2-bit field which specifies if the metadata stream containing the associated metadata service is carried in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, and if so, whether the associated metadata is carried in a transport stream or program stream. The coding of this field is defined in Table 2‑88.

Table 2-88 – MPEG\_carriage\_flags

|  |  |
| --- | --- |
| Value | Description |
| 0 | Carriage in the same transport stream where this metadata pointer descriptor is carried. |
| 1 | Carriage in a different transport stream from where this metadata pointer descriptor is carried. |
| 2 | Carriage in a program stream. This may or may not be the same program stream in which this metadata pointer descriptor is carried. |
| 3 | None of the above. |

**metadata\_locator\_record\_length**: The metadata\_locator\_record\_length is an 8-bit field that specifies the number of metadata\_locator\_record\_bytes immediately following. This field shall not be coded with the value 0.

**metadata\_locator\_record\_byte**: The metadata\_locator\_record\_byte is part of a string of one or more contiguous bytes that form the metadata locator record. This record specifies one or more locations outside of a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream. The format of the metadata locator record is defined by the metadata application signalled by the metadata\_application\_format field. The record may for example contain Internet URLs that specify where the metadata can be found, possibly in addition to their location(s) in the transport stream. If the MPEG\_carriage\_flags is coded with the value 0, 1 or 2 and the metadata locator record is present, then this signals alternative locations for the same metadata.

**program\_number**: The program\_number is a 16-bit field that identifies the program\_number of the MPEG-2 program in the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream in which associated metadata is carried. If the MPEG\_carriage\_flags have the value 0, then the transport stream is the current one, and if the MPEG\_carriage\_flags have the value 1, it is the transport stream signalled by the field's transport\_stream\_location and transport\_stream\_id.

**transport\_stream\_location**: The transport\_stream\_location is a 16-bit field that is defined privately. For example, this field may be used by applications to signal the original\_network\_id defined by ETSI.

**transport\_stream\_id**: The transport\_stream\_id is a 16-bit field that identifies the transport stream in which associated metadata is carried.

**private\_data\_byte**: The private\_data\_byte is an 8-bit field. The private\_data\_bytes represent data, the format of which is defined privately. These bytes can be used to provide additional information as deemed appropriate.

### 2.6.60 Metadata descriptor

The metadata descriptor (see Table 2-89) specifies parameters of a metadata service carried in an MPEG-2 TS or PS. In an MPEG-2 TS, the descriptor is included in the PMT in the descriptor loop for the elementary stream that carries the metadata service. The descriptor specifies the format of the associated metadata, and contains the value of the metadata\_service\_id to identify the metadata service to which the metadata descriptor applies. As needed, the descriptor can convey information to identify the metadata service from a collection of metadata transmitted in a DSM-CC carousel. Optionally metadata application format specific private data can be carried.

The metadata descriptor also signals whether decoder configuration is required and is able to carry the decoder configuration bytes, but this is only practical if the number of these bytes is small. If the decoder configuration information is too large to be carried by the descriptor, it shall be contained in a metadata service. This may be within the metadata service itself, or in another metadata service within the same program. Identification of the metadata service that contains the decoder configuration is provided by the metadata descriptor. If a DSM-CC carousel is used to carry the decoder configuration, then information can be provided how to retrieve the decoder configuration from the carousel.

| Table 2-89 – Metadata descriptor | | |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| Metadata\_descriptor () {  **descriptor\_tag**  **descriptor\_length**  **metadata\_application\_format**  if (metadata\_application\_format == 0xFFFF) {  **metadata\_application\_format\_identifier** } **metadata\_format**  if (metadata\_format== 0xFF){  **metadata\_format\_identifier** }  **metadata\_service\_id   decoder\_config\_flags  DSM-CC\_flag  reserved**  if (DSM-CC\_flag == '1'){  **service\_identification\_length** for(i=0; i<service\_identification\_length; i++) {  **service\_identification\_record\_byte** } }  if (decoder\_config\_flags == '001') {  **decoder\_config\_length** for(i=0; i<decoder\_config\_length; i++) {  **decoder\_config\_byte** } }  if (decoder\_config\_flags == '011') {  **dec\_config\_identification\_record\_length** for(i=0;i<dec\_config\_id\_record\_length;i++) {  **dec\_config\_identification\_record\_byte**  }}  if (decoder\_config\_flags == '100') {  **decoder\_config\_metadata\_service\_id**   }  if (decoder\_config\_flags == '101' || decoder\_config\_flags == '110') {  **reserved\_data\_length** for(i=0;i<reserved\_data\_length;i++) {  **reserved**  } }  for (i=0; i<N;i++) {  **private\_data\_byte**  } } | **8 8 16  32  8  32  8 3 1 4  8  8    8  8    8  8    8   8  8    8** | **uimsbf uimsbf uimsbf  uimsbf  uimsbf  uimsbf  uimsbf bslbf bslbf bslbf  uimsbf  bslbf    uimsbf  bslbf    uimsbf  bslbf    uimsbf   uimsbf  bslbf    bslbf** |

### 2.6.61 Semantic definition of fields in metadata descriptor

**metadata\_application\_format**: The metadata\_application\_format is a 16-bit field that specifies the application responsible for defining usage, syntax and semantics of the service\_identification\_record and any privately defined bytes in this descriptor. The coding of this field is defined in Table 2-84.

**metadata\_application\_format\_identifier**: The coding of this field is defined in 2.6.57.

**metadata\_format**: The coding of this field is defined in 2.6.59.

**metadata\_format\_identifier**: The coding of this field is defined in 2.6.59.

**metadata\_service\_id**. This 8-bit field identifies the metadata service to which this metadata descriptor applies.

**decoder\_config\_flags**: The decoder\_config\_flags is a 3-bit field which indicates whether and how decoder configuration information is conveyed. See Table 2-90.

Table 2-90 – decoder\_config\_flags

|  |  |
| --- | --- |
| Value | Description |
| '000' | No decoder configuration is needed. |
| '001' | The decoder configuration is carried in this descriptor in the decoder\_config\_byte field. |
| '010' | The decoder configuration is carried in the same metadata service as to which this metadata descriptor applies. |
| '011' | The decoder configuration is carried in a DSM-CC carousel. This value shall only be used if the metadata service to which this descriptor applies is using the same type of DSM-CC carousel. |
| '100' | The decoder configuration is carried in another metadata service within the same program, as identified by the decoder\_config\_metadata\_service\_id field in this metadata descriptor. |
| '101' .. '110' | Reserved. |
| '111' | Privately defined. |

**DSM-CC\_flag**: This is a one-bit flag that is set to '1' if the stream with which this descriptor is associated is carried in an ISO/IEC 13818-6 data or object carousel.

NOTE 1 – The use of the object or data carousel is indicated by the applied stream-type value for this metadata stream.

**service\_identification length**: This field specifies the number of service\_identification\_record\_bytes immediately following.

**service\_identification\_record\_byte**: This byte is part of a string of one or more contiguous bytes that specify the service\_identification\_record. This record contains data on retrieval of the metadata service from a DSM-CC carousel. The format of the metadata locator record is defined by the application indicated by the metadata application format. When a DSM-CC object carousel is used, the record may for example comprise the unique object identifier (the IOP:IOR() from 11.3.1 and 5.7.2.3 of ISO/IEC 13818-6 DSM-CC) for the metadata service. Similarly, in case of a DSM-CC data carousel, the record can for example provide the transaction\_id and the module\_id of the metadata service.

**decoder\_config\_length**: This field specifies the number of decoder\_config\_bytes immediately following.

**decoder\_config\_byte**: These bytes comprise the decoder configuration information. This sequence of bytes comprises the configuration information needed by the receiver to decode this service. It is intended that carriage in the metadata descriptor is only used when the configuration information is very small.

**dec\_config\_identification\_record\_length**: This field specifies the immediately following number of dec\_config\_identification\_record\_bytes.

**dec\_config\_identification\_record\_byte**: This byte is part of a string of one or more contiguous bytes that specify the dec\_config\_identification\_record. This record specifies how to retrieve the required decoder configuration from a DSM-CC carousel. The format of the metadata locator record is defined by the metadata application format. When a DSM-CC object carousel is used, the record may for example comprise the unique object identifier (the IOP:IOR() from 11.3.1 and 5.7.2.3 of ISO/IEC 13818-6 DSM-CC) for the decoder configuration. Similarly, in case of a DSM-CC data carousel, the record may for example provide the transaction\_id and the module\_id of the decoder configuration.

**decoder\_config\_metadata\_service\_id**: This is the value of the metadata\_service\_id that is assigned to the metadata service that contains the decoder configuration. The metadata service indicated by the decoder\_config\_metadata\_service\_id and the metadata service that uses that decoder configuration shall be in the same program. Hence in a transport stream, the metadata descriptors for both these metadata services shall be in the same PMT. The metadata descriptor of the metadata service indicated by the decoder\_config\_metadata\_service\_id shall have a decoder\_config\_flag field with a value of either '001', '010' or '011'.

**reserved\_data\_length**: This field specifies the number of reserved bytes immediately following.

**private\_data\_byte**: The private\_data\_byte is an 8-bit field. The private\_data\_bytes represent data, the format of which is defined privately. These bytes can be used to provide additional information as deemed appropriate.

### 2.6.62 Metadata STD descriptor

This descriptor defines parameters of the STD model (defined in 2.12.10) for the processing of the metadata stream to which this descriptor is associated. See Table 2-91.

Table 2-91 – Metadata STD descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| Metadata\_STD\_descriptor () {  **descriptor\_tag  descriptor\_length  reserved  metadata\_input\_leak\_rate  reserved  metadata\_buffer\_size  reserved  metadata\_output\_leak\_rate** } | **8 8 2 22 2 22 2 22** | **uimsbf uimsbf bslbf uimsbf bslbf uimsbf bslbf uimsbf** |

### 2.6.63 Semantic definition of fields in metadata STD descriptor

**metadata\_input\_leak\_rate**: The metadata\_input\_leak\_rate is a 22-bit field that specifies the leak rate for the associated metadata stream in the T-STD model out of the buffer TBn into buffer Bn. The leak rate is specified in units of 400 bits/s. For metadata carried in a program stream, the coding of the metadata\_input\_leak\_rate field is not specified, as the rate into Bn equals the rate of the program stream.

**metadata\_buffer\_size**: The metadata\_buffer\_size is a 22-bit field that specifies the size of buffer Bn in the STD model for the associated metadata stream. The size of Bn is specified in units of 1024 bytes.

**metadata\_output\_leak\_rate**: The metadata\_output\_leak\_rate is a 22-bit field that specifies for the associated metadata service the leak rate in the STD model out of buffer Bn to the decoder. The leak rate is specified in units of 400 bits/s. For metadata streams transported synchronously (stream-type 0x15 or 0x19), the metadata access units are instantaneously removed from Bn under the control of PTS timestamps and in that case the coding of the metadata\_output\_leak\_rate field is not specified.

### 2.6.64 AVC video descriptor

For AVC video streams, the AVC video descriptor provides basic information for identifying coding parameters of the associated AVC video stream, such as on profile and level parameters included in the SPS of an AVC video stream or in the subset SPS of an SVC video sub-bitstream.

For AVC video streams conforming to one or more profiles defined in Annex G, or Annex H or Annex I of Rec. ITU-T H.264 | ISO/IEC 14496 10, there may be one AVC video descriptor associated to each of the video sub-bitstreams, or MVC video subsets or MVCD video subsets identifying coding parameters of the associated re-assembled AVC video streams.

The AVC video descriptor also signals the presence of AVC still pictures, AVC 24-hour pictures as well as 3D rendering assistance SEIs such as frame packing arrangement SEI message or stereo video information SEI message in the AVC video stream. If this descriptor is not included in the PMT for an AVC video stream, a video sub-bitstream, or an MVC video sub-bitstream or an MVCD video sub-bitstream in a transport stream or in the PSM, if present, for an AVC video stream, a video sub bitstream or an MVC video sub-bitstream in a program stream, then such AVC video stream shall not contain AVC still pictures, shall not contain AVC 24-hour pictures and may or may not contain frame packing arrangement SEI message or stereo video information SEI message. (See Table 2-92.)

| Table 2-92 – AVC video descriptor | | |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| AVC\_video\_descriptor() {  **descriptor\_tag**  **descriptor\_length**  **profile\_idc**  **constraint\_set0\_flag**  **constraint\_set1\_flag**  **constraint\_set2\_flag**  **constraint\_set3\_flag**  **constraint\_set4\_flag**  **constraint\_set5\_flag**  **AVC\_compatible\_flags**  **level\_idc**  **AVC\_still\_present**  **AVC\_24\_hour\_picture\_flag**  **Frame\_Packing\_SEI\_not\_present\_flag**  **reserved**  **}** | **8**  **8**  **8**  **1**  **1**  **1**  **1**  **1**  **1**  **2**  **8**  **1**  **1**  **1**  **5** | **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf** |

### 2.6.65 Semantic definition of fields in AVC video descriptor

**profile\_idc, constraint\_set0\_flag, constraint\_set1\_flag, constraint\_set2\_flag, constraint\_set3\_flag, constraint\_set4\_flag, constraint\_set5\_flag and AVC\_compatible\_flags and level\_idc** – These fields, with the exception of AVC\_compatible\_flags, shall be coded according to the semantics for these fields defined in Rec. ITU‑T H.264 | ISO/IEC 14496-10. The semantics of AVC\_compatible\_flags are exactly equal to the semantics of the field(s) defined for the 2 bits between the *constraint\_set5 flag* and the *level\_idc* field in the sequence parameter set, as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10. The entire AVC video stream to which the AVC descriptor is associated shall conform to the profile, level and constraints signalled by these fields.

NOTE – In one or more sequences in the AVC video stream the level may be lower than the level signalled in the AVC video descriptor, while also a profile may occur that is a subset of the profile signalled in the AVC video descriptor. However, in the entire AVC video stream, only tools shall be used that are included in the profile signalled in the AVC video descriptor, if present. For example, if the main profile is signalled, then the baseline profile may be used in some sequences, but only using those tools that are in the main profile. If the sequence parameter sets in an AVC video stream signal different profiles, and no additional constraints are signalled, then the stream may need examination to determine which profile, if any, the entire stream conforms to. If an AVC video descriptor is to be associated with an AVC video stream that does not conform to a single profile, then the AVC video stream must be partitioned into two or more sub-streams, so that AVC video descriptors can signal a single profile for each such sub-stream.

**AVC\_still\_present** – This 1-bit field when set to '1' indicates that the AVC video stream may include AVC still pictures. When set to '0', then the associated AVC video stream shall not contain AVC still pictures.

**AVC\_24\_hour\_picture\_flag** – This 1-bit flag when set to '1' indicates that the associated AVC video stream may contain AVC 24-hour pictures. For the definition of an AVC 24-hour picture, see 2.1.2. If this flag is set to '0', the associated AVC video stream shall not contain any AVC 24-hour picture.

**Frame\_Packing\_SEI\_not\_present\_flag** – If this flag is set to '0', then the AVC video stream shall contain either the frame packing arrangement SEI message or stereo video information SEI message. If the AVC video descriptor is present and this flag is set to '1', then the presence of either of these SEI messages is unspecified.

### 2.6.66 AVC timing and HRD descriptor

The AVC timing and HRD descriptor provides timing and HRD parameters of the associated AVC video stream. For each AVC video stream and for each video sub-bitstream or MVC video sub-bitstream or MVCD video sub-bitstream carried in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, the AVC timing and HRD descriptor shall be included in the PMT or in the PSM, if PSM is present in the program stream, unless the AVC video stream, the video sub-bitstream or the MVC video sub-bitstream or MVCD video sub-bitstream carries VUI parameters with the timing\_info\_present\_flag set to '1':

• for each IDR picture or re-assembled IDR picture; and

• for each picture or re-assembled picture that is associated with a recovery point SEI message.

Absence of the AVC timing and HRD descriptor in the PMT for an AVC video stream or a re-assembled AVC video stream signals usage of the leak method in the T-STD for the transfer from MBn to EBn as defined:

• in 2.14.3.1 for an AVC video stream conforming to one or more profiles defined in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496-10;

• in 2.14.3.5 for video sub-bitstreams of an AVC video stream conforming to one or more profiles defined in Annex G of Rec. ITU-T H.264 | ISO/IEC 14496-10;

• in 2.14.3.7 for MVC video sub-bitstreams or MVCD video sub-bitstreams of an AVC video stream conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496-10.

But such usage can also be signalled by the hrd\_management\_valid\_flag set to '0' in the AVC timing and HRD descriptor. If the transfer rate into buffer EBn can be determined from HRD parameters contained in an AVC video stream or an AVC video stream re-assembled from video sub-bitstreams or MVC video sub-bitstreams or MVCD video sub-bitstreams, and if this transfer rate is used in the T-STD for the transfer between MBn to EBn, then the AVC timing and HRD descriptor with the hrd\_management\_valid\_flag set to '1' shall be included in the PMT for that AVC video stream or for the re‑assembled AVC video stream. (See Table 2-93.)

Table 2-93 – AVC timing and HRD descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| AVC\_timing\_and\_HRD\_descriptor () { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **hrd\_management\_valid\_flag** | **1** | **bslbf** |
| **reserved** | **6** | **bslbf** |
| **picture\_and\_timing\_info\_present** | **1** | **bslbf** |
| if (picture\_and\_timing\_info\_present) { |  |  |
| **90kHz\_flag** | **1** | **bslbf** |
| **reserved** | **7** | **bslbf** |
| if (90kHz\_flag = = '0') { |  |  |
| **N** | **32** | **uimsbf** |
| **K** | **32** | **uimsbf** |
| **}** |  |  |
| **num\_units\_in\_tick** | **32** | **uimsbf** |
| **}** |  |  |
| **fixed\_frame\_rate\_flag** | **1** | **bslbf** |
| **temporal\_poc\_flag** | **1** | **bslbf** |
| **picture\_to\_display\_conversion\_flag** | **1** | **bslbf** |
| **reserved** | **5** | **bslbf** |
| } |  |  |

### 2.6.67 Semantic definition of fields in AVC timing and HRD descriptor

**hrd\_management\_valid\_flag** – This 1-bit field is only defined for use in transport streams.

When the AVC timing and HRD descriptor is associated with an AVC video stream or a re-assembled AVC video stream carried in a transport stream, then the following applies. If the hrd\_management\_valid\_flag is set to '1', then Buffering Period SEI and Picture Timing SEI messages, as defined in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496‑10, shall be present in the associated AVC video stream or re-assembled AVC video stream. These Buffering Period SEI messages shall carry coded initial\_cpb\_removal\_delay and initial\_cpb\_removal\_delay\_offset values for the NAL HRD. If the hrd\_management\_valid\_flag is set to '1', then the transfer of each byte from MBn to EBn in the T-STD shall be according to the delivery schedule for that byte into the CPB in the NAL HRD, as determined from the coded initial\_cpb\_removal\_delay and initial\_cpb\_removal\_delay\_offset values for SchedSelIdx = cpb\_cnt\_minus1. When the hrd\_management\_valid\_flag is set to '0', the leak method for the transfer from MBn to EBn in the T-STD shall be used:

• as defined in 2.14.3.1 for AVC video streams conforming to one or more profiles defined in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496-10;

• as defined in 2.14.3.5 for video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex G of Rec. ITU-T H.264 | ISO/IEC 14496-10.

• as defined in 2.14.3.7 for MVC video sub-bitstreams or MVCD video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496‑10.

When the AVC timing and HRD descriptor is associated with an AVC video stream or a re-assembled AVC video stream carried in a program stream, then the meaning of the hrd\_management\_valid\_flag is not defined.

**picture\_and\_timing\_info\_present** – This 1-bit field when set to '1' indicates that the 90kHz\_flag and parameters for accurate mapping to 90-kHz system clock are included in this descriptor.

**90kHz\_flag, N, K** – The 90kHz\_flag when set to '1' indicates that the frequency of the AVC time base is 90 kHz. For an AVC video stream the frequency of the AVC time base is defined by the AVC parameter time\_scale in VUI parameters, as defined in Annex E of Rec. ITU-T H.264 | ISO/IEC 14496-10. The relationship between the AVC time\_scale and the STC shall be defined by the parameters N and K in this descriptor as follows.

****

where time\_scale denotes the exact frequency of the AVC time base, with K larger than or equal to N.

If the 90kHz\_flag is set to '1', then N equals 1 and K equals 300. If the 90kHz\_flag is set to '0', then the values of N and K are provided by the coded values of the N and K fields.

NOTE 1 – This allows mapping of time expressed in units of time\_scale to 90-kHz units, as needed for the calculation of PTS and DTS timestamps, for example in decoders for AVC access units for which no PTS or DTS is encoded in the PES header.

**num\_units\_in\_tick** – Coded exactly in the same way as the num\_units\_in\_tick field in VUI parameters in Annex E of Rec. ITU-T H.264 | ISO/IEC 14496-10. The information provided by this field shall apply to the entire AVC video stream to which the AVC timing and HRD descriptor is associated.

**fixed\_frame\_rate\_flag** – Coded exactly in the same way as the fixed\_frame\_rate\_flag in VUI parameters in Annex E of Rec. ITU-T H.264 | ISO/IEC 14496-10. When this flag is set to '1', it indicates that the coded frame rate is constant within the associated AVC video stream. When this flag is set to '0', no information about the frame rate of the associated AVC video stream is provided in this descriptor.

**temporal\_poc\_flag –** When the temporal\_poc\_flag is set to '1' and the fixed\_frame\_rate\_flag is set to '1', then the associated AVC video stream shall carry Picture Order Count (POC) information (PicOrderCnt) whereby pictures are counted in units of Δtfi,dpb( n ), where Δtfi,dpb( n ) is specified in equation E-10 of Rec. ITU-T H.264 | ISO/IEC 14496-10. When the temporal\_poc\_flag is set to '0', no information is conveyed regarding any potential relationship between the POC information in the AVC video stream and time.

NOTE 2 – This reduces the overhead necessary to signal timing for each access unit. An effective PTS and DTS can be calculated for access units for which no explicit PTS/DTS is carried. Repetition of most recently presented field of the appropriate parity (or frame) is implied when the difference between the PTSs of the current and the next picture is greater than 2 × Δtfi,dpb (or greater than Δtfi,dpb when *frame\_mbs\_only\_flag* is equal to 1).

**picture\_to\_display\_conversion\_flag** – This 1-bit field when set to '1' indicates that the associated AVC video stream may carry display information on coded pictures by providing the pic\_struct field in picture\_timing SEI messages (see Annex D of Rec. ITU-T H.264 | ISO/IEC 14496-10) and/or by providing the Picture Order Count (POC) information (PicOrderCnt), whereby pictures are counted in units of Δtfi,dpb( n ) (see also the semantics of temporal\_poc\_flag), so that timing information for a successive AVC access unit can be derived from the previous picture in decoding or presentation order.

When the picture\_to\_display\_conversion\_mode\_flag is set to '0', then picture timing SEI messages in the AVC video stream, if present, shall not contain the pic\_struct field, and hence the pic\_struct\_present\_flag shall be set to '0' in the VUI parameters in the AVC video stream.

### 2.6.68 MPEG-2 AAC audio descriptor

For individual ISO/IEC 13818-7 streams directly carried in PES packets, the MPEG-2 AAC audio descriptor defined in Table 2-94 provides basic information for identifying the coding parameters of such audio elementary streams.

Table 2-94 – MPEG-2 AAC\_audio\_descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| MPEG-2\_AAC\_audio\_descriptor () { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **MPEG-2\_AAC\_profile** | **8** | **uimsbf** |
| **MPEG-2\_AAC\_channel\_configuration** | **8** | **uimsbf** |
| **MPEG-2\_AAC\_additional\_information** | **8** | **uimsbf** |
| } |  |  |

### 2.6.69 Semantic definition of fields in MPEG-2 AAC audio descriptor

**MPEG-2\_AAC\_profile** – This 8-bit field indicates the AAC profile according to the index in Table 31 of ISO/IEC 13818‑7:2006.

**MPEG-2\_AAC\_channel\_configuration** – This 8-bit field indicates the number and configuration of audio channels presented to the listener by the AAC decoder for the specified program. Values in the range from 1 to 6 indicate number and configuration of audio channels as given for "Default bitstream index number" in Table 42 of ISO/IEC 13818‑7:2006. All other values indicate that the number and configuration of audio channels is undefined.

**MPEG-2\_AAC\_additional\_information** – This 8-bit field indicates whether or not bandwidth extension data as defined in ISO/IEC 13818-7:2006 is embedded in the AAC bitstream according to Table 2-95.

Table 2-95 – MPEG-2\_AAC\_additional\_information field values

|  |  |
| --- | --- |
| Value | Description |
| 0x00 | AAC data according to ISO/IEC 13818-7:2006 |
| 0x01 | AAC data with Bandwidth Extension data present according to ISO/IEC 13818‑7:2006 |
| 0x02 .. 0xFF | Reserved |

### 2.6.70 MPEG-4 text descriptor

The MPEG-4 text descriptor (see Table 2-96) carries textConfig() specified in ISO/IEC 14496-17 for the associated ISO/IEC 14496-17 text stream, thereby providing basic information needed for the decoding of the associated ISO/IEC 14496-17 stream. For each ISO/IEC 14496-17 text stream carried in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, the MPEG-4 text descriptor shall be included in the PMT or in the PSM, if PSM is present in the program stream.

Table 2-96 – MPEG-4 text descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| MPEG-4\_text\_descriptor () {  **descriptor\_tag**  **descriptor\_length**  textConfig()  } | **8**  **8** | **uimsbf**  **uimsbf** |

### 2.6.71 Semantic definition of fields in MPEG-4 text descriptor

**textConfig()** – This shall carry the TextConfig() of the associated ISO/IEC 14496-17 text stream, as defined in ISO/IEC 14496-17.

### 2.6.72 MPEG-4 audio extension descriptor

The MPEG-4 audio extension descriptor (see Table 2-97) carries zero or more audioProfileLevelIndication fields and zero or one audioSpecificConfig() field, both encoded as specified in ISO/IEC 14496-3. Note that for each ISO/IEC 14496-3 audio stream carried in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, it is required that the MPEG-4 audio descriptor be included in the PMT or in the PSM, if PSM is present in the program stream, while the MPEG-4 audio extension descriptor may be present too, providing additional information. If in the MPEG-4 audio descriptor the MPEG‑4\_audio\_profile\_and\_level field is encoded with the value 0xFF, indicating that the audio profile and level is not specified in the MPEG-4 audio descriptor, then the MPEG-4 audio extension descriptor shall be present in the same PMT or PSM as the MPEG-4 audio descriptor. Note that this descriptor allows to provide the audioSpecificConfig out of band, so as to allow receivers to retrieve information about the associated audio stream without accessing the stream itself. The descriptor also allows to associate an audioSpecificConfig to an audio stream.

Table 2-97 – MPEG-4 audio extension descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| MPEG-4\_audio\_extension\_descriptor () {  **descriptor\_tag**  **descriptor\_length**  **ASC\_flag**  **reserved**  **num\_of\_loops**  for (i=0; i<num\_of\_loops; i++) {  **audioProfileLevelIndication**  }  if (ASC\_flag == '1') {  **ASC\_size**  audioSpecificConfig()  }  } | **8**  **8**  **1**  **3**  **4**  **8**  **8** | **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf** |

### 2.6.73 Semantic definition of fields in MPEG-4 audio extension descriptor

**ASC\_flag** – A one-bit flag signalling the presence of the ASC\_size field in this descriptor.

**num\_of\_loops** – A 4-bit field specifying the number of immediately following audioprofileLevelIndication fields in this descriptor. This field may be encoded with the value zero.

**audioProfileLevelIndication** – The audio profile and level of the associated ISO/IEC 14496-3 audio stream, encoded as specified for the audioprofileLevelIndication field in 1.5.2.1 in ISO/IEC 14496-3. Note that a single ISO/IEC 14496‑3 audio stream may comply to more than one audio profile and level, and that this descriptor is designed to convey up to 15 different audioprofileLevelIndication values.

**ASC\_size** – The number of bytes of the immediately following AudioSpecificConfig().

**audioSpecificConfig()** – The audioSpecificConfig() of the associated ISO/IEC 14496-3 audio stream, as specified in 1.6.2.1 in ISO/IEC 14496-3.

### 2.6.74 Auxiliary video stream descriptor

The auxiliary video stream descriptor (see Table 2-98) specifies parameters for the decoding and interpretation of the auxiliary video stream to which the descriptor is associated. For each auxiliary video stream carried in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, the auxiliary video stream descriptor shall be included in the PMT or in the PSM, if PSM is present in the program stream.

Table 2-98 – Auxiliary video stream descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| Auxiliary\_video\_stream\_descriptor() {  **descriptor\_tag**  **descriptor\_length**  **aux\_video\_codedstreamtype**  si\_rbsp(descriptor\_length-1)  } | **8**  **8**  **8** | **uimsbf**  **uimsbf**  **uimsbf** |

### 2.6.75 Semantic definition of fields in auxiliary video stream descriptor

**aux\_video\_codedstreamtype** – An 8-bit unsigned integer that indicates the compression coding type of the auxiliary video stream. The value of aux\_video\_codedstreamtype shall match one of the stream types defined in Table 2-34 for video (for instance 0x02, 0x10 or 0x1B). In order to convey additional information such as profile/level, a descriptor that corresponds to the aux\_video\_codedstreamtype may also be included in the PMT or in the PSM, if PSM is present in the program stream, for the auxiliary video data stream.

NOTE – For example, if the auxiliary video is encoded using Rec. ITU-T H.264 | ISO/IEC 14496-10 Video, then the value of aux\_video\_codedstreamtype is 0x1B and an AVC video descriptor (descriptor\_tag = 40) can be optionally included.

**si\_rbsp()** – Supplemental information RBSP as defined in ISO/IEC 23002-3. It shall contain at least one auxiliary video supplemental information (AVSI) message (also defined in ISO/IEC 23002-3). The type of auxiliary video is inferred from si\_rbsp(). The total size of si\_rbsp() shall not exceed 254 bytes.

### 2.6.76 SVC extension descriptor

For video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex G of Rec. ITU‑T H.264 | ISO/IEC 14496-10, the SVC extension descriptor (see Table 2-99) provides information about the AVC video stream resulting from re-assembling (up to) the associated video sub-bitstream and provides information about scalability and re-assembly of the associated video sub-bitstream. There may be one SVC extension descriptor associated with any of the video sub-bitstreams of an AVC video stream conforming to one or more profiles defined in Annex G of Rec. ITU‑T H.264 | ISO/IEC 14496-10.

Table 2-99 – SVC extension descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| SVC\_extension\_descriptor() {  **descriptor\_tag**  **descriptor\_length**  **width**  **height**  **frame\_rate**  **average\_bitrate**  **maximum\_bitrate**  **dependency\_id**  **reserved**  **quality\_id\_start**  **quality\_id\_end**  **temporal\_id\_start**  **temporal\_id\_end**  **no\_sei\_nal\_unit\_present**  **reserved**  } | **8**  **8**  **16**  **16**  **16**  **16**  **16**  **3**  **5**  **4**  **4**  **3**  **3**  **1**  **1** | **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf** |

### 2.6.77 Semantic definition of fields in SVC extension descriptor

**width** – This 16-bit field indicates the maximum image width resolution, in pixels of the re-assembled AVC video stream.

**height** – This 16-bit field indicates the maximum image height resolution, in pixels of the re-assembled AVC video stream.

**frame\_rate** – This 16-bit field indicates the maximum frame rate, in frames/256 seconds of the re-assembled AVC video stream.

**average\_bitrate** – This 16-bit field indicates the average bit rate, in kbit per second, of the re-assembled AVC video stream.

**maximum\_bitrate** – This 16-bit field indicates the maximum bit rate, in kbit per second, of the re-assembled AVC video stream.

**dependency\_id** – This 3-bit field indicates the value of dependency\_id associated with the video sub-bitstream.

**quality\_id\_start** – This 4-bit field indicates the minimum value of the quality\_id of the NAL unit header syntax element of all the NAL units contained in the associated video sub-bitstream.

**quality\_id\_end** – This 4-bit field indicates the maximum value of the quality\_id of the NAL unit header syntax element of all the NAL units contained in the associated video sub-bitstream.

**temporal\_id\_start** – This 3-bit field indicates the minimum value of the temporal\_id of the NAL unit header syntax element of all the NAL units contained in the associated video sub-bitstream.

**temporal\_id\_end** – This 3-bit field indicates the maximum value of the temporal\_id of the NAL unit header syntax element of all the NAL units contained in the associated video sub-bitstream.

**no\_sei\_nal\_unit\_present** – This 1-bit flag when set to '1' indicates that no SEI NAL units are present in the associated video sub-bitstream.

NOTE – In case the no\_sei\_nal\_unit\_present flag is set to '1' for all SVC video sub-bitstreams and is not set to '1' or not present for the AVC video sub-bitstream of SVC, any SEI NAL units, if present, are included in the AVC video sub-bitstream of SVC. If the SVC extension descriptor is absent for all video sub-bitstreams, SEI NAL units may be present in any SVC dependency representation of an SVC video sub-bitstream, and may require re-ordering to the order of NAL units within an access unit as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10 before access unit re-assembling.

### 2.6.78 MVC extension descriptor

For MVC video sub-bitstreams of AVC video streams conforming to one or more profiles defined in Annex H of Rec. ITU‑T H.264 | ISO/IEC 14496-10, the MVC extension descriptor (see Table 2-100) provides information about the AVC video stream resulting from reassembling (up to) the associated MVC video sub-bitstream and provides information about the contained MVC video sub-bitstream and for the reassembly of the associated MVC video sub-bitstream. There may be one MVC extension descriptor associated with any of the MVC video sub-bitstreams (with stream\_type equal to 0x20) of an AVC video stream conforming to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496‑10. When the MVC video sub-bitstream is an MVC base view sub-bitstream, the MVC extension descriptor shall be present in the associated PMT or PSM for stream\_type equal to 0x1B.

This descriptor can also be used by applications that require association between stereoscopic MVC views and left or right eye using the syntax elements 'view\_association\_not\_present' and 'base\_view\_is\_left\_eyeview'.

Table 2-100 – MVC extension descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| MVC\_extension\_descriptor() { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **average\_bit\_rate** | **16** | **uimsbf** |
| **maximum\_bitrate** | **16** | **uimsbf** |
| **view\_association\_not\_present**  **base\_view\_is\_left\_eyeview**  **reserved** | **1**  **1**  **2** | **bslbf**  **bslbf**  **bslbf** |
| **view\_order\_index\_min** | **10** | **bslbf** |
| **view\_order\_index\_max** | **10** | **bslbf** |
| **temporal\_id\_start** | **3** | **bslbf** |
| **temporal\_id\_end** | **3** | **bslbf** |
| **no\_sei\_nal\_unit\_present** | **1** | **bslbf** |
| **no\_prefix\_nal\_unit\_present** | **1** | **bslbf** |
| } |  |  |

### 2.6.79 Semantics of fields in MVC extension descriptor

**average\_bitrate** – This 16-bit field indicates the average bit rate, in kbits per second, of the re-assembled AVC video stream. When set to 0, the average bit rate is not indicated.

**maximum\_bitrate** – This 16-bit field indicates the maximum bit rate, in kbits per second, of the re-assembled AVC video stream. When set to 0, the maximum bit rate is not indicated.

**view\_association\_not\_present** – This 1-bit flag when set to '0' indicates that the syntax element base\_view\_is\_left\_eyeview signals the association between base view and left or right eye. When this flag is set to '1' no such association is signalled.

**base\_view\_is\_left\_eyeview** – This flag shall be set to '1' when the view\_association\_not\_present\_flag is set to '1' and no view association is conveyed in the descriptor. When the view\_association\_not\_present\_flag is set to '0' and this flag is set to '1', it indicates that the base view is associated with the left eye view (or enhancement view is associated with the right eye view). When the view\_association\_not\_present\_flag is set to '0' and this flag is set to '0', it indicates that the base view is associated with the right eye view (or enhancement view is associated with the left eye view).

**view\_order\_index\_min** – This 10-bit field indicates the minimum value of the view order index of all the NAL units contained in the associated MVC video sub-bitstream.

**view\_order\_index\_max** – This 10-bit field indicates the maximum value of the view order index of all the NAL units contained in the associated MVC video sub-bitstream.

**temporal\_id\_start** – This 3-bit field indicates the minimum value of the temporal\_id of the NAL unit header syntax element of all the NAL units contained in the associated MVC video sub-bitstream.

**temporal\_id\_end** – This 3-bit field indicates the maximum value of the temporal\_id of the NAL unit header syntax element of all the NAL units contained in the associated MVC video sub-bitstream.

**no\_sei\_nal\_unit\_present** – This 1-bit flag when set to '1' indicates that no SEI NAL units are present in the associated video sub–bitstream.

NOTE – In case the no\_sei\_nal\_unit\_present flag is set to '1' for all MVC video sub-bitstreams and is not set to '1' or not present for the AVC video sub-bitstream of MVC, any SEI NAL units, if present, are included in the AVC video sub-bitstream of MVC. If the MVC extension descriptor is absent for all MVC video sub-bitstreams, SEI NAL units may be present in any MVC view‑component subset of an MVC video sub-bitstream, and may require re-ordering to the order of NAL units within an access unit as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10 before access unit re-assembling.

**no\_prefix\_nal\_unit\_present** – This 1-bit flag when set to '1' indicates that no prefix NAL units are present in either the AVC video sub-bitstream of MVC or MVC video sub-bitstreams. When this bit is set to '0', it indicates that prefix NAL units are present in the AVC video sub-bitstream of MVC only.

### 2.6.80 J2K video descriptor

For J2K video elementary streams conforming to one or more profiles defined in Rec. ITU-T T.800 | ISO/IEC 15444-1, the J2K video descriptor (see Table 2-101) provides information that may be present in each J2K access unit as well as for the J2K video sequence. In addition, it provides information to signal J2K still pictures. This descriptor shall be included for each J2K video elementary stream component in the PMT with stream\_type equal to 0x21.

| Table 2-101 – J2K video descriptor | | |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| J2K\_video\_descriptor() {  **descriptor\_tag**  **descriptor\_length**  **extended\_capability\_flag**  **profile\_and\_level**  **horizontal\_size**  **vertical\_size**  **max\_bit\_rate**  **max\_buffer\_size**  **DEN\_frame\_rate**  **NUM\_frame\_rate**  if (extended\_capability\_flag == '1') {  **stripe\_flag**  **block\_flag**  **mdm\_flag**  **reserved (all bits to be set to '0')**  } else {  **color\_specification**  }  **still\_mode**  **interlaced\_video**  **reserved**  if (extended\_capability\_flag == '1') {  **colour\_primaries**  **transfer\_characteristics**  **matrix\_coefficients**  **video\_full\_range\_flag**  **reserved**  if (stripe\_flag == '1') {  **strp\_max\_idx**  **strp\_height**  }  if (block\_flag == '1') {  **full\_horizontal\_size**  **full\_vertical\_size**  **blk\_width**  **blk\_height**  **max\_blk\_idx\_h**  **max\_blk\_idx\_v**  **blk\_idx\_h**  **blk\_idx\_v**  }  if (mdm\_flag == '1') {  **X\_c0, Y\_c0, X\_c1, Y\_c1, X\_c2, Y\_c2**  **X\_wp**  **Y\_wp**  **L\_max**  **L\_min**  **MaxCLL**  **MaxFALL**  }  }  for (i=0; i<N; i++) {  **private\_data\_byte**  }  } | **8**  **8**  **1**  **15**  **32**  **32**  **32**  **32**  **16**  **16**  **1**  **1**  **1**  **5**  **8**  **1**  **1**  **6**  **8**  **8**  **8**  **1**  **7**  **8**  **16**  **32**  **32**  **16**  **16**  **8**  **8**  **8**  **8**  **16x6**  **16**  **16**  **32**  **32**  **16**  **16**  **8** | **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf** |

### 2.6.81 Semantics of fields in J2K video descriptor

**extended\_capability\_flag** – This 1-bit field indicates that the J2K video stream uses extended color specification (through three bytes defining the chromaticity parameters, as described below), and that it might have one or several of the following capabilities enabled: stripes (through the J2K stripe mode), blocks (through the J2K block mode), or inclusion of mastering display metadata. The exact list of enabled capabilities is set through subsequent flags in the video descriptor (see below).

**profile\_and\_level** – This 15-bit field shall correspond to the 15 least significant bits of the 2-bytes Rsiz value included in all J2K codestream main headers of this J2K video stream. Rsiz values that are defined in Table A.10 of Rec. ITU‑T T.800 | ISO/IEC 15444-1 and do set to '0' their most significant bit are allowed.

NOTE – the combination of the extend\_capability\_flag and the profile\_and\_level field ensures backward and forward compatibility with legacy devices conforming to previous versions of this Recommendation | International Standard. Having the extended\_capability\_flag set to '1' leads indeed to a 16-bit value outside the range accepted by previous versions of this Recommendation | International Standard. This way, J2K video streams with extended capabilities can be unequivocally identified by both legacy and new devices.

**horizontal\_size –** This 32-bit field indicates the horizontal size of the frame (for progressive) or field (for interlaced) comprised in each J2K access unit. If J2K block mode is enabled, this frame or field corresponds to a spatial rectangular block of the entire video frame or field. It shall be coded the same as the Xsiz parameter found in all J2K codestream main headers of this J2K video stream, as defined in Annex A of Rec. ITU-T T.800 | ISO/IEC 15444-1.

**vertical\_size –** This 32-bit field indicates the vertical size of the frame (for progressive) or field (for interlaced) comprised in each J2K access unit. If J2K block mode is enabled, this frame or field corresponds to a spatial rectangular block of the entire video frame or field. If J2K stripe mode is disabled,it shall be coded the same as the Ysiz parameter found in all J2K codestream main headers of this J2K video stream. If J2K stripe mode is enabled, it shall be coded as the sum of the Ysiz parameters found in all J2K codestreams composing the frame (for progressive) or a field (for interlaced) comprised in each J2K access unit. Ysiz parameters are defined in Annex A of Rec. ITU-T T.800 | ISO/IEC 15444-1.

**max\_bit\_rate** **–** This field may be coded the same as the *brat\_max\_br* field specified in Table S.1 and shall not exceed the maximum compressed bit rate value for the profile and level specified in Table S.2. This field shall be set appropriately and signalled when profile\_and\_level = '000 0011 0000 0111', where no maximum bit rate is specified.

**max\_buffer\_size** – This field shall not exceed the Maximum buffer size value for the profile and level specified in Table S.2. When profile\_and\_level = '000 0011 0000 0111', the max\_buffer\_size shall be set appropriately and shall not exceed (max\_bit\_rate/1.60E5), where max\_bit\_rate is expressed in bit/s.

**DEN\_frame\_rate** **–** This field shall be coded the same as frat\_denominator field specified in Table S.1 (see Annex S).

**NUM\_frame\_rate** **–** This field shall be coded the same as frat\_numerator field specified in Table S.1 (see Annex S).

NOTE – J2K frame rate is derived from the DEN\_frame\_rate and NUM\_frame\_rate values. Table 2-102 lists examples of typical broadcast frame rates with associated values of DEN\_frame\_rate and NUM\_frame\_rate.

Table 2-102 – Example frame rates based on DEN\_frame\_rate and NUM\_frame\_rate values

|  |  |  |  |
| --- | --- | --- | --- |
| DEN\_frame\_rate | NUM\_frame\_rate | Frame rate ratio (decimal representation) | Frame rate |
| 0000 0000 0000 0000 |  |  | Forbidden |
| 0000 0011 1110 1001 | 0101 1101 1100 0000 | 24 000 / 1001 | 23.976 |
| 0000 0000 0000 0001 | 0000 0000 0001 1000 | 24 / 1 | 24.0 |
| 0000 0000 0000 0001 | 0000 0000 0001 1001 | 25 / 1 | 25.0 |
| 0000 0011 1110 1001 | 0111 0101 0011 0000 | 30 000 / 1001 | 29.97 |
| 0000 0000 0000 0001 | 0000 0000 0001 1110 | 30 / 1 | 30.0 |
| 0000 0000 0000 0001 | 0000 0000 0011 0010 | 50 / 1 | 50.0 |
| 0000 0011 1110 1001 | 1110 1010 0110 0000 | 60 000 / 1001 | 59.94 |
| 0000 0000 0000 0001 | 0000 0000 0011 1100 | 60 / 1 | 60.00 |

**stripe\_flag** – This 1-bit field is included only if the extended\_capability\_flag is set to '1'. It indicates whether the J2K video stream has J2K stripe mode enabled. When this flag is set to '1' the J2K access unit elementary stream header (see Table S.1) shall not include the syntax element j2k\_tcod, shall include the syntax element j2k\_strp, and the corresponding J2K access unit shall be made of a succession of J2K stripes. When this flag is set to '0', the J2K access unit elementary stream header shall include the syntax element j2k\_tcod, shall not include the syntax element j2k\_strp, and the corresponding J2K access unit shall be made of one J2K codestream in case of progressive content and two J2K codestreams in case of interlaced content.

**block\_flag** – This 1-bit field is included only if the extended\_capability\_flag is set to '1'. When set to '1', it indicates that the J2K video stream has J2K block mode enabled, meaning that this J2K video stream actually corresponds to a spatial rectangular block of the full video stream. Subdivision of each frame into rectangular independent blocks is further defined in Section S.3. When set to '0', then the associated J2K video stream shall not have J2K block mode enabled.

**mdm\_flag** – This 1-bit field is included only if the extended\_capability\_flag is set to '1'. When set to '1', it indicates that the J2K video descriptor contains the characteristics of the Mastering Display Metadata, as described in SMPTE ST 2086:2014 (see below corresponding fields). When set to '0', then the J2K video descriptor shall not contain the characteristics of the Mastering Display Metadata.

**color\_specification** – This 8-bit field is included only if the extended\_capability\_flag is set to '0' and corresponds to the legacy color specification method. It shall be coded the same as the bcol\_colcr 8-bit field of the j2k\_bcol box as specified in Table S.1 (see Annex S).

**still\_mode** – This 1-bit field, when set to '1', indicates that the J2K video stream may include J2K still pictures. When set to '0', then the associated J2K video stream shall not contain J2K still pictures.

**interlaced\_video** – This 1-bit field indicates whether the J2K video stream contains interlaced video. When this flag is set to '1' the J2K access unit elementary stream header (see Table S.1) shall include the syntax elements brat\_auf2, fiel\_box\_code, fiel\_fic and fiel\_fio. When this flag is set to '0', these syntax elements shall not be present in the J2K access unit elementary stream header.

**color\_primaries, transfer\_characteristics, matrix\_coefficients, video\_full\_range\_flag** – These four fields (three 8‑bit fields and one 1-bit field) are included only if the extended\_capability\_flag is set to '1' and correspond to a color specification method allowing a broader set of color code points than the legacy method (see color\_specification field above). These fields shall be coded according to the semantics with the same name defined in Rec. ITU-T H.273 | ISO/IEC 23001-8.

**strp\_max\_idx** – This 8-bit field is included only if J2K stripe mode is enabled. It shall be in the range 0x01 .. 0xff and indicates the maximum value of the stripe index. It corresponds to the number of stripes in the block/field/frame, minus one. Value 0x00 is forbidden as a minimum of 2 stripes is required (otherwise J2K stripe mode shall be disabled).

**strp\_height** – This 16-bit field is included only if J2K stripe mode is enabled. It indicates the default vertical size of a stripe. Depending on the vertical\_size field value, the last stripe might have a different height, as detailed in S.4.

**full\_horizontal\_size** – This 32-bit field is included only if J2K block mode is enabled. It indicates the horizontal size of the entire video frame of this J2K video stream.

**full\_vertical\_size** – This 32-bit field is included only if J2K block mode is enabled. It indicates the vertical size of the entire video frame of this J2K video stream.

**blk\_width** – This 16-bit field is included only if J2K block mode is enabled. It indicates the default width of a J2K block. Depending on the full\_horizontal\_size field value, the last block of a row might have a different width, as detailed in S.3.

**blk\_height** – This 16-bit field is included only if J2K block mode is enabled. It indicates the default height of a J2K block. Depending on the full\_vertical\_size field value, the last block of a column might have a different height, as detailed in S.3.

**max\_blk\_idx\_h** – This 8-bit field is included only if J2K block mode is enabled and indicates the maximum value of the horizontal block index for this video frame. It corresponds to the total number of blocks in the horizontal direction, minus one.

**max\_blk\_idx\_v** – This 8-bit field is included only if J2K block mode is enabled and indicates the maximum value of the vertical block index for this video frame. It corresponds to the total number of blocks in the vertical direction, minus one.

**blk\_idx\_h** – This 8-bit field is included only if J2K block mode is enabled and indicates the horizontal block index of the current block.

**blk\_idx\_v** – This 8-bit field is included only if J2K block mode is enabled and indicates the vertical block index of the current block.

The following fields X\_c0, Y\_c0, X\_c1, Y\_c1, X\_c2, Y\_c2, X**\_wp**, Y**\_wp**, L\_max and L\_min correspond to the fields defined in SMPTE ST2086:2014 "Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images". The fields MaxFALL and MaxCLL correspond to the fields defined in ANSI/CTA 861-G:2016 "A DTV Profile for Uncompressed High Speed Digital Interfaces". If these 12 fields have unknown values at the time of generating the stream, they shall not be included in the descriptor and the mdm\_flag shall be set to '0'.

**X\_c0, Y\_c0**, **X\_c1, Y\_c1, X\_c2, Y\_c2** – These 16-bit fields are included only if the mdm\_flag is set to '1'. They specify the normalized x and y chromaticity coordinates of the colour primary components of the mastering display in increments of 0.00002, according to the CIE 1931 definition of x and y as specified in ISO 11664-1 (see also ISO 11664-3 and CIE 15). For describing mastering displays that use red, green, and blue colour primaries, it is suggested that index value c0 should correspond to the green primary, c1 should correspond to the blue primary, and c2 should correspond to the red colour primary. The values of these 6 fields shall be in the range of 0 to 50 000, inclusive.

**X\_wp** and **Y\_wp –** These 16-bit fields are included only if the mdm\_flag is set to '1'. They specify the normalized x and y chromaticity coordinates of the white point of the mastering display in normalized increments of 0.00002, according to the CIE 1931 definition of x and y as specified in ISO 11664-1 (see also ISO 11664-3 and CIE 15). The values of X**\_wp** and Y**\_wp** shall be in the range of 0 to 50 000.

**L\_max** and **L\_min** – These 32-bit fields are included only if the mdm\_flag is set to '1'. They specify the nominal maximum and minimum display luminance, respectively, of the mastering display in units of 0.0001 candelas per square metre. L\_min shall be less than L\_max. At minimum luminance, the mastering display is considered to have the same nominal chromaticity as the white point.

**MaxCLL** – This 16-bit field is included only if the mdm\_flag is set to '1'. It specifies the Maximum Content Light Level and corresponds to the brightest pixel in the entire stream, in units of 1 cd/m2, where 0x0001 represents 1 cd/m2 and 0xFFFF represents 65535 cd/m2. It shall be calculated according to Annex P Calculation of MaxCLL and MaxFALL section P.1 in ANSI/CTA 861-G:2016 A DTV Profile for Uncompressed High Speed Digital Interfaces.

**MaxFALL** – This 16-bit field is included only if the mdm\_flag is set to '1'. It specifies the Maximum Frame Average Light Level and corresponds to the highest frame average brightness per frame in the entire stream, in units of 1 cd/m2, where 0x0001 represents 1 cd/m2 and 0xFFFF represents 65535 cd/m2. It shall be calculated according to Annex P Calculation of MaxCLL and MaxFALL section P.2 in ANSI/CTA 861-G:2016 "A DTV Profile for Uncompressed High Speed Digital Interfaces".

If for some reason the MaxCLL and/or MaxFALL values are unknown, the value 0x0000 shall be used.

### 2.6.82 MVC operation point descriptor

The MVC operation point descriptor (see Table 2-103) provides a method to indicate profile and level for one or more operation points each constituted by a set of one or more MVC video sub-bitstreams. If present, the MVC operation point descriptor shall be included in the group of data elements following immediately the program\_info\_length field in the program\_map\_section. If an MVC operation point descriptor is present within a program description, at least one hierarchy descriptor shall be present for each MVC video sub-bitstream present in the same program.

NOTE – In order to indicate different profiles, one MVC operation point descriptor per profile is needed.

Table 2-103 – MVC operation point descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| MVC\_operation\_point\_descriptor() {  **descriptor\_tag**  **descriptor\_length**  **profile\_idc**  **constraint\_set0\_flag**  **constraint\_set1\_flag**  **constraint\_set2\_flag**  **constraint\_set3\_flag**  **constraint\_set4\_flag**  **constraint\_set5\_flag**  **AVC\_compatible\_flags**  **level\_count**  for ( i = 0; i < level\_count; i++) {  **level\_idc**  **operation\_points\_count**  for ( j =0; j< operation\_points\_count; j++ ) {  **reserved**  **applicable\_temporal\_id**  **num\_target\_output\_views**  **ES\_count**  for ( k =0; k< ES\_count; k++ ) {  **reserved**  **ES\_reference**  }  }  }  } | **8**  **8**  **8**  **1**  **1**  **1**  **1**  **1**  **1**  **2**  **8**  **8**  **8**  **5**  **3**  **8**  **8**  **2**  **6** | **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **uimsbf** |

### 2.6.83 Semantic definition of fields in MVC operation point descriptor

**profile\_idc** – This 8-bit field indicates the profile, as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10, of all operation points described within this descriptor for the MVC bitstream.

**constraint\_set0\_flag, constraint\_set1\_flag, constraint\_set2\_flag,** **constraint\_set3\_flag, constraint\_set4\_flag, constraint\_set5\_flag** – These fields shall be coded according to the semantics for these fields defined in Rec. ITU‑T H.264 | ISO/IEC 14496-10.

**AVC\_compatible\_flags** – The semantics of AVC\_compatible\_flags are exactly equal to the semantics of the field(s) defined for the 2 bits between the constraint\_set2 flag and the level\_idc field in the sequence parameter set, as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10.

**level\_count** – This 8-bit field indicates the number of levels for which operation points are described.

**level\_idc** – This 8-bit field indicates the level, as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10, of the MVC bitstream for the operation points described by the following groups of data elements.

**operation\_points\_count** – This 8-bit field indicates the number of operation points described by the list included in the following group of data elements.

**applicable\_temporal\_id** – This 3-bit field indicates the highest value of the temporal\_id of the VCL NAL units in the re-assembled AVC video stream.

**num\_target\_output\_views** – This 8-bit field indicates the value of the number of the views, targeted for output for the associated operation point.

**ES\_count** – This 8-bit field indicates the number of ES\_reference values included in the following group of data elements. The elementary streams indicated in the following group of data elements together form an operation point of the MVC video bitstream. The value 0xff is reserved.

**ES\_reference** – This 6-bit field indicates the hierarchy layer index value present in the hierarchy descriptor which identifies a video sub-bitstream.

NOTE – The profile and level for a single operation point, e.g., the entire MVC video bitstream, can be signalled using the AVC video descriptor. Beyond that, MVC allows for decoding different view subsets which can require different profiles and/or levels. The specification of the MVC operation point descriptor supports the indication of different profiles and levels for multiple operation points.

### 2.6.84 MPEG2\_stereoscopic\_video\_ format\_ descriptor

The MPEG2\_stereoscopic\_video\_format\_descriptor (see Table 2-104) may be associated in the PMT for MPEG-2 video components (with stream\_type value equal to 0x02). When present, the descriptor shall be located in the loop following ES\_info\_length field in PMT. When the descriptor is included in the PMT, the associated MPEG-2 video elementary stream shall contain stereoscopic video format information in the user\_data extension as specified in Rec. ITU-T H.262 | ISO/IEC 13818‑2:2000/Amd.4. If the descriptor is not included for MPEG-2 video with stream\_type value equal to 0x02, then the associated MPEG-2 video elementary stream may or may not contain stereoscopic video format information.

Table 2-104 – MPEG2\_stereoscopic\_video\_format\_descriptor syntax

| Syntax | No. of bits | Format |
| --- | --- | --- |
| MPEG2\_stereoscopic\_video\_format\_descriptor() { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **stereo\_video\_arrangement\_type\_present** | **1** | **bslbf** |
| if (stereo\_video\_arrangement\_type\_present) { |  |  |
| **arrangement\_type** | **7** | **bslbf** |
| } |  |  |
| else { |  |  |
| **reserved**  } | **7** | **bslbf** |
| } |  |  |

### 2.6.85 Semantic definition of fields in the MPEG2\_stereoscopic\_video\_format\_descriptor

**stereo\_video\_arrangement\_type\_present**: When this bit is set to '1', then the following 7-bits indicate the type of stereo\_video\_format\_type included in the user\_data of associated MPEG-2 video elementary stream. If this bit is set to '0', then no such signalling is made available in this descriptor.

**arrangement\_type**: This field shall be set to the same value as arrangement\_type defined in Table L-1 of Rec. ITU‑T H.262 | ISO/IEC 13818-2:2000/Amd.4 and included in the user\_data extension of associated MPEG-2 video elementary stream.

### 2.6.86 Stereoscopic\_program\_info\_descriptor

Stereoscopic\_program\_info\_descriptor (see Table 2-105) specifies the identification of 2D-only (monoscopic), frame compatible stereoscopic 3D as well as service-compatible stereoscopic 3D services. This descriptor conveys information at a program level and assists decoders determine resources required to support the signalled services. When present, this descriptor shall be included in the loop following program\_info\_length field in the PMT. In addition, when this descriptor is present, stream\_type values and descriptors associated with the video components in the loop following ES\_info\_length in the PMT shall not conflict with the stereoscopic\_service\_type signalled in this descriptor.

Table 2-105 – Stereoscopic\_program\_info\_descriptor syntax

| Syntax | No. of bits | Format |
| --- | --- | --- |
| Stereoscopic\_program\_info\_descriptor() { |  |  |
| **descriptor\_tag**  **descriptor\_length**  **reserved**  **stereoscopic\_service\_type** | **8**  **8**  **5**  **3** | **uimsbf**  **uimsbf**  **bslbf**  **bslbf** |
| } |  |  |

### 2.6.87 Semantic definition of fields in the stereoscopic\_program\_info\_descriptor

**descriptor\_length**:This field shall be set to 0x01.

**stereoscopic\_service\_type**: This specifies the type of service that is provided through the associated components such as monoscopic, frame compatible stereoscopic or service-compatible stereoscopic according to Table 2-106.

Table 2-106 – Stereoscopic\_service\_type values

|  |  |
| --- | --- |
| Values | Description |
| '000' | unspecified |
| '001' | 2D-only (monoscopic) service (see Note 1) |
| '010' | Frame-compatible stereoscopic 3D service |
| '011' | Service-compatible stereoscopic 3D service (see Note 2) |
| '100' .. '111' | Rec. ITU-T H.222.0 | ISO/IEC 13818-1 reserved |
| NOTE 1 – 2D-only (monoscopic) service is used in 2D/3D mixed programs based on service-compatible stereoscopic 3D service, i.e., a service that has an arbitrary mixture of 2D and 3D video. 2D video can be coded independently using either MPEG-2 or AVC video to maintain stability of 3DTV broadcasting system.  NOTE 2 – Service-compatible stereoscopic 3D service is based on 'simulcast' of stereoscopic view sequences. Each view of the stereoscopic video sequences can be coded independently using either MPEG-2 or AVC video or any combination thereof. Base view video stream for service-compatible stereoscopic 3D services is signalled using stream\_type value of 0x02 for MPEG-2 video and stream\_type value of 0x1B for AVC video. Additional view video stream for service-compatible stereoscopic 3D services is signalled using stream\_type value of 0x22 for MPEG-2 video and stream\_type value of 0x23 for AVC video. | |

### 2.6.88 Stereoscopic\_video\_info\_descriptor

The stereoscopic\_video\_info\_descriptor (see Table 2-107) provides information related to service-compatible stereoscopic 3D services that carry left and right view in separate video streams. The two streams are called the "base view video stream" and the "additional view video stream". The base view video stream may be displayed on its own for a 2D video service. If signalled as specified below, the additional view video stream may also be displayed on its own for a 2D video service.

The stereoscopic\_video\_info\_descriptor is located in the loop following ES\_info\_length field in PMT. The stereoscopic\_video\_info\_descriptor shall be included for both the base view video component (stream\_type values 0x02 or 0x1B) and additional view video component (stream\_type values 0x22 or 0x23) in the PMT for programs that support service-compatible stereoscopic 3D video. The stereoscopic\_video\_info\_descriptor should not be associated for components with any other stream\_type values as its meaning is undefined for other stream\_type values.

Table 2-107 – Stereoscopic\_video\_info\_descriptor syntax

| Syntax | No. of bits | Format |
| --- | --- | --- |
| Stereoscopic\_ video\_info\_descriptor() {  **descriptor\_tag**  **descriptor\_length**  **reserved**  **base\_video\_flag**  if (base\_video\_flag) {  **reserved**  **leftview\_flag**  } else {  **reserved**  **usable\_as\_2D**  **horizontal\_upsampling\_factor**  **vertical\_upsampling\_factor**  }  } | **8**  **8**  **7**  **1**  **7**  **1**  **7**  **1**  **4**  **4** | **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf** |

### 2.6.89 Semantic definition of fields in the stereoscopic\_video\_info\_descriptor

**base\_video\_flag**: When the bit is set to '1', it indicates that the video stream is a base video stream. When the bit is set to '0', it indicates that the video stream is an additional view video stream. This bit shall be set to '1' for base view video stream components with stream\_type values of 0x02 and 0x1B and this bit shall be set to '0' for the additional view video stream components with stream\_type values of 0x22 and 0x23.

**leftview\_flag**: When the bit is set to '1', it indicates that the associated video stream component is the left view video stream. When the bit is set to '0', it indicates that the associated video stream component is the right view video stream.

**usable\_as\_2D**: When this bit is set to '1', it indicates that the additional view video stream may also be used for a 2D video service.

**horizontal\_upsampling\_factor and vertical\_upsampling\_factor**: These fields provide higher level information on any upsampling that may be required after the video component is decoded. The values and description of upsampling factors are defined in Table 2-108. These values are informational, and that the definitive values are those in the video elementary stream. When this syntax element is set to '0001' decoders are expected to use the information in the video elementary stream to determine appropriate upsampling factors (if needed).

Table 2-108 – Upsampling factor values

|  |  |
| --- | --- |
| Value | Description |
| '0000' | Forbidden |
| '0001' | unspecified |
| '0010' | Coded resolution is same as coded resolution of base view |
| '0011' | Coded resolution is ¾ coded resolution of base view |
| '0100' | Coded resolution is 2/3 coded resolution of base view |
| '0101' | Coded resolution is ½ coded resolution of base view |
| '0110' .. '1000' | reserved |
| '1001' .. '1111' | user\_private |

### 2.6.90 Extension descriptor

The extension descriptor (see Table 2-109) provides a mechanism to extend the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 descriptor range (see Table 2-45). The descriptors which are based on the extension descriptor are signalled using the extension descriptor with extension\_descriptor\_tag values defined in Table 2-110.

| Table 2-109 – Extension descriptor | | |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| Extension\_descriptor () { |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **extension\_descriptor\_tag** | **8** | **uimsbf** |
| if ( extension\_descriptor\_tag == 0x02) { |  |  |
| ObjectDescriptorUpdate() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x03) { |  |  |
| HEVC\_timing\_and\_HRD\_descriptor() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x04) { |  |  |
| af\_extensions\_descriptor () |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x05 ) { |  |  |
| HEVC\_operation\_point\_descriptor( ) |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x06 ) { |  |  |
| HEVC\_hierarchy\_extension\_descriptor( ) |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x07) { |  |  |
| Green\_extension\_descriptor () |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x08) { |  |  |
| MPEG-H\_3dAudio\_descriptor() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x09) { |  |  |
| MPEG-H\_3dAudio\_config\_descriptor() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x0A) { |  |  |
| MPEG-H\_3dAudio\_scene\_descriptor() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x0B) { |  |  |
| MPEG-H\_3dAudio\_text\_label\_descriptor() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x0C) { |  |  |
| MPEG-H\_3dAudio\_multi-stream\_descriptor() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x0D) { |  |  |
| MPEG-H\_3dAudio\_drc\_loudness\_descriptor() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x0E) { |  |  |
| MPEG-H\_3dAudio\_command\_descriptor() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x0F) { |  |  |
| Quality\_extension\_descriptor () |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x10) { |  |  |
| Virtual\_segmentation\_descriptor () |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x11) { |  |  |
| timed\_metadata\_extension\_descriptor() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x12) { |  |  |
| HEVC\_tile\_substream\_descriptor() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x13) { |  |  |
| HEVC\_subregion\_descriptor() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x14) { |  |  |
| JXS\_video\_descriptor() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x15) { |  |  |
| VVC\_timing\_and\_HRD\_descriptor() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x16) { |  |  |
| EVC\_timing\_and\_HRD\_descriptor() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x17) { |  |  |
| LCEVC\_video\_descriptor () |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x18) { |  |  |
| LCEVC\_linkage\_descriptor() |  |  |
| } |  |  |
| else if ( extension\_descriptor\_tag == 0x19) { |  |  |
| Media\_service\_kind\_descriptor() |  |  |
| } |  |  |
| else { |  |  |
| for ( i=0; i<N; i++ ) {  **reserved**  } | **8** | **bslbf** |
| } |  |  |
| } |  |  |

### 2.6.91 Semantic definition of fields in the extension descriptor

**descriptor\_tag** – The descriptor\_tag is an 8-bit field whose value is defined in Table 2-45.

**descriptor\_length** – The descriptor\_length is an 8-bit field specifying the number of bytes of the descriptor immediately following the descriptor\_length field.

**extension\_descriptor\_tag** – The extension\_descriptor\_tag is an 8-bit field which identifies each descriptor that uses this tag value. See Table 2-109 for the extension\_descriptor\_tag values.

**ObjectDescriptorUpdate()** – This structure is defined in section 8.5.5.2 of ISO/IEC 14496-1.

**HEVC\_timing\_and\_HRD\_descriptor()** – This structure is defined in 2.6.95 and 2.6.96.

**af\_extensions\_descriptor()** – This structure is defined in 2.6.99.

**HEVC\_operation\_point\_descriptor( )** – This structure is defined in 2.6.100 and 2.6.101.

**HEVC\_hierarchy\_extension\_descriptor( )** – This structure is defined in 2.6.102 and 2.6.103.

**Green\_extension\_descriptor()** – This structure is defined in 2.6.104 and 2.6.105.

**MPEG-H\_3dAudio\_descriptor()** – This structure is defined in 2.6.106 and 2.6.107.

**MPEG-H\_3dAudio\_config\_descriptor()** – This structure is defined in 2.6.108 and 2.6.109.

**MPEG-H\_3dAudio\_scene\_descriptor()** – This structure is defined in 2.6.110 and 2.6.111.

**MPEG-H\_3dAudio\_text\_label\_descriptor()** – This structure is defined in 2.6.112 and 2.6.113.

**MPEG-H\_3dAudio\_multi-stream\_descriptor()** – This structure is defined in 2.6.114 and 2.6.115.

**MPEG-H\_3dAudio\_drc\_loudness\_descriptor()** – This structure is defined in 2.6.116 and 2.6.117.

**MPEG-H\_3dAudio\_command\_descriptor()** – This structure is defined in 2.6.118.

**Quality\_extension\_descriptor()** – This structure is defined in 2.6.119.

**Virtual\_segmentation\_descriptor()** – This structure is defined in 2.6.120 and 2.6.121.

**timed\_metadata\_extension\_descriptor()** – This structure is defined in clause 8.2.3 of ISO/IEC 23001-13.

**HEVC\_tile\_substream\_descriptor() –** This structure is defined in 2.6.122 and 2.6.123.

**HEVC\_subregion\_descriptor()** – This structure is defined in 2.6.125 and 2.6.126.

**JXS\_video\_descriptor()** – This structure is defined in 2.6.127 and 2.6.128.

**VVC\_timing\_and\_HRD\_descriptor()** – This structure is defined in 2.6.131 and 2.6.132.

**EVC\_timing\_and\_HRD\_descriptor()** – This structure is defined in 2.6.135 and 2.6.136.

**LCEVC\_video\_descriptor() –** This structure is defined in 2.6.137 and 2.6.138.

**LCEVC\_linkage\_descriptor() –** This structure is defined in 2.6.139 and 2.6.140.

**Media\_service\_kind\_descriptor() –** This structure is defined in 2.6.141 and 2.6.142

Table 2-110 – Extension descriptor tag values

|  |  |  |  |
| --- | --- | --- | --- |
| **Extension\_descriptor\_tag** | **TS** | **PS** | **Identification** |
| 0 | n/a | n/a | Reserved |
| 1 | n/a | X | Forbidden |
| 2 | X | X | ODUpdate\_descriptor |
| 3 | X | n/a | HEVC\_timing\_and\_HRD\_descriptor() |
| 4 | X | n/a | af\_extensions\_descriptor() |
| 5 | X | n/a | HEVC\_operation\_point\_descriptor( ) |
| 6 | X | n/a | HEVC\_hierarchy\_extension\_descriptor( ) |
| 7 | X | n/a | Green\_extension\_descriptor() |
| 8 | X | n/a | MPEG-H\_3dAudio\_descriptor() |
| 9 | X | n/a | MPEG-H\_3dAudio\_config\_descriptor() |
| 0x0A | X | n/a | MPEG-H\_3dAudio\_scene\_descriptor() |
| 0x0B | X | n/a | MPEG-H\_3dAudio\_text\_label\_descriptor() |
| 0x0C | X | n/a | MPEG-H\_3dAudio\_multi-stream\_descriptor() |
| 0x0D | X | n/a | MPEG-H\_3dAudio\_drc\_loudness\_descriptor() |
| 0x0E | X | n/a | MPEG-H\_3dAudio\_command\_descriptor() |
| 0x0F | X | n/a | Quality\_extension\_descriptor() |
| 0x10 | X | n/a | Virtual\_segmentation\_descriptor() |
| 0x11 | X | n/a | timed\_metadata\_extension\_descriptor() |
| 0x12 | X | n/a | HEVC\_tile\_substream\_descriptor() |
| 0x13 | X | n/a | HEVC\_subregion\_descriptor() |
| 0x14 | X | n/a | JXS\_video\_descriptor() |
| 0x15 | X | n/a | VVC\_timing\_and\_HRD\_descriptor() |
| 0x16 | X | n/a | EVC\_timing\_and\_HRD\_descriptor() |
| 0x17 | X | X | LCEVC\_video\_descriptor() |
| 0x18 | X | X | LCEVC\_linkage\_descriptor() |
| 0x19 | X | X | Media\_service\_kind\_descriptor() |
| 0x1A .. 0xFF | n/a | n/a | Rec. ITU-T H.222.0 | ISO/IEC 13818-1 Reserved |

### 2.6.92 ODUpdate\_descriptor

The ODUpdate\_descriptor may be used to carry a set of ObjectDescriptors through an ObjectDescriptorUpdate, as a replacement or as a complement to ISO/IEC 14496 object descriptor streams defined in the IOD. If used, the ObjectDescriptorUpdate command shall be processed by the MPEG-4 terminal as defined in 7.2.5.5.2 of ISO/IEC 14496-1. The descriptors carried in the ODUpdate\_descriptor are in the same name scope as the scene description described in the InitialObjectDescriptor carried in the IOD descriptor.

When an ODUpdate\_descriptor is used within a transport stream, the ODUpdate\_descriptor shall be conveyed in the descriptor loop immediately following the program\_info\_length field in the program map table, and shall be included after an IOD descriptor.

When an ODUpdate\_descriptor is used within a program stream, the ODUpdate\_descriptor shall be conveyed in the descriptor loop immediately following the program\_stream\_info\_length field in the program stream map, and shall be included after an IOD descriptor.

If an ODUpdate\_descriptor is included before an IOD descriptor or if IOD descriptor is not present, then the ODUpdate descriptor shall be ignored. More than one ODUpdate\_descriptor may be included in a program map table or program stream map.

### 2.6.93 Transport\_profile\_descriptor

The Transport\_profile\_descriptor (see Table 2-111) may be associated in the PMT to signal a profile value of transport stream in the associated program. When present, the descriptor shall only be located in the loop following the program\_info\_length field in the PMT. If the descriptor is not included in the PMT, then the associated transport stream conforms to the complete profile.

Table 2-111 – Transport\_profile\_descriptor syntax

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| Transport\_profile\_descriptor{ |  |  |
| **descriptor\_tag** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| **transport\_profile** | **8** | **uimsbf** |
| for (i=0; i<N; i++) { |  |  |
| **private\_data** | **8** | **bslbf** |
| } |  |  |
| } |  |  |

### 2.6.94 Semantic definition of fields in the Transport\_profile\_descriptor

**transport\_profile** – This 8-bit profile value signals the use of constraints in the associated transport stream for the program. See Table 2-112.

Table 2-112 – Transport\_profile values

|  |  |
| --- | --- |
| Values | Description |
| 0x00 | unspecified |
| 0x01 | Complete profile a |
| 0x02 | Adaptive profile b |
| 0x03 .. 0x0E | reserved |
| 0x0F .. 0xFF | user\_private c |
| a  Transport streams using this profile conform to all the normative definitions for transport streams. These include conformant discontinuities, PCR jitter/accuracy, strict T-STD management, PCR interval conformance (less than 100 ms), as well as PTS/DTS interval (0.7 seconds) and compliance.  b  Transport streams using this profile conform to all the normative definitions for transport streams with the following exceptions:  – The PCR jitter may exceed the specified tolerance as applications that use this profile usually do not include null-PID packets. Clients that process these streams usually do not use the PCR to derive the decoder STC. However, the PCR value can be used in conjunction with the PTS and DTS for conformant STD management of all the media components in the associated program;  – the PCR interval occasionally exceeds 100 ms in applications that use this profile due to occasional bit rate variations in certain locations;  – conforming continuity counter errors and time base discontinuity may occur more frequently than in complete profile.  c  User private values of transport\_profile that need unique identification can use the MPEG registration\_descriptor with a unique format\_identifier value that is obtained from the Registration Authority. | |

### 2.6.95 HEVC video descriptor

For an HEVC video stream, the HEVC video descriptor (see Table 2-113) provides basic information for identifying coding parameters, such as profile and level parameters of that HEVC video stream. For an HEVC temporal video sub-bitstream or an HEVC temporal video subset, the HEVC video descriptor provides information such as the associated HEVC highest temporal sub-layer representation contained in the elementary stream to which it applies. This descriptor can also be used to indicate presence of WCG and HDR video components in the associated PID as well as additional parameters to assist decoders with HDR capability to render intended video data on HDR capable display devices. In addition, this can assist non-HDR capable decoders to use the information appropriately.

NOTE 1 – In case that the video characteristics change over time, care should be taken that the descriptor is updated accordingly.

This descriptor, when present, shall only be used for elementary streams with a stream\_type value of 0x24, 0x25 or 0x31. If more than one program element with a stream\_type value of 0x31 is carried in a program, an HEVC\_video\_descriptor may only be used for the first of these program elements. In this case, the information in the HEVC\_video\_descriptor applies to the combination of all HEVC tile substreams with a HEVC tile substream descriptor indicating SubstreamID values in the range of 0 up to TotalSubstreamIDs, as indicated in the HEVC subregion descriptor. When the program element for which the HEVC video descriptor is used is part of an HEVC layered video stream, i.e., the program contains at least one other program element with a stream\_type value in the range of 0x28 .. 0x2B, the semantics of HEVC\_still\_present\_flag, HEVC\_24hr\_picture\_present\_flag and sub\_pic\_hrd\_params\_not\_present\_flag shall apply to the whole HEVC layered video stream, i.e., also to all program elements with a stream\_type value in the range of 0x28 .. 0x2B.

NOTE 2 – For elementary streams with a stream\_type value in the range of 0x28 .. 0x2B, the applicable value of level\_idc can be ambiguous and depend on the output layer set, i.e., the combination with other elementary streams. This information is signalled by the HEVC operation point descriptor.

Table 2-113 – HEVC video descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| HEVC\_video\_descriptor() {  **descriptor\_tag**  **descriptor\_length**  **profile\_space**  **tier\_flag**  **profile\_idc**  **profile\_compatibility\_indication**  **progressive\_source\_flag**  **interlaced\_source\_flag**  **non\_packed\_constraint\_flag**  **frame\_only\_constraint\_flag**  **copied\_44bits**  **level\_idc**  **temporal\_layer\_subset\_flag**  **HEVC\_still\_present\_flag**  **HEVC\_24hr\_picture\_present\_flag**  **sub\_pic\_hrd\_params\_not\_present\_flag**  **reserved**  **HDR\_WCG\_idc**  if ( temporal\_layer\_subset\_flag == '1') {  **temporal\_id\_min**  **reserved**  **temporal\_id\_max**  **reserved**  }  } | **8**  **8**  **2**  **1**  **5**  **32**  **1**  **1**  **1**  **1**  **44**  **8**  **1**  **1**  **1**  **1**  **2**  **2**  **3**  **5**  **3**  **5** | **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf** |

### 2.6.96 Semantic definition of fields in HEVC video descriptor

**profile\_space, tier\_flag, profile\_idc, profile\_compatibility\_indication**, **progressive\_source\_flag, interlaced\_source\_flag, non\_packed\_constraint\_flag, frame\_only\_constraint\_flag, level\_idc** – When the HEVC video descriptor applies to an HEVC video stream or to an HEVC complete temporal representation, these fields shall be coded according to the semantics defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2 for *general\_profile\_space*, *general\_tier\_flag*, *general\_profile\_idc*, *general\_profile\_compatibility\_flag[i]*, *general\_progressive\_source\_flag*, *general\_interlaced\_‌source\_flag*, *general\_non\_packed\_constraint\_flag*, *general\_frame\_only\_constraint\_flag*, *general\_level\_idc*, respectively, for the corresponding HEVC video stream or HEVC complete temporal representation, and the entire HEVC video stream or HEVC complete temporal representation to which the HEVC video descriptor is associated shall conform to the information signalled by these fields.

When the HEVC video descriptor applies to an HEVC temporal video sub-bitstream or HEVC temporal video subset of which the corresponding HEVC highest temporal sub-layer representation is not an HEVC complete temporal representation, these fields shall be coded according to the semantics defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2 for *sub\_layer\_profile\_space*, *sub\_layer\_tier\_flag*, *sub\_layer\_profile\_idc*, *sub\_layer\_profile\_compatibility\_flag[i]*, *sub\_‌layer\_progressive\_source\_flag*, *sub\_layer\_interlaced\_source\_flag*, *sub\_layer\_non\_packed\_constraint\_flag*, *sub\_layer\_‌frame\_only\_constraint\_flag*, *sub\_layer\_level\_idc*, respectively, for the corresponding HEVC highest temporal sub-layer representation, and the entire HEVC highest temporal sub-layer representation to which the HEVC video descriptor is associated shall conform to the information signalled by these fields.

The 32 flags general\_profile\_compatibility\_flag[i] or sub\_layer\_profile\_compatibility\_flag[i], with index i ranging from 0 to 31, are arranged in the bit string profile\_compatibility\_indication with index i equal to 0 at the leftmost position and index i increasing from left to right in the bit string.

NOTE 1 – In one or more sequences in the HEVC video stream the level may be lower than the level signalled in the HEVC video descriptor, while also a profile may occur that is a subset of the profile signalled in the HEVC video descriptor. However, in the entire HEVC video stream, only subsets of the entire bitstream syntax shall be used that are included in the profile signalled in the HEVC video descriptor, if present. If the sequence parameter sets in an HEVC video stream signal different profiles, and no additional constraints are signalled, then the stream may need examination to determine which profile, if any, the entire stream conforms to. If an HEVC video descriptor is to be associated with an HEVC video stream that does not conform to a single profile, then the HEVC video stream should be partitioned into two or more sub-streams, so that HEVC video descriptors can signal a single profile for each such sub-stream.

**copied\_44bits** – When the HEVC video descriptor applies to an HEVC video stream or to an HEVC complete temporal representation, this bit field shall be coded according to the semantics of the syntax elements defined in Rec. ITU‑T H.265 | ISO/IEC 23008‑2 for the 44 bits found in the profile\_tier\_level() syntax element between *general\_frame\_only\_constraint\_flag* and *general\_level\_idc* for the corresponding HEVC video stream or HEVC complete temporal representation, and the entire HEVC video stream or HEVC complete temporal representation to which the HEVC video descriptor is associated shall conform to the information signalled by these fields.

When the HEVC video descriptor applies to an HEVC temporal video sub-bitstream or HEVC temporal video subset of which the corresponding HEVC highest temporal sub-layer representation is not an HEVC complete temporal representation, this bit field shall be coded according to the semantics of the syntax elements defined in Rec. ITU‑T H.265 | ISO/IEC 23008‑2 for the 44 bits found in the profile\_tier\_level() syntax element between *sub\_layer\_frame\_only\_constraint\_flag* and *sub\_layer\_level\_idc* for the corresponding HEVC highest temporal sub-layer representation, and the entire HEVC highest temporal sub-layer representation to which the HEVC video descriptor is associated shall conform to the information signalled by these fields.

**temporal\_layer\_subset\_flag** – This 1-bit flag, when set to '1', indicates that the syntax elements describing a subset of temporal layers are included in this descriptor. This field shall be set to 1 for HEVC temporal video subsets and for HEVC temporal video sub-bitstreams. When set to '0', the syntax elements temporal\_id\_min and temporal\_id\_max are not included in this descriptor.

**HEVC\_still\_present\_flag** – This 1-bit field, when set to '1', indicates that the HEVC video stream or the HEVC highest temporal sub-layer representation may include HEVC still pictures. When the HEVC\_still\_present\_flag is set to '0', the associated HEVC video stream shall not contain HEVC still pictures.

When the program element to which this descriptor applies is part of an HEVC layered video stream and the HEVC\_still\_present\_flag is set to '0', the whole HEVC layered video stream shall not contain HEVC still pictures.

NOTE 2 – According to Rec. ITU-T H.265 | ISO/IEC 23008‑2, IDR pictures are always associated with a TemporalId value equal to 0. Consequently, if the HEVC video descriptor applies to an HEVC temporal video subset, HEVC still pictures can only be present in the associated HEVC temporal video sub-bitstream.

**HEVC\_24\_hour\_picture\_present\_flag** – This 1-bit flag, when set to '1', indicates that the associated HEVC video stream or the HEVC highest temporal sub-layer representation may contain HEVC 24-hour pictures. For the definition of an HEVC 24-hour picture, see 2.1.49. When the HEVC\_24\_hour\_picture\_present\_flag is set to '0', the associated HEVC video stream shall not contain any HEVC 24-hour pictures.

When the program element to which this descriptor applies is part of an HEVC layered video stream and HEVC\_24\_hour\_picture\_present\_flag is set to '0', the whole HEVC layered video stream shall not contain any HEVC 24-hour pictures.

**sub\_pic\_hrd\_params\_not\_present\_flag** – This 1-bit field, when set to '0', indicates that the VUI in the HEVC video stream shall have the syntax element sub\_pic\_hrd\_params\_present\_flag set to '1'. When the sub\_pic\_hrd\_params\_not\_present\_flag is equal to '1', the associated HEVC video stream may not contain sub\_pic\_hrd\_params\_present\_flag in the VUI or the sub\_pic\_hrd\_params\_present\_flag may be set to '0'.

When the program element to which this descriptor applies is part of an HEVC layered video stream and sub\_pic\_hrd\_params\_not\_present\_flag is set to '0', the following apply:

* The HEVC timing and HRD descriptor shall be present in the program map table associated with the program.

NOTE 3 – If sub\_picture\_hrd\_params\_not\_present equals '0', HRD parameters can be expected to be present, though the hrd\_management\_valid\_flag is not mandated to be set to '1' in this case.

* The HRD parameter structures that are applicable for all program elements with stream\_type value of 0x24, 0x25, or in the range of 0x28 .. 0x2B, inclusively, shall be present in the HEVC video stream and the value of sub\_pic\_hrd\_params\_present\_flag in those HRD parameter structures shall be set to '1'.

**HDR\_WCG\_idc –** The value of this syntax element indicates the presence or absence of high dynamic range (HDR) and/or wide color gamut (WCG) video components in the associated PID according to Table 2-114. HDR is defined to be video that has high dynamic range if the video stream EOTF is higher than the Rec. ITU-R BT.1886 reference EOTF. This field also shall not be set to 2 unless bit\_depth\_luma\_minus8 as defined in Rec. ITU-T H.265 | ISO/IEC 23008-2 in the associated video is greater than or equal to 2. WCG is defined to be video that is coded using colour primaries with a colour gamut not contained within Rec. ITU-R BT.709-6. This field also shall not be set to 1 or 2 unless bit\_depth\_chroma\_minus8 as defined in Rec. ITU-T H.265 | ISO/IEC 23008-2 in the associated video is greater than or equal to 2.

Table 2-114 – Semantics of HDR\_WGC\_idc

| HDR\_WCG\_idc | Description |
| --- | --- |
| 0 | SDR (i.e., SDR video) is based on the Rec. ITU-R BT.709 OETF using BT.709 color primaries with a corresponding reference EOTF for flat panel displays as specified in BT.1886 (see Note 1) |
| 1 | WCG only, i.e., video color gamut in a Rec ITU-R BT.2020 container that exceeds Rec. ITU-R BT.709-6 (see Note 2) |
| 2 | Both HDR and WCG are to be indicated in the stream (see Note 3) |
| 3 | No indication made regarding HDR/WCG or SDR characteristics of the stream |
| NOTE 1 – An example where it would be desirable to set HDR\_WCG\_idc to 0 would be when the colour\_description\_present\_flag, as defined in Rec. ITU-T H.265 | ISO/IEC 23008-2, is set to '0', with colour\_primaries and transfer\_characteristics not present in the video stream.  NOTE 2 – An example where it would be desirable to set HDR\_WCG\_idc to 1 would be when colour\_primaries as defined in Rec. ITU-T H.265 | ISO/IEC 23008-2 is equal to 9 to indicate Rec. ITU-R BT.2020.  NOTE 3 – An example where it would be desirable to set HDR\_WCG\_idc to 2 would be when transfer\_characteristics as defined in Rec. ITU-T H.265 | ISO/IEC 23008-2 is equal to 16 to indicate BT.2100 PQ EOTF or equal to 18 to indicate BT.2100 HLG EOTF, and when colour\_primaries as defined in Rec. ITU-T H.265 | ISO/IEC 23008-2 is equal to 9 to indicate Rec. ITU-R BT.2020. | |

**temporal\_id\_min** – This 3-bit field indicates the minimum value of the *TemporalId*, as defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2, of all HEVC access units in the associated elementary stream.

**temporal\_id\_max** – This 3-bit field indicates the maximum value of the *TemporalId*, as defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2, of all HEVC access units in the associated elementary stream.

### 2.6.97 HEVC timing and HRD descriptor

For an HEVC video stream, an HEVC temporal video sub-bitstream or an HEVC temporal video subset, the HEVC timing and HRD descriptor (see Table 2-115) provides timing and HRD parameters, as defined in Annex C of Rec. ITU-T H.265 | ISO/IEC 23008-2, for the associated HEVC video stream or the HEVC highest temporal sub-layer representation thereof, respectively.

Table 2-115 – HEVC timing and HRD descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| HEVC\_timing\_and\_HRD\_descriptor() {  **hrd\_management\_valid\_flag**  **target\_schedule\_idx\_not\_present\_flag**  **target\_schedule\_idx**  **picture\_and\_timing\_info\_present\_flag**  if (picture\_and\_timing\_info\_present\_flag == '1') {  **90kHz\_flag**  **reserved**  if (90kHz\_flag = = '0') {  **N**  **K**  **}**  **num\_units\_in\_tick**  **}**  **}** | **1**  **1**  **5**  **1**  **1**  **7**  **32**  **32**  **32** | **bslbf**  **bslbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf** |

### 2.6.98 Semantic definition of fields in HEVC timing and HRD descriptor

**hrd\_management\_valid\_flag** – This 1-bit flag is only defined for use in transport streams. When the HEVC timing and HRD descriptor is associated with an HEVC video stream or with an HEVC highest temporal sub-layer representation carried in a transport stream, then the following rules apply.

When the value of hrd\_management\_valid\_flag is equal to '1', Buffering Period SEI and Picture Timing SEI messages, as defined in Annex C of Rec. ITU-T H.265 | ISO/IEC 23008‑2, shall be present in the associated HEVC video stream or HEVC highest temporal sub-layer representation. For HEVC layered video streams, each HEVC operation point signalled in the HEVC operation point descriptor shall have applicable Buffering Period SEI and Picture Timing SEI messages. All Buffering Period SEI messages shall carry coded nal\_initial\_cpb\_removal\_delay and nal\_initial\_cpb\_removal\_offset values and may additionally carry nal\_initial\_alt\_removal\_delay and nal\_initial\_alt\_cpb\_removal\_offset values for the NAL HRD. If the hrd\_management\_valid\_flag is set to '1', then the transfer of each byte from MBn to EBn in the T-STD as defined in 2.17.2 or the transfer from MBn,k to EBn in the T‑STD as defined in 2.17.3 or the transfer of each byte from MBl,k to EBl in the T-STD as defined in 2.17.4 shall be according to the delivery schedule for that byte into the CPB in the NAL HRD, as determined from the coded nal\_initial\_cpb\_removal\_delay and nal\_initial\_cpb\_removal\_offset or from the coded nal\_initial\_alt\_cpb\_removal\_delay and nal\_initial\_alt\_cpb\_removal\_offset values for SchedSelIdx equal to target\_schedule\_idx, as specified in Annex C of Rec. ITU-T H.265 | ISO/IEC 23008‑2. When the hrd\_management\_valid\_flag is set to '0', the leak method shall be used for the transfer from MBn to EBn in the T-STD as defined in 2.17.2 or the transfer from MBn,k to EBn in the T-STD as defined in 2.17.3 or the transfer from MBl,k to EBl in the T-STD as defined in 2.17.4.

**target\_schedule\_idx\_not\_present\_flag** – This 1-bit flag when set to '0' indicates that the following 5 bits represent the value target\_schedule\_idx as specified below. When set to '1', the following 5 bits are unspecified. When hrd\_management\_valid\_flag is equal to 0, then target\_schedule\_idx\_not\_present\_flag shall be set to '1'.

**target\_schedule\_idx** – When target\_schedule\_idx\_not\_present\_flag is equal to '0', this 5-bit field indicates the index of the delivery schedule which is assigned for SchedSelIdx. When the value of target\_schedule\_idx\_not\_present\_flag is equal to '1' and the value of hrd\_management\_valid\_flag is equal to '1', the value of target\_schedule\_idx is inferred to be equal to '0'.

**picture\_and\_timing\_info\_present\_flag** – This 1-bit flag when set to '1' indicates that the *90kHz\_flag* and parameters for accurate mapping to a 90-kHz system clock are included in this descriptor.

**90kHz\_flag** – This 1-bit flag when set to '1' indicates that the frequency of the HEVC time base is 90 kHz.

**N, K** – For an HEVC video stream or HEVC highest temporal sub-layer representation, the frequency of the HEVC time base is defined by the syntax element *vui\_time\_scale* in the VUI parameters, as defined in Annex E of Rec. ITU‑T H.265 | ISO/IEC 23008‑2. The relationship between the HEVC *time\_scale* and the STC shall be defined by the parameters N and K in this descriptor as follows.

*time\_scale* = (*N* x *system\_clock\_frequency*) / *K*

If the *90kHz\_flag* is set to '1', then N equals 1 and K equals 300. If the *90kHz\_flag* is set to '0', then the values of N and K are provided by the coded values of the N and K fields.

NOTE – This allows mapping of time expressed in units of *time\_scale* to 90 kHz units, as needed for the calculation of PTS and DTS timestamps, for example in decoders for HEVC access units for which no PTS or DTS is encoded in the PES header.

**num\_units\_in\_tick**– This 32-bit field is coded exactly in the same way as the *vui\_num\_units\_in\_tick* field in VUI parameters in Annex E of Rec. ITU-T H.265 | ISO/IEC 23008‑2. The information provided by this field shall apply to the entire HEVC video stream or HEVC highest temporal sub-layer representation to which the HEVC timing and HRD descriptor is associated.

### 2.6.99 AF extensions descriptor

The AF extensions descriptor (see Table 2-116) is used to signal that adaptation field descriptors could be present in the adaptation header of the component, as defined in 2.4.3.5.

NOTE – There may be AF descriptors in an adaptation field of a TS packet even though this descriptor is not set for the component.

Table 2-116 – Adaptation field extension descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| af\_extensions\_descriptor() {  } |  |  |

### 2.6.100 HEVC operation point descriptor

The HEVC operation point descriptor (see Table 2-117) provides a method to indicate profile and level for one or more HEVC operation points. When present, the HEVC operation point descriptor shall be included in the group of data elements which immediately follow the program\_info\_length field in the program\_map section.

NOTE – For some applications, the TS may not contain all operation points described in the HEVC operation point descriptor, or the HEVC operation point descriptor may not describe all operation points available in the TS. However, as far as matching elementary streams are found in the TS, the information provided in the descriptor should describe the operation points correctly.

Table 2-117 – HEVC operation point descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| HEVC\_operation\_point\_descriptor( ) {  **reserved**  **num\_ptl**  for ( i = 0; i < num\_ptl; i++, i++ ) {  **profile\_tier\_level\_info[i]**  }  **operation\_points\_count**  for ( i = 0; i < operation\_points\_count; i++ ) {  **target\_ols[i]**  **ES\_count[i]**  for ( j = 0; j < ES\_count[i]; j++ ) {  **reserved**  **prepend\_dependencies[i][j]**  **ES\_reference[i][j]**  }  **reserved**  **numEsInOp[i]**  for ( k = 0; k < NumESinOP[i]; k++ ) {  **necessary\_layer\_flag[i][k]**  **output\_layer\_flag[i][k]**  **ptl\_ref\_idx[i][k]**  }  **reserved**  **avg\_bit\_rate\_info\_flag[i]**  **max\_bit\_rate\_info\_flag[i]**  **constant\_frame\_rate\_info\_idc[i]**  **applicable\_temporal\_id[i]**  if ( constant\_frame\_rate\_info\_idc[i] > 0 ) {  **reserved**  **frame\_rate\_indicator[i]**  }  if ( avg\_bit\_rate\_info\_flag[i] == '1' ) {  **avg\_bit\_rate[i]**  }  if ( max\_bit\_rate\_info\_flag[i] == '1' ) {  **max\_bit\_rate[i]**  }  }  } | **2**  **6**  **96**  **8**  **8**  **8**  **1**  **1**  **6**  **2**  **6**  **1**  **1**  **6**  **1**  **1**  **1**  **2**  **3**  **4**  **12**  **24**  **24** | **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf**  **bslbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **uimsbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf** |

### 2.6.101 Semantic definition of fields in HEVC operation point descriptor

**num\_ptl** – This 6-bit field specifies the number of profile, tier and level structures signalled in this descriptor.

**profile\_tier\_level\_info[i]** – This 96-bit field shall be coded according to the syntax structure of profile\_tier\_level defined in 7.3.3 of Rec. ITU-T H.265 | ISO/IEC 23008-2 with the value of profilePresentFlag set equal to '1' and maxNumSubLayersMinus1 set equal to 6.

If multiple HEVC operation point descriptors are found for the same program, all profile\_tier\_level\_info[x] elements of all HEVC operation point descriptors for this program are aggregated in their order of occurrence into a common array, which is referenced in this specification as profile\_tier\_level\_array[]. If there is only a single HEVC operation point descriptor, profile\_tier\_level\_array[] contains the elements profile\_tier\_level\_info[x] in the order as found in that single descriptor.

**operation\_points\_count** – This 8-bit field indicates the number of HEVC operation points described by the list included in the following group of data elements.

**target\_ols[i]** – An 8-bit field that specifies the index into the list of output layer sets in the VPS, associated with the i-th HEVC operation point defined in this descriptor.

**ES\_count[i]** – This 8-bit field indicates the number of ES\_reference values included in the following group of data elements. The aggregation of elementary streams, according to the ordered list indicated in the following group of data elements, forms an HEVC operation point. The value 0xff is reserved.

Let OperationPointESList[i] be the list of elementary streams that are part of the i-th HEVC operation point.

**prepend\_dependencies[i][j]** – This flag if set to '1' specifies that the elementary stream indicated by ES\_reference[i][j], when not present yet in OperationPointESList[i], shall be added into OperationPointESList[i] and the elementary stream indicated by the syntax element hierarchy\_embedded\_layer\_index in the hierarchy descriptor, or all of the elementary streams indicated by the syntax element hierarchy\_ext\_embedded\_layer\_index in the HEVC hierarchy extension descriptor, with the hierarchy layer index value specified by the following syntax element ES\_reference[i][j], when not present yet in OperationPointESList[i], shall be added into OperationPointLayerList[i] immediately before the elementary stream signalled by the ES\_reference[i][j] in ascending order of the value of their associated hierarchy\_embedded\_layer\_index or hierarchy\_ext\_embedded\_layer\_index. When the value of prepend\_dependencies[i][j] is equal to '0', only the elementary stream indicated by ES\_reference[i][j], when not present yet in OperationPointESList[i], shall be added into OperationPointESList[i]. The elementary stream indicated by ES\_reference[i][m] shall be placed earlier (i.e., with a lower index) into OperationPointESList[i] than the elementary stream indicated with ES\_reference[i][n] when m is less than n. The order of elementary stream in the OperationPointESList[i] shall be in ascending order of their hierarchy\_layer\_index values.

**ES\_reference[i][j]** – This 6-bit field indicates the hierarchy layer index value present in the hierarchy descriptor or HEVC hierarchy extension descriptor which identifies an elementary stream. The value of ES\_reference[i][m] and ES\_reference[i][n] for m not equal to n shall not be the same.

**numEsInOp[i]** – This 6-bit field indicates the number of elementary streams in OperationPointESList[i] after all the ESs that are part of the i-th HEVC operation point have been included into OperationPointESList[i] (i.e., after parsing prepend\_dependencies[i][ES\_count[i] – 1]).

**necessary\_layer\_flag[i][k]** – This flag when set to '1' indicates that the k-th elementary stream in OperationPointESList[i] is a necessary layer, as defined in Annex F of Rec. ITU-T H.265 | ISO/IEC 23008-2, of the i-th operation point. This flag equal to '0' indicates that the k-th elementary stream in OperationPointESList[i] is not a necessary layer, as defined in Annex F of Rec. ITU-T H.265 | ISO/IEC 23008-2, of the i-th operation point.

**output\_layer\_flag[i][k]** – This flag when set to '1' indicates that the k-th elementary stream in OperationPointESList[i] is an output layer. Otherwise, when set to '0', it indicates that the k-th elementary stream in OperationPointESList[i] is not an output layer. When the value of necessary\_layer\_flag[i][k] is equal to '0', the value of output\_layer\_flag[i][k] shall be ignored.

**ptl\_ref\_idx[i][k]** – A 6-bit field that indicates the index x to the profile\_tier\_level\_info[x] element of the profile\_tier\_level\_array which applies to the k-th elementary stream in OperationPointESList[i]. When the value of necessary\_layer\_flag[i][k] is equal to '0', the value of ptl\_ref\_idx[i][k] shall be ignored.

**avg\_bit\_rate\_info\_flag[i]** – This flag indicates whether the syntax element avg\_bit\_rate[i] is present in this descriptor.

**max\_bit\_rate\_info\_flag[i]** – This flag indicates whether the syntax element max\_bit\_rate[i] is present in this descriptor.

**constant\_frame\_rate\_info\_idc[i]** – This 2-bit field, in combination with the syntax element frame\_rate\_indicator as specified below, indicates how the frame rate for the associated operation point j is determined. The value of 0 indicates that the frame rate is not specified for the i-th HEVC operation point and that the syntax element frame\_rate\_indicator is not present in this descriptor for the i-th HEVC operation point.

**applicable\_temporal\_id[i]** – This 3-bit field indicates the highest value of TemporalId of the VCL NAL units in the re-assembled HEVC video stream for operation point i.

**frame\_rate\_indicator[i]** – If constant\_frame\_rate\_info\_idc[i] is equal to 1, this 12-bit field indicates a constant number of ticks, as specified in the HEVC timing and HRD descriptor, for the distance in time between two pictures at the i-th HEVC operation point. If constant\_frame\_rate\_info\_idc[i] equals 2, this 12-bit field indicates the frame rate for the i-th operation point measured in frames per second. If constant\_frame\_rate\_info\_idc[i] equals 3, this 12-bit field indicates the frame rate for the i-th HEVC operation point measured in frames per 1.001 seconds.

**avg\_bit\_rate[i]** – This 24-bit field indicates the average bit rate, in 1000 bits per second, of the HEVC layered video stream corresponding to the i-th HEVC operation point.

**max\_bit\_rate[i]** – This 24-bit field indicates the maximum bit rate, in 1000 bits per second, of the HEVC layered video stream corresponding to the i-th HEVC operation point.

### 2.6.102 HEVC hierarchy extension descriptor

The HEVC hierarchy extension descriptor provides information to identify the program elements containing components of layered HEVC streams (see Table 2-118). When present, this descriptor shall only be used for elementary streams with the stream\_type value 0x28, 0x29, 0x2A or 0x2B.

Table 2-118 – HEVC hierarchy extension descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| HEVC\_hierarchy\_extension\_descriptor( ) { |  |  |
| **extension\_dimension\_bits** | **16** | **bslbf** |
| **hierarchy\_layer\_index** | **6** | **uimsbf** |
| **temporal\_id** | **3** | **uimsbf** |
| **nuh\_layer\_id** | **6** | **uimsbf** |
| **tref\_present\_flag** | **1** | **bslbf** |
| **reserved** | **2** | **bslbf** |
| **num\_embedded\_layers** | **6** | **uimsbf** |
| **reserved** | **2** | **bslbf** |
| **hierarchy\_channel** | **6** | **uimsbf** |
| for ( i = 0 ; i < num\_embedded\_layers ; i++ ) { |  |  |
| **reserved** | **2** | **bslbf** |
| **hierarchy\_ext\_embedded\_layer\_index[i]** | **6** | **uimsbf** |
| **}** |  |  |
| **}** |  |  |

### 2.6.103 Semantic definition of fields in HEVC hierarchy extension descriptor

When the HEVC hierarchy extension descriptor is present, it is used to specify the dependency of the associated elementary stream to other elementary streams in the same program.

**extension\_dimension\_bits** – A 16-bit field indicating the possible enhancement of the associated program element from the base layer resulting from the program element of the layer with nuh\_layer\_id equal to '0'.

The allocation of the bits to enhancement dimensions is given in Table 2-119.

Table 2-119 – Semantics of extension dimension bits

| Index to bits | Description |
| --- | --- |
| 0 | Multi-view enhancement |
| 1 | Spatial scalability, including SNR quality or fidelity enhancement |
| 2 | Depth enhancement |
| 3 | Temporal enhancement |
| 4 | Auxiliary enhancement |
| 5 .. 15 | Reserved |

The i-th bit equal to '1' indicates that the corresponding enhancement dimension is present. When the elementary stream contains auxiliary pictures as defined in Annex F of Rec. ITU-T H.265 | ISO/IEC 23008-2, the value of the 4th bit of extension\_dimension\_bits shall be set equal to '1', otherwise, it shall be set equal to '0'. When the elementary stream contains auxiliary pictures that are depth pictures, as defined in Annex F of Rec. ITU-T H.265 | ISO/IEC 23008-2, the value of both the 2nd and the 4th bits of extension\_dimension\_bits shall be set equal to '1'.

**hierarchy\_layer\_index** – A 6-bit field that defines a unique index of the associated program elements in a table of coding layer hierarchies. Indices shall be unique within a single program definition. For video sub-bitstreams of HEVC video streams conforming to one or more profiles defined in Annex F of Rec. ITU‑T H.265 | ISO/IEC 23008-2, this is the program element index, which is assigned in a way that the bitstream order will be correct if the associated dependency layers of the video sub-bitstreams of the same HEVC access unit are re-assembled in increasing order of hierarchy\_layer\_index.

**temporal\_id** – A 3-bit field that specifies the highest TemporalId of the NAL units in the elementary stream associated with this HEVC hierarchy extension descriptor.

**nuh\_layer\_id** – A 6-bit field that specifies the highest nuh\_layer\_id of the NAL units in the elementary stream associated with this HEVC hierarchy extension descriptor.

**tref\_present\_flag** – A 1-bit flag, which when set to '0' indicates that the TREF field may be present in the PES packet headers in the associated elementary stream. The value of '1' for this flag is reserved.

**num\_embedded\_layers** –A 6-bit field that specifies the number of direct dependent program elements that need to be accessed and be present in decoding order before decoding of the elementary stream associated with this HEVC hierarchy extension descriptor.

**hierarchy\_channel –** A 6-bit field that indicates the intended channel number for the associated program element in an ordered set of transmission channels. The most robust transmission channel is defined by the lowest value of this field with respect to the overall transmission hierarchy definition.

NOTE – A given hierarchy\_channel may at the same time be assigned to several program elements.

**hierarchy\_ext\_embedded\_layer\_index[i]** –A 6-bit field that defines the hierarchy\_layer\_index of the program element that needs to be accessed and be present in decoding order before decoding of the elementary stream associated with this HEVC hierarchy extension descriptor.

### 2.6.104 Green extension descriptor

The syntax of the green extension descriptor containing static metadata is shown in Table 2-120.

Table 2-120 – Green extension descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| Green\_extension\_descriptor() { |  |  |
| **num\_constant\_backlight\_voltage\_time\_intervals** | **2** | **uimsbf** |
| **reserved** | **6** | **bslbf** |
| for (i=0; i < num\_constant\_backlight\_voltage\_time\_intervals; i++) { |  |  |
| **constant\_backlight\_voltage\_time\_interval[i]** | **16** | **uimsbf** |
| } |  |  |
| **num\_max\_variations** | **2** | **uimsbf** |
| **reserved** | **6** | **bslbf** |
| for (j=0; j < num\_max\_variations; j++) { |  |  |
| **max\_variation[j]** | **16** | **uimsbf** |
| } |  |  |
| } |  |  |

### 2.6.105 Semantics for green extension descriptor

Semantics for all the syntax elements above are specified in 6.4 of ISO/IEC 23001-11.

### 2.6.106 MPEG-H 3D audio descriptor

The MPEG-H 3D audio descriptor (see Table 2-121) provides information on basic coding information in the associated ISO/IEC 23008-3 stream. This descriptor shall be present in the associated PMT for MPEG-H 3D audio content with stream\_type equal to 0x2D.

Table 2-121 – MPEG-H 3D audio descriptor

|  |  |  |  |
| --- | --- | --- | --- |
| Syntax | No. of bits | Mnemonic | |
| MPEG-H\_3dAudio\_descriptor() { |  |  | |
| **mpegh3daProfileLevelIndication** | **8** | **uimsbf** | |
| **interactivityEnabled** | **1** | **bslbf** | |
| **compatibleProfileSetsPresent** | **1** | **bslbf** |
| **reserved** | **8** | **bslbf** | |
| **referenceChannelLayout** | **6** | **uimsbf** | |
| if (compatibleProfileSetsPresent == '0') { |  |  | |
| **numCompatibleSets** | **8** | **uimsbf** | |
| for ( n = 0; n < numCompatibleSets; n++ ) { |  |  | |
| **CompatibleSetIndication** | **8** | **uimsbf** | |
| } |  |  | |
| } |  |  | |
| for (i=0; i<N; i++) { |  |  | |
| **reserved** | **8** | **bslbf** | |
| } |  |  | |
| } |  |  | |

### 2.6.107 Semantics for MPEG-H 3D audio descriptor

**mpegh3daProfileLevelIndication** – The audio profile and level of the associated ISO/IEC 23008-3 audio stream, encoded as specified for the mpegh3daProfileLevelIndication field in 5.3.2 in ISO/IEC 23008-3.

**referenceChannelLayout** – Reference channel configuration value as defined as "ChannelConfiguration" in ISO/IEC 23001-8 ("Codec Independent Code Points").

**interactivityEnabled** – If set to '1', this flag indicates that the 3D audio stream contains elements with associated metadata which enables user interactivity. If this flag is set to '0', no user interactivity of any kind is available. This flag may be used to determine the need for initializing the user interactivity interface in the Systems decoder.

**compatibleProfileSetsPresent** – A 1-bit flag, which when set to '0' indicates the presence of compatible profile sets.

**numCompatibleSets** – This field signals the number of compatible profile sets present in the current descriptor. The value of this field shall be equal to or higher than 1.

**CompatibleSetIndication** – See CompatibleSetIndication field in the CompatibleProfileLevelSet() config extension in ISO/IEC 23008-3.

### 2.6.108 MPEG-H 3D audio config descriptor

The MPEG-H 3D audio config descriptor (see Table 2-122) provides information on the complete configuration data of one ISO/IEC 23008-3 stream.

Table 2-122 – MPEG-H 3D audio config descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| MPEG-H\_3dAudio\_config\_descriptor() { |  |  |
| mpegh3daConfig() |  |  |
| } |  |  |

### 2.6.109 Semantics for MPEG-H 3D audio config descriptor

**mpegh3daConfig()** – The mpegh3daConfig() of the associated ISO/IEC 23008-3 audio stream, as specified in 5.2.2.1 in ISO/IEC 23008-3.

### 2.6.110 MPEG-H 3D audio scene descriptor

The MPEG-H 3D audio scene descriptor (see Table 2-123) provides information on user selectable and/or modifiable audio objects in an ISO/IEC 23008-3 stream.

| Table 2-123 – MPEG-H 3D audio scene descriptor | | |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| MPEG-H\_3dAudio\_scene\_descriptor() { |  |  |
| **groupDefinitionPresent** | **1** | **bslbf** |
| **switchGroupDefinitionPresent** | **1** | **bslbf** |
| **groupPresetDefinitionPresent** | **1** | **bslbf** |
| **reserved** | **5** | **bslbf** |
|  |  |  |
| **3dAudioSceneInfoID** | **8** | **bslbf** |
|  |  |  |
| if (groupDefinitionPresent) { |  |  |
| **reserved** | **1** | **bslbf** |
| **numGroups** | **7** | **uimsbf** |
| for ( i=0; i < numGroups; i++) { |  |  |
| **reserved** | **1** | **bslbf** |
| **mae\_groupID** | **7** | **uimsbf** |
| **reserved** | **3** | **bslbf** |
| **mae\_allowOnOff** | **1** | **bslbf** |
| **mae\_defaultOnOff** | **1** | **bslbf** |
| **mae\_allowPositionInteractivity** | **1** | **bslbf** |
| **mae\_allowGainInteractivity** | **1** | **bslbf** |
| **mae\_hasContentLanguage** | **1** | **bslbf** |
| **reserved** | **4** | **bslbf** |
| **mae\_contentKind** | **4** | **uimsbf** |
| if ( mae\_allowPositionInteractivity ) { |  |  |
| **reserved** | **1** | **bslbf** |
| **mae\_interactivityMinAzOffset** | **7** | **uimsbf** |
| **reserved** | **1** | **bslbf** |
| **mae\_interactivityMaxAzOffset** | **7** | **uimsbf** |
| **reserved** | **3** | **bslbf** |
| **mae\_interactivityMinElOffset** | **5** | **uimsbf** |
| **reserved** | **3** | **bslbf** |
| **mae\_interactivityMaxElOffset** | **5** | **uimsbf** |
| **mae\_interactivityMinDistOffset** | **4** | **uimsbf** |
| **mae\_interactivityMaxDistOffset** | **4** | **uimsbf** |
| } |  |  |
| if ( mae\_allowGainInteractivity ) { |  |  |
| **reserved** | **2** | **bslbf** |
| **mae\_interactivityMinGain** | **6** | **uimsbf** |
| **reserved** | **3** | **bslbf** |
| **mae\_interactivityMaxGain** | **5** | **uimsbf** |
| } |  |  |
| if ( mae\_hasContentLanguage ) { |  |  |
| **mae\_contentLanguage** | **24** | **uimsbf** |
| } |  |  |
| } |  |  |
| } |  |  |
|  |  |  |
| if (switchGroupDefinitionPresent) { |  |  |
| **reserved** | **3** | **bslbf** |
| **numSwitchGroups** | **5** | **uimsbf** |
| for ( i=0; i < numSwitchGroups; i++) { |  |  |
| **reserved** | **1** | **bslbf** |
| **mae\_switchGroupID** | **5** | **uimsbf** |
| **mae\_switchGroupAllowOnOff** | **1** | **bslbf** |
| **mae\_switchGroupDefaultOnOff** | **1** | **bslbf** |
| **reserved** | **3** | **bslbf** |
| **mae\_bsSwitchGroupNumMembers** | **5** | **uimsbf** |
| for ( i = 0; i < mae\_bsSwitchGroupNumMembers + 1; i++ ) { |  |  |
| **reserved** | **1** | **bslbf** |
| **mae\_switchGroupMemberID** | **7** | **uimsbf** |
| } |  |  |
| **reserved** | **1** | **bslbf** |
| **mae\_switchGroupDefaultGroupID** | **7** | **uimsbf** |
| } |  |  |
| } |  |  |
|  |  |  |
| if (presetGroupDefinitionPresent) { |  |  |
| **reserved** | **3** | **bslbf** |
| **mae\_numGroupPresets** | **5** | **uimsbf** |
| for ( i = 0; i < mae\_numGroupPresets; i++ ) { |  |  |
| **reserved** | **3** | **bslbf** |
| **mae\_groupPresetID** | **5** | **uimsbf** |
| **reserved** | **3** | **bslbf** |
| **mae\_groupPresetKind** | **5** | **uimsbf** |
| **reserved** | **4** | **bslbf** |
| **mae\_numGroupPresetConditions** | **4** | **uimsbf** |
| for ( j = 0; j < mae\_numGroupPresetConditions+1; j++ ) { |  |  |
| **mae\_groupPresetGroupID** | **7** | **uimsbf** |
| **mae\_groupPresetConditionOnOff** | **1** | **bslbf** |
| if (mae\_groupPresetConditionOnOff ) { |  |  |
| **reserved** | **4** | **bslbf** |
| **mae\_groupPresetDisableGainInteractivity** | **1** | **bslbf** |
| **mae\_groupPresetGainFlag** | **1** | **bslbf** |
| **mae\_groupPresetDisablePositionInteractivity** | **1** | **bslbf** |
| **mae\_groupPresetPositionFlag** | **1** | **bslbf** |
| if ( mae\_groupPresetGainFlag ) { |  |  |
| **mae\_groupPresetGain** | **8** | **uimsbf** |
| } |  |  |
| if( mae\_groupPresetPositionFlag ){ |  |  |
| **mae\_groupPresetAzOffset** | **8** | **uimsbf** |
| **reserved** | **2** | **bslbf** |
| **mae\_groupPresetElOffset** | **6** | **uimsbf** |
| **reserved** | **4** | **bslbf** |
| **mae\_groupPresetDistFactor** | **4** | **uimsbf** |
| } |  |  |
| } |  |  |
| } |  |  |
| } |  |  |
| } |  |  |
|  |  |  |
| for (i=0; i<N; i++) { |  |  |
| **reserved** | **8** | **bslbf** |
| } |  |  |
| } |  |  |

### 2.6.111 Semantic definition of fields in MPEG-H 3D audio scene descriptor

**groupDefinitionPresent** – A one-bit flag signalling the presence of interactivity information of one group in this descriptor.

**groupContentDataPresent** – A one-bit flag signalling the presence of content information of one group in this descriptor.

**switchGroupDefinitionPresent** – A one-bit flag signalling the presence of switch group information in this descriptor.

**presetGroupDefinitionPresent** – A one-bit flag signalling the presence of preset group information in this descriptor.

**3dAudioSceneInfoID** – See 'mae\_audioSceneInfoID' in 15.3 of ISO/IEC 23008-3.

**numGroups** – This field signals the number of groups in the audio scene description. This field can take values between 1 and 127, and shall be less or equal to the value of mae\_numGroups present in the associated ISO/IEC 23008‑3 stream.

**mae\_groupID** – See 15.3 of ISO/IEC 23008-3.

**mae\_allowOnOff** – See 15.3 of ISO/IEC 23008-3.

**mae\_defaultOnOff** – See 15.3 of ISO/IEC 23008-3.

**mae\_allowPositionInteractivity** – See 15.3 of ISO/IEC 23008-3.

**mae\_allowGainInteractivity** – See 15.3 of ISO/IEC 23008-3.

**mae\_hasContentLanguage** – See 15.3 of ISO/IEC 23008-3.

**mae\_contentKind** – See 15.3 of ISO/IEC 23008-3.

**mae\_interactivityMinAzOffset** – See 15.3 of ISO/IEC 23008-3.

**mae\_interactivityMaxAzOffset** – See 15.3 of ISO/IEC 23008-3.

**mae\_interactivityMinElOffset** – See 15.3 of ISO/IEC 23008-3.

**mae\_interactivityMaxElOffset** – See 15.3 of ISO/IEC 23008-3.

**mae\_interactivityMinDistOffset** – See 15.3 of ISO/IEC 23008-3.

**mae\_interactivityMaxDistOffset** – See 15.3 of ISO/IEC 23008-3.

**mae\_interactivityMinGain** – See 15.3 of ISO/IEC 23008-3.

**mae\_interactivityMaxGain** – See 15.3 of ISO/IEC 23008-3.

**mae\_contentLanguage** – See 15.3 of ISO/IEC 23008-3.

**numSwitchGroups** – This field signals the number of switch groups minus one in the overall scene. This field can take values between 0 and 31, resulting in a maximum number of 32 switch groups. It shall be less or equal to the value of mae\_numSwitchGroups present in the associated ISO/IEC 23008-3 stream.

**mae\_switchGroupID** – See ISO/IEC 23008-3.

**mae\_switchGroupAllowOnOff** – See 15.3 of ISO/IEC 23008-3.

**mae\_switchGroupDefaultOnOff** – See 15.3 of ISO/IEC 23008-3; if mae\_switchGroupAllowOnOff is '0', then mae\_switchGroupDefaultOnOff shall be set to '0'.

**mae\_bsSwitchGroupNumMembers** – See 15.3 of ISO/IEC 23008-3.

**mae\_switchGroupMemberID** – See 15.3 of ISO/IEC 23008-3.

**mae\_switchGroupDefaultGroupID** – See 15.3 of ISO/IEC 23008-3.

**mae\_numGroupPresets** – See 15.3 of ISO/IEC 23008-3.

**mae\_groupPresetID** – See 15.3 of ISO/IEC 23008-3.

**mae\_groupPresetKind** – See 15.3 of ISO/IEC 23008-3.

**mae\_groupPresetNumConditions** – See 15.3 of ISO/IEC 23008-3.

**mae\_groupPresetGroupID** – See 15.3 of ISO/IEC 23008-3.

**mae\_groupPresetConditionOnOff** – See 15.3 of ISO/IEC 23008-3.

**mae\_groupPresetDisableGainInteractivity** – See 15.3 of ISO/IEC 23008-3.

**mae\_groupPresetGainFlag** – See 15.3 of ISO/IEC 23008-3.

**mae\_groupPresetDisablePositionInteractivity** – See 15.3 of ISO/IEC 23008-3.

**mae\_groupPresetPositionFlag** – See 15.3 of ISO/IEC 23008-3.

**mae\_groupPresetGain** – See 15.3 of ISO/IEC 23008-3.

**mae\_groupPresetAzOffset** – See 15.3 of ISO/IEC 23008-3.

**mae\_groupPresetElOffset** – See 15.3 of ISO/IEC 23008-3.

**mae\_groupPresetDistFactor** – See 15.3 of ISO/IEC 23008-3.

Data fields provided both in this descriptor and as in-band information in the ISO/IEC 23008-3 stream shall be set to the same value.

### 2.6.112 MPEG-H 3D audio text label descriptor

The MPEG-H 3D audio scene descriptor provides text labels for the audio objects and presets in an ISO/IEC 23008-3 stream. See Table 2-124.

| Table 2-124 – MPEG-H 3D audio text label descriptor | | |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| MPEG-H\_3dAudio\_text\_label\_descriptor() { |  |  |
| **3dAudioSceneInfoID** | **8** | **uimsbf** |
| **reserved** | **4** | **bslbf** |
| **numDescLanguages** | **4** | **uimsbf** |
| for (i=0; i< numDescLanguage; i++) { |  |  |
| **descriptionLanguage** | **24** | **uimsbf** |
| **reserved** | **1** | **bslbf** |
| **numGroupDescriptions** | **7** | **uimsbf** |
| for ( n = 0; n < numGroupDescriptions; n++ ) { |  |  |
| **reserved** | **1** | **bslbf** |
| **mae\_descriptionGroupID;** | **7** | **uimsbf** |
| **groupDescriptionDataLength** | **8** | **uimsbf** |
| for ( c = 0; c < groupDescriptionDataLength; c++ ) { |  |  |
| **groupDescriptionData** | **8** | **uimsbf** |
| } |  |  |
| } |  |  |
| **reserved** | **3** | **bslbf** |
| **numSwitchGroupDescriptions** | **5** | **uimsbf** |
| for ( n = 0; n < numSwitchGroupDescriptions; n++ ) { |  |  |
| **reserved** | **3** | **bslbf** |
| **mae\_descriptionSwitchGroupID;** | **5** | **uimsbf** |
| **switchGroupDescriptionDataLength** | **8** | **uimsbf** |
| for ( c = 0; c < switchGroupDescriptionDataLength; c++ ) { |  |  |
| **switchGroupDescriptionData** | **8** | **uimsbf** |
| } |  |  |
| } |  |  |
| **reserved** | **3** | **bslbf** |
| **numGroupPresetsDescriptions** | **5** | **uimsbf** |
| for ( n = 0; n < numGroupPresetsDescriptions; n++ ) { |  |  |
| **reserved** | **3** | **bslbf** |
| **mae\_descriptionGroupPresetID** | **5** | **uimsbf** |
| **groupPresetDescriptionDataLength** | **8** | **uimsbf** |
| for ( c = 0; c < groupPresetDescriptionLength; c++ ) { |  |  |
| **groupPresetDescriptionData** | **8** | **uimsbf** |
| } |  |  |
| } |  |  |
| } |  |  |
| for (i=0; i<N; i++) { |  |  |
| **reserved** | **8** | **bslbf** |
| } |  |  |
| } |  |  |

### 2.6.113 Semantic definition of fields in MPEG-H 3D audio text label descriptor

**3dAudioSceneInfoID** – See 'mae\_audioSceneInfoID' in 15.3 ISO/IEC 23008-3.

**numDescLanguages** – The number of available languages for description text.

**descriptionLanguage** – Identifies the language or languages used by the description text of a metadata element group. It contains a 3-character code as specified by ISO 639-2. Both ISO 639-2/B and ISO 639-2/T may be used. Each character is coded into 8 bits according to ISO/IEC 8859-1 and inserted in order into the 24-bit field.

**numGroupDescriptions** – The number of available descriptions for groups.

**mae\_descriptionGroupID** – See 15.3 of ISO/IEC 23008-3.

**groupDescriptionDataLength** – The length, specified in bytes, of the following group description.

**groupDescriptionData** – This field contains a description of a metadata element group, i.e., a string describing the content by a high-level description. The format shall follow UTF-8 according to ISO/IEC 10646.

**numSwitchGroupDescriptions** – The number of available descriptions for switch groups.

**mae\_descriptionSwitchGroupID** – See 15.3 of ISO/IEC 23008-3.

**switchGroupDescriptionDataLength** – The length, specified in bytes, of the following switch group description.

**switchGroupDescriptionData** – This field contains a description of a switch group, i.e., a string describing the content by a high-level description. The format shall follow UTF-8 according to ISO/IEC 10646.

**numGroupPresetsDescriptions** – The number of available descriptions for group presets.

**mae\_descriptionGroupPresetID** – See 15.3 of ISO/IEC 23008-3.

**groupPresetDescriptionDataLength** – The length, specified in bytes, of the following group preset description.

**groupPresetDescriptionData** – This field contains a description of a metadata element group, i.e., a string describing the content by a high-level description. The format shall follow UTF-8 according to ISO/IEC 10646.

Data fields provided both in this descriptor and as in-band information in the IS/IEC 23008-3 stream shall be set to the same value.

### 2.6.114 MPEG-H 3D audio multi-stream descriptor

The MPEG-H 3D audio multi-stream descriptor (see Table 2-125) provides information on the location of each mae\_groupID in case of transmission over multiple streams.

In combination with this descriptor, an MPEG-H 3D audio scene descriptor explaining the actual representation of the overall audio scene is required to be present in the descriptor loop of the main stream.

| Table 2-125 – MPEG-H 3D audio multi-stream descriptor | | |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| MPEG-H\_3dAudio\_multi-stream\_descriptor() { |  |  |
| **thisIsMainStream** | **1** | **bslbf** |
| **thisStreamID** | **7** | **uimsbf** |
|  |  |  |
| if ( thisIsMainStream ) { |  |  |
| **reserved** | **1** | **slbf** |
| **numAuxiliaryStreams** | **7** | **uimsbf** |
| **reserved** | **1** | **bslbf** |
| **mae\_numGroups** | **7** | **uimsbf** |
| for (i=0; i< mae\_numGroups; i++) { |  |  |
| **mae\_groupID** | **7** | **uimsbf** |
| **isInMainStream** | **1** | **bslbf** |
| if ( thisIsMainStream == '0' ) { |  |  |
| **isInTS** | **1** | **bslbf** |
| **auxiliaryStreamID** | **7** | **uimsbf** |
| } |  |  |
| } |  |  |
| } |  |  |
|  |  |  |
| for (i=0; i<N; i++) { |  |  |
| **reserved** | **8** | **bslbf** |
| } |  |  |
| } |  |  |

### 2.6.115 Semantic definition of fields in MPEG-H 3D audio multi-stream descriptor

**thisIsMainStream** – If this flag is '1', the stream is a main stream, otherwise it is an auxiliary stream.

**thisStreamID** – This integer provides a unique ID of all available ISO/IEC 23008-3 Audio streams with MHAS transport syntax, both main and auxiliary streams (stream\_type 0x2D and 0x2E).

**numAuxiliaryStreams** – This integer provides information on how many auxiliary stream are available.

**mae\_numGroups** – This field signals the number of groups in the overall audio scene (complete number of groups in the main stream plus all possible additional streams). This field can take values between 1 and 127. It shall be set to the same value as the corresponding field in the associated ISO/IEC 23008-3 stream.

**mae\_groupID** – This integer provides information on the mae\_groupID (as described in ISO/IEC 23008-3, section 15) the loop instance refers to.

**isInMainStream** – If this flag is set to '1', the audio data related to the group (as indicated through mae\_groupID) is present in the main stream, otherwise the encoded data is transmitted in an auxiliary stream.

**isInTS** – If this flag is set to '1', the audio data related to the group (as indicated through mae\_groupID) is present in the same transport stream. If this flag is set to 0, the data must be retrieved from an external source.

**auxiliaryStreamID** – In case of transmission of encoded audio data as identified by groupID in an auxiliary stream, this integer identifies the used auxiliary stream.

The location of the 'external source' may be signalled using the TEMI location descriptor and/or the TEMI BaseURL\_descriptor as defined in section U.3.

When the TEMI descriptor(s) are conveyed in a TEMI\_AU as a separate elementary stream on a separated PID, an MPEG-H\_3dAudio\_extStreamID\_descriptor() shall be present in the associated descriptor loop of the TEMI elementary stream which provides an ID in the field "auxiliaryStreamID" that is matching the "auxiliaryStreamID" provided in the MPEG-H\_3dAudio\_multi-stream\_descriptor() of the main stream.

### 2.6.116 MPEG-H 3D audio DRC and Loudness descriptor

The MPEG-H 3D audio dynamic range control (DRC) and Loudness descriptor (see Table 2-126) provides information on DRC and Loudness information contained in an ISO/IEC 23008-3 stream.

| Table 2-126 – MPEG-H 3D audio DRC and Loudness descriptor() | | |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| MPEG-H\_3dAudio\_drc\_loudness\_descriptor () {  **reserved**  **mpegh3daDrcAndLoudnessInfoPresent** | **7**  **1** | **bslbf**  **bslbf** |
| if (mpegh3daDrcAndLoudnessInfoPresent) {  **reserved**  **drcInstructionsUniDrcCount**  **reserved**  **loudnessInfoCount**  **reserved**  **downmixIdCount** | **2**  **6**  **2**  **6**  **3**  **5** | **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf** |
| for (i=0; i<drcInstructionsUniDrcCount; i++) {  **reserved**  **drcInstructionsType**  if (drcInstructionsType == 2) {  **reserved**  **mae\_groupID**  } else if (drcInstructionsType == 3) {  **reserved**  **mae\_groupPresetID**  } | **6**  **2**  **1**  **7**  **3**  **5** | **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf** |
| **reserved**  **drcSetId**  **reserved**  **downmixId**  **reserved**  **additionalDownmixIdCount**  **limiterPeakTargetPresent**  **drcSetTargetLoudnessPresent** | **2**  **6**  **1**  **7**  **3**  **3**  **1**  **1** | **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf**  **bslbf** |
| for (j=0; j<additionalDownmixIdCount; j++) {  **reserved**  **additionalDownmixId**  } | **1**  **7** | **bslbf**  **uimsbf** |
| **drcSetEffect**  if ( limiterPeakTargetPresent ) {  **bsLimiterPeakTarget**  } | **16**  **8** | **uimsbf**  **uimsbf** |
| if ( drcSetTargetLoudnessPresent ) {  **reserved**  **bsDrcSetTargetLoudnessValueUpper**  **drcSetTargetLoudnessValueLowerPresent**  if(drcSetTargetLoudnessValueLowerPresent) {  **reserved**  **bsDrcSetTargetLoudnessValueLower**  } | **1**  **6**  **1**  **2**  **6** | **bslbf**  **uimsbf**  **bslbf**  **bslbf**  **uimsbf** |
| }  **reserved**  **dependsOnDrcSet**  if (dependsOnDrcSet == 0) {  **noIndependentUse**  } else {  **reserved**  }  } | **1**  **6**  **1**  **1** | **bslbf**  **uimsbf**  **bslbf**  **bslbf** |
| for (i=0; i<loudnessInfoCount; i++) {  **reserved**  **loudnessInfoType**  if (loudnessInfoType == 1 || loudnessInfoType == 2) {  **reserved**  **mae\_groupID**  } else if (loudnessInfoType == 3) {  **reserved**  **mae\_groupPresetID**  }  **loudnessInfo\_size**  loudnessInfo()  } | **6**  **2**  **1**  **7**  **3**  **5**  **8** | **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **uimsbf** |
| for (i=0; i<downmixIdCount; i++) {  **reserved**  **downmixId**  **downmixType**  **CICPspeakerLayoutIdx**  }  } | **1**  **7**  **2**  **6** | **bslbf**  **uimsbf**  **uimsbf**  **uimsbf** |
| for (i=0; i<N; i++) {  **reserved**  }  } | **8** | **bslbf** |

### 2.6.117 Semantic definition of fields in MPEG-H 3D audio DRC and Loudness descriptor

**mpegh3daDrcAndLoudnessInfoPresent** – A one-bit flag signalling the presence of dynamic range control and loudness information in this descriptor.

**drcInstructionsUniDrcCount** – This field signals the number of DRC sets in the stream. This field can take values between 0 and 63, resulting in a maximum number of 63 DRC sets.

**loudnessInfoCount** – This field signals the number of loudness info blocks in the stream. This field can take values between 0 and 63, resulting in a maximum number of 63 loudness info blocks.

**downmixIdCount** – This field signals the number of downmixId definitions in the stream. This field can take values between 0 and 31, resulting in a maximum number of 31 downmixId definitions.

**drcInstructionsType** – See ISO/IEC 23008-3.

**mae\_groupID** – See ISO/IEC 23008-3.

**mae\_groupPresetID** – See ISO/IEC 23008-3.

**drcSetId** – See ISO/IEC 23003-4.

**downmixId** – See ISO/IEC 23003-4 and ISO/IEC 23008-3.

**additionalDownmixIdCount** – See ISO/IEC 23003-4.

**limiterPeakTargetPresent** – See ISO/IEC 23003-4.

**drcSetTargetLoudnessPresent** – See ISO/IEC 23003-4.

**additionalDownmixId** – See ISO/IEC 23003-4.

**drcSetEffect** – See ISO/IEC 23003-4.

**bsLimiterPeakTarget** – See ISO/IEC 23003-4.

**bsDrcSetTargetLoudnessValueUpper** – See ISO/IEC 23003-4.

**drcSetTargetLoudnessValueLowerPresent** – See ISO/IEC 23003-4.

**bsDrcSetTargetLoudnessValueLower** – See ISO/IEC 23003-4.

**dependsOnDrcSet** – See ISO/IEC 23003-4.

**noIndependentUse** – See ISO/IEC 23003-4.

**loudnessInfoType** – See ISO/IEC 23008-3.

**loudnessInfo\_size** – The number of bytes of the immediately following loudnessInfo().

**loudnessInfo()** – One loudnessInfo() structure as defined in ISO/IEC 23003-4.

**downmixType** – See ISO/IEC 23008-3.

**CICPspeakerLayoutIdx** – See ISO/IEC 23008-3.

Data fields provided both in this descriptor and as in-band information in the IS/IEC 23008-3 stream shall be set to the same value.

### 2.6.118 MPEG-H 3D audio command descriptor

The MPEG-H 3D audio command descriptor (see Table 2-127) encapsulates an MHAS packet of the type PACTYP\_USERINTERACTION that contains an interaction command or type PACTYP\_AUDIOSCENEINFO that contains the audio scene information. Both types are supported by a MPEG-H 3D audio decoder.

By inserting the MHAS packet contained in this descriptor into the MHAS elementary stream, a receiver can send this command to the MPEG-H 3D audio decoder. As the descriptor contains a complete MHAS packet including MHAS header, nothing else needs to be added when multiplexing these bytes, consecutively, on an MHAS packet boundary in the MHAS stream.

Table 2-127 – MPEG-H 3D audio command descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| MPEG-H\_3dAudio\_command\_descriptor() { |  |  |
| for (i = 0; i < N; i++) { |  |  |
| **data** | **8** | **bslbf** |
| } |  |  |
| } |  |  |

**data –** Data bytes shall be contiguous bytes of data from a complete MHAS packet (including the MHAS header) of the type PACTYP\_USERINTERACTION.

### 2.6.119 Quality extension descriptor

The quality extension descriptor shall be sent once per event or program and hence is signalled using a descriptor in the program map table. This descriptor shall appear in the elementary stream loop of the PID for which quality information is provided.

Dynamic quality metadata is stored in access units and is associated with one or more video frames. These access units are encapsulated in MPEG sections identified by stream\_type value of 0x2F.

The quality extension descriptor describes metrics that are present in each Quality Access Unit, and the constant field size that is used for the values. The quality metrics are defined in 4.3 of ISO/IEC 23001-10.

The syntax of quality extension descriptor containing static metadata is shown in Table 2-128.

Table 2-128 – Quality extension descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| Quality\_extension\_descriptor() { |  |  |
| **field\_size\_bytes** | **8** | **uimsbf** |
| **metric\_count** | **8** | **uimsbf** |
| for (i=0; i < metric\_count; i++) { |  |  |
| **metric\_code[i]** | **32** | **uimsbf** |
| } |  |  |
| } |  |  |

**field\_size\_bytes** – The constant size in byte of the value for a metric in each Quality Access Unit.

**metric\_count** – The number of metrics for quality values in each Quality Access Unit

**metric\_code** – The code name of the metrics in the Quality Access Unit.

Semantics for all the syntax elements above are specified in 4.2 of ISO/IEC 23001-10. The quality metrics to be signalled can also be found in 4.3 of ISO/IEC 23001-10 and include:

• PSNR = Peak Signal to Noise Ratio

• SSIM = Structural Similarity Index

• MS-SSIM = Multi-Scale Structural Similarity Index

• VQM = Video Quality Metric

• PEVQ = Perceptual Evaluation of Video Quality

• MOS = Mean Opinion Score

• FSIG = Frame significance

### 2.6.120 Virtual segmentation descriptor

The virtual segmentation descriptor (see Table 2-129) appears in the elementary stream descriptor loop in the PMT and is used to indicate that the current elementary stream is virtually segmented using boundary descriptors (see U.3.11). This segmentation may come in a set of partitions – e.g., one partition demarcates the stream into 10-s virtual segments, while another creates 2-s virtual segments. If the boundary descriptor carried in transport stream packets appears in the elemental stream declared by the PMT stream description carrying the virtual segmentation descriptor, it is an explicit indication of segment boundary point, otherwise a reference PID shall be defined in the virtual segmentation descriptor; that reference PID indicates that segment boundary points are indicated in another elemental stream, for which the boundary descriptor shall be present in the PMT. The virtual segmentation for elementary streams for which no virtual segmentation descriptor is present in the PMT is undefined.

Table 2-129 – Virtual segmentation descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| Virtual\_segmentation\_descriptor(){ |  |  |
| if (descriptor\_length > 1) {  **num\_partitions**  **timescale\_flag**  **reserved**  if ( timescale\_flag == 1 ) {  **ticks\_per\_second**  **maximum\_duration\_length\_minus\_1 (MDL)**  **reserved**  }  for ( i = 0; i < num\_partitions; i++ ) {  **explicit\_boundary\_flag**  **partition\_id**  **reserved**  **SAP\_type\_max**  if (explicit\_boundary\_flag == 0 ) {  **reserved**  **boundary\_PID**  **reserved**  }  else {  **maximum\_duration**  }  }  }  } | **3**  **1**  **4**  **21**  **2**  **1**  **1**  **3**  **4**  **3**  **5**  **13**  **3**  **MDL\*8 + 5** | **uimsbf**  **bslbf**  **bslbf**  **uismbf**  **uismbf**  **bslbf**  **bslbf**  **uimsbf**  **bslbf**  **uismbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf** |

### 2.6.121 Semantic definition of fields in virtual segmentation descriptor

**timescale\_flag**: If set to '1', timescale information is present. If set to '0', ticks\_per\_second is inferred to be 1, and MDL is inferred to be 0 (i.e., maximum\_duration\_length\_minus\_1=−1). The value of '0' allows maximum segment duration of up to 31 seconds, expressed in integer seconds.

**ticks\_per\_second**: Precision, in ticks per second, of the maximum\_duration field, e.g., 0.1 second precision is 10 ticks/s, 0.01 second precision with 100 ticks/s, etc.

**maximum\_duration\_length\_minus\_1**: Length, in bytes (minus one), of the maximum\_duration field variable byte length. This provides additional bytes in addition to the 5 bits pre-allocated to the maximum\_duration field.

**num\_partitions**: Number of partitions described in the virtual segmentation descriptor.

**explicit\_boundary\_flag**: If set to '0', this elementary stream is a dependent stream, and boundary data for it is provided on a reference partition on a different PID, specified by boundary\_PID; otherwise, the current PID carries boundary descriptors.

**partition\_id**: ID of the partition described in the boundary descriptor.

**boundary\_PID**: PID carrying boundary\_descriptor() that is used by this partition of this elementary stream.

**SAP\_type\_max**: Maximum possible value of SAP in this partition. If SAP\_type\_max value is 0, any SAP value may appear in the stream.

**maximum\_duration**: Maximum virtual segment duration for a segment on partition partition\_id, expressed in units of ticks\_per\_second. For consecutive virtual segments S(i) and S(i+1) on the above partition, if PTS(i) stands for the earliest PTS in segment S(i), and PTS(i+1) stands for the earliest PTS in segment S(i+1) this duration equals (PTS(i+1) – PTS(i))\*ticks\_per\_second/90000. If set to 0, virtual segment duration is unlimited.

### 2.6.122  HEVC tile substream descriptor

For each ES containing an HEVC tile substream (i.e., a tile or fixed set of subsequent tiles), an HEVC tile substream descriptor (see Table 2-130) assigns a SubstreamID to that HEVC tile substream. It optionally contains additional SubstreamIDs needed to form a subregion or an index to a pattern that indicates these additional SubstreamIDs by an array of offsets found in the HEVC subregion descriptor. Type and length of the descriptor are provided by header bytes that are not shown in the following syntax table.

Table 2-130 – HEVC tile substream descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| HEVC\_tile\_substream\_descriptor() {  **ReferenceFlag**  **SubstreamID**  if ( descriptor\_length > 1 ) {  if ( ReferenceFlag == '1' ) {  **PreambleFlag**  **PatternReference**  }  else {  for ( i=0; i<N1; i++ ) {  **Flag[i]**  **AdditionalSubstreamID[i]**  }  }  }  } | **1**  **7**  **1**  **7**  **1**  **7** | **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf** |
| Note: The value N1 is implied by the descriptor size. It equals the value of descriptor\_length minus 1. | | |

The HEVC tile substream descriptor can be used in three different ways, each signalling the SubstreamID:

1. If its size is only one byte (preceding header bytes excluded), ReferenceFlag is reserved. It implicitly signals a value of 0 for the PatternReference (referring to SubstreamOffset[k][0][i] in the HEVC subregion descriptor as specified below)
2. If ReferenceFlag is set to '1', it specifies the index of the pattern to be used to calculate additional SubstreamIDs (other than index 0)
3. If ReferenceFlag is set to '0', it specifies the additional SubstreamIDs directly

### 2.6.123   Semantic definition of the fields in the HEVC tile substream descriptor

**ReferenceFlag** – When this bit is set to '1', this descriptor indicates the index j > 0 of the pattern signalled by the HEVC subregion descriptor to be used to calculate additional SubstreamIDs. If the value of the field descriptor\_length specifying the size of this descriptor excluding preceding header bytes is 1, the value of ReferenceFlag is reserved.

**SubstreamID –** This 7-bit field assigns a number in the range of 1 to TotalSubstreamIDs (as found) to the HEVC tile substream, which is unique among all ESs with stream type equal to 0x31 or 0x24 (the latter in case that multiple substreams are carried in a single ES) that belong to the same program. This value is used to identify each HEVC tile substream. The value 0 is used for substreams that contain information applicable to multiple HEVC tile substreams.

**PreambleFlag** – When this bit is set to '1' and each HEVC tile substream is carried in its own ES (signalled by the SubstreamMarkingFlag in the HEVC subregion descriptor set to '1'), an access unit carried in the ES signalled by the value of PreambleSubstreamID in the HEVC subregion descriptor is prepended before an access unit carried in this ES as specified in 2.17.5.2. In case SubstreamMarkingFlag in the HEVC subregion descriptor set to '0', the semantics of this flag is reserved.

**PatternReference** – This 7-bit field assigns a number in the range of 1 to PatternCount (as found in the HEVC subregion descriptor) to the HEVC tile substream. This number is used as an index j to the array SubstreamOffset[k][j][i] when the reassembly process is executed.

**Flag[i] –** When Flag[0] is set to '1' and each HEVC tile substream is carried in its own ES (signalled by the SubstreamMarkingFlag in the HEVC subregion descriptor set to '1'), an additional ES signalled by the value of PreambleSubstreamID in the HEVC subregion descriptor is prepended before this ES. In other cases, the semantics of Flag[0] is reserved. Flag[i] for values of i > 0 is reserved.

NOTE – If ReferenceFlag is set to '0', Flag[0] has the same semantics as PreambleFlag for the case that ReferenceFlag is set to '1'.

**AdditionalSubstreamID –** This array of 7 bit fields indicates additional SubstreamIDs that belong to the subregion when the reassembly process according to 2.17.5.2 or 2.17.5.3 is executed.

### 2.6.124 HEVC tile substream af\_descriptor

When multiple HEVC tile substreams are carried in the same ES, HEVC tile substream selection shall be based on signalization using HEVC tile substream af\_descriptors. The syntax of this descriptor is specified in Table U.15.

### 2.6.125 HEVC subregion descriptor

One HEVC subregion descriptor is associated to the whole program. The HEVC subregion descriptor signals patterns of SubstreamIDs that belong to a subregion. Its syntax is given in Table 2-131, the semantics are specified in 2.6.126. It can signal different layouts, which e.g. consist of different numbers of HEVC tile substreams, and indicate the level value of the profile as specified in ISO/IEC 23008‑2 that applies for each pattern. It also indicates how HEVC tile substreams are identified.

NOTE – As the subregion comprises less pixels than the whole panorama, the level value for the subregion is less than or equal to the level value that applies for the whole panorama.

If multiple HEVC tile substreams, e.g. the whole panorama, are contained in a single ES, the motion-constrained tile sets extraction information set SEI message shall be present at all random access points.

Table 2-131 — HEVC subregion descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| HEVC\_subregion\_descriptor() {  **SubstreamMarkingFlag**  **SubstreamIDsPerLine**  **TotalSubstreamIDs**  **LevelFullPanorama**  for ( i=0; i<N1; i++ ) {  if ( SubstreamMarkingFlag == 1 ) {  **reserved**  **PreambleSubstreamID[i]**  }  **SubstreamCountMinus1[i]**  **Level[i]**  **PictureSizeHor[i]**  **PictureSizeVert[i]**  **reserved**  **PatternCount[i]**  for ( j=0; j<PatternCount[i]; j++ ) {  for ( k=0; k<SubstreamCountMinus1[i]; k++ ) {  **SubstreamOffset[k][j][i]**  }  }  }  } | **1**  **7**  **8**  **8**  **1**  **7**  **8**  **8**  **16**  **16**  **1**  **7**  **8** | **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **uimsbf**  **tcimsbf** |
| NOTE: The value N1 is implied by the descriptor size. It equals the number of different subregion layouts, indexed by i, that can be selected from the whole panorama. | | |

### 2.6.126 Semantic definition of the fields in the HEVC subregion descriptor

**SubstreamMarkingFlag** – This 1 bit flag, when set to '1' indicates that each HEVC tile substream is carried in a separate ES. When set to '0', multiple HEVC tile substreams are carried in a common ES and HEVC tile substreams are identified by HEVC tile substream af\_descriptors.

**SubstreamIDsPerLine** – This 7-bit value indicates the number of HEVC tile substreams that are coded representations of tiles that are arranged horizontally and span the width of the whole panorama.

**TotalSubstreamIDs –** This 8-bit value indicates the total number of HEVC tile substreams that represent tiles for the whole panorama. It does not include any additional HEVC tile substreams that provide slice headers and parameter sets.

**LevelFullPanorama –** This 8-bit value indicates the level value of the profile as specified in ISO/IEC 23008‑2 that applies for the whole panorama.

**PreambleSubstreamID[i] –** The value of this 7-bit field indicates the SubstreamID of the ES to be prepended to the ES to which this descriptor applies when the reassembly process is executed. The value can be either 0 or in the range of TotalSubstreamIDs +1 to 127. This field is only available when each HEVC tile substream is carried in its own ES.

**SubstreamCountMinus1[i]** – The value of this 8-bit field in the range of 0 to (TotalSubstreamIDs – 1) indicates

* in case that SubstreamMarkingFlag is set to '1': the number of HEVC tile substreams to be appended to the ES to which this descriptor applies when the reassembly process according to 2.17.5.2 is executed;
* in case that SubstreamMarkingFlag is set to '0': the number of HEVC tile substreams to be added to the list of SubstreamIDs that belong to the subregion when the reassembly process according to 2.17.5.3 is executed;

**Level[i] –** This 8-bit field indicates the value of the profile as specified in ISO/IEC 23008‑2 that applies that applies to the subregion layout indexed by i.

**PictureSizeHor[i] –** This 16-bit field indicates the horizontal subregion dimension, measured in pixels.

**PictureSizeVert[i]** – This 16-bit field indicates the vertical subregion dimension, measured in pixels.

**PatternCount[i]** – This 7-bit field indicates the number of different subregion layouts.

**SubstreamOffset[k][j][i]** – This array of 8 bit fields indicates the offset that has to be added

* in case that SubstreamMarkingFlag is set to '1': to the SubstreamID of the ES to which this descriptor applies. The resulting value indicates the SubstreamID of the ES to be appended when the reassembly process is executed;
* in case that SubstreamMarkingFlag is set to '0': to the SubstreamIDx that relates to the upper left HEVC tile substream of the region to be displayed. The resulting value indicates the SubstreamID to be added to the list of SubstreamIDs that belong to the subregion.

NOTE – Examples are shown in Annex V.

### 2.6.127 JPEG XS video descriptor

For JPEG XS video elementary streams conforming to ISO/IEC 21122-1 and to one or more profiles defined in ISO/IEC 21122-2, the JPEG XS video descriptor (see Table 2-132) provides information that may be present in each JPEG XS access unit as well as for the JPEG XS video sequence. In addition, it provides information to signal JPEG XS still pictures. This descriptor shall be included for each JPEG XS video elementary stream component in the PMT with stream\_type equal to 0x32.

| Table 2-132 – JPEG XS video descriptor | | |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| JXS\_video\_descriptor() {  **descriptor\_version**  **horizontal\_size**  **vertical\_size**  **brat**  **frat**  **schar**  **Ppih**  **Plev**  **max\_buffer\_size**  **buffer\_model\_type**  **colour\_primaries**  **transfer\_characteristics**  **matrix\_coefficients**  **video\_full\_range\_flag**  **reserved**  **still\_mode**  **mdm\_flag**  **zero\_bits**  if (mdm\_flag == '1') {  **X\_c0, Y\_c0, X\_c1, Y\_c1, X\_c2, Y\_c2**  **X\_wp**  **Y\_wp**  **L\_max**  **L\_min**  **MaxCLL**  **MaxFALL**  }  for (i=0; i<N; i++) {  **private\_data\_byte**  }  } | **8**  **16**  **16**  **32**  **32**  **16**  **16**  **16**  **32**  **8**  **8**  **8**  **8**  **1**  **7**  **1**  **1**  **6**  **16x6**  **16**  **16**  **32**  **32**  **16**  **16**  **8** | **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf** |

### 2.6.128 Semantics of fields in JPEG XS video descriptor

**descriptor\_version** – This 8-bit field specifies the version of the JPEG XS video descriptor in case additional capabilities are added in the future. Value of this field shall be '0000 0000' and all other values are reserved for future use.

**horizontal\_size** – This field shall be coded the same as Wf parameter found in the JPEG XS codestream picture header, asdefined in ISO/IEC 21122-1.

**vertical\_size** –This field shall be coded the same as Hf parameter found in the JPEG XS codestream picture header, asdefined in ISO/IEC 21122-1.

**brat** – This parameter specifies the maximum bitrate of the elementary stream in Mbit per second. A detailed definition can be found in ISO/IEC 21122-3.

**frat** – This parameter specifies the frame rate of the elementary stream and also contains a flag indicating if the stream is interlaced or progressive. A detailed definition can be found in ISO/IEC 21122-3.

**schar** – This parameter specifies the image sample characteristics and the sampling structure. A detailed definition can be found in ISO/IEC 21122-3.

**Ppih** – This field specifies the Profile of the elementary stream. It shall be coded the same as the Ppih parameter defined in ISO/IEC 21122-1 using the values defined in ISO/IEC 21122-2.

**Plev** – This field specifies the level and sublevel of the elementary stream. It shall be coded the same as the Plev parameter defined in ISO/IEC 21122-1 using the values defined in ISO/IEC 21122-2.

**max\_buffer\_size** – This 32-bit field specifies the size of the elementary stream buffer (EB in Figure W.2) required to ensure decoding without overflow nor underflow. This value, expressed in Mbytes, shall not exceed (max\_rate/160), where max\_rate corresponds to the maximum encoded rate allowed by the combination of the selected level and sublevel, as defined in ISO/IEC 21122-2, expressed in Mbits/s. When level or sublevel is unrestricted, max\_buffer\_size shall not exceed (brat/160).

**buffer\_model\_type** – This 8-bit field specifies the smoothing buffer model type to which the carried video stream is conforming to. Currently, only value 2 is allowed, other values are reserved for future ISO/IEC use.

NOTE – Streams generated with buffer model type 1 are also compliant with type 2 and can therefore be signalled with this field set to '2'.

**colour\_primaries, transfer\_characteristics, matrix\_coefficients, video\_full\_range\_flag** – These four fields (three 1‑byte integers and one 1-bit flag) shall be coded according to the semantics with the same name defined in ISO/IEC 23091-2.

**still\_mode** – This 1-bit field, when set to '1', indicates that the JPEG XS video stream may include JPEG XS still pictures, as described in W.2. Whenset to '0', then the associated JPEG XS video stream shall not contain JPEG XS still pictures.

**mdm\_flag** – When set to '1', this 1-bit field indicates that the JPEG XS video descriptor contains the characteristics of the Mastering Display Metadata, as described in ISO/IEC 21122-3 (see below corresponding fields). When set to '0', then the JPEG-XS video descriptor shall not contain the characteristics of the Mastering Display Metadata.

**zero\_bits** – These 6 bits shall be all set to '0' and are reserved for future use.

The following fields X\_c0, Y\_c0, X\_c1, Y\_c1, X\_c2, Y\_c2, X\_wp, Y\_wp, L\_max and L\_min correspond to the fields defined in SMPTE ST2086:2014 "Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images". The fields MaxFALL and MaxCLL correspond to the fields defined in ANSI/CTA 861-G:2016 "A DTV Profile for Uncompressed High Speed Digital Interfaces". If these 12 fields have unknown values at the time of generating the stream, they shall not be included in the descriptor and the mdm\_flag shall be set to '0'.

**X\_c0, Y\_c0**, **X\_c1, Y\_c1, X\_c2, Y\_c2** – These 16-bit fields are included only if the mdm\_flag is set to '1'. They specify the normalized x and y chromaticity coordinates of the colour primary components of the mastering display in increments of 0.00002, according to the CIE 1931 definition of x and y as specified in ISO 11664-1 (see also ISO 11664-3 and CIE 15). For describing mastering displays that use red, green, and blue colour primaries, it is suggested that index value c0 should correspond to the green primary, c1 should correspond to the blue primary, and c2 should correspond to the red colour primary. The values of these 6 fields shall be in the range of 0 to 50 000, inclusive.

**X\_wp** and **Y\_wp** – These 16-bit fields are included only if the mdm\_flag is set to '1'. They specify the normalized x and y chromaticity coordinates of the white point of the mastering display in normalized increments of 0.00002, according to the CIE 1931 definition of x and y as specified in ISO 11664-1 (see also ISO 11664-3 and CIE 15). The values of X\_wp and Y\_wpshall be in the range of 0 to 50 000.

**L\_max** and **L\_min** – These 32-bit fields are included only if the mdm\_flag is set to '1'. They specify the nominal maximum and minimum display luminance, respectively, of the mastering display. The minimum luminance of the mastering display is computed by L\_min×0.0001 cd/m2, the maximum luminance by L\_max×0.0001 cd/m2. L\_min shall be less than L\_max. At minimum luminance, the mastering display is considered to have the same nominal chromaticity as the white point.

**MaxCLL** – This 16-bit field is included only if the mdm\_flag is set to '1'. It specifies the Maximum Content Light Level and corresponds to the brightest pixel in the entire stream, in units of 1 cd/m2, where 0x0001 represents 1 cd/m2 and 0xFFFF represents 65535 cd/m2. It shall be calculated according to Annex P Calculation of MaxCLL and MaxFALL section P.1 in ANSI/CTA 861-G:2016 A DTV Profile for Uncompressed High Speed Digital Interfaces.

**MaxFALL** – This 16-bit field is included only if the mdm\_flag is set to '1'. It specifies the Maximum Frame Average Light Level and corresponds to the highest frame average brightness per frame in the entire stream, in units of 1 cd/m2, where 0x0001 represents 1 cd/m2 and 0xFFFF represents 65535 cd/m2. It shall be calculated according to Annex P Calculation of MaxCLL and MaxFALL section P.2 in ANSI/CTA 861-G:2016 "A DTV Profile for Uncompressed High Speed Digital Interfaces".

If for some reason the MaxCLL and/or MaxFALL values are unknown, the value 0x0000 shall be used.

### 2.6.129 VVC video descriptor

For a VVC video stream, the VVC video descriptor (see Table 2-133) provides basic information for identifying coding parameters, such as profile and level parameters, of that VVC video stream. For a VVC temporal video sub-bitstream or a VVC temporal video subset, the VVC video descriptor provides information such as the associated VVC highest temporal sublayer representation contained in the elementary stream to which it applies.

Table 2-133 – VVC video descriptor

|  |  |  |
| --- | --- | --- |
| **Syntax** | **No. of bits** | **Mnemonic** |
| VVC\_video\_descriptor() {  **descriptor\_tag**  **descriptor\_length**  **profile\_idc**  **tier\_flag**  **num\_sub\_profiles**  for ( i=0; i<num\_sub\_profiles; i++) {  **sub\_profile\_idc[ i ]**  }  **progressive\_source\_flag**  **interlaced\_source\_flag**  **non\_packed\_constraint\_flag**  **frame\_only\_constraint\_flag**  **reserved\_zero\_4bits**  **level\_idc**  **temporal\_layer\_subset\_flag**  **VVC\_still\_present\_flag**  **VVC\_24hr\_picture\_present\_flag**  **reserved**  **HDR\_WCG\_idc**  **reserved**  **video\_properties\_tag**  if ( temporal\_layer\_subset\_flag == '1') {  **reserved**  **temporal\_id\_min**  **reserved**  **temporal\_id\_max**  }  } | **8**  **8**  **7**  **1**  **8**  **32**  **1**  **1**  **1**  **1**  **4**  **8**  **1**  **1**  **1**  **5**  **2**  **2**  **4**  **5**  **3**  **5**  **3** | **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf** |

### 2.6.130 Semantic definition of fields in VVC video descriptor

**profile\_idc, tier\_flag, num\_sub\_profiles, sub\_profile\_idc[ i ] –** These fields shall be coded according to the semantics defined in Rec. ITU-T H.266 | ISO/IEC 23090-3 for general\_profile\_idc, general\_tier\_flag, ptl\_num\_sub\_profiles, general\_sub\_profile\_idc[i], respectively, for the corresponding VVC video stream or VVC complete temporal representation. When the corresponding VVC video stream or VVC complete temporal representation contains a DCI, the values of these fields shall be coded exactly the same as the corresponding fields in the DCI.

**progressive\_source\_flag, interlaced\_source\_flag, non\_packed\_constraint\_flag –** These fields shall be coded according to the semantics defined in Rec. ITU-T H.274 | ISO/IEC 23002-7 for vui\_progressive\_source\_flag, vui\_interlaced\_source\_flag, vui\_non\_packed\_constraint\_flag, respectively, for the corresponding VVC video stream or VVC complete temporal representation.

NOTE 1 – Potentially, the corresponding VVC video stream or VVC complete temporal representation contains an SPS with a vui\_payload() syntax structure that indicates the values of these fields. The semantics defined in Rec. ITU-T H.274 | ISO/IEC 23002-7 allows to indicate that the source scan type of the pictures should be interpreted as unknown or unspecified or specified by external means or that the presence of frame packing arrangement SEI messages is not constraint (unknown or unspecified).

**frame\_only\_constraint\_flag –** This field shall be coded according to the semantics defined in Rec. ITU-T H.266 | ISO/IEC 23090-3 for ptl\_frame\_only\_constraint\_flag for the corresponding VVC video stream or VVC complete temporal representation. When the corresponding VVC video stream or VVC complete temporal representation contains a DCI, the values of this field shall be coded the same as the corresponding field in the DCI.

**reserved\_zero\_4bits –** This field shall be set to '0000'.

**level\_idc –** When the VVC video descriptor applies to a VVC video stream or to a VVC complete temporal representation, this field shall be coded according to the semantics defined in Rec. ITU-T H.266 | ISO/IEC 23090-3 for general\_level\_idc, for the corresponding VVC video stream or VVC complete temporal representation. When the VVC video descriptor applies to a VVC temporal video sub-bitstream or VVC temporal video subset of which the corresponding VVC highest temporal sublayer representation is not a VVC complete temporal representation, this field shall be coded according to semantics defined in Rec. ITU-T H.266 | ISO/IEC 23090-3 for sublayer\_level\_idc, for the corresponding VVC highest temporal sublayer representation. When the corresponding VVC video stream or VVC complete temporal representation contains a DCI, the values of this field shall be coded exactly the same as the corresponding fields in the DCI.

When the VVC video descriptor applies to a VVC video stream or to a VVC complete temporal representation, the entire VVC video stream or VVC complete temporal representation to which the VVC video descriptor is associated shall conform to the information signaled by these fields. When the VVC video descriptor applies to a VVC temporal video sub-bitstream or VVC temporal video subset of which the corresponding VVC highest temporal sublayer representation is not a VVC complete temporal representation, the entire VVC highest temporal sublayer representation to which the VVC video descriptor is associated shall conform to the information signalled by these fields.

NOTE 2 – In one or more sequences in the VVC video stream the level may be lower than the level signalled in the VVC video descriptor, while also a profile may occur that is a subset of the profile signalled in the VVC video descriptor. However, in the entire VVC video stream, only subsets of the entire bitstream syntax shall be used that are included in the profile signalled in the VVC video descriptor, if present. When the VVC video stream contains a DCI NAL unit, the profile-tier-level indication signals the profile and level information to which all sequences in the VVC video stream conforms. When not, if the sequence parameter sets in a VVC video stream signal different profiles, and no additional constraints are signalled, then the stream may need examination to determine which profile, if any, the entire stream conforms to.

**temporal\_layer\_subset\_flag** – This 1-bit flag, when set to '1', indicates that the syntax elements describing a subset of temporal layers are included in this descriptor. This field shall be set to 1 for VVC temporal video subsets and for VVC temporal video sub-bitstreams. When set to ‘0’, the syntax elements temporal\_id\_min and temporal\_id\_max are not included in this descriptor.

**VVC\_still\_present\_flag** – This 1-bit field, when set to '1', indicates that the VVC video stream or the VVC highest temporal sublayer representation may include VVC still pictures. When set to '0', then the associated VVC video stream shall not contain VVC still pictures.

NOTE 3 – According to Rec. ITU-T H.266 | ISO/IEC 23090-3, IDR pictures are always associated to a TemporalId value equal to 0, Consequently, if the VVC video descriptor applies to a VVC temporal video subset, VVC still pictures can only be present in the associated VVC temporal video sub-bitstream.

**VVC\_24\_hour\_picture\_present\_flag** – This 1-bit flag, when set to '1', indicates that the associated VVC video stream or the VVC highest temporal sublayer representation may contain VVC 24-hour pictures. For the definition of a VVC 24-hour picture, see 2.1.151. If this flag is set to '0', the associated VVC video stream shall not contain any VVC 24-hour picture.

**HDR\_WCG\_idc** – This 2-bit field indicates the presence or absence of high dynamic range (HDR) and/or wide color gamut (WCG) video components in the associated PID according to Table 2-134. This field shall not be set to 2 unless sps\_bit\_depth\_minus8 as defined in Rec. ITU-T H.266 | ISO/IEC 23090-3 in the associated video is greater than or equal to 2.

Table 2-134 – Semantics of HDR\_WGC\_idc

| **HDR\_WCG\_idc** | **Description** |
| --- | --- |
| 0 | SDR (i.e., SDR video) is based on the Rec. ITU-R BT.709 OETF using BT.709 color primaries with a corresponding reference EOTF for flat panel displays as specified in BT.1886 (see Note 1) |
| 1 | WCG only, i.e., video color gamut in a Rec. ITU-R BT.2020 container that exceeds Rec. ITU-R BT.709-6 (see Note 2) |
| 2 | Both HDR and WCG are to be indicated in the stream (see Note 3) |
| 3 | No indication is made regarding HDR/WCG or SDR characteristics of the stream |
| NOTE 1 – An example where it would be desirable to set HDR\_WCG\_idc to 0 would be when the colour\_description\_present\_flag, as defined in Rec. ITU-T H.274 | ISO/IEC 23002-7, is set to '0', with colour\_primaries and transfer\_characteristics not present in the video stream.  NOTE 2 – An example where it would be desirable to set HDR\_WCG\_idc to 1 would be when colour\_primaries as defined in Rec. ITU-T H.274 | ISO/IEC 23002-7 (which refers to the ColourPrimaries parameter in Rec. ITU-T H.273 | ISO/IEC 23091-2) is equal to 9 to indicate Rec. ITU-R BT.2020.  NOTE 3 – An example where it would be desirable to set HDR\_WCG\_idc to 2 would be when transfer\_characteristics as defined in Rec. ITU-T H.274 | ISO/IEC 23002-7 (which refers to the TransferCharacteristics parameter in Rec. ITU-T H.273 | ISO/IEC 23091-2) is equal to 16 to indicate BT.2100 PQ EOTF or equal to 18 to indicate BT.2100 HLG EOTF, and when colour\_primaries as defined in Rec. ITU-T H.274 | ISO/IEC 23002-7 is equal to 9 to indicate Rec. ITU-R BT.2020. | |

**video\_properties\_tag** – This 4-bit field is used to indicate specific video property CICP combinations for SDR, WCG, or HDR/WCG streams depending on the value of HDR\_WCG\_idc. The CICP values are defined in Rec. ITU-T H.273 | ISO/IEC 23091-2. When HDR\_WCG\_idc is equal to 0, Table 2-135 applies. When HDR\_WCG\_idc is equal to 1, Table 2-136 applies. When HDR\_WCG\_idc is equal to 2, Table 2-137 applies. When HDR\_WCG\_idc is equal to 3, Table 2­138 applies.

NOTE 4 – The video\_properties\_tag descriptor can indicate widely used combinations that are documented by. ITU-T H.Sup 19 | ISO/IEC TR 23091-4. The system identifiers found in that Supplement | Technical Report are mentioned in Tables 2-135 to 2-138 for convenience.

Table 2-135 – SDR widely used video property combinations

|  |  |
| --- | --- |
| **video\_properties\_tag** | **CICP Values - System Identifier**  **[ColourPrimaries, TransferCharacteristics, MatrixCoefficients, VideoFullRangeFlag]** |
| 0 | Video property CICP combination not specified or unknown |
| 1 | [1,1,1,0]- BT709\_YCC |
| 2 | [1,1,0,0]- BT709\_RGB |
| 3 | [6,6,6,0]- BT601\_525 |
| 4 | [5,6,5,0]- BT601\_625 |
| 5 | [1,1,0,1]- FR709\_RGB |
| 6-15 | Reserved |

Table 2-136 – WCG widely used video property combinations

|  |  |
| --- | --- |
| **video\_properties\_tag** | **CICP Values - System Identifier**  **[ColourPrimaries, TransferCharacteristics, MatrixCoefficients, VideoFullRangeFlag]** |
| 0 | Video property CICP combination not specified or unknown |
| 1 | [9,14,9,0]- BT2020\_YCC\_NCL |
| 2 | [9,14,0,0]- BT2020\_RGB |
| 3 | [9,14,0,1]- FR2020\_RGB |
| 4 | [12,1,6,1]- FRP3D65\_YCC |
| 5-15 | Reserved |

Table 2-137 – HDR/WCG widely used video property combinations

|  |  |
| --- | --- |
| **video\_properties\_tag** | **CICP Values - System Identifier**  **[ColourPrimaries, TransferCharacteristics, MatrixCoefficients, VideoFullRangeFlag]** |
| 0 | Video property CICP combination not specified or unknown |
| 1 | [9,16,9,0]- BT2100\_PQ\_YCC |
| 2 | [9,18,9,0]- BT2100\_HLG\_YCC |
| 3 | [9,16,14,0]- BT2100\_PQ\_ICTCP |
| 4 | [9,16,0,0]- BT2100\_PQ\_RGB |
| 5 | [9,18,0,0]- BT2100\_HLG\_RGB |
| 6-15 | Reserved |

Table 2-138 – No Indication

|  |  |
| --- | --- |
| **video\_properties\_tag** | **CICP Values - System Identifier**  **[ColourPrimaries, TransferCharacteristics, MatrixCoefficients, VideoFullRangeFlag]** |
| 0 | Video property CICP combination not specified or unknown |
| 1-15 | Reserved or private |

**temporal\_id\_min** – This 3-bit field indicates the minimum value of the TemporalId, as defined in Rec. ITU-T H.266 | ISO/IEC 23090-3, of all VVC access units in the associated elementary stream.

**temporal\_id\_max** – This 3-bit field indicates the maximum value of the TemporalId, as defined in Rec. ITU-T H.266 | ISO/IEC 23090-3, of all VVC access units in the associated elementary stream.

NOTE 5 – These values indicate maximum and minimum values, but they do not guarantee that all temporal layers are present at all times, because the number of temporal layers may change eg. when splicing occurs. However, if the temporal level goes beyond the range given by min and max values, a new descriptor with appropriate values needs to be inserted.

### 2.6.131 VVC timing and HRD descriptor

For a VVC video stream, a VVC temporal video sub-bitstream or a VVC temporal video subset, the VVC timing and HRD descriptor (see Table 2-139) provides timing and HRD parameters, as defined in Annex C of Rec. ITU-T H.266 | ISO/IEC 23090-3, for the associated VVC video stream or the VVC highest temporal sublayer representation thereof, respectively.

Table 2-139 – VVC timing and HRD descriptor

|  |  |  |
| --- | --- | --- |
| **Syntax** | **No. of bits** | **Mnemonic** |
| VVC\_timing\_and\_HRD\_descriptor() {  **hrd\_management\_valid\_flag**  **reserved**  **picture\_and\_timing\_info\_present\_flag**  if (picture\_and\_timing\_info\_present\_flag == '1') {  **90kHz\_flag**  **reserved**  if (90kHz\_flag = = '0') {  **N**  **K**  **}**  **num\_units\_in\_tick**  **}**  **}** | **1**  **6**  **1**  **1**  **7**  **32**  **32**  **32** | **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf** |

### 2.6.132 Semantic definition of fields in VVC timing and HRD descriptor

**hrd\_management\_valid\_flag** – This 1-bit flag is only defined for use in transport streams. When the VVC timing and HRD descriptor is associated to a VVC video stream or to a VVC highest temporal sublayer representation carried in a transport stream, then the following applies.

If the hrd\_management\_valid\_flag is set to '1', then Buffering Period SEI and Picture Timing SEI messages, as defined in Annex C of Rec. ITU-T H.266 | ISO/IEC 23090-3, shall be present in the associated VVC video stream or VVC highest temporal sublayer representation. These Buffering Period SEI messages shall carry coded nal\_initial\_cpb\_removal\_delay and nal\_initial\_cpb\_removal\_delay\_offset values and may additionally carry nal\_initial\_alt\_removal\_delay and nal\_initial\_alt\_cpb\_removal\_delay\_offset values for the NAL HRD. If the hrd\_management\_valid\_flag is set to '1', then the transfer of each byte from MBn to EBn in the T-STD as defined in 2.23.2 or the transfer from MBn,k to EBn in the T-STD as defined in 2.23.3 shall be according to the delivery schedule for that byte into the CPB in the NAL HRD, as determined from the coded nal\_initial\_cpb\_removal\_delay and nal\_initial\_cpb\_removal\_delay\_offset or from the coded nal\_initial\_alt\_cpb\_removal\_delay and nal\_initial\_alt\_cpb\_removal\_delay\_offset values for SchedSelIdx equal to cpb\_cnt\_minus1 as specified in Annex C of Rec. ITU-T H.266 | ISO/IEC 23090-3. When the hrd\_management\_valid\_flag is set to '0', the leak method shall be used for the transfer from MBn to EBn in the T-STD as defined in 2.23.2 or the transfer from MBn,k to EBn in the T-STD as defined in 2.23.3.

**picture\_and\_timing\_info\_present\_flag** – This 1-bit flag when set to '1' indicates that the 90kHz\_flag and parameters for accurate mapping to 90-kHz system clock are included in this descriptor.

**90kHz\_flag** – This 1-bit flag when set to '1' indicates that the frequency of the VVC time base is 90 kHz.

**N, K** – For a VVC video stream or VVC highest temporal sublayer representation, the frequency of the VVC time base is defined by the syntax element time\_scale in the SPS or VPS, as defined in 7.3.5 of Rec. ITU-T H.266 | ISO/IEC 23090-3. The relationship between the VVC time\_scale and the STC shall be defined by the parameters N and K in this descriptor as follows.

*time\_scale* = (*N* x *system\_clock\_frequency*) / *K*

If the 90kHz\_flag is set to '1', then N equals 1 and K equals 300. If the 90kHz\_flag is set to '0', then the values of N and K are provided by the coded values of the N and K fields.

NOTE – This allows mapping of time expressed in units of time\_scale to 90 kHz units, as needed for the calculation of PTS and DTS timestamps, for example in decoders for VVC access units for which no PTS or DTS is encoded in the PES header.

**num\_units\_in\_tick** – This 32-bit field is coded exactly in the same way as the num\_units\_in\_tick field in SPS or VPS as defined in 7.3.5 of Rec. ITU-T H.266 | ISO/IEC 23090-3. The information provided by this field shall apply to the entire VVC video stream or VVC highest temporal sublayer representation to which the VVC timing and HRD descriptor is associated.

### 2.6.133 EVC video descriptor

For an EVC video stream, the EVC video descriptor (see Table 2-140) provides basic information for identifying coding parameters, such as profile and level parameters, of that EVC video stream. For an EVC temporal video sub-bitstream, the EVC video descriptor provides information such as the associated EVC highest temporal sub-layer representation contained in the elementary stream to which it applies.

Table 2-140 – EVC video descriptor

|  |  |  |
| --- | --- | --- |
| **Syntax** | **No. of bits** | **Mnemonic** |
| EVC\_video\_descriptor() {  **descriptor\_tag**  **descriptor\_length**  **profile\_idc**  **level\_idc**  **toolset\_idc\_h**  **toolset\_idc\_l**  **progressive\_source\_flag**  **interlaced\_source\_flag**  **non\_packed\_constraint\_flag**  **frame\_only\_constraint\_flag**  **reserved**  **temporal\_layer\_subset\_flag**  **EVC\_still\_present\_flag**  **EVC\_24hr\_picture\_present\_flag**  **HDR\_WCG\_idc**  **reserved**  **video\_properties\_tag**  if ( temporal\_layer\_subset\_flag == '1') {  **reserved**  **temporal\_id\_min**  **reserved**  **temporal\_id\_max**  }  } | **8**  **16**  **8**  **8**  **32**  **32**  **1**  **1**  **1**  **1**  **1**  **1**  **1**  **1**  **2**  **2**  **4**  **5**  **3**  **5**  **3** | **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf** |

### 2.6.134 Semantic definition of fields in EVC video descriptor

**profile\_idc, level\_idc, toolset\_idc\_h, toolset\_idc\_l** – These fields shall be coded according to the semantics defined in SPS parameters in ISO/IEC 23094-1 for the corresponding EVC video stream.

**progressive\_source\_flag, interlaced\_source\_flag, non\_packed\_constraint\_flag** – These fields shall be coded according to the semantics sei picture timing message defined in ISO/IEC 23094-1 for as indicated by ct\_type.

**frame\_only\_constraint\_flag** – This field shall be coded according to the semantics defined in ISO/IEC 23094-1 for field\_seq\_flag in the VUI parameters for the corresponding EVC video stream.

**temporal\_layer\_subset\_flag** – This 1-bit flag, when set to '1', indicates that the syntax elements describing a subset of temporal layers are included in this descriptor. This field shall be set to 1 for EVC temporal video subsets. When set to ‘0’, the syntax elements temporal\_id\_min and temporal\_id\_max are not included in this descriptor.

**EVC\_still\_present\_flag** – This 1-bit field, when set to '1', indicates that the EVC video stream or the EVC highest temporal sublayer representation may include EVC still pictures. When set to '0', then the associated EVC video stream shall not contain VVC still pictures.

**EVC\_24hr\_picture\_present\_flag** – This flag when set equal to 0 specifies that no long term reference picture (LTRP) is used for inter prediction of any coded picture in the coded video stream (CVS). When set equal to 1, LTRPs may be used for inter prediction of one or more coded pictures in the CVS.

**HDR\_WCG\_idc** – This 2-bit field indicates the presence or absence of high dynamic range (HDR) and/or wide color gamut (WCG) video components in the associated PID according to Table 2-133. This field shall not be set to 2 unless bit\_depth\_luma\_minus8 and bit\_depth\_chroma\_minus8 as defined in | ISO/IEC 23094-1 in the associated video is greater than or equal to 2. This information is defined in VUI parameters of Rec. ISO/IEC 23094-1 for the corresponding EVC video stream.

**video\_properties\_tag** – This 4-bit field is used to indicate specific video property CICP combinations for SDR, WCG, or HDR/WCG streams depending on the value of HDR\_WCG\_idc. The CICP values are defined in Rec. ITU-T H.273 | ISO/IEC 23091-2. When HDR\_WCG\_idc is equal to 0, Table 2-134 applies. When HDR\_WCG\_idc is equal to 1, Table 2-135 applies. When HDR\_WCG\_idc is equal to 2, Table 2-136 applies. When HDR\_WCG\_idc is equal to 3, Table 2‑137 applies. This information is defined in the VUI parameters of Rec. ISO/IEC 23094-1 for the corresponding EVC video stream.

NOTE 1 – The video\_properties\_tag descriptor can indicate widely used combinations that are documented by ITU-T H.Sup 19 | ISO/IEC TR 23091-4. The system identifiers found in that Supplement | Technical Report are mentioned in Tables 2-135 to 2-138 for convenience

**temporal\_id\_min** – This 3-bit field indicates the minimum value of the TemporalId, as defined in ISO/IEC 23094-1, of all EVC access units in the associated elementary stream.

**temporal\_id\_max** – This 3-bit field indicates the maximum value of the TemporalId, as defined in ISO/IEC 23094-1, of all EVC access units in the associated elementary stream.

NOTE 2 – These values of temporal\_id\_min and temporal\_id\_max indicate maximum and minimum values, respectively, but they do not guarantee that all temporal layers are present at all times, because the number of temporal layers may change eg. when splicing occurs. However, if the temporal level goes beyond the range given by min and max values, a new descriptor with appropriate values needs to be inserted.

### 2.6.135 EVC timing and HRD descriptor

For an EVC video stream or an EVC temporal video sub-bitstream, the EVC timing and HRD descriptor (see Table 2­141) provides timing and HRD parameters, as defined in Annex C of Rec. ISO/IEC 23094-1, for the associated EVC video stream or the EVC highest temporal sub-layer representation thereof, respectively.

Table 2-141 – EVC timing and HRD descriptor

|  |  |  |
| --- | --- | --- |
| **Syntax** | **No. of bits** | **Mnemonic** |
| EVC\_timing\_and\_HRD\_descriptor() {  **hrd\_management\_valid\_flag**  **reserved**  **picture\_and\_timing\_info\_present\_flag**  if (picture\_and\_timing\_info\_present\_flag == '1') {  **90kHz\_flag**  **reserved**  if (90kHz\_flag = = '0') {  **N**  **K**  }  **num\_units\_in\_tick**  }  } | **1**  **6**  **1**  **1**  **7**  **32**  **32**  **32** | **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf** |

### 2.6.136 Semantic definition of fields in EVC timing and HRD descriptor

**hrd\_management\_valid\_flag** – This 1-bit flag is only defined for use in transport streams. When the EVC timing and HRD descriptor is associated to an EVC video stream or to an EVC highest temporal sub-layer representation carried in a transport stream, then the following applies.

If the hrd\_management\_valid\_flag is set to '1', then Buffering Period SEI and Picture Timing SEI messages, as defined in Annex C of ISO/IEC 23094-1, shall be present in the associated EVC video stream or EVC highest temporal sub-layer representation. These Buffering Period SEI messages shall carry coded nal\_initial\_cpb\_removal\_delay and nal\_initial\_cpb\_removal\_delay\_offset values and may additionally carry nal\_initial\_alt\_removal\_delay and nal\_initial\_alt\_cpb\_removal\_delay\_offset values for the NAL HRD. If the hrd\_management\_valid\_flag is set to '1', then the transfer of each byte from MBn to EBn in the T-STD as defined in 2.23.2 or the transfer from MBn,k to EBn in the T-STD as defined in 2.23.3 shall be according to the delivery schedule for that byte into the CPB in the NAL HRD, as determined from the coded nal\_initial\_cpb\_removal\_delay and nal\_initial\_cpb\_removal\_delay\_offset or from the coded nal\_initial\_alt\_cpb\_removal\_delay and nal\_initial\_alt\_cpb\_removal\_delay\_offset values for SchedSelIdx equal to cpb\_cnt\_minus1 as specified in Annex C of ISO/IEC 23094-1. When the hrd\_management\_valid\_flag is set to '0', the leak method shall be used for the transfer from MBn to EBn in the T-STD as defined in 2.24.2 or the transfer from MBn,k to EBn in the T-STD as defined in 2.24.3.

**picture\_and\_timing\_info\_present\_flag** – This 1-bit flag when set to '1' indicates that the 90kHz\_flag and parameters for accurate mapping to 90-kHz system clock are included in this descriptor.

**90kHz\_flag** – This 1-bit flag when set to '1' indicates that the frequency of the EVC time base is 90 kHz.

**N, K** – For an EVC video stream or EVC highest temporal sub-layer representation, the frequency of the EVC time base is defined by the syntax element time\_scale in the SPS, as defined in Annex E of ISO/IEC 23094-1. The relationship between the EVC time\_scale and the STC shall be defined by the parameters N and K in this descriptor as follows.

*time\_scale* = (*N* x *system\_clock\_frequency*) / *K*

If the 90kHz\_flag is set to '1', then N equals 1 and K equals 300. If the 90kHz\_flag is set to '0', then the values of N and K are provided by the coded values of the N and K fields.

NOTE – This allows mapping of time expressed in units of time\_scale to 90 kHz units, as needed for the calculation of PTS and DTS timestamps, for example in decoders for EVC access units for which no PTS or DTS is encoded in the PES header.

**num\_units\_in\_tick** – This 32-bit field is coded exactly in the same way as the num\_units\_in\_tick field in SPS as defined in ISO/IEC 23094-1. The information provided by this field shall apply to the entire EVC video stream or EVC highest temporal sub-layer representation to which the EVC timing and HRD descriptor is associated.

### 2.6.137 LCEVC video descriptor

Table 2-142 provides a description of the LCEVC video descriptor.

| Table 2-142 – LCEVC video descriptor | | |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| LCEVC\_video\_descriptor() { |  |  |
| **lcevc\_stream\_tag** | **8** | **uimsbf** |
| **profile\_idc** | **4** | **uimsbf** |
| **level\_idc** | **4** | **uimsbf** |
| **sublevel\_idc** | **2** | **uimsbf** |
| **processed\_planes\_type\_flag** | **1** | **bslbf** |
| **picture\_type\_bit\_flag** | **1** | **bslbf** |
| **field\_type\_bit\_flag** | **1** | **bslbf** |
| **reserved** | **3** | **bslbf** |
| **HDR\_WCG\_idc** | **2** | **uimsbf** |
| **reserved\_zero\_2bit** | **2** | **bslbf** |
| **video\_properties\_tag** | **4** | **uimsbf** |
| } |  |  |

### 2.6.138 Semantic definition of fields of LCEVC video descriptor

**lcevc\_stream\_tag –** This is an 8-bit field specifying the identifier of an association between base and enhancement streams of an LCEVC encoding.

**profile\_idc, level\_idc, sublevel\_idc –** These fields shall be coded according to the semantics defined for profile\_idc, level\_idc, sublevel\_idc in clause 7.3.4 (SC, sequence\_configuration) of ISO/IEC 23094-2.

**processed\_planes\_type\_flag** – This field shall be coded according to the semantics defined for processed\_planes\_type\_flag in clause 7.3.5 (GC, global\_configuration) of ISO/IEC 23094-2.

**picture\_type\_bit\_flag, field\_type\_bit\_flag** – These fields shall be coded according to the semantics defined for picture\_type\_bit\_flag, field\_type\_bit\_flag in clause 7.3.6 (PC, picture\_configuration) of ISO/IEC 23094-2.

**HDR\_WCG\_idc** – This 2-bit field indicates the presence or absence of high dynamic range (HDR) and/or wide colour gamut (WCG) video components in the associated PID according to Table 2-134.

**video\_properties\_tag** – This 4-bit field is used to indicate specific widely used video property CICP combinations as indicated by ITU-T H.Sup. 19 | ISO/IEC 23091-4 for SDR, WCG, or HDR/WCG streams depending on the value of HDR\_WCG\_idc. When HDR\_WCG\_idc is equal to 0, Table 2-135 applies. When HDR\_WCG\_idc is equal to 1, Table 2‑136 applies. When HDR\_WCG\_idc is equal to 2, Table 2-137 applies. When HDR\_WCG\_idc is equal to 3, Table 2­138 applies.

NOTE 1 – Since the LCEVC enhancement stream and the base video stream that is enhanced are encoded separately, there can be separate values for HDR\_WCG\_idc and HDR\_WCG\_idc in the base and in the enhancement streams, and such values are not necessarily identical.

NOTE 2 – There is no Buffering Period SEI and Picture Timing SEI messages defined for LCEVC because the ones from the base encoding are used. For this reason, LCEVC timing and HRD descriptors are not needed.

### 2.6.139 LCEVC linkage descriptor

Table 2-143 provides a description of the LCEVC linkage descriptor.

Table 2-143 – LCEVC linkage descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| LCEVC\_linkage\_descriptor() { |  |  |
| **num\_lcevc\_stream\_tags**  for (i=0; i<num\_lcevc\_stream\_tags; i++) {  **lcevc\_stream\_tag**  } | **8**  **8** | **uimsbf**  **uimsbf** |
| } |  |  |

### 2.6.140 Semantic definition of fields of LCEVC linkage descriptor

**num\_lcevc\_stream\_tags** **–** The number of lcevc\_stream\_tag items that will follow.

**lcevc\_stream\_tag** – Tag value that allows indicating this video elementary stream as the base of an LCEVC video stream that carries the same tag value in its LCEVC video descriptor.

### 2.6.141 Media service kind descriptor

The media\_service\_kind\_descriptor can read on the program or PID ES stream itself or can read on auxiliary PID ES that are dependent on the principal PID media component ES (i.e., in most cases the principal media component is video). The media service kind descriptor can also read upon the entire program to describe content with a single principal media component. Multiple data elements (N) can be included in the descriptor, each with one or more language code/media\_service\_type pairs.

If the same media component is described at a program level and at an elementary level, then the elementary stream media\_service\_kind description shall take precedence.

Table 2-144 provides a description of the media service kind descriptor. The descriptor uses two loops as indicated in Table 2-144. The outer loop can indicate the number of language\_pairs and an unique identifier for the descriptor. The inner loop indicates the language code and its purpose(s).

Table 2-144 – Media service kind descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| Media\_service\_kind\_descriptor() { |  |  |
| for (i=0; i<N; i++) { |  |  |
| **media\_description\_flag**  **identifier\_flag**  **lang\_pairs**  **media\_type\_idc**  **reserved**  if (identifier\_flag) {  **ID\_length\_code**  **ID\_type**  if (ID\_length\_code = = '7') {  **ID\_len** }  **media\_ID\_field** } | **1**  **1**  **3**  **2**  **1**  **3**  **13**  **8**  **ID\_len\*8** | **bslbf**  **bslbf**  **uimbsf**  **uimbsf**  **bslbf**  **uimbsf**  **uimbsf**  **uimbsf**  **bslbf** |
| for (j= 0; j<lang\_pairs; j++) {  **configuration\_type**  **lang\_purpose\_cnt**  **lang\_len\_idc**  **reserved**  if (lang\_len\_idc = = 0) {  **lang\_len**  } | **2**  **3**  **2**  **1**  **8** | **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **uimsbf** |
| **IETF\_BCP\_47\_language\_code** | **lang\_len\*8** | **bslbf** |
| for (k=0; k<lang\_purpose\_cnt; k++) {  **media\_service\_type**  } | **8** | **uimsbf** |
| }  } |  |  |
| } |  |  |

### 2.6.142 Semantic definition of fields in media service kind descriptor

**media\_description\_flag –** This flag indicates if the specific entry in the descriptor applies to the elementary stream/program itself or indicates an entry for another stream associated with the elementary stream/program as described in Table 2-145 which is needed to produce a specific purpose of the program.

Table 2-145 – media\_description\_flag

|  |  |
| --- | --- |
| **media\_description\_flag** | **media\_description\_flag value** |
| 0 | self |
| 1 | associate |

**identifier\_flag –** A flag to indicate if the media\_service\_kind descriptor data element will carry an identifier.

**lang\_pairs –** This field indicates the number of language code/media\_service\_type pairs carried with the data element.

**media\_type\_idc –** This indicator would have the values "0-unknown", "1-Video", "2-Audio" or "3-Text", as indicated in Table 2-146.

Table 2-146 – Media type indicator

|  |  |
| --- | --- |
| media\_type\_idc | Media type value |
| 0 | unknown |
| 1 | video |
| 2 | audio |
| 3 | text/data |

**ID\_length\_code** – This parameter assigns to ID\_len the number of bytes from the list of potential byte numbers needed for the carriage of the optional identifier. When the parameter has a value of 7 then the ID\_len can be configured to other byte length numbers. See Table 2-147.

Table 2-147 – ID\_length\_code

|  |  |
| --- | --- |
| Value | Length, bytes (*ID\_len)* |
| 0 | 1 |
| 1 | 2 |
| 2 | 4 |
| 3 | 8 |
| 4 | 12 |
| 5 | 16 |
| 6 | 20 |
| 7 | Explicit value provided in field ID\_len |

**ID\_type –** Parameter to indicate the type of identifier used according to Table 2-148.

NOTE 1 – This parameter can originate from a list of SCTE UPID identifiers list. Alternatively, identifiers can be reserved outside of the UPID identifiers list.

Table 2-148 – ID\_type

|  |  |
| --- | --- |
| Range | Definition |
| 0x0000 .. 0x01FF | Reserved for MPEG standardization |
| 0x0200 .. 0x02FF | Value defined in ANSI/SCTE 35 Table 22 (segmentation\_upid\_type) + 0x200 |
| 0x0300 .. 0x0FFF | Reserved for MPEG standardization |
| 0x1000 .. 0x1FFF | User Private types |

**ID\_len –** The number of bytes needed to carry an optional identifier in the data element of the descriptor**.**

**media\_ID\_field –** The parameter carrying the identifier in the data element of the descriptor.

**configuration\_type –** This parameter, as defined in Table 2-149, indicates if the elementary stream or associate stream is complete (value= 0 e.g., a 5.1 audio stream) and playable as an independent media component or is partial (value= 1 e.g., M&E or dialogue only) and not playable alone as the media component stream. A complete combination (value= 2 e.g., 5.1/2.0 stream with multiple dialogue languages) contains several variants with each version playable as an independent media component. Internal coding configurations if defined are deferred to the elementary specific format metadata.

Table 2-149 – configuration type values

|  |  |
| --- | --- |
| configuration\_type | configuration type value |
| 0 | complete |
| 1 | partial |
| 2 | complete combination |
| 3 | reserved |

configuration\_type value semantics

**complete**: Describes a media component stream that is a single version playable video/audio/text experience otherwise known as complete or complete main.

**partial**: Describes a media component stream that is not a playable video/audio/text experience without additional streams.

**complete combination**: Describes a media component stream that contains multiple playable video/audio/text experiences which can be extracted through different parameter settings.

**lang\_purpose\_cnt –** This parameter indicates the number of media\_service\_types (up to 6) associated with a single language code. If the count is zero it means to use the default media\_service\_type of "und" and if no "main" exists for the track, then this defaults to "main". A value of 7 is reserved for future purposes.

**lang\_len\_idc –** Indicator to determine if a default length is used to indicate the IETF\_BCP\_47\_language\_code field length. If indicator is 0, then char length is assignable by a byte value. When set to 1 or 2, the value of lang\_len is given in Table 2-150. A value of 3 is reserved for future use.

Table 2-150 – lang\_len\_indicator

|  |  |
| --- | --- |
| lang\_length\_idc | lang\_len |
| 0 | No default length. Explicit value provided in field lang\_len |
| 1 | 2 |
| 2 | 3 |
| 3 | reserved |

**IETF\_BCP47\_language\_code –** Identifies the language or languages, including sign language, used by the associated program element. The IETF\_BCP47\_language\_code contains a multi-character code as specified by IETF BCP 47 where the number of characters is determined by the value in the preceding lang\_len field. IETF BCP 47 incorporates ISO 639‑3 language code with additional description. Each character is coded into 8 bits according to ISO 8859-1 and inserted in order into this field. In the case of multilingual audio streams, the sequence of ISO\_639\_language\_code fields shall reflect the content of the audio stream or designated as 'mul' three-character subtag for dynamic multilingual cases but in the case that the descriptor needs to remain static. In the case of video components that are not sign language,'zxx' can be used as a three-character primary language subtag.

**media\_service\_type –** This 8-bit field specifies the purpose(s) of the stream(s) described in the descriptor as defined in Table 2-151.

| Table 2-151 – Media service type values | |
| --- | --- |
| Value | Description |
| 0x00 | undefined |
| 0x01 | main |
| 0x02 | alternate |
| 0x03 | supplementary |
| 0x04 | emergency |
| 0x05 | description |
| 0x06 | enhanced-audio-intelligibility |
| 0x07 | dub |
| 0x08 | primary commentary |
| 0x09 | primary |
| 0x0A | native |
| 0x0B | Music and effects |
| 0x0C | dialogue |
| 0x0D | voice-over |
| 0x0E | sign |
| 0x0F | multi-view |
| 0x10 | karaoke |
| 0x11 | caption |
| 0x12 | subtitle |
| 0x13 | forced-subtitle |
| 0x14 | metadata |
| 0x15 | non-primary |
| 0x16 | substitution |
| 0x17 | alternate commentary |
| 0x18 | stadium sound |
| 0x19 .. 0xEF | Reserved |
| 0xF0 .. 0xFF | User private |

media\_service\_type value semantics

[General]

**undefined**: This is a default value that indicates that the purpose of the stream is unknown. It is different from the language code "und", which would indicate that the language of the audio is unknown. If "main" is not assigned in the descriptor for the stream, then this defaults to "main".

**main**: Main media component(s) that is/are intended for presentation if no other information is provided.

**alternate**: Media content component(s) that is/are an alternative to (a) main media content component(s) of the same media component type.

**supplementary**: Media content component that is supplementary to a media content component of a different media component type.

**emergency**: Experience that provides information about a current emergency. It is intended to enable the protection of life, health, safety and property, and may also include critical details regarding the emergency and how to respond to the emergency.

**substitution**: Indicates that the media content component is continuous but may have a substitution in language or media\_service\_type that defaults to the main type for that media content component when not available. This often occurs in linear channels where Spanish or audio description may not continuously exist for all programs on the channel, but the linear channel track does need to be continuous.

[Video specific]

**sign**: Visual media component representing a sign-language interpretation of an audio component.

**multi-view**: This value indicates that the video stream is an alternate view of the main video feed. This can be used for different sports angle feed in a football game.

**caption**: Textual representation of the audio track usually including both dialogue and all sounds. This text is often embedded in the video track intended for an audience with loss or partial loss of hearing and is in the same language as the dialogue.

**forced-subtitle**: Textual information meant for display when no other text representation is selected. It is used to clarify dialogue, alternative languages, texted graphics or location/person IDs that are not otherwise covered in the dubbed/localized audio.

[Audio specific]

**description**: This value indicates that the referenced program element is prepared for the visually impaired viewer. Alternate labels for this are "audio descriptive services" (ADS) or "descriptive video services" (DVS) or "visual impaired commentary".

**enhanced-audio-intelligibility**: Experience containing an element for improved intelligibility of the dialogue. An alternative label for this is "hearing impaired".

**dub**: Experience that contains an element that is presented in a language different from the original (e.g., dubbed audio, translated captions).

**primary commentary**: This value indicates the referenced program element is prepared with narration in addition to the natural sound of the content. This narration can be an announcer to a sporting event or the director's comments on a film or program, and is the default commentary for the content.

**alternate commentary**: This value indicates the referenced program element is prepared as bonus or alternate narration in addition to the natural sound of the content. This additional commentary is of the same language as the primary commentary and can be used for cases such as an additional away announcer to a sporting event or producer's comments of the film or program.

**primary**: This value indicates the default recommended language of the channel or program. This is a separate value from the language code itself. If the language code is Welsh with an undefined media\_service\_type, it means "this is Welsh". When the language code is welsh with media\_service\_type 0x09, it means "this is Welsh, and it is the primary language of this program."

**non-primary**: This value indicates a secondary language of the channel or program. This is a separate value from the language code itself, and indicates a secondary priority of language in the channel or program. Secondary languages may have substitutions in its track when that language is not available in order to provide a continuous track for the channel.

**substitution**: This value indicates the language in the channel or program may be replaced by another language if the language is sometimes unavailable, but the track needs to be continuous with the program. An example of this is a Spanish SAP (secondary audio program) where the commercial does not have a Spanish language track so an English language track is substituted until the advertisement is done.

**native**: This value indicates the language in which the program was produced. An absence of this value indicates that it is either undefined or a dubbed language.

**Music and effects**: This value indicates that the referenced program element has no language. This can also be known as "clean effects".

**dialogue**: An isolated language track intended to be combined with a music and effects track.

**stadium sound**: Audio track for an event (e.g., sporting event) devoid of any additional commentary. It would be the natural sound one would hear if physically attending the event at the stadium. An alternative term for this would be natural sound.

**voice-over**: An isolated narration track intended to be combined with audio that reflects the natural sound of the scene content.

[Text/Data specific]

**dub**: Experience that contains an element that is presented in a language different from the original (e.g., dubbed audio, translated captions).

**commentary**: Experience that contains a commentary (e.g., director's commentary, sports announcer).

**primary**: This value indicates the default recommended textual language of the channel or program. This is a separate value from the language code itself. If the language code is Welsh with an undefined media\_service\_type, it means "this is Welsh subtitles". When the language code is Welsh with media\_service\_type 0x09, it means "this is Welsh, and it is the primary subtitles language of this program".

**non-primary**: This value indicates a secondary textual language of the channel or program. This is a separate value from the language code itself and indicates a secondary priority of textual language in the channel or program. Secondary textual languages may have substitutions in its track when that textual language is not available in order to provide a continuous track for the channel.

**native**: This value indicates the language in which the program was produced. An absence of this value indicates that it is either an undefined or a dubbed language.

**karaoke**: Textual representation of a song's lyrics, usually in the same language as the associated songs.

**subtitle**: Textual representation of the audio track. It usually represents only the dialogue and often in a language other than the audio track dialogue. This is intended for a foreign language audience.

**metadata**: Media component containing information intended to be processed by application specific elements.

NOTE 2 – Similar corresponding description values can be found in the definition of the ISO\_639\_language descriptor and in other standards/specifications ISO/IEC 23009-1 (MPEG-DASH), SMPTE ST 2067-2 (IMF Core Constraints), or DASH-IF interoperability points.

Figure 2-4 and Figure 2-5 show examples of the Media Service Kind Descriptor.

|  |  |
| --- | --- |
| // Program – Media Service Kind Descriptor  i=0 // Identify Program  media\_description\_flag 🡪self  identifier\_flag🡪 true  lang\_pairs 🡪 0  media\_type\_idc 🡪 content  ID\_length\_code🡪 4 (ID\_len is 12)  ID\_type 🡪 0x20A (EIDR-Content)  media\_ID\_field 🡪 10.5240/0E4F-892E-442F-6BD4-15B0-1  i=1 // Identify Video  media\_description\_flag 🡪 self  identifier\_flag🡪 false  lang\_pairs 🡪2  media\_type\_idc 🡪 video  j=0  configuration\_type 🡪 complete  lang\_purpose\_cnt🡪0  lang\_len\_idc🡪2 (lang\_len = 3)  BCP\_47\_language\_code. 🡪zxx  default // media\_service\_type 🡪 und 🡪 main  j=1 // identify Closed Captioning English  configuration\_type🡪complete combination  lang\_purpose\_cnt🡪1  lang\_len\_idc🡪2 (lang\_len = 3)  BCP\_47\_language\_code. 🡪 eng  k=1 // media\_service\_type 🡪 caption | i=2 // Identify Main Audio Track of Linear Channel  media\_description\_flag 🡪 self  identifier\_flag 🡪 false  lang\_pairs 🡪 1  media\_type\_idc 🡪 audio  j=0 // Identify as English Main-Primary-Native  configuration\_type 🡪 complete  lang\_purpose\_cnt 🡪 3  len\_idc 🡪 2 (lang\_len = 3)  BCP\_47\_language\_code. 🡪eng  k=1 // media\_service\_type 🡪 main  k=2 // media\_service\_type 🡪 primary  k=3 // media\_service\_type 🡪 native  i=3 // Identify 2nd Audio Track as Spanish  media\_description\_flag 🡪 self  identifier\_flag🡪 false  lang\_pairs 🡪1  media\_type\_idc 🡪 audio  j=0 // Identify as Spanish Main  configuration\_type 🡪 complete  lang\_purpose\_cnt🡪0  len\_idc🡪2 (lang\_len = 3)  BCP\_47\_language\_code 🡪 spa  default // media\_service\_type 🡪 und 🡪 main |

Figure 2-4 – Media Service Kind Descriptor semantics at program level

|  |
| --- |
| // An Audio elementary stream demuxed that has an association to a video track or content  i=0 // Identify Audio Elementary Stream  media\_description\_flag 🡪 self  identifier\_flag 🡪 false  lang\_pairs 🡪 1  media\_type\_idc 🡪 audio  j=0 // Identify as English Main-Native  configuration\_type 🡪 complete  lang\_purpose\_cnt 🡪 2  len\_idc🡪2 (lang\_len = 3)  BCP\_47\_language\_code 🡪 eng  k=1//media\_service\_type 🡪 main  k=2//media\_service\_type 🡪 native  i=1 // Associate to Video via identifier  media\_description\_flag 🡪 associate  identifier\_flag 🡪 true  lang\_pairs 🡪 0  media\_type\_idc 🡪 video  ID\_length\_code🡪 4 (ID\_len is 12)  ID\_type 🡪 0x20A (EIDR-Video)  media\_ID\_field 🡪 10.5239/C370-DCA5 |

Figure 2-5 – Media Program Kind Descriptor semantics at elementary stream level

## 2.7 Restrictions on the multiplexed stream semantics

### 2.7.1 Frequency of coding the system clock reference

The program stream shall be constructed such that the time interval between the bytes containing the last bit of system\_clock\_reference\_base fields in successive packs shall be less than or equal to 0.7 s. Thus:



for all i and i′ where i and i′ are the indexes of the bytes containing the last bit of consecutive system\_clock\_reference\_base fields.

### 2.7.2 Frequency of coding the program clock reference

The transport stream shall be constructed such that the time interval between the bytes containing the last bit of program\_clock\_reference\_base fields in successive occurrences of the PCRs in transport stream packets of the PCR\_PID for each program shall be less than or equal to 0.1 s. Thus:



for all i and i′ where i and i′ are the indexes of the bytes containing the last bit of consecutive program\_clock\_reference\_base fields in the transport stream packets of the PCR\_PID for each program.

There shall be at least two (2) PCRs, from the specified PCR\_PID within a transport stream, between consecutive PCR discontinuities (refer to 2.4.3.4) to facilitate phase locking and extrapolation of byte delivery times.

### 2.7.3 Frequency of coding the elementary stream clock reference

The program stream and transport stream shall be constructed such that if the elementary stream clock reference field is coded in any PES packets containing data of a given elementary stream the time interval in the PES\_STD between the bytes containing the last bit of successive ESCR\_base fields shall be less than or equal to 0.7 s. In PES Streams the ESCR encoding is required with the same interval. Thus:



for all i and i′ where i and i′ are the indexes of the bytes containing the last bits of consecutive ESCR\_base fields.

NOTE – The coding of elementary stream clock reference fields is optional; they need not be coded. However, if they are coded, this constraint applies.

### 2.7.4 Frequency of presentation timestamp coding

The program stream and transport stream shall be constructed so that the maximum difference between coded presentation timestamps referring to each elementary video or audio stream is 0.7 s. Thus:



for all n, k, and k″ satisfying:

• Pn(k) and Pn(k″) are presentation units for which presentation timestamps are coded;

• k and k″ are chosen so that there is no presentation unit, Pn(k′) with a coded presentation timestamp and with k < k′ < k″; and

• No decoding discontinuity exists in elementary stream n between Pn(k) and Pn(k″).

The 0.7-s constraint does not apply in the case of:

• still pictures as defined in 2.1;

• AVC still pictures;

• AVC access units with a very low frame rate, where the presentation time of subsequent access units differs by more than 0.7 s. In this particular case, the VUI parameters num\_units\_in\_tick and time\_scale shall be present either in the AVC video stream or in an AVC-timing and HRD descriptor associated with the AVC video stream.

NOTE – The presentation time of an AVC access unit is equivalent to the DPB output time to,dpb(n) defined in Annex C of Rec. ITU‑T H.264 | ISO/IEC 14496-10.

### 2.7.5 Conditional coding of timestamps

For each elementary stream of a program stream or transport stream, a presentation timestamp (PTS) shall be encoded for the first access unit.

A decoding discontinuity exists at the start of an access unit An(j) in an elementary stream n if the decoding time tdn(j) of that access unit is greater than the largest value permissible given the specified tolerance on the system\_clock\_frequency. For video, except when trick mode status is true or when low\_delay flag is '1', this is allowed only at the start of a video sequence. If a decoding discontinuity exists in any elementary video or audio stream in the transport stream or program stream, then a PTS shall be encoded referring to the first access unit after each decoding discontinuity except when trick mode status is true.

When low\_delay is '1' a PTS shall be encoded for the first access unit after an EBn or Bn underflow.

A PTS may only be present in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 video or audio elementary stream PES packet header if the first byte of a picture start code or the first byte of an audio access unit is contained in the PES packet.

A decoding\_timestamp (DTS) shall appear in a PES packet header if and only if the following two conditions are met:

• a PTS is present in the PES packet header;

• the decoding time differs from the presentation time.

For each AVC 24-hour picture, no explicit PTS and DTS value shall be encoded in the PES header. For such AVC access unit, decoders shall infer the presentation time from the parameters within the AVC video stream. Therefore, each AVC video stream that contains one or more AVC 24-hour picture(s):

• shall either carry picture timing SEI messages with coded values of cpb\_removal\_delay and dpb\_output\_delay; or

• shall carry VUI parameters with the fixed\_frame\_rate\_flag set to '1' and shall carry Picture Order Count (POC) information (PicOrderCnt) whereby pictures are counted in units of Δtfi,dpb( n ), where Δtfi,dpb( n ) is specified in equation E-10 of Rec. ITU-T H.264 | ISO/IEC 14496-10.

NOTE 1 – The requirements in the second bullet are met if an AVC timing and HRD descriptor is associated with the AVC video stream with the fixed\_frame\_rate\_flag set to '1' and the temporal\_poc\_flag set to '1'.

The following applies to AVC access units in an AVC video stream carried in a Rec. ITU-T H.222.0 | ISO/IEC 13818‑1 stream. For each AVC access unit that does not represent an AVC 24-hour picture, a PES header with a coded PTS and, if applicable, DTS value shall be provided, unless all conditions expressed under one of the following four bullets are true:

• In the AVC video sequence the following SEI messages are present, as signalled by VUI parameters:

a) picture timing SEI messages providing the cpb\_removal\_delay and the dpb\_output\_delay parameters; and

b) buffering period SEI messages providing the initial\_cpb\_removal\_delay and the initial\_cpb\_removal\_delay\_offset parameters.

NOTE 2 – When picture timing SEI messages are present in the AVC video sequence, then these messages are present for each AVC access unit, as required by Rec. ITU-T H.264 | ISO/IEC 14496-10. When buffering period SEI messages are present in the AVC video sequence, then these messages shall be present for each IDR access unit and for each access unit that is associated with a recovery point SEI message, as required by Rec. ITU-T H.264 | ISO/IEC 14496-10.

• An AVC timing and HRD descriptor is associated with the AVC video stream and in this descriptor the fixed\_frame\_rate\_flag is set to '1' and the temporal\_poc\_flag is set to '1'.

• An AVC timing and HRD descriptor is associated with the AVC video stream and in this descriptor the fixed\_frame\_rate\_flag is set to '1', the picture\_to\_display\_conversion\_flag is set to '1', the temporal\_poc\_flag is set to '0' and in the AVC video sequence picture timing SEI messages with the pic\_struct field are present.

NOTE 3 – In this specific case the pic\_struct field is used to determine subsequent PTS values.

• An AVC timing and HRD descriptor is associated with the AVC video stream and in this descriptor the fixed\_frame\_rate\_flag is set to '1' and the temporal\_poc\_flag is set to '0' and the picture\_to\_display\_conversion\_flag is set to '0'.

NOTE 4 – In this case the POC information in the AVC video stream is used to determine the subsequent PTS values.

### 2.7.6 Timing constraints for scalable coding

If an audio sequence is coded using an extension bitstream, such as specified in ISO/IEC 13818-3, then corresponding decoding/presentation units in the two layers shall have identical PTS values.

If a video sequence is coded as an SNR enhancement of another sequence, such as specified in 7.8 of Rec. ITU-T H.262 | ISO/IEC 13818-2, then the set of presentation times for both sequences shall be the same.

If a video sequence is coded as two partitions, such as specified in 7.10 of Rec. ITU-T H.262 | ISO/IEC 13818-2, then the set of presentation times for both partitions shall be the same.

If a video sequence is coded as a spatial scalable enhancement of another sequence, such as specified in 7.7 of Rec. ITU‑T H.262 | ISO/IEC 13818-2, then the following shall apply:

• If both sequences have the same frame rate, the set of presentation times for both sequences shall be the same.

NOTE – This does not imply that the picture coding type is the same in both layers.

• If the sequences have different frame rates, the set of presentation times shall be such that as many presentation times as possible shall be common to both sequences.

• The picture from which the spatial prediction is made shall be one of the following:

– the coincident or most recently decoded lower layer picture;

– the coincident or most recently decoded lower layer picture that is an I- or P-picture;

– the second most recently decoded lower layer picture that is an I- or P-picture, and provided that the lower layer does not have the low\_delay flag set to '1'.

If a video sequence is coded as a temporally scalable enhancement of another sequence, such as specified in 7.9 of Rec. ITU-T H.262 | ISO/IEC 13818-2, then the following lower layer pictures may be used as the reference. Times are relative to presentation times of:

• the coincident or most recently presented lower layer picture;

• the next lower layer picture to be presented.

For AVC video streams conforming to one or more profiles defined in Annex G of Rec. ITU-T H.264 | ISO/IEC 14496‑10, there are no timing constraints on SVC dependency representations or re-assembled AVC access units of video sub-bitstreams of an AVC video stream.

### 2.7.7 Frequency of coding P-STD\_buffer\_size in PES packet headers

In a program stream, the P-STD\_buffer\_scale and P-STD\_buffer\_size fields shall occur in the first PES packet of each elementary stream and again whenever the value changes. They may also occur in any other PES packet.

### 2.7.8 Coding of system header in the program stream

In a program stream, the system header may be present in any pack, immediately following the pack header. The system header shall be present in the first pack of a program stream. The values encoded in all the system headers in the program stream shall be identical.

### 2.7.9 Constrained system parameter program stream

A program stream is a "Constrained System Parameters Stream" (CSPS) if it conforms to the bounds specified in this subclause. Program streams are not limited to the bounds specified by the CSPS. A CSPS may be identified by means of the CSPS\_flag defined in the system header in 2.5.3.5. The CSPS is a subset of all possible program streams.

Packet rate

In the CSPS, the maximum rate at which packets shall arrive at the input to the P-STD is 300 packets per second if the value encoded in the rate\_bound field (refer to 2.5.3.6) is less than or equal to 4 500 000 bits/s if the packet\_rate\_restriction\_flag is set to '1', and less than or equal to 2 000 000 bits/s if the packet\_rate\_restriction\_flag is set to '0'. For higher bit rates the CSPS packet rate is bounded by a linear relation to the value encoded in the rate\_bound field.

Specifically, for all packs p in the program stream when the packet\_rate\_restriction\_flag (refer to 2.5.3.5) is set to a value of '1',

 (2-27)

and if the packet\_rate\_restriction\_flag is set to a value of '0'

 (2-28)

where:

 (2-29)

NP is the number of packet\_start\_code\_prefixes and system\_header\_start\_codes between adjacent pack\_start\_codes or between the last pack\_start\_code and the MPEG\_program\_end\_code as defined in Table 2-37 and semantics in 2.5.3.2.

t(i) is the time, measured in seconds, encoded in the SCR of pack p.

t(i′) is the time, measured in seconds, encoded in the SCR for pack p + 1, immediately following pack p, or in the case of the final pack in the program stream, the time of arrival of the byte containing the last bit of the MPEG\_program\_end\_code.

Decoder buffer size

In the case of a CSPS the maximum size of each input buffer in the system target decoder is bounded. Different bounds apply for video elementary streams and audio elementary streams.

In the case of a Rec. ITU-T H.262 | ISO/IEC 13818-2 or ISO/IEC 11172-2 video elementary stream in a CSPS, the following applies:

BSn has a size which is equal to the sum of the size of the Video Buffer Verifier (VBV) as specified in the Rec. ITU‑T H.262 | ISO/IEC 13818-2 or ISO/IEC 11172-2 stream, respectively, and an additional amount of buffering BSadd. BSadd is specified as:

*BSadd* ≤ *MAX* [6 × 1024, *Rvmax* × 0.001] bytes

where Rvmax is the maximum bit rate of the Rec. ITU-T H.262 | ISO/IEC 13818-2 or ISO/IEC 11172-2 video elementary stream.

In the case of a Rec. ITU-T H.264 | ISO/IEC 14496-10 video elementary stream in a CSPS, the following applies:

BSn has a size which is equal to the sum of cpb\_size and an additional amount of buffering BSadd. BSadd is specified as:

*BSadd* ≤ *MAX* [6 × 1024, *Rvmax* × 0.001] bytes

where Rvmax is the maximum video bit rate of the AVC video stream, and

where cpb\_size is the CpbSize[ cpt\_cnt\_minus1 ] size of the CPB for the byte stream format signalled in the NAL hrd\_parameters() in the AVC video stream. If the NAL hrd\_parameters() are not present in the AVC video stream, then the cpb\_size shall be the size defined as 1200 × MaxCPB in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496-10 for the applied level.

In the case of an audio elementary stream in a CSPS, unless otherwise specified below, the following applies:

*BSn* ≤ 4096 bytes

In the case of ISO/IEC 13818-7 ADTS audio elementary stream in a CSPS the following applies to support 8 channels:

*BSn* ≤ 8976 bytes

### 2.7.10 Transport stream

Sample rate locking in transport streams

In the transport stream there shall be a specified constant rational relationship between the audio sampling rate and the system clock frequency in the system target decoder, and likewise a specified rational relationship between the video frame rate and the system clock frequency. The system\_clock\_frequency is defined in 2.4.2.1. The video frame rate is specified in Rec. ITU-T H.262 | ISO/IEC 13818-2 or in ISO/IEC 11172-2. The audio sampling rate is specified in ISO/IEC 13818-3 or in ISO/IEC 11172-3. For all presentation units in all audio elementary streams in the transport stream, the ratio of system\_clock\_frequency to the actual audio sampling rate, SCASR, is constant and equal to the value indicated in the following table at the nominal sampling rate indicated in the audio stream.

 (2-30)

The notation  denotes real division.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Nominal audio sampling frequency (kHz) | 16 | 32 | 22.05 | 44.1 | 24 | 48 |
| SCASR | 27 000 000 ------------- 16 000 | 27 000 000 ------------- 32 000 | 27 000 000 ------------- 22 050 | 27 000 000 ------------- 44 100 | 27 000 000 ------------- 24 000 | 27 000 000 ------------- 48 000 |

For all presentation units in each ISO/IEC 11172-2 video and Rec. ITU-T H.262 | ISO/IEC 13818-2 video stream in the transport stream, the ratio of system\_clock\_frequency to the actual video frame rate, SCFR, is constant and equal to the value indicated in the following table at the nominal frame rate indicated in the video stream.

 (2-31)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nominal frame rate (Hz) | 23.976 | 24 | 25 | 29.97 | 30 | 50 | 59.94 | 60 |
| SCFR | 1 126 125 | 1 125 000 | 1 080 000 | 900 900 | 900 000 | 540 000 | 450 450 | 450 000 |

The values of the SCFR are exact. The actual frame rate differs slightly from the nominal rate in cases where the nominal rate is 23.976, 29.97, or 59.94 frames per second.

For ISO/IEC 14496-2 video streams carried in a transport stream, the time base of the ISO/IEC 14496-2 video stream, as defined by vop\_time\_increment\_resolution, shall be locked to the STC and shall be exactly equal to N times system\_clock\_frequency divided by K, with N and K integers that have a fixed value within each visual object sequence, with K greater than or equal to N.

For Rec. ITU-T H.264 | ISO/IEC 14496-10 video streams, the time base of the Rec. ITU-T H.264 | ISO/IEC 14496-10 video stream shall be locked to the system clock frequency. The frequency of the AVC time base is defined by the AVC parameter time\_scale, and this frequency shall be exactly equal to N times system\_clock\_frequency divided by K, with N and K integers that have a fixed value within each AVC video sequence and K greater than or equal to N. For example, if the time\_scale is set to 90 000, then the frequency of the AVC time base is exactly equal to system\_clock\_frequency divided by 300.

## 2.8 Compatibility with ISO/IEC 11172

The program stream of this Recommendation | International Standard is defined to be forward compatible with ISO/IEC 11172-1. Decoders of the program stream as defined in this Recommendation | International Standard shall also support decoding of ISO/IEC 11172-1.

## 2.9 Registration of copyright identifiers

### 2.9.1 General

Parts 1, 2 and 3 of ISO/IEC 13818 provide support for the management of audiovisual works copyrighting. In Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 this is by means of a copyright descriptor, while Rec. ITU-T H.262 | ISO/IEC 13818-2 and ISO/IEC 13818-3 contain fields for identifying copyright holders through syntax fields in the elementary stream syntax. This Recommendation | International Standard presents the method of obtaining and registering copyright identifiers in Rec. ITU-T H.222.0 | ISO/IEC 13818-1.

Rec. ITU-T H.222.0 | ISO/IEC 13818-1 specifies a unique 32-bit copyright\_identifier which is a work type code identifier (such as ISBN, ISSN, ISRC, etc.) carried in the copyright descriptor. The copyright\_identifier enables identification of a wide number of Copyright Registration Authorities. Each Copyright Registration Authority may specify a syntax and semantic for identifying the audiovisual works or other copyrighted works within that particular copyright organization through appropriate use of the variable length additional\_copyright\_info field which contains the copyright number.

In the following subclause and Annexes L, M and N, the benefits and responsibilities of all parties to the registration of copyright\_identifier are outlined.

### 2.9.2 Implementation of a Registration Authority (RA)

ISO/IEC JTC 1 shall call for nominations for an international organization which will serve as the Registration Authority for the **copyright\_identifier** as defined in 2.6.24. The selected organization shall serve as the Registration Authority. The so-named Registration Authority shall execute its duties in compliance with Annex H/JTC 1 Directives. The registered copyright\_identifier is hereafter referred to as the Registered Identifier (RID).

Upon selection of the Registration Authority, JTC 1 shall require the creation of a Registration Management Group (RMG) which will review appeals filed by organizations whose request for a RID to be used in conjunction with Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 has been denied by the Registration Authority.

Annexes L, M and N provide information on the procedure for registering a unique copyright identifier.

## 2.10 Registration of private data format

The registration descriptor of Rec. ITU-T H.222.0 | ISO/IEC 13818-1 is provided by this text in order to enable users of this Specification to unambiguously carry data when its format is not recognized by this Specification. This provision will permit this Specification to carry all types of data while providing for a method of unambiguous identification of the characteristics of the underlying private data.

### 2.10.1 General

In the following subclause and Annexes O and P, the benefits and responsibilities of all parties to the registration of private data format are outlined.

### 2.10.2 Implementation of a Registration Authority (RA)

ISO/IEC JTC 1/SC 29 shall call for nominations from member bodies of ISO or National Committees of IEC which will serve as the Registration Authority for the **format\_identifier** as defined in 2.6.8 and 2.6.9. The selected organization shall serve as the Registration Authority. The so-named Registration Authority shall execute its duties in compliance with Annex H/JTC 1 Directives. The registered private data format\_identifier is hereafter referred to as the Registered Identifier (RID).

Upon selection of the Registration Authority, JTC 1 shall require the creation of a Registration Management Group (RMG) which will review appeals filed by organizations whose request for an RID to be used in conjunction with this Specification has been denied by the Registration Authority.

Annexes O and P provide information on the procedures for registering a unique format identifier.

## 2.11 Carriage of ISO/IEC 14496 data

### 2.11.1 Introduction

A Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream may carry individual ISO/IEC 14496-2 and 14496-3 elementary streams as well as ISO/IEC 14496-1 audiovisual scenes with its associated streams. Typically, the ISO/IEC 14496 streams will be elements of a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, as defined by the PMT in a transport stream and the PSM in a program stream.

For the carriage of ISO/IEC 14496 data in transport streams and program streams, distinction is made between individual elementary streams and an ISO/IEC 14496-1 audiovisual scene with its associated streams. For carriage of individual ISO/IEC 14496-2 and 14496-3 elementary streams, only system tools from Rec. ITU-T H.222.0 | ISO/IEC 13818-1 are used, as defined in 2.11.2. For carriage of an audiovisual ISO/IEC 14496-1 scene and associated ISO/IEC 14496 elementary streams, contained in ISO/IEC 14496-1 SL\_packetized streams or M4Mux streams, tools from both Rec. ITU-T H.222.0 | ISO/IEC 13818-1 and from ISO/IEC 14496-1 are used, as defined in 2.11.3.

Carriage of Rec. ITU-T H.264 | ISO/IEC 14496-10 video over Rec. ITU-T H.222.0 | ISO/IEC 13818-1 streams is specified in 2.14.

Carriage of ISO/IEC 14496-17 text streams over Rec. ITU-T H.222.0 | ISO/IEC 13818-1 streams is specified in 2.15.

### 2.11.2 Carriage of individual ISO/IEC 14496-2 and 14496-3 Elementary Streams in PES packets

#### 2.11.2.1 Introduction

Individual ISO/IEC 14496-2 and 14496-3 elementary streams may be carried in PES packets as PES\_packet\_data\_bytes. For PES packetization no specific data alignment constraints apply. For synchronization PTSs and, when appropriate, DTSs are encoded in the header of the PES packet that carries the ISO/IEC 14496 elementary stream data; for PTS and DTS encoding the same constraints apply as for ISO/IEC 13818 elementary streams. See Table 2-152 for an overview of how to carry individual ISO/IEC 14496 streams within a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream.

Table 2-152 – Carriage of individual ISO/IEC 14496 streams in Rec. ITU-T H.222.0 | ISO/IEC 13818-1

|  |  |  |  |
| --- | --- | --- | --- |
| ISO/IEC 14496-2 visual | Carriage in PES packets | Stream\_type = 0x10 | Stream\_id = '1110 xxxx' |
| ISO/IEC 14496-3 audio | Carriage in PES packets | Stream\_type = 0x11 | Stream\_id = '110x xxxx' |

If a PTS or DTS is present in the PES packet header it shall refer to the visual object that follows either the first VOP start code or the first still texture object startcode that commences in the PES packet. Each ISO/IEC 14496-2 video stream carried by Rec. ITU-T H.222.0 | ISO/IEC 13818-1 shall contain the information required to decode the ISO/IEC 14496‑2 video stream; consequently the stream shall contain Visual Object Sequence Headers, Visual Object Headers and Video Object Layer Headers.

In the case of an ISO/IEC 14496-3 elementary stream signalled by a stream\_type value of 0x11, before PES packetization the elementary stream data shall be first encapsulated in the LATM transport syntax defined in ISO/IEC 14496-3. In such a case, if a PTS is present in the PES packet header it shall refer to the first audio frame that follows the first syncword that commences in the payload of the PES packet.

In the case of an ISO/IEC 14496-3 elementary stream signalled by a stream\_type value of 0x1C, the first byte of each audio frame shall be the first byte of the payload of a PES packet; prior to PES packetization no encapsulation in any additional transport syntax shall be applied. An audio frame from an ISO/IEC 14496-3 elementary stream signalled by a stream\_type value of 0x1C may be fragmented for carriage in multiple PES packets. In the PES packet header the data\_alignment\_indicator shall be set to '1' in each PES packet that carries a complete such audio frame or the first fragment thereof. The data\_alignment\_indicator shall be set to '0' for PES packets carrying subsequent (non-first) fragments of an audio frame.

Carriage of individual ISO/IEC 14496-2 and ISO/IEC 14496-3 elementary streams in PES packets shall be identified by appropriate stream\_id and stream\_type values, indicating the use of ISO/IEC 14496-2 Visual or 14496-3 Audio. In addition, such carriage shall be signalled by the MPEG-4\_video descriptor or MPEG-4\_audio descriptor, respectively. These descriptors shall be conveyed in the descriptor loop for the respective elementary stream entry in the Program Map Table in case of a transport stream or in the Program Stream Map, when present, in case of a program stream. Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 does not specify presentation of ISO/IEC 14496-2 and ISO/IEC 14496-3 elementary streams in the context of a program.

Carriage of an individual ISO/IEC 14496-17 text streams in PES packets shall be identified by appropriate stream\_id and stream\_type values, indicating the use of ISO/IEC 14496-17 text.

#### 2.11.2.2 STD extensions for individual ISO/IEC 14496 elementary streams

The T-STD model includes a transport buffer TBn and a multiplex buffer Bn prior to decoding of each individual ISO/IEC 14496 elementary stream n. Note that in the T-STD the single multiplex buffer Bn is also applied for ISO/IEC 14496-2 video, as indicated in Figure 2-6, instead of the approach with two buffers MBn and EBn used for ISO/IEC 13818-2 video in the T-STD. For buffers TBn and Bn and the rate Rxn between TBn and Bn the following constraints apply.



Figure 2-6 – T-STD model extensions for individual ISO/IEC 14496 elementary streams

In case of carriage of an ISO/IEC 14496-2 stream:

Size BSn of Buffer Bn:

BSn = BSmux + BSoh + VBVmax[profile,level]

where:

BSoh, packet overhead buffering, is defined as:

BSoh = (1/750) seconds × max{Rmax[profile,level], 2 000 000 bit/s}

and:

BSmux, additional multiplex buffering, is defined as:

BSmux = 0.004 seconds × max{Rmax[profile,level], 2 000 000 bit/s}

Rate Rxn:

Rxn = 1.2 × Rmax[profile,level]

where:

VBVmax[profile,level] and Rmax[profile,level] are defined in ISO/IEC 14496-2 for each profile and level. For profiles and levels for which no VBVmax value is specified, the size of Bn and the rate Rxn are user defined.

For carriage of an ISO/IEC 14496-3 audio stream the following applies.

Size BSn of Buffer Bn, whereby BSn = BSmux + BSdec + BSoh:

For ISO/IEC 14496-3 audio, except for ISO/IEC 14496-3 DST, ALS and SLS:

BSn = 3584 bytes if 1-2 channels

Here, the size of the access unit decoding buffer BSdec, and the PES packet overhead buffer BSoh are constrained by: BSdec + BSoh ≤ 2848 bytes; a portion (736 bytes) of the 3584 byte buffer is allocated for buffering to allow multiplexing. The rest, 2848 bytes, are shared for access unit buffering BSdec, BSoh and additional multiplexing.

BSn = 8976 bytes if 3-8 channels

BSn = 12804 bytes if 9-12 channels

BSn = 51216 bytes if 13-48 channels

For ISO/IEC 14496-3 DST-64, ALS and SLS audio:

if number of audio channels <= 8 then BSn = 1 600 000 bytes,

else BSn = 200 000 × (number of audio channels) bytes.

For ISO/IEC 14496-3 DST-128 audio:

BSn = 400 000 × (number of audio channels) bytes.

For ISO/IEC 14496-3 DST-256 audio:

BSn = 800 000 × (number of audio channels) bytes.

Rate Rxn:

For ISO/IEC 14496-3 audio, except for ISO/IEC 14496-3 DST, ALS and SLS:

Rxn = 2 000 000 bit/s if 1-2 channels

Rxn = 5 529 600 bit/s if 3-8 channels

Rxn = 8 294 400 bit/s if 9-12 channels

Rxn = 33 177 600 bit/s if 13-48 channels

For ISO/IEC 14496-3 DST-64, ALS and SLS audio:

if number of audio channels <= 8 then Rxn = 30 000 000 bit/s,

else Rxn = 120 000 000 bit/s.

For ISO/IEC 14496-3 DST-128 and DST-256 audio:

Rxn = 120 000 000 bit/s.

The P-STD model includes a multiplex buffer Bn prior to decoding of each individual ISO/IEC 14496 elementary stream n. The size BSn of buffer Bn in the P-STD is defined by the P-STD\_buffer\_size field in the PES packet header.

### 2.11.3 Carriage of audiovisual ISO/IEC 14496-1 scenes and associated ISO/IEC 14496 streams

#### 2.11.3.1 Introduction

This clause describes the encapsulation and signalling when an audiovisual scene represented by ISO/IEC 14496 data is carried in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program stream or transport stream. ISO/IEC 14496 content consists of the initial object descriptor and a variable number of streams such as object descriptor streams, scene description streams (carrying either BIFS-Command or BIFS-Anim access units), IPMP streams, OCI streams and audiovisual streams. Each of the ISO/IEC 14496 streams shall be contained in an SL-packetized stream and may optionally be multiplexed into a M4Mux stream, both defined in ISO/IEC 14496-1. For carriage in Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 program stream or transport stream, these SL-packetized streams and M4Mux streams shall contain encoded Object Clock Reference (OCR) and M4Mux Clock Reference (FCR) fields as specified in 2.11.3.4 and in 2.11.3.5, respectively. The SL-packetized streams or M4Mux streams are then encapsulated either in PES packets or in ISO\_IEC\_14496\_sections prior to transport stream packetization and multiplexing or program stream multiplexing. ISO\_IEC\_14496\_sections are built on the long format of H.222.0 | ISO/IEC 13818-1 sections.

Additionally, an ISO/IEC 14496 audiovisual scene may refer to non SL-Packetized streams carried in an Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 transport stream using a "pid://PID\_NUMBER" URL scheme instead of a "od://OD\_ID" URL scheme.

ISO/IEC 14496 streams may derive their time base from the PCR of the program through the OCR\_ES\_ID mechanism.

#### 2.11.3.2 Assignment of ES\_ID values

An ISO/IEC 14496-1 scene carried over a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream may associate a number of ISO/IEC 14496, ISO/IEC 13818 and other streams by the use of the ES\_ID parameter. The scene and the associated streams may be carried over the same Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, but a scene may also reference streams carried elsewhere, for example over an IP network. How to identify such other means is not defined in this Specification.

ISO/IEC 14496-1 defines name scoping rules for identifiers. These rules allow the same ES\_ID value to be used for two different streams within ISO/IEC 14496 content. When one or multiple ISO/IEC 14496-1 scenes are carried in a Rec. ITU‑T H.222.0 | ISO/IEC 13818‑1 Program, duplicate ES\_ID values shall not occur within the program such that each ISO/IEC 14496 SL-packetized stream or ISO/IEC 14496-1 M4Mux channel has a unique ES\_ID value in the program.

#### 2.11.3.3 Timing of ISO/IEC 14496 scenes and associated streams

When carried over a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, the object time base of each ISO/IEC 14496 stream shall be locked to the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 STC, that is:

If X(t) = fstc(t)/fobject(t)

then the value of X(t) shall be constant at any time t.

where:

fstc(t) denotes the intended frequency of the STC at time t, i.e., 27 000 000 Hz

fobject(t) denotes the frequency of the object time base at time t

The object time base of ISO/IEC 14496 streams carried over a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream is conveyed as follows:

– The object time base of an SL-packetized stream carried in PES packets without the use of the M4Mux shall be conveyed by coded OCRs in the SL packet header of that stream. See 2.11.3.4.

– The object time base of SL-packetized streams carried in PES packets within a M4Mux stream shall be conveyed by FCRs in that M4Mux stream. See 2.11.3.5. Consequently, all ISO/IEC 14496 streams contained within the same M4Mux stream share the same object time base.

– The object time base of an SL-packetized stream carried in sections shall be conveyed by another ISO/IEC 14496 stream within the transport stream or program stream as indicated by the OCR\_ES\_ID field in the ES descriptor for that stream.

– The object time base of an SL-packetized stream whose OCR\_ES\_ID identifies a non SL-packetized stream with a PID equal to the PCR PID is fstc(t) / 300

The following constraints shall apply for encoding of OCRs and FCRs in SL-packetized streams and M4Mux streams carried over a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream:

– The OCRs and FCRs in each SL-packetized stream and each M4Mux stream associated with the same scene shall have the same resolution.

– The resolution of OCRs and FCRs for a scene, fcr, shall have a value smaller than or equal to 90 000 Hz.

– The ratio (fstc(t)/300)/fcr, shall be an integer value larger than or equal to one. Consequently the resolution of the OCR and FCR syntax elements may only take values such as 90 000 Hz, 45 000 Hz, 30 000 Hz, 22 500 Hz, 18 000 Hz, etc.

Within the above constraints and the ISO/IEC 14496-1 constraint that the resolution fcr shall represent an integer number of cycles per second, fcr can be selected as appropriate for the scene.

The ISO/IEC 14496 time stamps coded in the SL packet header shall refer to instants of the object time base of the stream carried in the SL packet. The resolution of each such time stamp shall be of a factor 2k smaller than the resolution of the OCRs or FCRs associated with the stream, with k a positive integer larger than or equal to zero. To achieve the same wrap around, the length of the time stamp fields, TimeStampLength, shall be k bit smaller than the length of the OCR or FCR field, OCRLength and FCRLength, respectively. Hence for each stream the following conditions shall apply for encoding of time stamps:

– TimeStampResolution = (OCRResolution or FCRResolution respectively)/2k, with k a positive integer larger than or equal to zero. ISO/IEC 14496-1 requires TimeStampResolution to represent an integer number of cycles per second.

– TimeStampLength = OCRLength or FCRLength respectively – k.

For SL-packetized streams inheriting their object time base from the PCR PID, the following considerations apply:

– TimeStampResolution = 90000 / 2k, with k a positive integer larger than or equal to zero.

– TimeStampLength = 33–k.

For SL-packetized streams carrying an OCR, the relationship between a value of the STC and the corresponding value of the object time base of a stream is established by associating PTS fields in PES packet headers with the OCR or FCR in SL packet headers and M4Mux Stream packets respectively, as specified in 2.11.3.6 and 2.11.3.7.

For SL-packetized streams inheriting their time base from the PCR, the object time base of such a stream is fstc(t) / 300.

#### 2.11.3.4 Delivery timing of SL-packetized streams

To carry ISO/IEC 14496 content in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, ISO/IEC 14496-1 SL-packetized streams are used. In each SL-packetized stream carried in a PES packet without the use of M4Mux, the objectClockReference field shall be encoded as follows:

1) An objectClockReference (OCR) field shall be present in the first SL packet header of a SL-packetized stream.

2) The SL-packetized stream shall be constructed such that the time interval between the bytes containing the last bit of successive OCR fields shall be less than or equal to 0.7 s. Thus:



for all i′ and i″ where i' and i″ are the indexes of the bytes containing the last bit of consecutive OCR fields in the SL-packetized stream.

If an objectClockReference is encoded in an SL packet header, also the instantBitrate field shall be coded.

#### 2.11.3.5 Delivery timing of M4Mux streams

Next to SL-packetized streams also the ISO/IEC 14496-1 M4Mux tool may be used to carry ISO/IEC 14496 content in Rec. ITU-T H.222.0 | ISO/IEC 13818‑1 streams. The payload of M4Mux packets shall consist of SL packets as specified in ISO/IEC 14496‑1. In each M4Mux stream carried in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream the fmxClockReference field shall be encoded as follows:

1) An fmxClockReference (FCR) field shall be present in the first M4Mux packet of a M4Mux stream.

2) The M4Mux stream shall be constructed such that the time interval between the bytes containing the last bit of successive FCR fields shall be less than or equal to 0.7 s. Thus:



for all i′ and i″ where i′ and i″ are the indexes of the bytes containing the last bit of consecutive FCR fields in the M4Mux stream.

3) All ISO/IEC 14496 time stamps within the SL-packetized streams carried within a M4Mux stream shall refer to instants of the object time base conveyed by the FCR fields in the M4Mux stream. The SL‑packetized streams carried in M4Mux packets need not carry OCR fields. If OCR fields are present, they may be ignored.

#### 2.11.3.6 Carriage of SL-packetized streams in PES packets

A single ISO/IEC 14496-1 SL-packetized stream may be mapped into a single PES stream. One and only one SL packet from an SL-packetized stream shall constitute the payload of one PES packet. PES packets that carry an SL-packetized stream shall be identified by stream\_id = 0xFA in the PES packet header.

When an OCR field is coded in the SL packet header, a PTS shall be encoded in the header of the PES packet that carries such SL packet header. This PTS shall be encoded with the 33-bit value of the 90-kHz portion of the STC that corresponds to the value of the object time base at the instant in time indicated by the OCR.

The ES\_ID associated with the SL-packetized stream shall be signalled by an SL descriptor as specified in 2.6.46.

#### 2.11.3.7 Carriage of M4Mux streams in PES packets

PES packets with a payload consisting of M4Mux packets shall be identified by stream\_id = 0xFB in the PES packet header. An integer number of M4Mux packets shall constitute the payload of one PES packet, i.e., the payload of a PES packet carrying a M4Mux stream shall start with a M4Mux packet header and shall end with the last byte of a M4Mux packet.

If an fmxClockReference (FCR) field is encoded in one of the M4Mux packets contained in a PES packet, then a PTS shall be encoded in the header of the PES packet that contains such M4Mux packet. This PTS shall be encoded with the 33-bit value of the 90-kHz portion of the STC that corresponds to the value of the object time base of the M4Mux stream at the instant in time indicated by the FCR. In case multiple M4Mux packets with an encoded FCR field are contained in a PES packet, the PTS shall correspond to the time indicated by the FCR in the first such M4Mux packet encountered in the payload of the PES packet.

The ES\_IDs associated with each SL-packetized stream conveyed in the M4Mux stream shall be signalled by an FMC descriptor as specified in 2.6.44.

#### 2.11.3.8 Carriage of SL packets and M4Mux packets in sections

For transport of ISO/IEC 14496 content in sections, ISO\_IEC\_14496\_sections are defined. Any ISO/IEC 14496 stream may be carried over ISO\_IEC\_14496\_sections. A single ISO\_IEC\_14496\_section shall contain either an entire SL packet of an SL-packetized stream or an integer number of M4Mux packets each carrying an SL packet of the same ISO/IEC 14496-1 elementary stream.

Table 2-153 shows the syntax of ISO\_IEC\_14496\_sections defined to convey ISO/IEC 14496-1 elementary streams, qualified by the table\_id as either object descriptor stream data, scene description stream data or any other ISO/IEC 14496 stream data. Object descriptor stream data consists of an Object Descriptor Table that comprises a number of object descriptors. The Object Descriptor Table may be transmitted in multiple ISO\_IEC\_14496\_sections. Scene description data consists of a Scene Description Table that may comprise a number of BIFS commands. The Scene Description Table may be transmitted in multiple ISO\_IEC\_14496\_sections. It is not required that a complete table be received in order to process its payload. However, the payload of sections shall be processed in the correct order, as indicated by the value of the section\_number field in the ISO\_IEC\_14496\_section header bytes. Other ISO/IEC 14496 stream data consists of an ISO/IEC 14496 table. The ISO/IEC 14496 table may be transmitted in multiple ISO\_IEC\_14496\_sections.

Table 2-153 – Section syntax for transport of ISO/IEC 14496 stream

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| ISO\_IEC\_14496\_section() { |  |  |
| **table\_id** | **8** | **uimsbf** |
| **section\_syntax\_indicator** | **1** | **bslbf** |
| **private\_indicator** | **1** | **bslbf** |
| **reserved** | **2** | **bslbf** |
| **ISO\_IEC\_14496\_section\_length** | **12** | **uimsbf** |
| **table\_id\_extension** | **16** | **uimsbf** |
| **reserved** | **2** | **bslbf** |
| **version\_number** | **5** | **uimsbf** |
| **current\_next\_indicator** | **1** | **bslbf** |
| **section\_number** | **8** | **uimsbf** |
| **last\_section\_number** | **8** | **uimsbf** |
| if (PMT\_has\_SL\_descriptor(current\_PID)) { |  |  |
| SL\_Packet() |  |  |
| } |  |  |
| else if (PMT\_has\_FMC\_descriptor(current\_PID)) { |  |  |
| for (i = 1; i < N1; i++) |  |  |
| M4MuxPacket() |  |  |
| } |  |  |
| else { |  |  |
| for (i = 1; i < N2; i++) |  |  |
| **reserved** | **8** | **bslbf** |
| } |  |  |
| **CRC\_32** | **32** | **rpchof** |
| } |  |  |

**table\_id** – This 8-bit field shall be set to 0x04, 0x05, or 0x08, in case of an ISO\_IEC\_14496\_section. A value of 0x04 indicates an ISO\_IEC\_14496\_section that carries an ISO/IEC 14496-1 scene description stream. A value of 0x05 indicates an ISO\_IEC\_14496\_section that carries an ISO/IEC 14496-1 object descriptor stream. A value of 0x08 indicates an ISO\_IEC\_14496\_section that carries other ISO/IEC 14496 streams.

**section\_syntax\_indicator** – This 1-bit field shall be set to '1'.

**private\_indicator** – This 1-bit field shall not be specified by this Specification.

**ISO\_IEC\_14496\_section\_length** – This 12-bit field shall specify the number of remaining bytes in the section immediately following the ISO\_IEC\_14496\_section\_length field up to the end of the ISO\_IEC\_14496\_section. The value of this field shall not exceed 4093 (0xFFD).

**table\_id\_extension** – This 16-bit field shall not be specified by this Specification; its use and value are defined by the user.

**version\_number** – This 5-bit field shall represent the version number of the Object Descriptor Table or Scene Description Table respectively. The version number shall be incremented by 1 modulo 32 with each new version of the table. Version control is at the discretion of the application.

**current\_next\_indicator** – This 1-bit field shall be set to 1.

**section\_number** – This 8-bit field shall represent the number of the ISO\_IEC\_14496\_section. The section\_number field of the first ISO\_IEC\_14496\_section of the Object Descriptor Table or the Scene Description Table shall have a value equal to 0x00. The value of section\_number shall be incremented by 1 with each additional section in the table.

**last\_section\_number** – This 8-bit field shall specify the number of the last section of the Object Descriptor Table or Scene Description Table of which this section is a part.

**PMT\_has\_SL\_descriptor(current\_PID**) – A pseudo function that shall be true if an SL descriptor is contained in the descriptor loop in the Program Map Table for the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program element that conveys this ISO\_IEC\_14496\_section.

**SL\_Packet()** – A sync layer packet as specified in 10.2.2 of ISO/IEC 14496-1.

**PMT\_has\_FMC\_descriptor(current\_PID**) – A pseudo function that shall be true if an FMC descriptor is contained in the descriptor loop in the Program Map Table for the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program element that conveys this ISO\_IEC\_14496\_section.

**M4MuxPacket()** – A M4Mux packet as specified in 7.4.2.4 of ISO/IEC 14496-1.

**CRC\_32** – This 32-bit field shall contain the CRC value that gives a zero output of the registers in the decoder defined in Annex A after processing the entire ISO\_IEC\_14496\_section.

#### 2.11.3.9 T-STD extensions

##### 2.11.3.9.1 T-STD Model for 14496 content

Figure 2-7 shows extensions of the Transport System Target Decoder for delivery of ISO/IEC 14496 program elements encapsulated in Rec. ITU-T H.222.0 | ISO/IEC 13818-1 transport streams.



Figure 2-7 – T-STD model for ISO/IEC 14496 content

The following notation is used in Figure 2-7 and its description:

TBn is the transport buffer.

MBn is the multiplex buffer for M4Mux stream n or for SL-packetized stream n.

FBnp is the M4Mux buffer for the elementary stream in M4Mux channel p of M4Mux stream n.

DBnp is the decoder buffer for the elementary stream in M4Mux channel p of M4Mux stream n.

DBn is the decoder buffer for elementary stream n.

Dnp is the decoder for the elementary stream in M4Mux channel p of M4Mux stream n.

Dn is the decoder for elementary stream n.

Rxn is the rate at which data are removed from TBn.

Rbxn is the rate at which data are removed from MBn.

Anp(j) is the jth access unit in elementary stream in M4Mux channel p of M4Mux stream n. Anp(j) is indexed in decoding order.

An(j) is the jth access unit in elementary stream n. An(j) is indexed in decoding order.

Tdnp(j) is the decoding time, measured in seconds, in the system target decoder of the jth access unit in elementary stream in M4Mux channel p of M4Mux stream n.

Tdn(j) is the decoding time, measured in seconds, in the system target decoder of the jth access unit in elementary stream n.

Cnp(k) is the kth composition unit in elementary stream in M4Mux channel p of M4Mux stream n. Cnp(k) results from decoding Anp(j). Cnp(k) is indexed in composition order.

Cn(k) is the kth composition unit in elementary stream n. Cn(k) results from decoding An(j). Cn(k) is indexed in composition order.

tcnp(k) is the composition time, measured in seconds, in the system target decoder of the kth composition unit in elementary stream in M4Mux channel p of M4Mux stream n.

tcn(k) is the composition time, measured in seconds, in the system target decoder of the kth composition unit in elementary stream n.

t(i) indicates the time in seconds at which the ith byte of the transport stream enters the system target decoder.

##### 2.11.3.9.2 Processing of M4Mux streams

Complete transport stream packets containing data from M4Mux stream n are passed to the transport buffer for M4Mux stream n, TBn. The size of TBn is fixed at 512 bytes. All bytes that enter TBn are removed from TBn at a rate Rxn, specified by the TB\_leak\_rate field in the MultiplexBuffer descriptor associated with M4Mux stream n. When there is no data in buffer TBn, rate Rxn is equal to zero. Duplicate transport stream packets are not delivered to MBn.

In case of carriage in PES packets, the PES packet header and payload data bytes are delivered to buffer MBn; all other bytes leaving TBn do not enter MBn, and may be used to control the system. In case of carriage in ISO\_IEC\_14496\_sections, the section header, payload and CRC-32 data bytes are delivered to buffer MBn; all other bytes do not enter MBn and may be used to control the system. In either case, the size of MBn shall be specified by the MB\_buffer\_size field in the MultiplexBuffer descriptor.

The M4Mux Stream packet bytes in buffer MBn are all delivered to their associated M4Mux buffer at the rate specified by the field fmxRate encoded in the M4Mux stream and in compliance with the M4Mux buffer model defined in 11.2.9 of ISO/IEC 14496-1. Only M4Mux packet payload data bytes in M4Mux channel p of M4Mux stream n enter buffer FBnp. M4Mux packet header bytes in M4Mux channel p of M4Mux stream n are discarded and may be used to control the system. The rate specified by the fmxRate field shall be applicable for all M4Mux packets in the stream immediately following the M4Mux Clock Reference channel packet up to the next encountered M4Mux Clock Reference channel packet. When there is no M4Mux stream data present in MBn, no data is removed from MBn. Bytes from the PES packet header or from the ISO\_IEC\_14496\_section header that immediately precede a M4Mux header are instantaneously removed and discarded and may be used to control the system. Bytes from the ISO\_IEC\_14496\_section CRC-32 fields that immediately follow the last M4Mux Stream packet in the section payload are removed instantaneously and discarded and may be used to verify the integrity of the data. Bytes from the M4Mux Clock Reference channel are instantaneously removed and discarded and may be used to lock the ISO/IEC 14496 object time base to the STC. When there is no PES packet or section payload data bytes, respectively present in MBn, no data is removed from MBn. All data that enters MBn leaves it. All PES packet payload bytes of stream n enter the M4Mux demultiplexer instantaneously upon leaving MBn.

##### 2.11.3.9.3 Definition of M4Mux Buffer, FBnp

For each channel p of a M4Mux stream n, the size of M4Mux buffer FBnp is defined using the M4MuxBufferSize descriptor. M4Mux packet payload bytes are transferred from buffer FBnp to decoder buffer DBnp in compliance with the M4Mux buffer model defined in 11.2.9 of ISO/IEC 14496-1. Only SL packet payload bytes in M4Mux channel p of M4Mux stream n enter buffer DBnp. The SL packet header bytes in M4Mux channel p of M4Mux stream n are discarded and may be used to control the system.

##### 2.11.3.9.4 Processing of SL-packetized streams

Complete transport stream packets containing data from SL-packetized stream n are passed to the transport buffer for SL-packetized stream n, TBn. All bytes that enter TBn are removed at a rate Rxn, specified by the TB\_leak\_rate field in the MultiplexBuffer descriptor. When there is no data in buffer TBn, rate Rxn is equal to zero. Duplicate transport stream packets are not delivered to MBn.

In case of carriage in PES packets, the PES packet header and payload data bytes are delivered to buffer MBn; all other bytes leaving TBn do not enter MBn, and may be used to control the system. In case of carriage in ISO\_IEC\_14496\_sections, the section header, payload and CRC-32 data bytes are delivered to buffer MBn; all other bytes do not enter MBn and may be used to control the system. In either case the size of MBn is specified by the MB\_buffer\_size field in the MultiplexBuffer descriptor.

The SL-packetized stream bytes in buffer MBn are all delivered to the decoder buffer DBn at the rate specified by the field instantBitRate encoded in the SL-packetized stream and in compliance with the System Decoder Model defined in 7.4 of ISO/IEC 14496-1. The rate specified by the instantBitRate field shall be applicable for all data bytes in the SL‑packetized stream immediately following the instantBitRate field in the SL packet header up to the next encountered instantBitRate field. If there are no SL-packetized stream bytes in MBn, no bytes are removed from MBn. Bytes from the PES packet header or from the ISO\_IEC\_14496\_section header that immediately precede a SL packet header are instantaneously removed and discarded and may be used to control the system. Bytes from the ISO\_IEC\_14496\_section CRC-32 fields that immediately follow the last SL packet payload byte in the section are removed instantaneously and discarded and may be used to verify the integrity of the data. When there are no PES packet or section payload data bytes, respectively present in MBn, no data is removed from MBn. All data that enters MBn leaves it. All PES packet payload bytes of stream n enter buffer DBn instantaneously upon leaving MBn, with the exception of the SL packet headers. Bytes from the SL packet headers do not enter DBn and may be used to control the system. The size of decoder buffer DBn is given by the bufferSizeDB of the DecoderConfigDescriptor defined in ISO/IEC 14496-1.

##### 2.11.3.9.5 Buffer management

Transport streams shall be constructed so that conditions defined in this subclause are satisfied.

TBn shall not overflow and shall be empty at least once every second. MBn shall not overflow. FBnp shall not overflow. DBnp and DBn shall neither underflow nor overflow. Underflow of DBnp occurs when one or more bytes of an access unit are not present in DBnp at the decoding time associated with this access unit. Underflow of DBn occurs when one or more bytes of an access unit are not present in DBn at the decoding time associated with this access unit.

#### 2.11.3.10 Carriage within a transport stream

###### 2.11.3.10.1 Overview

A transport stream may contain one or more programs, each described by a Program Map Table. ISO/IEC 14496 content can be conveyed in addition to the already defined stream types for such a program. Elements of the ISO/IEC 14496 content may be conveyed in one or more Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program elements referenced by a unique PID value within a transport stream. As a special case, it is possible that a program within a transport stream consists only of ISO/IEC 14496 program elements. ISO/IEC 14496 content associated with a program and carried in the transport stream shall be referenced in the Program Map Table of that program. An initial object descriptor shall be used to define an ISO/IEC 14496-1 scene; the use of this descriptor is specified in 2.11.3.10.2.

Carriage of ISO/IEC 14496 content in a PID is signalled by a stream\_type value of 0x12 or 0x13 in the Program Map Table in association with that PID value. A value of 0x12 indicates carriage in PES packets. The stream\_id field in the PES packet header signals whether the PES packet contains a single SL packet or a number of M4Mux packets. A stream\_type value of 0x13 in the Program Map Table indicates that the program element carries an object descriptor stream or a BIFS-Command stream contained in sections. In this case the table\_id in the section header indicates whether an object descriptor stream is carried in the sections or a BIFS-Command stream. See also Table 2-154. The section contains either a single SL packet or a number of M4Mux packets, as indicated by the presence of an SL descriptor or a FMC descriptor respectively in the descriptor loop of the Program Map Table for the Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 program element that carries the sections. When ISO/IEC 14496 content is carried, the SL descriptor and the FMC descriptor shall specify the ES\_ID for each encapsulated ISO/IEC 14496 stream. When the assignment of ES\_ID values changes, the Program Map Table shall be updated and the version\_number of the PMT shall be incremented by 1 modulo 32. An example of a content access procedure for ISO/IEC 14496 program components within a transport stream is given in Annex R.

| Table 2-154 – ISO/IEC defined options for carriage of an ISO/IEC 14496 scene and  associated streams in Rec. ITU-T H.222.0 | ISO/IEC 13818-1 | | | | |
| --- | --- | --- | --- | --- |
| ISO/IEC 14496-1 object descriptor streams | Encapsulation in SL packets | Carriage in PES packets | Stream\_type = 0x12 | Stream\_id = '1111 1010' |
| Carriage in ISO\_IEC\_ 14496\_sections | Stream\_type = 0x13 | Table\_id = 0x05 |
| Encapsulation in SL packets followed by Multiplex into M4Mux packets | Carriage in PES packets | Stream\_type = 0x12 | Stream\_id = '1111 1011' |
| Carriage in ISO\_IEC\_ 14496\_sections | Stream\_type = 0x13 | Table\_id = 0x05 |
| ISO/IEC 14496-1 scene description streams | Encapsulation in SL packets | Carriage in PES packets | Stream\_type = 0x12 | Stream\_id = '1111 1010' |
| Carriage in ISO\_IEC\_ 14496\_sections | Stream\_type = 0x13 | Table\_id = 0x04 |
| Encapsulation in SL packets followed by Multiplex into M4Mux packets | Carriage in PES packets | Stream\_type = 0x12 | Stream\_id = '1111 1011' |
| Carriage in ISO\_IEC\_ 14496\_sections | Stream\_type = 0x13 | Table\_id = 0x04 |
| All other ISO/IEC 14496 streams | Encapsulation in SL packets | Carriage in PES packets | Stream\_type = 0x12 | Stream\_id = '1111 1010' |
|  | Encapsulation in SL packets followed by Multiplex into M4Mux packets | Carriage in PES packets | Stream\_type = 0x12 | Stream\_id = '1111 1011' |

###### 2.11.3.10.2 Initial Object Descriptor

In case of carriage of an ISO/IEC 14496-1 scene, the ISO/IEC 14496-1 initial object descriptor serves as the initial access point to all associated streams. The initial object descriptor shall be conveyed in the IOD descriptor located in the descriptor loop immediately following the program\_info\_length field in the Program Map Table of the program to which the scene is associated. It contains ES\_Descriptors identifying the scene description and object descriptor streams that form part of this program. It may also contain ES\_Descriptors identifying one or more associated IPMP or OCI streams. Identification of streams is done by means of ES\_IDs as specified in clauses 7.2.1 and 7.2.6.5 of ISO/IEC 14496-1.

#### 2.11.3.11 P-STD Model for 14496 content

Figure 2-8 shows the STD model when ISO/IEC 14496 systems data are carried in a program stream.



Figure 2-8 – P-STD model for ISO/IEC 14496 Systems stream

The following notation is used in Figure 2-8 and its description:

Bn is the input buffer for M4Mux stream n or for SL-packetized stream n.

FBnp is the M4Mux buffer for the elementary stream in M4Mux channel p of M4Mux stream n.

DBnp is the decoder buffer for the elementary stream in M4Mux channel p of M4Mux stream n.

DBn is the decoder buffer for elementary stream n.

Dnp is the decoder for elementary stream in M4Mux channel p of M4Mux stream n.

Dn is the decoder for elementary stream n.

Anp(j) is the jth access unit in elementary stream in M4Mux channel p of M4Mux stream n. Anp(j) is indexed in decoding order.

An(j) is the jth access unit in elementary stream n. An(j) is indexed in decoding order.

Tdnp(j) is the decoding time, measured in seconds, in the system target decoder of the jth access unit in elementary stream in M4Mux channel p of M4Mux stream n.

Tdn(j) is the decoding time, measured in seconds, in the system target decoder of the jth access unit in elementary stream n.

Cnp(k) is the kth composition unit in elementary stream in M4Mux channel p of M4Mux stream n. Cnp(k) results from decoding Anp(j). Cnp(k) is indexed in composition order.

Cn(k) is the kth composition unit in elementary stream n. Cn(k) results from decoding An(j). Cn(k) is indexed in composition order.

tcnp(k) is the composition time, measured in seconds, in the system target decoder of the kth composition unit in elementary stream in M4Mux channel p of M4Mux stream n.

tcn(k) is the composition time, measured in seconds, in the system target decoder of the kth composition unit in elementary stream n.

t(i) indicates the time in seconds at which the ith byte of the program stream enters the system target decoder.

###### 2.11.3.11.1 Processing of M4Mux streams

At the input of the STD each byte in the payload of PES packets carrying a M4Mux stream n is transferred instantaneously to buffer Bn. The i-th byte enters Bn at time t(i). PES packet header bytes do not enter buffer Bn and may be used to control the system. The size of Bn is specified by the P-STD\_buffer\_size field in the header of the PES packet that carries stream n.

The M4Mux stream packet bytes in buffer Bn are all delivered to their associated M4Mux buffer at the rate specified by the field fmxRate encoded in the M4Mux stream and in compliance with the M4Mux buffer model defined in 7.4.2.9 of ISO/IEC 14496-1. Only M4Mux packet payload data bytes in M4Mux channel p of M4Mux stream n enter buffer FBnp. M4Mux packet header bytes in M4Mux channel p of M4Mux stream n are discarded and may be used to control the system. The rate specified by the fmxRate field shall be applicable for all M4Mux packets in the stream up to the next encountered M4Mux Clock Reference channel packet. Bytes from the M4Mux Clock Reference channel are instantaneously removed and discarded and may be used to lock the ISO/IEC 14496 object time base to the STC. When there is no PES packet payload data present in Bn, no data is removed from Bn. All data that enters Bn leaves it. All PES packet payload bytes of stream n enter the M4Mux demultiplexer instantaneously upon leaving Bn.

###### 2.11.3.11.2 Definition of M4Mux Buffer, FBnp

For each channel p of a M4Mux stream n, the size of M4Mux buffer FBnp is defined using the M4MuxBufferSize descriptor if a Program Stream Map is present in the program stream. M4Mux packet payload bytes are transferred from buffer FBnp to decoder buffer DBnp in compliance with the M4Mux buffer model defined in 7.4.2.9 of ISO/IEC 14496-1. Only SL packet payload bytes in M4Mux channel p of M4Mux stream n enter buffer DBnp. The SL packet header bytes in M4Mux channel p of M4Mux stream n are discarded and may be used to control the system

###### 2.11.3.11.3 Processing of SL-packetized streams

At the input of the STD each byte in the payload of PES packets carrying an SL-packetized stream n is transferred instantaneously to buffer Bn. The i-th byte enters Bn at time t(i). PES packet header bytes do not enter buffer Bn and may be used to control the system. The size of Bn is specified by the P-STD\_buffer\_size field in the header of the PES packet that carries stream n. The SL-packetized stream bytes in buffer Bn are delivered to the decoder buffer DBn at the rate specified by the field instantBitRate encoded in the SL-packetized stream and in compliance with the System Decoder Model defined in 7.1.4.1 of ISO/IEC 14496-1. The rate specified by the instantBitRate field shall be applicable for all data bytes in the SL-packetized stream up to the next encountered instantBitRate field. When there is no PES packet payload data present in Bn, no data is removed from Bn. All data that enters Bn leaves it. All bytes of stream n enter buffer DBn instantaneously upon leaving Bn, with the exception of the SL packet headers. Bytes from the SL packet headers do not enter DBn and may be used to control the system. The size of decoder buffer DBn is given by the bufferSizeDB of the DecoderConfigDescriptor defined in ISO/IEC 14496-1.

###### 2.11.3.11.4 Buffer management

Program streams shall be constructed so that Bn does not overflow. FBnp shall not overflow. DBnp and DBn shall neither underflow nor overflow. Underflow of DBnp occurs when one or more bytes of an access unit are not present in DBnp at the decoding time associated with this access unit. Underflow of DBn occurs when one or more bytes of an access unit are not present in DBn at the decoding time associated with this access unit.

#### 2.11.3.12 Carriage within a program stream

###### 2.11.3.12.1 Overview

A program stream contains only one program. ISO/IEC 14496 data can be conveyed in addition to the already defined stream types for such a program. As a special case, it is also possible that a program stream carries only ISO/IEC 14496 data. If a Program Stream Map is present, ISO/IEC 14496 content carried in the program stream shall be referenced as follows. Carriage of ISO/IEC 14496-1 scenes and associated ISO/IEC 14496 streams in SL and M4Mux packets is indicated by the appropriate stream\_id and by an initial object descriptor; the use of this descriptor is specified in 2.11.3.12.2. For each carried ISO/IEC 14496 stream the SL descriptor and the FMC descriptor shall specify the ES\_ID. When the assignment of ES\_ID values changes, the Program Stream Map, if present, shall be updated and the program\_stream\_map\_version shall be incremented by 1 modulo 32. Note that in a Program Stream the ISO/IEC 14496 content may also be referenced by private means.

For an example of a content access procedure for ISO/IEC 14496 program components within a program stream, see Annex R.

###### 2.11.3.12.2 Initial object descriptor

In case of carriage of an ISO/IEC 14496-1 scene, the ISO/IEC 14496 initial object descriptor serves as the initial access point to all associated streams. If a Program Stream Map is present in the program stream, the initial object descriptor shall be conveyed in the IOD descriptor that is located in the descriptor loop immediately following the program\_stream\_info\_length field. It contains ES\_Descriptors identifying the scene description and object descriptor streams of the scene that form part of this program. It may also contain ES\_Descriptors identifying one or more associated IPMP or OCI streams. Identification of streams is done by means of ES\_IDs as specified in clause 7.2.1 and 7.2.6.5 of ISO/IEC 14496‑1. In a program stream, the initial object descriptor may also be conveyed by private means.

## 2.12 Carriage of metadata

### 2.12.1 Introduction

A Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream can carry metadata. The format of the metadata may be defined by ISO or by any other authority. This subclause defines how to carry the metadata; transport mechanisms are defined as well as metadata related-signalling, the applied metadata timing model and extensions of the STD model for decoding of metadata.

A metadata service is defined to be a coherent set of metadata of the same format delivered to a receiver for a specific purpose. Metadata services are contained in metadata streams; each metadata stream carries one or more metadata services. This Specification assumes the notion of metadata Access Units within a metadata service. The definition of a Metadata Access Unit is metadata format specific, but each metadata service is assumed to represent a concatenation (or a collection) of metadata Access Units.

When transporting a metadata service over a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, a unique metadata service id is assigned to each such service. A metadata service id references uniquely a metadata service among all the metadata services available on the same transport or program stream, and *not* unique *solely* within a metadata stream. The metadata service identifier is used to retrieve the metadata service and all the information needed to decode it.

Decoding of metadata may require the availability of decoder configuration data. If a metadata service carried in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream requires decoder configuration data for decoding, then this metadata decoder configuration data shall be carried within the same program of the same Rec. ITU-T H.222.0 | ISO/IEC 13818‑1 stream.

Clause 2.12.2 discusses metadata timing, while 2.12.3 provides an overview of tools that are defined for transport of metadata over a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream. The use of available transport tools is specified in 2.12.4 up to 2.12.8, and 2.12.9 specifies metadata related signalling. Finally, the STD model for metadata decoding is specified in 2.12.10.

Since many forms of metadata may be carried, it is essential to signal both the precise format and encoding of the metadata, and the semantic meaning the metadata conveys. The former is signalled by the metadata format, while the latter is signalled by the metadata application format. In other words, the metadata format conveys how the metadata shall be decoded, while the metadata application format conveys how to use the metadata, essentially which application uses the metadata. This division is important since it separates the encoding or representation of the metadata from its meaning, thereby allowing an application to be agnostic of the means by which its metadata is conveyed.

### 2.12.2 Metadata time-line model

Metadata may refer to time codes associated with the content, for example to indicate the beginning of a content segment. Each time indication made in the metadata refers to a certain metadata content time line specific to the actual metadata format and/or metadata application format. For example, one metadata (application) format may use UTC, while another metadata application format may use SMPTE time codes. To allow for transport of the content at any time over any media, the metadata content time line is expected but not required to be transport agnostic.

When transporting content and the associated metadata over Rec. ITU-T H.222.0 | ISO/IEC 13818-1 streams, accurate time references from the metadata to the content are to be maintained. The same is needed if the metadata is delivered over other means. To achieve this, the time line model of Figure 2-9 is assumed in this Specification.



Figure 2-9 – Timing model for delivery of content and metadata

Metadata is associated with the audiovisual content, usually in a transport agnostic way, at production or any other stage prior to transport. Where needed, time information is embedded in the metadata to indicate for example specific segments within the content, using the metadata content time line used in the metadata. For example UTC or SMPTE time codes may be used. The time line format is independent of any time code that may or may not be embedded in the audiovisual stream itself. For example, the metadata time line may utilize UTC, while SMPTE time codes are embedded in the video stream.

The following requirements shall be met for each metadata stream:

• no time discontinuities shall occur in the metadata content time line;

• the metadata content time line shall be locked to the sampling clock of the content;

• each time reference in the metadata stream refers to the same metadata content time line.

At transport, a transport-specific timing is associated with the content; this is the delivery time line. In the case of transport over a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, the delivery time line is provided by the System Time Clock, the STC. The content may be delivered as a contiguous piece of information, but it is also possible to interrupt the delivery of the content, for example in the case of news-flash interruptions of a program; in such and other cases time line discontinuities may occur.

When time references are used in the metadata, in the System Target Decoder (STD) these time references are to be associated unambiguously with time values in the received content. To achieve this, a receiver content time line is required. The STC can be used as the receiver content time line, but due to STC discontinuities that may occur, the STC does not necessarily offer an unambiguous time association. Therefore the NPT (Normal Play Time) concept from ISO/IEC 13818-6 DSM-CC is also available for use as the receiver content time line. In any playback mode, such as normal, reverse, slow motion, fast forward, fast backward and still picture, the NPT provides an unambiguous time association, independent of STC discontinuities, and independent of insertions of other content. Note that a new NPT\_reference\_descriptor needs to be transmitted when the STC rolls over.

To maintain the accurate time references from metadata to the content, information is needed how to map a metadata time, MT, defined on the metadata content time line to the corresponding receiver time, RT, of the receiver content time line. This is achieved by providing the offset in time (in 90-kHz units) between the metadata content time line and the receiver content time line. The offset is provided in the content labelling descriptor. The offset conveys the value of the metadata time base at the instant in time at which the receiver content time base reaches a specified value. See also Figure 2-9.

The timing in metadata systems may refer to a specific picture or audio frame, for example using SMPTE time codes. The offset in time between the metadata content time line and the receiver content time line is expressed in units of 90 kHz, and consequently the metadata time reference will translate into a 90-kHz value in receivers. To accommodate for inaccuracies, receivers shall assume that when reference is made to a picture or audio frame the closest match shall be used. For example, the translated 90-kHz metadata time reference shall be matched with the picture or frame whose PTS value is closest to the translated value.

When using NPT, during playback in any mode at any point in time the offset remains constant between the metadata time base and the NPT time base. As long as neither STC discontinuities nor insertions with other content occur, the same is true for the offset in time between the metadata time base and the STC time base, but only in normal playback mode. For privately defined time lines the offset is also required to be constant, but possibly within constraints not defined in this Specification.

When synchronous transport of metadata is applied in PES packets or by using the synchronized DSM-CC download protocol, PTSs are assigned to the metadata. Such PTS may for example indicate the point in time at which the metadata becomes valid. This implies *a priori* knowledge of how to associate the metadata to the delivery timing. However, synchronously transported metadata may also contain time references, which are to be mapped from the metadata content time line to the receiver content time line using the specified offset between both time lines. See also Figure 2-10.



Figure 2-10 – Delivery of metadata in PES packets

### 2.12.3 Options for transport of metadata

To acknowledge the very diverse characteristics of metadata, a variety of tools is defined to transport the metadata over a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream.

This Specification defines two tools for synchronous delivery of the metadata:

• carriage in PES packets;

• use of DSM-CC synchronized download protocol.

In addition, this Specification defines three tools for asynchronous delivery of metadata:

• carriage in metadata sections;

• use of DSM-CC data carousels;

• use of DSM-CC object carousels.

Note that some of the asynchronous transport options support carousels and file structures. The choice of transport tool depends on the requirements that apply to the delivery of the metadata, and the requirements of the tools, as described in the following subclauses.

Metadata may also be carried by private means such as PES packets with stream id value 0xBD or 0xBF (private\_stream\_id\_1 or private\_stream\_id\_2) or private sections. This Specification does not specify how to use private means for carriage of metadata, but allows for signalling of such metadata using the descriptors defined in 2.6.56 up to 2.6.63.

The basic referencing of metadata services is the same for all tools, using the metadata service id. However, there are differences per tool. When PES packets, metadata sections, or synchronized DSM-CC download sections are used, data from each metadata service is explicitly signalled within a metadata stream, using the metadata\_service\_id field. However, when using DSM-CC carousels, this signalling is left at the discretion of metadata applications. Note that this Specification allows for carriage of a metadata service in a DSM-CC carousel, but does not constrain how many metadata services can be carried in one DSM-CC carousel.

Metadata decoder configuration data is signalled explicitly when carried in a metadata descriptor, in PES packets with stream\_type 0x15 and stream\_id 0xFC, in metadata sections or in synchronized DSM-CC download sections. When metadata decoder configuration data is carried in a DSM-CC carousel, the signalling of such data is required, but not defined by this Specification; instead, such signalling is left at the discretion of applications.

### 2.12.4 Use of PES packets to transport metadata

#### 2.12.4.1 General

PES packets provide a mechanism for synchronous transport of metadata. By means of the PTS in the PES packet header the metadata access units are associated with a certain instant of the STC, without the need for time references in the metadata. This implies *a priori* knowledge of how to associate the metadata to the delivery timing. Specific stream\_id and stream\_type values are assigned to signal PES packets carrying metadata; see 2.12.9.

When using PES packets with a stream\_type of 0x15 and a stream\_id of 0xFC to transport the metadata, a Metadata Access Unit Wrapper shall be used as the tool to align PES packets and the metadata Access Units, using metadata\_AU\_cells. This allows random access indication, whose meaning depends on the format of the metadata, and a cell sequence counter to identify loss of metadata\_AU\_cells. Each metadata Access Unit is carried and, if appropriate, fragmented in one or more metadata\_AU\_cells. In each PES packet that carries metadata, the first PES\_packet\_data\_byte shall be the first byte of a Metadata\_AU\_cell. For each metadata Access Unit contained in the same PES packet, the PTS in the PES header applies. The PTS signals the time at which the metadata Access Units are decoded instantaneously and removed from buffer Bn in the STD. Note that the relationship between a decoded metadata Access Unit and audiovisual content is beyond the scope of this Specification.

A PES packet may contain a single metadata\_AU\_cell. This is useful if a metadata Access Unit does not fit into a single PES packet, in which case the fragmentation of the metadata Access Unit is handled by the metadata\_AU\_cell.

When metadata is carried by PES packets in a program stream, and if a Program Stream Map is applied in that program stream, then the Program Stream Map shall specify which PES packets contain the associated metadata.

#### 2.12.4.2 Metadata Access Unit Wrapper

The metadata Access Unit Wrapper (see Table 2-155) shall be used when carrying metadata Access Units in PES packets with a stream\_type of 0x15 and a stream\_id value of 0xFC or in synchronized DSM-CC download sections of stream\_type 0x19. The wrapper defines a structure consisting of a concatenated number of Metadata\_AU\_cells (see Table 2-156). By coding the size of the contained metadata in each metadata\_AU\_cell, metadata agnostic parsing is possible in receivers: the parser can retrieve the metadata and provide it to a metadata decoder without *a priori* knowledge on any detail of the metadata. The Metadata\_AU\_cell shall be aligned with the transport; that is the first byte of the payload of the PES packet or synchronized DSM-CC download section shall be the first byte of a Metadata\_AU\_cell.

If a metadata Access Unit does not fit entirely into a metadata\_AU\_cell, then the metadata Access Unit shall be fragmented into multiple metadata\_AU\_cells, where the fragmentation\_indication in each such metadata\_AU\_cell signals that the metadata\_AU\_cell contains a fragment.

To each Metadata\_AU\_cell that is contained in the same PES packet or synchronized download section, the PTS as coded in the header of the PES packet or synchronized download section, respectively, applies.

Table 2-155 – Metadata Access Unit Wrapper

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| Metadata\_AU\_wrapper () {  for (i = 0; i < N; i++){ Metadata\_AU\_cell ()  } } |  |  |

Table 2-156 – Metadata AU cell

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| Metadata\_AU\_cell () {  **metadata\_service\_id**  **sequence\_number  cell\_fragment\_indication**  **decoder\_config\_flag   random\_access\_indicator**  **reserved  AU\_cell\_data\_length**  for (i = 0; I < AU\_cell\_data\_length; i++){  **AU\_cell\_data\_byte**  } } | **8 8 2 1 1 4 16  8** | **uimsbf uimsbf bslbf bslbf bslbf bslbf uimsbf  bslbf** |

**metadata\_service\_id**: This 8-bit field identifies the metadata service associated with the metadata Access Unit carried in this metadata AU cell.

**sequence\_number**: This 8-bit field specifies the sequence number of the metadata\_AU\_cell. This number increments by one for each successive metadata\_AU\_cell constituting the metadata\_AU\_wrapper, independent of the coded value of the metadata\_service\_id.

**cell\_fragment\_indication**: This 2-bit field conveys information on the metadata Access Unit carried in this metadata\_AU\_cell, corresponding to Table 2-157.

Table 2-157 – Cell fragment indication

|  |  |
| --- | --- |
| Value | Description |
| '11' | A single cell carrying a complete metadata Access Unit. |
| '10' | The first cell from a series of cells with data from one metadata Access Unit. |
| '01' | The last cell from a series of cells with data from one metadata Access Unit. |
| '00' | A cell from a series of cells with data from one metadata Access Unit, but neither the first nor the last one. |

**random\_access\_indicator**: This 1-bit field, when coded with the value '1', indicates that the metadata carried in this metadata\_AU\_cell represents an entry point to the metadata service where decoding is possible without information from previous metadata\_AU\_cells. The meaning of a random access point is defined by the format of the metadata.

**decoder\_config\_flag**: This 1-bit field signals the presence of decoder configuration information in the carried metadata Access Unit. Note that this does not preclude the presence of metadata in the Access Unit next to decoder configuration data.

**AU\_cell\_data\_length**: This 16-bit field specifies the number of AU\_cell\_data\_bytes immediately following.

**AU\_cell\_data\_byte**: This 8-bit field contains contiguous bytes from a metadata Access Unit.

### 2.12.5 Use of the DSM-CC synchronized download protocol to transport metadata

For synchronized transport, in addition to PES packets, the DSM-CC synchronized download protocol can be used. When using synchronized DSM-CC download sections to transport the metadata, the Metadata Access Unit Wrapper defined in 2.12.4.2 shall be used as the tool to encapsulate metadata Access Units. This allows random access indication, whose meaning depends on the format of the metadata, and a cell sequence counter to identify loss of metadata\_AU\_cells. In each DSM-CC synchronized download section that carries metadata, the first byte of the payload shall be the first byte of a Metadata\_AU\_cell. For each metadata Access Unit contained in the same DSM-CC synchronized download section, the PTS in the section header applies. The PTS signals the time at which the metadata Access Units are decoded instantaneously and removed from buffer Bn in the STD. Note that the relationship between a decoded metadata Access Unit and audiovisual content is beyond the scope of this Specification. A specific stream\_type value (as detailed in Table 2-34) is assigned to signal carriage of metadata in DSM-CC synchronized download sections.

### 2.12.6 Use of metadata sections to transport metadata

If asynchronous transport of metadata Access Units without a carousel delivery mechanism is needed, metadata sections can be utilized. The syntax and semantics of metadata sections are defined in this subclause. Each metadata section shall carry either one complete metadata Access Unit or a single part of one metadata Access Unit, as signalled by the section\_fragment\_indication field (see Table 2-158).

For transport in metadata sections, the metadata Access Units are structured in one or more Metadata Tables. Each Metadata Table contains one or more complete metadata Access Units from one or more metadata services. Conceptually, the transport mechanism of Metadata Tables is comparable to the transport mechanism of Program Map Tables and Program Association Tables. Each Metadata Table may be made up of multiple metadata sections. Each Metadata Table may contain metadata from multiple metadata services.

Specific stream\_type and table\_id values are assigned to metadata sections. Metadata decoder configuration data can also be carried in sections, signalled by a metadata description value, as assigned by the metadata decoder configuration descriptor.

Table 2-158 – Section syntax for transport of metadata

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| Metadata\_section() {  **table\_id  section\_syntax\_indicator  private\_indicator  random\_access\_indicator  decoder\_config\_flag  metadata\_section\_length  metadata\_service\_id  reserved  section\_fragment\_indication  version\_number  current\_next\_indicator  section\_number  last\_section\_number**  for (i = 1; i < N; i++){  **metadata\_byte**  }  **CRC\_32** } | **8 1 1 1 1 12 8 8 2 5 1 8 8  8  32** | **uimsbf bslbf bslbf bslbf bslbf uimsbf uimsbf bslbf bslbf uimsbf bslbf uimsbf uimsbf  bslbf  rpchof** |

**table\_id**: The table\_id is an 8-bit field that shall be set to '0x06' for each metadata section.

**section\_syntax\_indicator**: This 1-bit field shall be set to '1'.

**private\_indicator**: This 1-bit field is not specified by this Specification.

**random\_access\_indicator**: This 1-bit field, when coded with the value '1', indicates that the metadata carried in this metadata section represents an access point to the metadata service where decoding is possible without information from previous metadata sections. The meaning of a random access point is defined by the format of the metadata.

**decoder\_config\_flag**: This 1-bit field, when coded with the value '1', indicates that decoder configuration information is present in the metadata Access Unit carried in this metadata section.

**metadata\_section\_length**: This 12-bit field shall specify the number of remaining bytes in the section immediately following the metadata\_section\_length field, and including the CRC. The value of this field shall not exceed 4093 (0xFFD).

**metadata\_service\_id**: This 8-bit field identifies the metadata service associated with the metadata Access Unit carried in this metadata section. Each Metadata Table may contain metadata from multiple metadata services.

**section\_fragment\_indication**: This 2-bit field conveys information on the fragmentation of the metadata Access unit carried in this metadata section, corresponding to Table 2-159.

Table 2-159 – Section fragment indication

|  |  |
| --- | --- |
| Value | Description |
| '11' | A single metadata section carrying a complete metadata Access Unit. |
| '10' | The first metadata section from a series of metadata sections with data from one metadata Access Unit. |
| '01' | The last metadata section from a series of metadata sections with data from one metadata Access Unit. |
| '00' | A metadata section from a series of metadata sections with data from one metadata Access Unit, but neither the first nor the last one. |

**version\_number**: This 5-bit field is the version number of the whole Metadata Table. The version number shall be incremented by 1 modulo 32 whenever the information contained within the Metadata Table changes. When the current\_next\_indicator is set to '1', then the version\_number shall be that of the currently applicable Metadata Table. When the current\_next\_indicator is set to '0', then the version\_number shall be that of the next applicable Metadata Table.

**current\_next\_indicator**: A 1-bit field, which when set to '1' indicates that the Metadata Table sent is currently applicable. When the bit is set to '0', it indicates that the Metadata Table sent is not yet applicable and shall be the next Metadata Table to become valid.

**section\_number**: This 8-bit field gives the number of the metadata section. The section\_number of the first section in a Metadata Table shall be 0x00. The section\_number shall be incremented by 1 with each additional section in this Metadata Table.

**last\_section\_number**: This 8-bit field specifies the number of the last section (that is, the section with the highest section\_number) of the complete Metadata Table of which this section is a part.

**metadata\_byte**: This 8-bit contains contiguous bytes from a metadata Access Unit.

**CRC\_32**: This 32-bit field shall contain the CRC value that gives a zero output of the registers in the decoder defined in Annex A after processing the entire metadata\_section.

### 2.12.7 Use of the DSM-CC data carousel to transport metadata

The DSM-CC tools as defined in ISO/IEC 13818-6 for Data Carousels can be used if a carousel delivery mechanism is required without the need to express the hierarchical organization of the metadata structure in the transport mechanism. Information on the carousel in which the metadata is contained, is included in the metadata descriptor defined in 2.6.60 and 2.6.62. A specific stream\_type value is assigned to signal carriage of metadata in the DSM-CC data carousel. Note that signalling of metadata services within a DSM-CC data carousel is required, but not defined by this Specification.

### 2.12.8 Use of the DSM-CC object carousel to transport metadata

If a carousel delivery mechanism is required with the capability to express the hierarchical organization of the metadata structure in the transport, then the DSM-CC tools and file structures as defined in ISO/IEC 13818-6 for User to User Object Carousels can be used. These file structures provide the tools to structure the metadata as deemed appropriate for efficient parsing of the metadata and for expressing the hierarchical organization of the metadata. Information needed to identify the carousel in which the metadata is contained, is included in the metadata descriptor defined in 2.6.60 and 2.6.61. This may be the IOP:IOR() as defined in 11.3.1 and 5.7.2.3 of ISO/IEC 13818-6 DSM-CC. A specific stream\_type value is assigned to signal carriage of metadata in the DSM-CC object carousel. Note that signalling of metadata services within a DSM-CC object carousel is required, but not defined by this Specification.

### 2.12.9 Metadata-related signalling

#### 2.12.9.1 General

Metadata-related signalling covers four distinct areas:

• signalling of metadata services and streams;

• signalling of content for use by a metadata system;

• association of metadata to content; and

• signalling of decoder configuration data.

#### 2.12.9.2 Signalling of metadata services and streams

Carriage of metadata is signalled by a stream\_type value in the inclusive range from 0x15 to 0x19, specifying which of the five methods described in 2.12.4 to 2.12.8 is used to transport the metadata, and if appropriate, by a stream\_id value of 0xFC indicating a metadata stream.

To uniquely identify a metadata service a metadata\_service\_id value is assigned to each such service by the transport; the assigned value shall be unique within the transport or program stream carrying the metadata service. If the metadata is carried in PES packets with a stream\_id of 0xFC, or in metadata sections, or in ISO/IEC 13818-6 synchronized download sections, the assigned metadata\_service\_id value is signalled explicitly in the header of the metadata\_AU\_cell or the metadata section. If a ISO/IEC 13818-6 carousel is used to carry the metadata, then the signalling of metadata services is left to the application. The metadata descriptor specifies the format of the metadata and provides information on the decoder configuration data, and is linked to the metadata service by carrying information on the metadata service it is associated with.

#### 2.12.9.3 Signalling of content for use by a metadata system

In 2.6.56 and 2.6.57, a content labelling descriptor is defined that can be used to assign a metadata application format specific reference, the content\_reference\_id\_record, to audiovisual or any other content carried over an MPEG-2 transport stream or program stream. The content\_reference\_id\_record can be used by the metadata system as a label to refer to such content. The content may represent, for example, a program or a stream or segments thereof. The content labelling descriptor also provides information on the content time base used for time referencing from the metadata, including the constant offset in time between the metadata time base and the applied content time base. The descriptor allows carriage of private data. The metadata\_application\_format may define constraints on the content\_reference\_record, such as constraints on the time period during which it is valid.

#### 2.12.9.4 Association of metadata to content

In 2.6.58 and 2.6.59 the metadata pointer descriptor is defined to associate a single metadata service to audiovisual or any other content in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream. The metadata is associated with the content within the context as defined by the location of the descriptor. In a transport stream, the descriptor may be located in the PMT in the descriptor loop for either the program or an elementary stream, but may also be located in tables not defined in this Specification, such as tables describing bouquets of broadcast services.

The metadata pointer descriptor points from the content's context to the metadata service associated with that content. The descriptor provides the value of the metadata\_service\_id that is assigned to the associated metadata service, as well as one or more locations of the associated metadata. The location may for example be within the same transport stream as the content, or within another transport stream, but also at a non-Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream location such as the Internet.

#### 2.12.9.5 Signalling decoder configuration data

Decoding of metadata may require the availability of metadata decoder configuration data. If needed, decoder configuration data shall be contained in one of the metadata services in the same program in the same Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 stream as the metadata service. If decoder configuration data is needed to decode a metadata service, then the metadata descriptor either carries such data or provides the information on retrieval of the decoder configuration data from the same or another metadata service. In a transport stream such other service can be found by searching in the PMT for a metadata\_descriptor with the metadata\_service\_id as specified in the decoder\_config\_metadata\_service\_id field (and with the same metadata\_format and the same metadata\_application format).

#### 2.12.9.6 Overview of metadata signalling

Figure 2-11 provides an example of metadata signalling, in which a single program, the "content program", carries the content (or essence) while the metadata is carried in a separate program, the "metadata program". In this example, the metadata program and the content program exist on the same transport stream.



Figure 2-11 – Metadata signalling and referencing

In the content program there are two metadata-related descriptors, the content\_labeling descriptor and the metadata\_pointer descriptor. The content\_labeling descriptor associates a label, illustrated in the diagram by "content label" and encoded in the descriptor in the content\_reference\_id fields, with the content. The label can then be used by the metadata service to refer to the essence, either in whole, in part, or by a time-described segment. For example, the content\_labeling descriptor could provide the label "News of 1/1/02", and the metadata could then refer to a specific story item in the "News of 1/1/02", for example by providing the specific timing of the story item.

The metadata pointer descriptor provides information of where the metadata service can be found for the given content. In this example, the metadata is carried in a separate program, but it would be equally valid to have the metadata carried in the same program as the content, or provided by some means beyond the scope of this Specification, for instance from a URL. This descriptor also provides the metadata service id value that is assigned to the metadata service. This is required since a metadata stream could carry multiple metadata services for many different programs and each program needs to be able to uniquely identify its own metadata service.

In the metadata program, the metadata descriptor signals to which metadata service within a metadata stream it applies. If used, the metadata descriptor provides details of where to find the decoder configuration information.

Upon identifying a metadata pointer descriptor in the PMT by a receiver decoding the content program, the receiver retrieves the metadata descriptor from the metadata program. If needed first the decoder configuration data is retrieved, then the decoder is configured accordingly, after which the metadata service can start being decoded.

#### 2.12.10 STD model for metadata

The STD model specifies normative constraints on Rec. ITU-T H.222.0 | ISO/IEC 13818-1 streams that carry metadata. For decoding of metadata in the STD, the regular T-STD and P-STD models are applicable with buffer Bn, input rate Rxn of the metadata into Bn and output rate Rmetadata out of Bn and into Dmetadata, the metadata decoder. See Figure 2-12.



Figure 2-12 – Metadata decoding in the STD

The metadata enters buffer Bn at rate Rxn. In the P-STD, rate Rxn equals the rate of the program stream. In the T-STD, rate Rxn is the rate out of TBn and equal to the rate defined by the metadata\_input\_leak\_rate field in the metadata STD descriptor. The size BSn of buffer Bn is equal to the size defined in the metadata\_buffer\_size field in the metadata STD descriptor. In case of synchronous delivery, metadata decoding is instantaneous and controlled by PTSs. At decode time, that is when the STC equals the PTS, the associated metadata is removed instantaneously from Bn. In case of asynchronous delivery, the metadata is removed from Bn at a rate Rmetadata equal to the rate defined by the metadata\_output\_leak\_rate field in the metadata STD descriptor. Buffer Bn shall not overflow.

Note that the STD model defines constraints on the delivery of the metadata, without specifying any constraint on the timing used in the metadata.

## 2.13 Carriage of ISO 15938 data

### 2.13.1 Introduction

Carriage of metadata over a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream as defined in 2.12 allows for carriage of ISO 15938 data by appropriate coding of the metadata\_format field. In this subclause, for the purpose to transport ISO 15938 data, a specific instance is defined. Carriage of ISO 15938 data shall meet each requirement defined in 2.12, but in addition the requirements defined in this subclause shall apply for transport of ISO 15938 data.

### 2.13.2 ISO 15938 decoder configuration data

Decoding of ISO 15938 data requires the availability of decoder configuration data. Consequently, when ISO 15938 data is carried in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, then the metadata descriptor shall signal carriage of associated decoder configuration data in the same Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream by coding a value of the decoder\_config\_flags of either '001' or '010' or '011' or '100'.

## 2.14 Carriage of Rec. ITU-T H.264 | ISO/IEC 14496-10 video

### 2.14.1 Introduction

This Specification defines the carriage of Rec. ITU-T H.264 | ISO/IEC 14496-10 elementary stream within Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 systems, both for program and transport streams. Typically, a Rec. ITU‑T H.264 | ISO/IEC 14496-10 stream will be an element of a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, as defined by the PMT in a transport stream and the PSM in a program stream. The carriage and buffer management of AVC video streams is defined using existing parameters from this Recommendation | International Standard such as PTS and DTS, as well as information present within an AVC video stream.

Carriage of AVC video streams in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream defines accurate mapping between STD parameters and HRD parameters that may be present in an AVC video stream. Requirements are defined for the presence of HRD parameters in the AVC video stream, to ensure that it can be verified whether each STD requirement is met for each AVC video stream carried in a transport stream or a program stream.

NOTE 1 – Though the timing information present in the AVC video stream may not use a 90-kHz clock, the PTS and DTS timestamps need to be expressed in units of 90 kHz.

When a Rec. ITU-T H.264 | ISO/IEC 14496-10 stream is carried in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, the Rec. ITU-T H.264 | ISO/IEC 14496-10 coded data shall be contained in PES packets. The Rec. ITU-T H.264 | ISO/IEC 14496-10 coded data shall comply with the byte stream format defined in Annex B of Rec. ITU-T H.264 | ISO/IEC 14496‑10, with the following constraints:

• Each AVC access unit shall contain an access unit delimiter NAL Unit;

NOTE 2 – Rec. ITU-T H.264 | ISO/IEC 14496-10 requires that an access unit delimiter NAL Unit, if present, is the first NAL Unit within an AVC access unit. Access unit delimiter NAL Units simplify the ability to detect the boundary between pictures; they avoid the need to process the content of slice headers, and they are particularly useful for the Baseline and Extended profiles where slice order can be arbitrary.

• All Sequence and Picture Parameter Sets (SPS and PPS) necessary for decoding the AVC video stream shall be present within that AVC video stream.

NOTE 3 – Rec. ITU-T H.264 | ISO/IEC 14496-10 also allows delivery of SPS and PPS by external means. This Specification does not provide support for such delivery, and therefore requires SPS and PPS to be carried within the AVC video stream.

• Each AVC video sequence that contains hrd\_parameters() with the low\_delay\_hrd\_flag set to '1', shall carry VUI parameters in which the timing\_info\_present\_flag shall be set to '1'.

NOTE 4 – If the low\_delay\_hrd\_flag is set to '1', then buffer underflow is allowed to occur in the STD model; see 2.14.3 and 2.14.4. Setting the timing\_info\_present\_flag to '1' ensures that the AVC video stream contains sufficient information to determine the DPB output time and the CPB removal time of AVC access units, also in case of underflow.

To provide display specific information such as aspect\_ratio, it is strongly recommended that each AVC video stream carries VUI parameters with sufficient information to ensure that the decoded AVC video stream can be displayed correctly by receivers.

When an AVC video stream conforms to one or more profiles defined in Annex G of Rec. ITU-T H.264 | ISO/IEC 14496-10, the following constraints additionally apply:

• The AVC video sub-bitstream of SVC as defined in 2.1.10 shall be an element of a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program and the stream\_type for this elementary stream shall be equal to 0x1B.

• For each SVC video sub-bitstream as defined in 2.1.140 that is an element of the same Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 program, the stream\_type for this elementary stream shall be equal to 0x1F.

• All subset Sequence Parameter Sets and Picture Parameter Sets necessary for decoding an SVC video sub-bitstream shall be present within the elementary stream carrying the SVC video sub-bitstream.

• In each elementary stream with stream\_type equal to 0x1F, exactly one VDRD\_drd\_nal\_unit as defined in 2.14.3.3 may precede all the NAL units of the same SVC dependency representation.

NOTE 5 – If any VDRD\_drd\_nal\_unit is included in any SVC dependency representation then the HRD model should include this VDRD\_drd\_nal\_unit in the buffer model as additional non-VCL NAL units. The NAL unit type 24 may be used in a different way by other specifications out of scope of this Specification. When carrying AVC base and SVC enhancement layers in different elementary streams, usage of VDRD is strongly recommended if access units are not aligned with PES packets.

• The TREF field as defined in 2.4.3.7 may be present in the PES headers of elementary streams with stream\_type equal to 0x1F. The TREF field shall be set and shall be present in the PES headers as specified in 2.14.3.5 and 2.14.3.6 respectively.

NOTE 6 – Currently the presence of TREF is only specified for elementary streams with stream\_type equal to 0x1F.

• When a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program includes more than one SVC video sub‑bitstream, or more than one AVC video sub-bitstream of SVC and at least one SVC video sub‑bitstream, a hierarchy descriptor as defined in 2.6.7 shall be used to indicate the dependencies of the related video sub-bitstreams.

• All NAL units of a re-assembled AVC access unit shall be passed to the decoder in the order of NAL units within an access unit as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10.

NOTE 7 – If SEI NAL units are present in any SVC dependency representation of an SVC video sub-bitstream, these NAL units may require re-ordering to the order of NAL units within an access unit as defined in Rec. ITU‑T H.264 | ISO/IEC 14496-10 before access unit re-assembling.

• The profile and level limitations indicated by profile\_idc and level\_idc syntax elements in the AVC\_video\_descriptor, if present, and the Type II HRD parameters in the AVC\_timing\_and\_HRD\_descriptor, if present, for an AVC video stream resulting from re-assembling (up to) the video sub-bitstream associated with the descriptors shall include NAL units with nal\_unit\_type syntax element equal to 14 in the AVC video sub-bitstream of SVC and, if present in the SVC video sub-bitstream, NAL units with nal\_unit\_type syntax element equal to 24.

When an AVC video stream conforms to one or more profiles defined in Annex H of Rec. ITU-T H.264 | ISO/IEC 14496‑10, the following constraints additionally apply:

• The AVC video sub-bitstream of MVC or MVC base view sub-bitstream, as defined in 2.1.9 and 2.1.101, shall be an element of a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program and the stream\_type for this elementary stream shall be equal to 0x1B.

• For each MVC video sub-bitstream, as defined in 2.1.104, that is an element of the same Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 program, the stream\_type for this elementary stream shall be equal to 0x20.

• Each MVC video sub-bitstream shall be associated with one or more consecutive view order index values.

NOTE – According to the definitions in 2.1.104 and 2.1.101, respectively, an MVC video sub-bitstream or MVC base view sub-bitstream does not necessarily include view components for all view\_id values included in one MVC view\_id subset if one or more views are not required for decoding the transmitted views. As an example, consider a MVC bitstream having 4 views V1, V2, V3, and V4 in ascending order of view order index, where view V1 is the base view, view V2 is depending directly on V1, V3 is depending directly on V1 and V2, and V4 is depending directly on V2. Using such encoded views, two MVC sub-bitstreams M1 and M2 may be created as follows: M1 is associated with the output views V1 and V2, and M2 with the output view V4. In this example, it is possible that only M1 and M2 are transmitted to a receiver, thus sub-bitstream for V3 is not required to be transmitted since a combination of both sub-bitstreams M1 and M2 refers to the set of views V1, V2 and V4, and can be decoded without the presence of V3.

• Each view order index value shall be associated with exactly one MVC view\_id subset.

NOTE 8 – This restriction greatly simplifies the re-assembly of any decodable sub-bitstream.

• When a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program includes more than one MVC video sub‑bitstream or more than one AVC video sub-bitstream of MVC and at least one MVC video sub‑bitstream, one or more hierarchy descriptors as defined in 2.6.6 and 2.6.7 shall be used to indicate the dependencies of the related video sub-bitstreams. If more than one hierarchy descriptor is present for one elementary stream, the value of the syntax element hierarchy\_layer\_index shall be the same within the same elementary stream. The syntax element hierarchy\_type shall be set to the value 9 or 15.

NOTE – Provided an MVC video sub-bitstream B depends on video sub-bitstream A and this dependency is indicated using a hierarchy descriptor, further an MVC video sub-bitstream C depends on B and this dependency is also indicated using a second hierarchy descriptor, then this implicitly indicates a dependency of C on A and no third hierarchy descriptor is needed.

• The subset sequence parameter sets and picture parameter sets necessary for decoding an MVC video sub-bitstream shall be present within the elementary stream carrying the MVC video sub-bitstream.

• In each elementary stream with stream type equal to 0x20 exactly one VDRD\_NAL\_unit, as defined in 2.14.3.3, may precede all the NAL units of the same MVC view-component subset.

NOTE 9 – If any VDRD\_nal\_unit is included in any MVC view component subset, then the HRD model should include this VDRD\_nal\_unit in the buffer model as additional non-VCL NAL units. The NAL unit type 24 may be used in a different way by other specifications out of scope of this Specification.

• All NAL units of a re-assembled AVC access unit shall be passed to the decoder in the order of NAL units within an access unit, as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10.

NOTE 10 – If SEI NAL units are present in any MVC view-component subset of an MVC video sub-bitstream, these NAL units may require re-ordering to the order of NAL units within an access unit, as defined in Rec. ITU‑T H.264 | ISO/IEC 14496-10 before access unit re-assembling.

• The profile\_idc and level\_idc indication in the AVC\_video\_descriptor, if present, and the Type II HRD parameters in the AVC\_timing\_and\_HRD\_descriptor, if present, for an AVC video stream resulting from re-assembling (up to) the MVC video sub-bitstream associated with the descriptors shall include NAL units with nal\_unit\_type syntax element equal to 14, if present, in the AVC video sub-bitstream of MVC or MVC base view sub-bitstream and, if present, in the MVC video sub-bitstream, NAL units with nal\_unit\_type syntax element equal to 20 and 24.

When an AVC video stream conforms to one or more profiles defined in Annex I of Rec. ITU-T H.264 | ISO/IEC 14496‑10, the following constraints additionally apply:

• The AVC video sub-bitstream of MVCD or MVCD base view sub-bitstream, as defined in 2.1.10 and 2.1.106, shall be an element of a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program and the stream\_type for this elementary stream shall be equal to 0x1B.

• For each MVCD video sub-bitstream, as defined in 2.1.101, that is an element of the same Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 program, the stream\_type for this elementary stream shall be equal to 0x26.

• Each MVCD video sub-bitstream shall be associated with one or more consecutive view order index values.

• Each view order index value shall be associated to exactly one MVCD view\_id subset.

NOTE 11 – This restriction greatly simplifies the re-assembly of any decodable sub-bitstream.

• The subset sequence parameter sets and picture parameter sets necessary for decoding an MVCD video sub-bitstream shall be present within the elementary stream carrying the MVCD video sub-bitstream.

• In each elementary stream with stream type equal to 0x26 exactly one VDRD\_ nal\_unit, as defined in 2.14.3.3, may precede all the NAL units of the same MVCD view-component subset.

NOTE 12 – If any VDRD\_ nal\_unit is included in any MVCD view component subset, then the HRD model should include this VDRD\_nal\_unit in the buffer model as additional non-VCL NAL units. The NAL unit type 24 may be used in a different way by other specifications out of scope of this Specification.

• All NAL units of a re-assembled AVC access unit shall be passed to the decoder in the order of NAL units within an access unit, as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10.

NOTE 13 – If SEI NAL units are present in any MVCD view-component subset of an MVCD video sub-bitstream, these NAL units may require re-ordering to the order of NAL units within an access unit, as defined in Rec. ITU‑T H.264 | ISO/IEC 14496‑10 before access unit re-assembling.

### 2.14.2 Carriage in PES packets

Rec. ITU-T H.264 | ISO/IEC 14496-10 Video is carried in PES packets as PES\_packet\_data\_bytes, using one of the 16 stream\_id values assigned to video, while signalling the Rec. ITU-T H.264 | ISO/IEC 14496-10 Video stream by means of the assigned stream-type value in the PMT or PSM (see Table 2-34). The highest level that may occur in an AVC video stream as well as a profile that the entire stream conforms to should be signalled using the AVC video descriptor. If an AVC video descriptor is associated with an AVC video stream, then this descriptor shall be conveyed in the descriptor loop for the respective elementary stream entry in the Program Map Table in case of a transport stream or in the Program Stream Map, when PSM is present, in case of a program stream. This Recommendation | International Standard does not specify presentation of Rec. ITU-T H.264 | ISO/IEC 14496-10 streams in the context of a program.

For PES packetization, no specific data alignment constraints apply. For synchronization and STD management, PTSs and, when appropriate, DTSs are encoded in the header of the PES packet that carries the Rec. ITU-T H.264 | ISO/IEC 14496-10 video elementary stream data. For PTS and DTS encoding, the constraints and semantics apply as defined in 2.4.3.7 and 2.7.

### 2.14.3 STD extensions

#### 2.14.3.1 T-STD extensions

The T-STD model includes a transport buffer TBn and a multiplex buffer MBn prior to buffer EBn for decoding of each AVC video elementary stream n conforming to one or more profiles defined in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496-10 video elementary stream n. See Figure 2-13.



Figure 2-13 – T-STD model extensions for Rec. ITU-T H.264 | ISO/IEC 14496-10 video

NOTE – For carriage of AVC streams according to High, Progressive High, Constrained High, High 10, High 4:2:2, High 4:4:4 Predictive, High 10 Intra, High 4:2:2 Intra, High 4:4:4 Intra, or CAVLC 4:4:4 Intra profiles, bitrate and CPB buffer size limitations of this specification are more restrictive than the limits specified in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496 10.

DPBn buffer management

Carriage of an AVC video stream over Rec. ITU-T H.222.0 | ISO/IEC 13818-1 does not impact the size of buffer DPBn. For decoding of an AVC video stream in the STD the size of DPBn is as defined in Rec. ITU-T H.264 | ISO/IEC 14496‑10. The DPB buffer shall be managed as specified in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496‑10 (C.2 and C.4). A decoded AVC access unit enters DPBn instantaneously upon decoding of the AVC access unit, hence at the CPB removal time of the AVC access unit. A decoded AVC access unit is presented at the DPB output time. If the AVC video stream provides insufficient information to determine the CPB removal time and the DPB output time of AVC access units, then these time instants shall be determined in the STD model from PTS and DTS timestamps as follows:

1) The CPB removal time of AVC access unit n is the instant in time indicated by DTS(n) where DTS(n) is the DTS value of AVC access unit n.

2) The DPB output time of AVC access unit n is the instant in time indicated by PTS(n) where PTS(n) is the PTS value of AVC access unit n.

NOTE 1 – AVC video sequences in which the low\_delay\_hrd\_flag in hrd parameters() is set to 1 carry sufficient information to determine the DPB output time and the CPB removal time of each AVC access unit. Hence for AVC access units for which STD underflow may occur, the CPB removal time and the DPB output time are defined by HRD parameters, and not by DTS and PTS timestamps.

TBn, MBn and EBn buffer management

The input to buffer TBn and its size TBSn are specified in 2.4.2.4. For buffers MBn and EBn, and for the rate Rxn between TBn and MBn and the rate Rbxn between MBn and EBn the following constraints apply for carriage of a Rec. ITU-T H.264 | ISO/IEC 14496-10 stream:

Size EBSn of buffer EBn:

EBSn = cpb\_size

Where cpb\_size is the size CpbSize[ cpb\_cnt\_minus1 ] of the CPB for the byte stream format signalled in the NAL hrd\_parameters() carried in VUI parameters in the AVC video stream. If NAL hrd\_parameters() are not present in the AVC video stream, then the cpb\_size shall be the size defined as 1200 × MaxCPB in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496-10 for the level of the AVC video stream.

Size MBSn of Buffer MBn:

MBSn = BSmux + BSoh + 1200 × MaxCPB[level] – cpb\_size

where BSoh, packet overhead buffering, is defined as:

BSoh = (1/750) seconds × max{1200 × MaxBR[level], 2 000 000 bit/s}

and BSmux, additional multiplex buffering, is defined as:

BSmux = 0.004 seconds × max{1200 × MaxBR[level], 2 000 000 bit/s}

where MaxCPB[level] and MaxBR[level] are defined for the byte stream format in Table A.1 (Level Limits) in Rec. ITU-T H.264 | ISO/IEC 14496-10 for the level of the AVC video stream, and

where cpb\_size is the size CpbSize[ cpb\_cnt\_minus1 ] of the CPB for the byte stream format signalled in the NAL hrd\_parameters() carried in VUI parameters in the AVC video stream. If NAL hrd\_parameters() are not present in the AVC video stream, then the cpb\_size shall be the size 1200 × MaxCPB defined in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496-10 for the level of the AVC video stream.

Rate Rxn:

when there is no data in TBn then Rxn is equal to zero.

Otherwise: Rxn = bit\_rate

where bit\_rate is 1.2 × BitRate[ SchedSelIdx ] of data flow into the CPB for the byte stream format and BitRate[ SchedSelIdx ] is as defined in Annex E of Rec. ITU-T H.264 | ISO/IEC 14496-10.

NOTE 2 – Annex E specifies the values for BitRate[ SchedSelIdx] when NAL\_hrd\_parameters() is present in the VUI parameters of the AVC video stream and default values for BitRate[ SchedSelIdx ] based on profile and level when NAL\_hrd\_parameters() is not present.

***Transfer between MBn and EBn***

If the AVC\_timing\_and\_HRD\_descriptor is present with the hrd\_management\_valid\_flag set to '1', then the transfer of data from MBn to EBn shall follow the HRD defined scheme for data arrival in the CPB as defined in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496-10.

Otherwise, the leak method shall be used to transfer data from MBn to EBn as follows:

Rate Rbxn:

Rbxn = 1200 × MaxBR[level]

where MaxBR[level] is defined for the byte stream format in Table A.1 (Level Limits) in Rec. ITU‑T H.264 | ISO/IEC 14496-10 for each level.

If there is PES packet payload data in MBn, and buffer EBn is not full, the PES packet payload is transferred from MBn to EBn at a rate equal to Rbxn. If EBn is full, data are not removed from MBn. When a byte of data is transferred from MBn to EBn, all PES packet header bytes that are in MBn and precede that byte, are instantaneously removed and discarded. When there is no PES packet payload data present in MBn, no data is removed from MBn. All data that enters MBn leaves it. All PES packet payload data bytes enter EBn instantaneously upon leaving MBn.

*Removal of AVC access units from EBn*

Each AVC access unit An(j) that is present in EBn is removed instantaneously at time tdn(j). The decoding time tdn(j) is specified by the DTS or from the CPB removal time, as derived from information in the AVC video stream.

STD delay

The total delay of any Rec. ITU-T H.264 | ISO/IEC 14496-10 data other than AVC still picture data through the System Target Decoders buffers TBn, MBn, and EBn shall be constrained by tdn(j) – t(i) ≤ 10 seconds for all j, and all bytes i in AVC access unit An(j).

The delay of any AVC still picture data through the System Target Decoders buffers TBn, MBn, and EBn shall be constrained by tdn(j) – t(i) ≤ 60 seconds for all j, and all bytes i in AVC access unit An(j).

Buffer management conditions

Transport streams shall be constructed so that the following conditions for buffer management are satisfied:

• TBn shall not overflow and shall be empty at least once every second.

• MBn, EBn, and DPBn shall not overflow.

• EBn shall not underflow, except when VUI parameters are present for the AVC video sequence with the low\_delay\_hrd\_flag set to '1'. Underflow of EBn occurs for AVC access unit An(j) when one or more bytes of An(j) are not present in EBn at the decoding time tdn(j).

NOTE 3 – An AVC video stream may carry information to determine compliance of the AVC video stream to the HRD, as specified in Annex C of Rec. ITU-T H.264 | ISO/IEC 14496-10. The presence of this information can be signalled in a transport stream using the AVC timing and HRD descriptor with the hrd\_management\_valid\_flag set to '1'. Irrespective of the presence of this information, compliance of an AVC video stream to the T-STD ensures that HRD buffer management requirements for CPBn are met when each byte in the AVC video stream is delivered to and removed from CPBn in the HRD at exactly the same instant in time at which the byte is delivered to and removed from EBn in the T-STD.

#### 2.14.3.2 P-STD extensions

The P-STD model for the decoding of an AVC video elementary stream n conforming to one or more profiles defined in Annex A of Rec. ITU-T H.264 | ISO/IEC 14496-10 elementary stream includes a multiplex buffer Bn and a decoder Dn followed by a buffer DPBn (see Figure 2-14). For each AVC video stream n, the size BSn of buffer Bn in the P-STD is defined by the P-STD\_buffer\_size field in the PES packet header.



Figure 2-14 – P-STD model extensions for Rec. ITU-T H.264 | ISO/IEC 14496-10 video

DPBn buffer management

Buffer DPBn shall be managed in exactly the same way as in the T-STD; see 2.14.3.1.

Bn buffer management

The AVC access unit data enters buffer Bn as specified in 2.5.2.2. At time tdn(j), AVC access unit An(j) is decoded and instantaneously removed from Bn. The decoding time tdn(j) is specified by the DTS or by the CPB removal time, derived from information in the AVC video stream. Upon decoding, the AVC access unit instantaneously enters DPBn or is output without entry into DPBn, according to the rules specified in Rec. ITU-T H.264 | ISO/IEC 14496-10.

STD delay

The total delay of any Rec. ITU-T H.264 | ISO/IEC 14496-10 data other than AVC still picture data through the System Target Decoders buffer Bn shall be constrained by tdn(j) – t(i) ≤ 10 seconds for all j, and all bytes i in AVC access unit An(j).

The delay of any AVC still picture data through the System Target Decoders buffer Bn shall be constrained by tdn(j) – t(i) ≤ 60 seconds for all j, and all bytes i in AVC access unit An(j).

Buffer management conditions

Program streams shall be constructed so that the following conditions for buffer management are satisfied:

• Bn shall not overflow.

• Bn shall not underflow, except when VUI parameters are present for the AVC video sequence with the low\_delay\_hrd\_flag set to '1' or when trick\_mode status is true. Underflow of Bn occurs for AVC access unit An(j) when one or more bytes of An(j) are not present in Bn at the decoding time tdn(j).

#### 2.14.3.3 View and dependency representation delimiter NAL unit

See Table 2-160.

Table 2-160 – View and dependency representation delimiter NAL unit

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| VDRD\_nal\_unit() { |  |  |
| **forbidden\_zero\_bit** | **1** | **bslbf** |
| **nal\_ref\_idc** | **2** | **bslbf** |
| **nal\_unit\_type** | **5** | **bslbf** |
| } |  |  |

#### 2.14.3.4 Semantics of view and dependency representation delimiter NAL unit

**forbidden\_zero\_bit** – shall be equal to 0x0

**nal\_ref\_idc** – shall be equal to 0x0

**nal\_unit\_type** – shall be equal to 0x18

#### 2.14.3.5 T-STD extensions for SVC

The T-STD model described in 2.14.3.1 is applied if the received elementary stream is a video sub-bitstream of stream\_type 0x1B, i.e., only the AVC video sub-bitstream of SVC is received and decoded.

When there is a set of received video sub-bitstreams in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, for which dependencies may be signalled in the hierarchy\_descriptor as defined in 2.6.6, and when there is at least one of the video sub-bitstreams in the set of received elementary streams having the value of stream\_type equal to 0x1F, the T‑STD model as described in 2.14.3.1 is extended as illustrated in Figure 2-15 and as specified below.



Figure 2-15 – T-STD model extensions for Rec. ITU-T H.264 | ISO/IEC 14496-10  
Video with scalable video sub-bitstreams

The following additional notations are used to describe the T-STD extensions and are illustrated in Figure 2-15 above.

ESn is the received elementary stream associated with dependency\_id value equal to n

ESH is the received elementary stream associated with the highest value H of dependency\_id present in the set of received elementary streams

j is an index to the re-assembled access units

jn is an index to the SVC dependency representations of the elementary stream associated with dependency\_id value equal to n

DRn(jn) is the jn-th SVC dependency representation of video sub-bitstream associated with dependency\_id value equal to n

AH(j) is the j-th access unit resulting from re-assembling (up to) the jH-th SVC dependency representations with dependency\_id value equal to H

tdn(jn) is the decoding time, measured in seconds, in the system target decoder of the jn-th SVC dependency representation of the video sub-bitstream associated with dependency\_id value equal to n

tdH(j) is the decoding time, measured in seconds, in the system target decoder of the j-th access unit AH(j) resulting from re-assembling (up to) the jH-th SVC dependency representations DRH(jH)

tref n(jn) is equal to the decoding time value tdn-1(jn-1) associated with the jn-1-th SVC dependency representation DRn-1(jn-1), indicated by the TREF field of the PES header of the jn-th SVC dependency representation DRn(jn) of the same access unit

TBn is the transport buffer for elementary stream associated with dependency\_id value equal to n

TBSn is the size of the transport buffer TBn, measured in bytes

MBn is the multiplexing buffer for elementary stream associated with dependency\_id value equal to n

MBSn is the size of the multiplexing buffer MBn, measured in bytes

DRBn is the dependency representation buffer for elementary stream ESn

DRBSn is the size of dependency representation buffer DRBn, measured in bytes

EBH is the elementary stream buffer for all video sub-bitstreams

EBSH is the size of elementary stream buffer EBH, measured in bytes

Rxn  is the transfer rate from TBn to MBn as specified below

Rbxn is the transfer rate from MBn to DRBn as specified below

Carriage in PES packets

For correct re-assembling of the SVC dependency representations to an AVC access unit, if there is both an SVC dependency representation with any dependency\_id equal to n and an SVC dependency representation with dependency\_id equal to (n+1) in the same AVC access unit, the following applies:

• a PES packet per SVC dependency representation start shall be present, i.e., at most one SVC dependency representation may commence in the same PES packet;

• the PTS and, if applicable, the DTS value shall be provided in the PES header of each SVC dependency representation;

• if the DTS value of the SVC dependency representation with dependency\_id equal to n is different from the DTS value of the SVC dependency representation with dependency\_id equal to (n+1) of the same access unit, the TREF field as defined in 2.4.3.7 shall be present in the PES header extension of the SVC dependency representation with dependency\_id equal to (n+1) and the TREF field value shall be equal to the DTS value of the SVC dependency representation with dependency\_id equal to n.

DPB buffer management

The DPB buffer management for the re-assembled AVC video stream shall conform to 2.14.3.1 using AVC access unit timing values, as DTS or CPB removal time, and PTS or DPB removal time, associated with the SVC dependency representations of the video sub-bitstream in elementary stream ESH.

TBn, MBn, EBn buffer management

The following applies:

• There is exactly one transport buffer TB as defined in 2.14.3.1 for each received elementary stream in the set of received video sub-bitstreams contained in elementary streams as shown in Figure 2-13.

• There is exactly one multiplexing buffer MB0 for the AVC video sub-bitstream of SVC in elementary stream ES0, where the size of the multiplexing buffer MBS0 is constrained as follows:

MBS0 = BSmux,0 + BSoh,0 + 1200 × MaxCPB[level]0 – cpb\_size0

where BSmux,0, BSoh,0 are defined in 2.14.3.1 for the AVC video sub-bitstream of SVC in elementary stream ES0;

where MaxCPB[level]0 and cpb\_size0 for the elementary stream ES0 are defined as in 2.14.3.1.

NOTE 1 – If HRD parameters are present in at least one of the video sub-bitstreams, those parameters have to be carefully handled in order to not unnecessarily increase the multiplexing buffers allocation.

• There is exactly one multiplexing buffer MBn for each received elementary stream associated with dependency\_id value not equal to 0, where the size of each multiplexing buffer MBSn is constrained as follows:

MBSn = BSmux,n + BSoh,n

where BSmux,n, BSoh,n are defined in 2.14.3.1 for the AVC video stream resulting from re-assembling (up to) the SVC video sub-bitstream in elementary stream ESn.

• There is exactly one elementary stream buffer EBH for all the elementary streams in the set of received elementary streams as shown in Figure 2-13, of which the size EBSH has the following value:

EBSH = cpb\_sizeH

where cpb\_sizeH is the cpb\_size for the SVC video sub-bitstream in elementary stream ESH as defined in 2.14.3.1 for the re-assembled AVC video stream.

• There is exactly one dependency representation buffer DRBn for each elementary stream in the set of received elementary streams as shown in Figure 2-13, where each dependency representation buffer DRBn in the set of received elementary streams is allocated within EBH. Even though the size DRBSn of individual DRBn is not constrained, the sum of the sizes DRBSn is constrained as follows:

EBSH = ∑n (DRBSn)

• Transfer from TBn to MBn is applied as follows:

When there is no data in TBn then Rxn is equal to zero. Otherwise:

Rxn = bit\_rate

where bit\_rate is 1.2 x BitRate[ SchedSelIdx ] of data flow into the CPB for the byte stream format and BitRate[ SchedSelIdx ] is as defined in Annex E of Rec. ITU-T H.264 | ISO/IEC 14496-10 when NAL hrd\_parameters() is present in the VUI parameters of the SVC video sub-bitstream.

NOTE 2 – Annex E also specifies default values for BitRate[ SchedSelIdx] based on profile and level when NAL HRD parameters are not present in the VUI. The SVC video sub-bitstream level is determined by the level of AVC video stream resulting from re-assembling (up to) the associated video sub-bitstream n in elementary stream ESn.

• Transfer from MBn to DRBn is applied as follows:

If the AVC\_timing\_and\_HRD\_descriptor is present with the hrd\_management\_valid\_flag set to '1' for elementary stream ESH, then the transfer of data from MBn to DRBn shall follow the HRD defined scheme for data arrival in the CPB of elementary stream ESH as defined in Annex C of Rec. ITU‑T H.264 | ISO/IEC 14496-10.

Otherwise, the leak method shall be used to transfer data from MBn to DRBn as follows:

Rate Rbxn:

Rbxn = 1200 × MaxBR[level]n

where MaxBR[level]n is defined for the byte stream format in Table A.1 (Level Limits) in Rec. ITU‑T H.264 | ISO/IEC 14496-10 for the level of the AVC video stream resulting from re‑assembling (up to) the associated video sub-bitstream n in elementary stream ESn. If there is PES packet payload data in MBn, and buffer EBH is not full, the PES packet payload is transferred from MBn to DRBn at a rate equal to Rbxn. If EBH is full, data are not removed from MBn. When a byte of data is transferred from MBn to DRBn, all PES packet header bytes that are in MBn and precede that byte are instantaneously removed and discarded. When there is no PES packet payload data present in MBn, no data is removed from MBn. All data that enters MBn leaves it. All PES packet payload data bytes enter DRBn instantaneously upon leaving MBn.

Access unit re-assembling and EB removal

The following specifies the access unit re-assembling that results in AVC access unit AH(j):

i) Collect all SVC dependency representations DRn(jn) starting with the highest value of dependency\_id n, equal to H, to the lowest value of dependency\_id n, equal to m, present in access unit AH(j) following the rule below:

• For each two corresponding DRy+1(jy+1) and DRy(jy) of the SVC dependency representations collected for access unit AH(j), if TREF field is present for DRy+1(jy+1), the TREF value of DRy+1(jy+1) trefy+1(jy+1) shall be equal to the DTS value tdy(jy) of DRy(jy), otherwise the DTS value of tdy+1(jy+1) of DRy+1(jy+1) shall be equal to DTS value tdy(jy) of DRy(jy).

ii) Assemble the SVC dependency representations in consecutive order of dependency\_id n starting from 'm' to 'H' for the j-th access unit AH(j). If SEI NAL units are present in any SVC dependency representation with dependency\_id not equal to 0, these NAL units shall be re-ordered to the order of NAL units within an access unit as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10 before access unit re-assembling.

NOTE 2 – The number of SVC dependency representations for each access unit AH(j) may vary depending on value of j. If m represents the lowest value of dependency\_id for access unit AH(j), then the collected and assembled SVC dependency representations include DRm(jm), DRm+1(jm+1), …, DRH(jH).

The following specifies the removal of access unit AH(j) from buffer EBH:

At time tdH(j) the AVC access unit AH(j) shall be re-assembled and available for removal from buffer EBH. The decoding time tdH(j) is specified by the DTS or by the CPB removal time that is associated with the SVC dependency representations in elementary stream ESH, as derived from information in the re-assembled AVC video stream.

STD delay

The STD delay for re-assembled AVC access units shall follow the constraints specified in 2.14.3.1.

Buffer management conditions

Transport streams shall be constructed so that the following conditions for buffer management are satisfied:

• Each TBn shall not overflow and shall be empty at least once every second.

• Each MBn, EBH, and DPB shall not overflow.

• EBH shall not underflow, except when VUI parameters are present for the AVC video sequence of the re-assembled AVC video stream with the low\_delay\_hrd\_flag set to '1'. Underflow of EBH occurs for AVC access unit AH(j) when one or more bytes of AH(j) are not present in EBH at the decoding time tdH(j).

#### 2.14.3.6 P-STD extensions for SVC

The P-STD model described in 2.14.3.2 is applied if the decoded elementary stream is a video sub-bitstream of stream\_type 0x1B, i.e., only the AVC video sub-bitstream of SVC is decoded.

When there is a set of decoded video sub-bitstreams in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, for which dependencies may be signalled in the hierarchy\_descriptor as defined in 2.6.6, and when there is at least one of the video sub-bitstreams in the set of decoded elementary streams having the value of stream\_type equal to 0x1F, the P‑STD model as described in 2.14.3.2 is extended as illustrated in Figure 2-16 and as specified below.



Figure 2-16 – P-STD model extensions for Rec. ITU-T H.264 | ISO/IEC 14496-10  
Video with scalable video sub-bitstreams

The following additional notations are used to describe the P-STD extensions and are illustrated in Figure 2-16 above.

ESn is the decoded elementary stream associated with dependency\_id value equal to n

ESH  is the decoded elementary stream associated with the highest value H of dependency\_id in the set of decoded elementary streams

j is an index to the re-assembled access units

jn  is an index to the SVC dependency representations of the elementary stream associated with dependency\_id value equal to n

DRn(jn) is the jn-th SVC dependency representation of video sub-bitstream associated with dependency\_id value equal to n

AH(j) is the j-th access unit resulting from re-assembling (up to) the jH-th SVC dependency representations with dependency\_id value equal to H

tdn(jn) is the decoding time, measured in seconds, in the system target decoder of the jn-th SVC dependency representation of the video sub-bitstream associated with dependency\_id value equal to n

tdH(j) is the decoding time, measured in seconds, in the system target decoder of the j-th access unit AH(j) resulting from re-assembling (up to) the jH-th SVC dependency representations DRH(jH)

tref n(jn) is equal to the decoding time value tdn-1(jn-1) associated with the jn-1-th SVC dependency representation DRn-1(jn-1), indicated by the TREF field of the PES header of the jn-th SVC dependency representation DRn(jn) of the same access unit

BH  is the input buffer for all decoded video sub-bitstreams

BSH is the size of the input buffer BH, measured in bytes

DRBn is the dependency representation buffer for elementary stream ESn

DRBSn is the size of dependency representation buffer DRBn, measured in bytes

Carriage in PES packets

For correct re-assembling of the SVC dependency representations to an AVC access unit, if there is both an SVC dependency representation with any dependency\_id equal to n and an SVC dependency representation with dependency\_id equal to (n+1) in the same AVC access unit, the following applies:

• a PES packet per SVC dependency representation start shall be present, i.e., at most one SVC dependency representation may commence in the same PES packet;

• the PTS and, if applicable, the DTS value shall be provided in the PES header of each SVC dependency representation;

• if the DTS value of the SVC dependency representation with dependency\_id equal to n is different from the DTS value of the SVC dependency representation with dependency\_id equal to (n+1) of the same access unit, the TREF field as defined in 2.4.3.7 shall be present in the PES header extension of the SVC dependency representation with dependency\_id equal to (n+1) and the TREF value shall be equal to the DTS value of the SVC dependency representation with dependency\_id equal to n.

DPB buffer management

The DPB buffer management for the re-assembled AVC video stream shall conform to 2.14.3.1 using AVC access unit timing values, as DTS or CPB removal time, and PTS or DPB removal time, associated with the SVC dependency representations of the video sub-bitstream in elementary stream ESH.

Bn buffer management

The following applies:

• There is exactly one elementary stream buffer BH for all the elementary streams in the set of decoded elementary streams as shown in Figure 2-16, where the size of BSH is defined by the P‑STD\_buffer\_size field in the PES packet header of elementary stream ESH.

• There is exactly one dependency representation buffer DRBn for each elementary stream in the set of decoded elementary streams as shown in Figure 2-16, where each dependency representation buffer DRBn in the set of decoded elementary streams is allocated within BSH. Even though the size DRBSn of individual DRBn is not constrained, the sum of the sizes DRBSn is constrained as follows:

BSH = ∑n (DRBSn)

where BSH is the size of the input buffer for the SVC video sub-bitstream in elementary stream ESH as defined in 2.14.3.2 for the re-assembled AVC video stream.

Access unit re-assembling and B removal

The following specifies the access unit re-assembling that results in AVC access unit AH(j):

i) Collect all SVC dependency representations DRn(jn) starting with the highest value of dependency\_id n, equal to H, to the lowest value of dependency\_id n, equal to m, present in access unit AH(j) following the rule below:

• For each two corresponding DRy+1(jy+1) and DRy(jy) of the SVC dependency representations collected for access unit AH(j), if TREF field is present for DRy+1(jy+1), the TREF value of DRy+1(jy+1) trefy+1(jy+1) shall be equal to the DTS value tdy(jy) of DRy(jy), otherwise the DTS value of tdy+1(jy+1) of DRy+1(jy+1) shall be equal to DTS value tdy(jy) of DRy(jy).

ii) Assemble the SVC dependency representations in consecutive order of dependency\_id n starting from 'm' to 'H' for the j-th access unit AH(j). If SEI NAL units are present in any SVC dependency representation with dependency\_id not equal to 0, these NAL units shall be re-ordered to the order of NAL units within an access unit as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10 before access unit re-assembling.

NOTE – The number of SVC dependency representations for each access unit AH(j) may vary depending on value of j. If m represents the lowest value of dependency\_id for access unit AH(j), then the collected and assembled SVC dependency representations include DRm(jm), DRm+1(jm+1), …, DRH(jH).

The following specifies the removal of access unit AH(j) from buffer BH:

At time tdH(jH) the AVC access unit AH(jH) shall be re-assembled and available for removal from buffer BH. The decoding time tdH(j) is specified by the DTS or by the CPB removal time that is associated with the SVC dependency representations in elementary stream ESH, as derived from information in the re-assembled AVC video stream.

STD delay

The STD delay for the re-assembled AVC access units shall follow the constraints specified in 2.14.3.2.

Buffer management conditions

Program streams shall be constructed so that the following conditions for buffer management are satisfied:

• BH shall not overflow.

• BH shall not underflow, except when VUI parameters are present for the AVC video sequence of the re‑assembled AVC video stream with the low\_delay\_hrd\_flag set to '1' or when trick\_mode status is true. Underflow of BH occurs for AVC access unit AH(j) when one or more bytes of AH(j) are not present in BH at the decoding time tdH(j).

#### 2.14.3.7 T-STD extensions for MVC and MVCD

The T-STD model described in 2.14.3.1 is applied if the received elementary stream is a video sub-bitstream of stream\_type 0x1B, i.e., only the AVC video sub-bitstream of MVC or MVC base view sub-bitstream is received and decoded.

When there is a set of received video sub-bitstreams and MVC video sub-bitstreams in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, of which dependencies may be signalled in the hierarchy descriptor, as defined in 2.6.7, and when there is at least one of the MVC video sub-bitstreams in the set of received elementary streams having the value of stream\_type equal to 0x20, the T-STD model as described in 2.14.3.1 is extended as illustrated in Figure 2-17 and as specified below.



Figure 2-17 – T-STD model extensions for Rec. ITU-T H.264 | ISO/IEC 14496-10 Video  
with MVC video sub-bitstreams

The following additional notations are used to describe the T-STD extensions and are illustrated in Figure 2-17 above.

ESn is the received elementary stream associated with the n-th MVC video sub-bitstream, where n is the index to the MVC view\_id subsets starting with value 0 for the MVC view\_id subset containing the base view and ordered according to the minimum view order index contained in each MVC video sub-bitstream

ESH is the received elementary stream associated with the H-th MVC video sub-bitstream which includes the view components with the highest view order index present in all MVC video sub-bitstreams of received elementary streams

j is an index to the re-assembled access units

jn is an index to the MVC view-component subsets of the elementary stream ESn associated with the n-th MVC video sub-bitstream

VSn(jn) is the jn-th MVC view-component subset of the MVC video sub-bitstream associated with ESn

AH(j) is the j-th access unit resulting from re-assembling (up to) the H-th MVC view-component subset associated with ESH

tdn(jn) is the decoding time, measured in seconds, in the system target decoder of the MVC view-component subset VSn(jn)

tdH(j) is the decoding time, measured in seconds, in the system target decoder of the j-th access unit AH(j) resulting from re-assembling (up to) the MVC view-component subset VSH(jH)

TBn is the transport buffer for elementary stream ESn

TBSn is the size of the transport buffer TBn, measured in bytes

MBn is the multiplexing buffer for elementary stream ESn

MBSn is the size of the multiplexing buffer MBn, measured in bytes

VSBn is the view component subset buffer for elementary stream ESn

VSBSn is the size of view component subset buffer VSBn, measured in bytes

EBH is the elementary stream buffer for the AVC video sub-bitstream of MVC and all MVC video sub-bitstreams

EBSH is the size of elementary stream buffer EBH, measured in bytes

Rxn transfer rate from TBn to MBn, as specified below

Rbxn transfer rate from MBn to VSBn, as specified below

Carriage in PES packets

For correct re-assembling of the MVC view-component subsets to an AVC access unit, the following applies:

• a PES packet per MVC view-component subset start shall be present, i.e., at most one MVC view‑component subset may commence in the same PES packet;

• the PTS and, if applicable, the DTS value shall be provided in the PES header of each MVC view‑component subset.

DPB buffer management

The DPB buffer management for the re-assembled AVC video stream shall conform to 2.14.3.1 using AVC access unit timing values, as DTS or CPB removal time, and PTS or DPB removal time, associated with the MVC view-component subsets of the MVC video sub-bitstream in elementary stream ESH.

TBn, MBn, EBn buffer management

The following applies:

• There is exactly one transport buffer TB, as defined in 2.14.3.1, for each received elementary stream in the set of received MVC video sub-bitstreams, including AVC video sub-bitstream of MVC, contained in elementary streams as shown in Figure 2-17.

• There is exactly one multiplexing buffer MB0 for the AVC video sub-bitstream of MVC in elementary stream ES0, where the size of the multiplexing buffer MBS0 is constrained as follows:

MBS0 = BSmux,0 + BSoh,0 + 1200 × MaxCPB[level]0 – cpb\_size0

where BSmux,0, BSoh,0 are defined in 2.14.3.1 for the AVC video sub-bitstream of MVC in elementary stream ES0.

where MaxCPB[level]0 and cpb\_size0 for the elementary stream ES0 are defined, as in 2.14.3.1.

NOTE 1 – If HRD parameters are present in at least one of the MVC video sub-bitstreams, those parameters have to be carefully handled in order to not unnecessarily increase the multiplexing buffers allocation.

• There is exactly one multiplexing buffer MBn for each received elementary stream associated with view order index value not equal to 0, where the size of each multiplexing buffer MBSn in the set of received elementary streams is constrained as follows:

MBSn = BSmux,n + BSoh,n

where BSmux,n, BSoh,n are defined in 2.14.3.1 for the AVC video stream resulting from re-assembling (up to) the MVC video sub-bitstream in elementary stream ESn.

• There is exactly one elementary stream buffer EBH for all the elementary streams in the set of received elementary streams as shown in Figure 2-17, of which the size EBSH has the following value:

EBSH = cpb\_sizeH

where cpb\_sizeH is the cpb\_size for the MVC video sub-bitstream in elementary stream ESH as defined in 2.14.3.1 for the re-assembled AVC video stream.

• There is exactly one view component subset buffer VSBn for each elementary stream in the set of received elementary streams as shown in Figure 2-17, where each view component subset buffer VSBn in the set of received elementary streams is allocated within EBH. Even though the size VSBSn of individual VSBn is not constrained, the sum of the sizes VSBSn is constrained as follows:

EBSH = ∑n (VSBSn)

• Transfer from TBn to MBn is applied as follows:

Rate Rxn:

when there is no data in TBn, then Rxn is equal to zero.

Otherwise: Rxn = bit\_rate

where bit\_rate is 1.2 × BitRate[ SchedSelIdx ] of data flow into the CPB for the byte stream format and BitRate[ SchedSelIdx ] is as defined in Annex E of Rec. ITU-T H.264 | ISO/IEC 14496-10 when NAL hrd\_parameters() is present in the VUI parameters of the MVC video sub-bitstream.

NOTE 2 – Annex E also specifies default values for BitRate[ SchedSelIdx] based on profile and level when NAL HRD parameters are not present in the VUI. The MVC video sub-bitstream level is determined by the level of AVC video stream resulting from re-assembling (up to) the associated MVC video sub-bitstream n in elementary stream ESn.

• Transfer from MBn to VSBn is applied as follows:

If the AVC\_timing\_and\_HRD\_descriptor is present with the hrd\_management\_valid\_flag set to '1' for elementary stream ESH, then the transfer of data from MBn to VSBn shall follow the HRD defined scheme for data arrival in the CPB of elementary stream ESH as defined in Annex C of Rec. ITU‑T H.264 | ISO/IEC 14496-10.

Otherwise, the leak method shall be used to transfer data from MBn to VSBn as follows:

Rate Rbxn:

Rbxn = 1200 × MaxBR[level]n

where MaxBR[level]n is defined for the byte stream format in Table A.1 (Level limits) in Rec. ITU‑T H.264 | ISO/IEC 14496-10 for the level of the AVC video stream resulting from re‑assembling (up to) the associated MVC video sub-bitstream n in elementary stream ESn. If there is PES packet payload data in MBn, and buffer EBH is not full, the PES packet payload is transferred from MBn to VSBn at a rate equal to Rbxn. If EBH is full, data are not removed from MBn. When a byte of data is transferred from MBn to VSBn, all PES packet header bytes that are in MBn and precede that byte are instantaneously removed and discarded. When there is no PES packet payload data present in MBn, no data is removed from MBn. All data that enters MBn leaves it. All PES packet payload data bytes enter VSBn instantaneously upon leaving MBn.

Access unit re-assembling and EB removal

The following specifies the access unit re-assembling that results in AVC access unit AH(j):

i) Assemble the MVC view-component subsets for the j-th access unit AH(j) following the rule below:

• For each two corresponding MVC view-component subsets VSy+1(jy+1) and VSy(jy) collected for access unit AH(j), where VSy is associated with a program element identified by the hierarchy\_layer\_index indicated in the associated hierarchy descriptor, and VSy+1 indicates the hierarchy\_layer\_index of VSy as the hierarchy\_embedded\_layer\_index in the hierarchy descriptor associated with program element associated with VSy+1, the DTS value of tdy+1(jy+1) of VSy+1(jy+1) shall be equal to DTS value tdy(jy) of VSy(jy).

NOTE 3 – If no hierarchy descriptor is present, VSy is associated with the AVC sub-bitstream and VSy+1 is associated with the MVC sub-bitstream.

ii) If SEI NAL units are present in any MVC view-component subset with view order index not equal to 0, these NAL units shall be re-ordered to the order of NAL units within an access unit, as defined in Rec. ITU‑T H.264 | ISO/IEC 14496-10, before access unit re-assembling.

The following specifies the removal of access unit AH(j) from buffer EBH:

At the decoding time tdH(j), the AVC access unit AH(j) shall be re-assembled and available for removal from buffer EBH. The decoding time tdH(j) is specified by the DTS or by the CPB removal time that is associated with the MVC view-component subsets in elementary stream ESH, as derived from information in the re‑assembled AVC video stream.

STD delay

The STD delay for re-assembled AVC access units shall follow the constraints specified in 2.14.3.1.

Buffer management conditions

Transport streams shall be constructed so that the following conditions for buffer management are satisfied:

• Each TBn shall not overflow and shall be empty at least once every second.

• Each MBn, EBH, and DPB shall not overflow.

• EBH shall not underflow, except when VUI parameters are present for the AVC video sequence of the re-assembled AVC video stream with the low\_delay\_hrd\_flag set to '1'. Underflow of EBH occurs for AVC access unit AH(j) when one or more bytes of AH(j) are not present in EBH at the decoding time tdH(j).

Buffer management for MVCD video sub-bitstream follows the same extensions as specified for MVC video in this clause. Carriage in PES packets for MVCD video sub-bitstream is specified below:

Carriage of MVCD sub-bitstream in PES packets

For correct re-assembling of the MVCD view-component subsets to an AVC access unit, the following applies:

• a PES packet per MVCD view-component subset start shall be present, i.e., at most one MVCD view‑component subset may commence in the same PES packet;

• the PTS and, if applicable, the DTS value shall be provided in the PES header of each MVCD view‑component subset.

#### 2.14.3.8 P-STD extensions for MVC

The P-STD model described in 2.14.3.2 is applied if the decoded elementary stream is a video sub-bitstream of stream\_type 0x1B, i.e., only the AVC video sub-bitstream of MVC or MVC base view sub-bitstream is decoded.

When there is a set of decoded MVC video sub-bitstreams in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, of which view order index values may be signalled in the MVC\_extension\_descriptor, as defined in 2.6.78, and when there is at least one of the MVC video sub-bitstreams in the set of decoded elementary streams having the value of stream\_type equal to 0x20, the P-STD model, as described in 2.14.3.2, is extended as illustrated in Figure 2-18 and as specified below.



Figure 2-18 – P-STD model extensions for Rec. ITU-T H.264 | ISO/IEC 14496-10 Video   
with MVC video sub-bitstreams

The following additional notations are used to describe the P-STD extensions and are illustrated in Figure 2-18 above.

ESn is the received elementary stream associated with the n-th MVC video sub-bitstream, where n is the index to the MVC view\_id subsets starting with value 0 for the MVC view\_id subset containing the base view and ordered according to the minimum view order index contained in each MVC view\_id subset

ESH is the received elementary stream associated with the H-th MVC video sub-bitstream which includes the view components with the highest view order index present in all MVC view\_id subsets of received elementary streams

j is an index to the re-assembled access units

jn is an index to the MVC view-component subsets of the elementary stream associated with the n-th MVC view\_id subset

VSn(jn) is the jn-th MVC view-component subset of the MVC video sub-bitstream associated with ESn

AH(j) is the j-th access unit resulting from re-assembling (up to) the H-th MVC view-component subset associated with ESH

tdn(jn) is the decoding time, measured in seconds, in the system target decoder of the MVC view‑component subset VSn(jn)

tdH(j) is the decoding time, measured in seconds, in the system target decoder of the j-th access unit AH(j) resulting from re-assembling (up to) the MVC view-component subset VSH(jH)

BH is the input buffer for all decoded MVC video sub-bitstreams

BSH is the size of the input buffer BH, measured in bytes

VSBn is the view component subset buffer for elementary stream ESn

VSBSn is the size of view component subset buffer VSBn, measured in bytes

Carriage in PES packets

For correct re-assembling of the MVC view-component subsets to an AVC access unit, the following applies:

• a PES packet per MVC view-component subset start shall be present, i.e., at most one MVC view‑component subset may commence in the same PES packet;

• the PTS and, if applicable, the DTS value shall be provided in the PES header of each MVC view‑component subset.

DPB buffer management

The DPB buffer management for the re-assembled AVC video stream shall conform to 2.14.3.1 using AVC access unit timing values, as DTS or CPB removal time, and PTS or DPB removal time, associated with the MVC view-component subsets of the MVC video sub-bitstream in elementary stream ESH.

Bn buffer management

The following applies:

• There is exactly one elementary stream buffer BH for all the elementary streams in the set of decoded elementary streams as shown in Figure 2-18, where the size of BSH is defined by the P‑STD\_buffer\_size field in the PES packet header of elementary stream ESH.

• There is exactly one view component subset buffer VSBn for each elementary stream in the set of decoded elementary streams as shown in Figure 2-18, where each view component subset buffer VSBn in the set of decoded elementary streams is allocated within BSH. Even though the size VSBSn of individual VSBn is not constrained, the sum of the sizes VSBSn is constrained as follows:

BSH = ∑n (VSBSn)

where BSH is the size of the input buffer for the MVC video sub-bitstream in elementary stream ESH, as defined in 2.14.3.2, for the re-assembled AVC video stream.

Access unit re-assembling and B removal

The following specifies the access unit re-assembling that results in AVC access unit AH(j):

i) Assemble the MVC view-component subsets for the j-th access unit AH(j) following the rule below:

• For each two corresponding MVC view-component subsets VSy+1(jy+1) and VSy(jy) collected for access unit AH(j), where VSy is associated with a program element identified by the hierarchy\_layer\_index indicated in the associated hierarchy descriptor, and VSy+1 references VSy by the hierarchy\_embedded\_layer\_index indicated in the hierarchy descriptor associated with program element associated with VSy+1, the DTS value of tdy+1(jy+1) of VSy+1(jy+1) shall be equal to DTS value tdy(jy) of VSy(jy).

NOTE – If no hierarchy descriptor is present, VSy is associated with the AVC sub-bitstream and VSy+1 is associated with the MVC sub-bitstream.

ii) If SEI NAL units are present in any MVC view-component subset with view order index not equal to 0, these NAL units shall be re-ordered to the order of NAL units within an access unit, as defined in Rec. ITU-T H.264 | ISO/IEC 14496-10, before access unit re-assembling.

The following specifies the removal of access unit AH(j) from buffer BH:

At the decoding time tdH(jH), the AVC access unit AH(jH) shall be re-assembled and available for removal from buffer BH. The decoding time tdH(j) is specified by the DTS or by the CPB removal time that is associated with the MVC view-component subsets in elementary stream ESH, as derived from information in the re‑assembled AVC video stream.

STD delay

The STD delay for the re-assembled AVC access units shall follow the constraints specified in 2.14.3.2.

Buffer management conditions

Program streams shall be constructed so that the following conditions for buffer management are satisfied:

• BH shall not overflow.

• BH shall not underflow, except when VUI parameters are present for the AVC video sequence of the re‑assembled AVC video stream with the low\_delay\_hrd\_flag set to '1', or when trick\_mode status is true. Underflow of BH occurs for AVC access unit AH(j) when one or more bytes of AH(j) are not present in BH at the decoding time tdH(j).

## 2.15 Carriage of ISO/IEC 14496-17 text streams

### 2.15.1 Introduction

This subclause defines the carriage of ISO/IEC 14496-17 elementary text streams within Rec. ITU-T H.222.0 | ISO/IEC 13818-1 systems, both for program and transport streams. Typically, an ISO/IEC 14496-17 text stream will be an element of an ISO/IEC 13818-1 program, as defined by the PMT in a transport stream and the PSM in a program stream. The carriage and buffer management of ISO/IEC 14496-17 text streams is defined using existing parameters from Rec. ITU-T H.222.0 | ISO/IEC 13818-1 such as PTS and DTS, as well as information from the ISO/IEC 14496-17 text stream. For this purpose, 2.15.3 specifies the decoding of ISO/IEC 14496-17 streams within the T-STD and P‑STD models, using the Hypothetical Text Decoder (HTD) defined in ISO/IEC 14496-17.

When an ISO/IEC 14496-17 text stream is carried in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, the ISO/IEC 14496-17 coded data shall be contained in PES packets, as defined in 2.15.2. Information needed to decode the ISO/IEC 14496-17 text stream is provided by the MPEG-4 text descriptor, specified in 2.6.70.

Applications that wish to transmit font streams as specified in ISO/IEC 14496-18, can reference the streamed fonts as specified in Annex A.1 of ISO/IEC 14496-17. To transmit ISO/IEC 14496-18 font streams, applications can use the data or object carousel as defined in ISO/IEC 13818-6.

### 2.15.2 Carriage in PES packets

ISO/IEC 14496-17 text streams are carried in PES packets, using one of the assigned 14 stream\_id\_extension values. If the textFormat field in TextConfig() is coded with the value 0x01, indicating Timed Text as specified in 3GPP TS 26.245, then the text stream consists of a concatenation of 3GPP Text Access Units in display order, whereby each Text Access Units consists of one or more so-called TTUs. For carriage in PES packets, alignment between PES packets and TTUs is required; i.e., the first byte in the payload of a PES packet carrying an ISO/IEC 14496-17 text stream shall be the first byte of a TTU, which is the first byte of the TTU header specified in 7.4.2 in ISO/IEC 14496-17. There is no further need for alignment; hence, the first TTU in a PES packet may contain for example a non-first fragment of a Text Access unit or a Sample Description.

A PTS shall be coded in the PES packet header of each PES packet carrying an ISO/IEC 14496-17 text stream. The PTS shall refer to the first Text Access Unit that commences in the PES packet. A Text Access Unit commences in a PES packet if a TTU[1] with a complete text sample or the TTU[2] with the first fragment of a text sample is present in the PES packet. For identifying whether a TTU[2] carries the first or a subsequent text sample fragment, see 7.3.2.2 and 7.4.5 in ISO/IEC 14496-17.

### 2.15.3 STD extensions

#### 2.15.3.1 T-STD extensions

The T-STD model includes a transport buffer TBn and a multiplex buffer Bn prior to the Hypothetical Text Decoder defined in ISO/IEC 14496-17. transport stream packets with ISO/IEC 14496-17 data, as indicated by its PID, enter the buffer TBn. Bytes are removed from TBn to enter Buffer Bn at the rate Rxn. Delivery of bytes from Bn to enter the TTU Decoder in the HTD is at rate Rbxn. See also Figure 2-19.



Figure 2-19 – T-STD model extensions for ISO/IEC 14496-17 text streams

Buffer management

The size TBSn of TBn is equal to 512 bytes. For ISO/IEC 14496-17 text streams, the size BSn of Bn is equal to 4096 bytes. The rate Rxn is equal to 2 000 000 bits/s. The rate Rbxn meets the delivery schedule defined for ISO/IEC 14496-17 data at the input of the HTD in clause 7.7 of ISO/IEC 14496-17, i.e., Rbxn = R[profile, level] if the textSampleBuffer is not full and Rbxn = 0 if the textSampleBuffer is full, where R[profile, level] is defined as the profile and level specific rate R in clause 7.8 of ISO/IEC 14496-17.

Each of the following requirements shall apply:

– Buffer TBn shall not overflow.

– Buffer Bn shall not overflow.

– Each HTD requirement specified in clause 7.7 of ISO/IEC 14496-17 shall be met.

#### 2.15.3.2 P-STD extensions

The P-STD model for the decoding of an ISO/IEC 14496-17 text stream includes a multiplex buffer Bn prior to the Hypothetical Text Decoder defined in ISO/IEC 14496-17. For each ISO/IEC 14496-17 text stream n, the size BSn of buffer Bn in the P-STD is defined by the P-STD\_buffer\_size field in the PES packet header.

Each of the following requirements shall apply:

– Buffer Bn shall not overflow.

– Each HTD requirement specified in clause 7.7 of ISO/IEC 14496-17 shall be met.

## 2.16 Carriage of auxiliary video streams

ISO/IEC 23002-3 specifies auxiliary video streams. ISO/IEC 23002-3 auxiliary video streams can be carried over Rec. ITU-T H.222.0 | ISO/IEC 13818-1 streams as follows:

• in Table 2-27, 16 stream-id\_extension values are assigned to signal auxiliary video streams;

• in Table 2-34, one stream-type value is assigned to signal an auxiliary video stream;

• in Table 2-45, one descriptor tag is assigned to indicate an auxiliary video stream descriptor;

• in 2.6.74 the auxiliary video stream descriptor is specified;

• the auxiliary video stream descriptor is associated with each auxiliary video stream.

Auxiliary video streams provide additional information about a conventional primary video sequence, as specified in ISO/IEC 23002-3. The auxiliary video stream shall be synchronized with its primary video counterpart through the use of timestamps in the associated PES header based on the same PCR clock.

In case a program contains multiple video streams, it will be up to the application to specify the association between the video component and auxiliary video streams.

## 2.17 Carriage of HEVC

### 2.17.1 Constraints for the transport of HEVC

For HEVC video streams, HEVC temporal video sub-bitstreams or HEVC temporal video subsets, the following constraints additionally apply:

• Each HEVC access unit shall contain an access unit delimiter NAL unit.

NOTE 1 – HEVC requires that an access unit delimiter NAL unit, if present, is the first NAL unit within an HEVC access unit. Access unit delimiter NAL units simplify the ability to detect the boundary between HEVC access units.

• An HEVC video stream or HEVC temporal video sub-bitstream of an HEVC video stream conforming to one or more profiles defined in Annex A of Rec. ITU-T H.265 | ISO/IEC 23008-2 shall be an element of a Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 program and the stream\_type for this elementary stream shall be equal to 0x24.

NOTE 2 – Such a stream can be the HEVC base sub-partition of an HEVC video stream conforming to one or more profiles defined in Annex G or Annex H of Rec. ITU-T H.265 | ISO/IEC 23008-2.

• The video parameter sets, sequence parameter sets and picture parameter sets, as specified in Rec. ITU‑T H.265 | ISO/IEC 23008‑2, that are necessary for decoding an HEVC video stream or HEVC temporal video sub-bitstream shall be present within the elementary stream carrying that HEVC video stream or HEVC temporal video sub-bitstream.

• For each HEVC temporal video subset of an HEVC video stream conforming to one or more profiles defined in Annex A of Rec. ITU-T H.265 | ISO/IEC 23008‑2 that is an element of the same Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, the stream\_type for this elementary stream shall be equal to 0x25.

NOTE 3 – Such a stream can be an HEVC sub-partition of an HEVC video stream conforming to one or more profiles defined in Annex G or Annex H of Rec. ITU-T H.265 | ISO/IEC 23008-2.

• An HEVC enhancement sub-partition of an HEVC video stream conforming to one or more profiles defined in Annex G or Annex H of Rec. ITU-T H.265 | ISO/IEC 23008-2 shall be an element of an Rec. ITU-T H.222.0 | ISO/IEC 13818 1 program. The stream\_type for this elementary stream shall be set according to Table 2-34.

• The video parameter sets, sequence parameter sets, and picture parameter sets, as specified in Rec. ITU-T H.265 | ISO/IEC 23008-2, necessary for decoding an HEVC video stream, HEVC temporal video sub‑bitstream, or HEVC base sub-partition shall be present within the elementary stream carrying that HEVC video stream, HEVC temporal video sub-bitstream, or HEVC base sub-partition.

• When a Rec. ITU-T H.222.0 | ISO/IEC 13818 1 program includes one or more elementary streams with stream\_type equal to 0x28, 0x29, 0x2A or 0x2B, at least one HEVC operation point descriptor shall be present in the program map table associated with the program.

• When a Rec. ITU-T H.222.0 | ISO/IEC 13818 1 program includes more than one elementary stream with the same stream\_type value of 0x24, 0x25 or in the range of 0x28 .. 0x2B and hierarchy cannot be implied as specified in Table 2-137, one hierarchy descriptor as defined in 2.6.7 shall be present for each elementary stream with a stream\_type value of 0x24, 0x25 and one HEVC hierarchy extension descriptor as defined in 2.6.102 shall be present for each elementary stream with a stream\_type value in the range of 0x28 .. 0x2B.

NOTE 4 – Hierarchy descriptors or HEVC hierarchy extension descriptors are needed to assign a hierarchy layer index to each elementary stream if hierarchy cannot be implied, as specified in Table 2-137.

• In each elementary stream with stream\_type equal to 0x24 with a hierarchy descriptor, the hierarchy\_type in the hierarchy descriptor shall be equal to 15.

• In each elementary stream with stream\_type equal to 0x25 with a hierarchy descriptor, the hierarchy\_type in the hierarchy descriptor shall be equal to 3.

• The video parameter sets, sequence parameter sets and picture parameter sets, as specified in Rec. ITU‑T H.265 | ISO/IEC 23008‑2, that are necessary for decoding the HEVC highest temporal sub-layer representation of an HEVC temporal video subset shall be present within the elementary stream carrying the HEVC temporal video sub-bitstream associated by a hierarchy descriptor.

• The aggregation of the HEVC temporal video sub-bitstream with associated HEVC temporal video subsets according to the hierarchy descriptors, as specified in 2.17.3, shall result in a valid HEVC video stream.

NOTE 5 – The resulting HEVC video stream contains a set of temporal sub-layers, as specified in Rec. ITU-T H.265 | ISO/IEC 23008‑2, with TemporalId values forming a contiguous range of integer numbers.

• The aggregation, as specified in 2.17.4, of an HEVC enhancement sub-partition with all HEVC sub‑partitions according to the HEVC operation signalled in the HEVC operation point descriptor shall result in a valid layered HEVC video stream.

NOTE 6 – The resulting HEVC video stream is the HEVC operation point of that HEVC enhancement sub-partition.

• Each HEVC picture with nuh\_layer\_id larger than 0 shall be contained within an elementary stream with a stream\_type equal to either 0x28, 0x29, 0x2A or 0x2B.

• An elementary stream ESte with stream\_type equal to either 0x29 or 0x2B shall satisfy the following:

– The elementary stream ESte contains an HEVC temporal video subset which is a temporal enhancement for exactly one reference elementary stream ESref, i.e., the layer Li present in the reference elementary stream ESref shall also be present in the elementary stream ESte.

NOTE 7 – The reference elementary stream ESref is either an HEVC temporal video sub-bitstream with a stream\_type equal to either 0x28 or 0x2A or another HEVC temporal video subset with stream\_type equal to either 0x29 or 0x2B, respectively, for which the same applies.

Carriage in PES packets

Rec. ITU-T H.265 | ISO/IEC 23008‑2 video is carried in PES packets as PES\_packet\_data\_bytes, using one of the 16 stream\_id values assigned to video, while signalling the Rec. ITU-T H.265 | ISO/IEC 23008‑2 video stream, by means of the assigned stream-type value in the PMT (see Table 2-34). The highest level that may occur in an HEVC video stream, as well as a profile and tier that the entire stream conforms to should be signalled using the HEVC video descriptor. If an HEVC video descriptor is associated with an HEVC video stream, an HEVC temporal video sub-bitstream, an HEVC temporal video subset or an HEVC enhancement sub-partition, then this descriptor shall be conveyed in the descriptor loop for the respective elementary stream entry in the program map table. This Recommendation | International Standard does not specify the presentation of Rec. ITU-T H.265 | ISO/IEC 23008‑2 streams in the context of a program stream.

For PES packetization, no specific data alignment constraints apply. For synchronization and STD management, PTSs and, when appropriate, DTSs are encoded in the header of the PES packet that carries the ITU-T H.265 | ISO/IEC 23008‑2 video elementary stream data. For PTS and DTS encoding, the constraints and semantics apply as defined in 2.4.3.7 and 2.7.

DPB buffer management

Carriage of an HEVC video stream, an HEVC temporal video sub-stream, an HEVC temporal video subset or an HEVC enhancement sub-partition over Rec. ITU-T H.222.0 | ISO/IEC 13818-1 does not impact the size of the buffer DPB. For decoding of an HEVC video stream, an HEVC temporal video sub-bitstream or an HEVC temporal video sub-bitstream and its associated HEVC temporal video subsets in the STD, the size of DPB is as defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2. The DPB shall be managed as specified in Annex C of Rec. ITU-T H.265 | ISO/IEC 23008‑2 (clauses C.3 and C.5). A decoded HEVC access unit enters the DPB instantaneously upon decoding the HEVC access unit, hence at the CPB removal time of the HEVC access unit. A decoded HEVC access unit is presented at the DPB output time. If the HEVC video stream, HEVC temporal video sub-bitstream or HEVC temporal video subset provides insufficient information to determine the CPB removal time and the DPB output time of HEVC access units, then these time instants shall be determined in the STD model from PTS and DTS timestamps as follows:

1) The CPB removal time of HEVC access unit n is the instant in time indicated by DTS(n) where DTS(n) is the DTS value of HEVC access unit n.

2) The DPB output time of HEVC access unit n is the instant in time indicated by PTS(n) where PTS(n) is the PTS value of HEVC access unit n.

NOTE 8 – HEVC video sequences in which the *low\_delay\_hrd\_flag* in the syntax structure *hrd\_parameters()* is set to 1 carry sufficient information to determine the DPB output time and the CPB removal time of each HEVC access unit. Hence for HEVC access units for which STD underflow may occur, the CPB removal time and the DPB output time are defined by HRD parameters, and not by DTS and PTS timestamps.

NOTE 9 – An HEVC video stream may carry information to determine compliance of the HEVC video stream to the HRD, as specified in Annex C of Rec. ITU-T H.265 | ISO/IEC 23008‑2. The presence of this information can be signalled in a transport stream using the HEVC timing and HRD descriptor with the *hrd\_management\_valid\_flag* set to '1'. Irrespective of the presence of this information, compliance of an HEVC video stream to the T-STD ensures that HRD buffer management requirements for the CPB are met when each byte in the HEVC video stream is delivered to and removed from the CPB in the HRD at exactly the same instant in time at which the byte is delivered to and removed from EBn in the T-STD.

### 2.17.2 T-STD Extensions for single layer HEVC

When there is an HEVC video stream or HEVC temporal video sub-bitstream in an ITU-T H.222.0 | ISO/IEC 13818-1 program and there is no HEVC temporal video subset associated with this elementary stream of stream\_type 0x24 in the same ITU-T H.222.0 | ISO/IEC 13818-1 program, the T-STD model as described in 2.4.2.1 is extended as illustrated in Figure 2-20 and as specified below.



Figure 2-20 – T-STD model extensions for single layer HEVC

TBn, MBn, EBn buffer management

The following additional notations are used to describe the T-STD extensions and are illustrated in Figure 2-20.

t(i) indicates the time in seconds at which the i-th byte of the transport stream enters the system target decoder

TBn is the transport buffer for elementary stream n

TBS is the size of the transport buffer TBn, measured in bytes

MBn is the multiplexing buffer for elementary stream n

MBSn is the size of the multiplexing buffer MBn, measured in bytes

EBn is the elementary stream buffer for the HEVC video stream

j is an index to the HEVC access unit of the HEVC video stream

An(j) is the j-th access unit of the HEVC video bitstream

tdn (j) is the decoding time of An(j), measured in seconds, in the system target decoder

Rxn is the transfer rate from the transport buffer TBn to the multiplex buffer MBn as specified below.

Rbxn is the transfer rate from the multiplex buffer MBn to the elementary stream buffer EBn as specified below.

The following apply:

• There is exactly one transport buffer TBn for the received HEVC video stream or HEVC temporal video sub-bitstream where the size TBS is fixed to 512 bytes.

• There is exactly one multiplexing buffer MBn for the HEVC video stream or HEVC temporal video sub-bitstream, where the size MBSn of the multiplexing buffer MBis constrained as follows:

MBSn = BSmux + BSoh + CpbNalFactor × MaxCPB[tier, level] – cpb\_size

where BSoh, packet overhead buffering, is defined as:

BSoh = (1/750) seconds × max{ BrNalFactor × MaxBR[tier, level], 2 000 000 bit/s}

and BSmux, additional multiplex buffering, is defined as:

BSmux = 0.004 seconds × max{ BrNalFactor × MaxBR[tier, level], 2 000 000 bit/s}

MaxCPB[tier, level] and MaxBR[tier, level] are taken from Annex A of Rec. ITU-T H.265 | ISO/IEC 23008‑2 for the tier and level of the HEVC video stream or HEVC temporal video sub-bitstream where rates are expressed in bit/s. cpb\_size is taken from the HRD parameters, as specified in Annex E of Rec. ITU-T H.265 | ISO/IEC 23008‑2, included in the HEVC video stream or HEVC temporal video sub-bitstream, where the size is expressed in bits. Implicit conversion is carried out according to Note 2 in 2.4.2.4.

• There is exactly one elementary stream buffer EBn for all the elementary streams in the set of received elementary streams associated by hierarchy descriptors, with a total size EBSn

EBSn = cpb\_size

where cpb\_size is taken from the HRD parameters, as specified in Annex E of Rec. ITU-T H.265 | ISO/IEC 23008‑2, included in the HEVC video stream or the HEVC temporal video sub-bitstream, where the size is expressed in bits. Implicit conversion is carried out according to Note 2 in 2.4.2.4.

• Transfer from TBn to MBn is applied as follows:

When there is no data in TBn then Rxn is equal to zero. Otherwise:

Rxn = bit\_rate

where bit\_rate is BrNalFactor/BrVlcFactor x BitRate[ SchedSelIdx ] of data flow into the CPB for the byte stream format and BitRate[ SchedSelIdx ] is as defined in Annex E of Rec. ITU-T H.265 | ISO/IEC 23008‑2 when NAL hrd\_parameters() is present in the VUI parameters of the HEVC video stream.

NOTE – Annex E also specifies default values for BitRate[ SchedSelIdx] based on profile, tier and level when NAL HRD parameters are not present in the VUI.

• Transfer from MBn to EBn is applied as follows:

If the *HEVC\_timing\_and\_HRD\_descriptor* is present with the *hrd\_management\_valid\_flag* set to '1' for the elementary stream, then the transfer of data from MBn to EBn shall follow the HRD defined scheme for data arrival in the CPB of the elementary stream as defined in Annex C of Rec. ITU-T H.265 | ISO/IEC 23008‑2.

Otherwise, the leak method shall be used to transfer data from MBn to EBn as follows:

Rbxn = BrNalFactor × MaxBR[tier, level]

where MaxBR[tier, level] is taken from Annex A of Rec. ITU-T H.265 | ISO/IEC 23008‑2 for the tier and level of the HEVC video stream or HEVC temporal video sub-bitstream.

If there is PES packet payload data in MBn, and buffer EBn is not full, the PES packet payload is transferred from MBn to EBn at a rate equal to Rbxn. If EBn is full, data are not removed from MBn. When a byte of data is transferred from MBn to EBn, all PES packet header bytes that are in MBn and precede that byte are instantaneously removed and discarded. When there is no PES packet payload data present in MBn, no data is removed from MBn. All data that enters MBn leaves it. All PES packet payload data bytes enter EBn instantaneously upon leaving MBn.

STD delay

The STD delay of any ITU-T H.265 | ISO/IEC 23008‑2 data other than HEVC still picture data through the system target decoders buffers TBn, MBn, and EBn shall be constrained by tdn(j) – t(i) ≤ 10 seconds for all j, and all bytes i in access unit An(j).

The delay of any HEVC still picture data through the system target decoders TBn, MBn, and EBn shall be constrained by tdn(j) – t(i) ≤ 60 seconds for all j, and all bytes i in access unit An(j).

Buffer management conditions

Transport streams shall be constructed so that the following conditions for buffer management are satisfied:

• Each TBn shall not overflow and shall be empty at least once every second.

• Each MBn, EBn and DPB shall not overflow.

• EBn shall not underflow, except when VUI parameters are present for the HEVC video sequence with the *low\_delay\_hrd\_flag* set to '1'. Underflow of EBn occurs for HEVC access unit An(j) when one or more bytes of An(j) are not present in EBn at the decoding time tdn(j).

### 2.17.3 T-STD extensions for layered transport of HEVC temporal video subsets

When there is an HEVC video sub-bitstream and at least one associated elementary stream of type 0x25 in an ITU‑T H.222.0 | ISO/IEC 13818-1 program, the T-STD model as described in 2.4.2.1 is extended as illustrated in Figure 2‑21 and as specified below.



Figure 2-21 – T-STD model extensions for layered transport of HEVC temporal video subsets

The following additional notations are used to describe the T-STD extensions and are illustrated in Figure 2-21.

t(i) indicates the time in seconds at which the i-th byte of the transport stream enters the system target decoder.

H is the number of received HEVC temporal video subsets, associated by hierarchy descriptors with the same HEVC temporal video sub-bitstream.

k is an index identifying the H+1 received elementary streams which contain exactly one HEVC temporal video sub-bitstream and H HEVC temporal video subsets associated by hierarchy descriptors. The index value k equal to 0 identifies the elementary stream which contains the HEVC temporal video sub-bitstream and index values k ranging from 1 up to H identify the associated HEVC temporal video subsets.

ESn,k is the received elementary stream which contains the k-th HEVC temporal video subset or the HEVC temporal video sub-bitstream if k equals 0.

ESn,H is the received elementary stream containing the highest HEVC temporal video subset present in the set of received elementary streams.

PIDH is the packet identifier value which identifies ESn,H.

j is an index to the output access units.

An(j) is the j-th access unit of the HEVC complete temporal representation.

tdn(j) is the decoding time of An(j) in the system target decoder.

TBn,k is the transport buffer for elementary stream k.

TBSn,k is the size of the transport buffer TBn,k, measured in bytes.

MBn,k is the multiplexing buffer for elementary stream k.

MBSn,k is the size of the multiplexing buffer MBn,k, measured in bytes.

EBn is the elementary stream buffer for the received HEVC temporal video sub-bitstream ESn,0 and the received HEVC temporal video subsets ESn,1 to ESn,H.

EBSn is the size of elementary stream buffer EBn, measured in bytes.

Rxn,k is the transfer rate from the k-th transport buffer TBn,k to the k-th multiplex buffer MBn,k as specified below.

Rbxn,k is the transfer rate from the k-th multiplex buffer MBn,k to the elementary stream buffer EBn as specified below.

NOTE – The index n, where used, indicates that the received elementary streams and associated buffers belong to a certain HEVC temporal video sub-bitstream and its associated HEVC temporal video subsets, distinguishing these elementary streams and associated buffers from other elementary streams and buffers, maintaining consistency with the notation in Figure 2-20.

TBn,k, MBn,k, EBn buffer management

The following apply:

• There is one transport buffer TBn,k for each received elementary stream ESn,k, where the size TBSn,k is fixed to 512 bytes.

• There is one multiplex buffer MBn,k for each received elementary stream ESn,k, where the size MBSn,k of the multiplex buffer MBn,k is constrained as follows:

MBSn,k = BSmux + BSoh + CpbNalFactor × MaxCPB[tier, level] – cpb\_size

where

BSoh, packet overhead buffering, and BSmux, additional multiplex buffering, are as specified in 2.17.2;

MaxCPB[tier, level] and MaxBR[tier, level] are taken from the tier and level specification of HEVC for the tier and level of the HEVC highest temporal sub-layer representation associated with ESn,k;

cpb\_size is taken from the HRD parameters, as specified in Annex E of Rec. ITU-T H.265 | ISO/IEC 23008‑2, included in the HEVC highest temporal sub-layer representation associated with ESn,k. In the HRD parameters, cpb\_size is expressed in bits, and its value is implicitly converted into a value expressed in bytes according to Note 2 in 2.4.2.4.

• There is exactly one elementary stream buffer EBn for the H + 1 elementary streams in the set of received elementary streams ESn,0 to ESn,H, with a total size EBSn

EBSn = cpb\_size

where cpb\_size is taken from the HRD parameters, as specified in Annex E of Rec. ITU-T H.265 | ISO/IEC 23008‑2, included in the HEVC highest temporal sub-layer representation associated with ESn,H. In the HRD parameters, cpb\_size is expressed in bits, and its value is implicitly converted into a value expressed in bytes according to Note 2 in 2.4.2.4.

• Transfer from TBn,k to MBn,k is applied as follows:

When there is no data in TBn,k then Rxn,k is equal to zero. Otherwise:

Rxn,k = bit\_rate

where bit\_rate is as specified in 2.17.2.

• Transfer from MBn,k to EBn is applied as follows:

If the HEVC\_timing\_and\_HRD\_descriptor is present with the hrd\_management\_valid\_flag set to '1' for the HEVC video sub-bitstream, then the transfer of data from MBn,k to EBn shall follow the HRD defined scheme for data arrival in the CPB of elementary stream ESn,H as defined in Annex C of Rec. ITU-T H.265 | ISO/IEC 23008‑2.

Otherwise, the leak method shall be used to transfer data from MBn,k to EBn as follows:

Rbxn,k = BrNalFactor × MaxBR[tier, level]

where MaxBR[tier, level] is defined for the byte stream format in Annex A of Rec. ITU-T H.265 | ISO/IEC 23008‑2 for the tier and level of the HEVC video stream or the HEVC highest temporal sub-layer representation associated with ESn,k.

If there is PES packet payload data in MBn,k, and EBn is not full, the PES packet payload is transferred from MBn,k to EBn at a rate equal to Rbxn,k. If EBn is full, data are not removed from MBn,k. When a byte of data is transferred from MBn,k to EBn,, all PES packet header bytes that are in MBn,k and precede that byte are instantaneously removed and discarded. When there is no PES packet payload data present in MBn,k, no data is removed from MBn,k. All data that enters MBn,k leaves it. All PES packet payload data bytes enter EBn instantaneously upon leaving MBn,k.

At the output of the elementary stream buffer EBn, the elementary streams are aggregated by removing all HEVC access units in ascending DTS order and transferring them to the HEVC decoder DH, irrespective of which elementary stream ESn,k each HEVC access unit belongs to.

STD delay

The STD delay of any ITU-T H.265 | ISO/IEC 23008‑2 data other than HEVC still picture data through the system target decoders buffers TBn,k, MBn,k, and EBn shall be constrained by tdn(j) – t(i) ≤ 10 seconds for all k, all j, and all bytes i in access unit An(j).

The delay of any HEVC still picture data through the system target decoders TBn,k, MBn,k, and EBn shall be constrained by tdn(j) – t(i) ≤ 60 seconds for all k, all j, and all bytes i in access unit An(j).

Buffer management conditions

Transport streams shall be constructed so that the following conditions for buffer management are satisfied:

• Each TBn,k shall not overflow and shall be empty at least once every second.

• Each MBn,k, EBn, and DPB shall not overflow.

• EBn shall not underflow, except when VUI parameters are present for the HEVC video sequence with the *low\_delay\_hrd\_flag* set to '1'. Underflow of EBn occurs for HEVC access unit An(j) when one or more bytes of An(j) are not present in EBn at the decoding time tdn(j).

### 2.17.4 T-STD extensions for layered transport of HEVC sub-partitions with bitstream-partition-specific CPB operation

When there is at least one elementary stream with stream\_type value in the range of 0x28 to 0x2B in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, the T-STD model as described in 2.4.2.1 is extended for elementary streams with a stream\_type value in the range of 0x28 to 0x2B as illustrated in Figure 2-22 and as specified below.

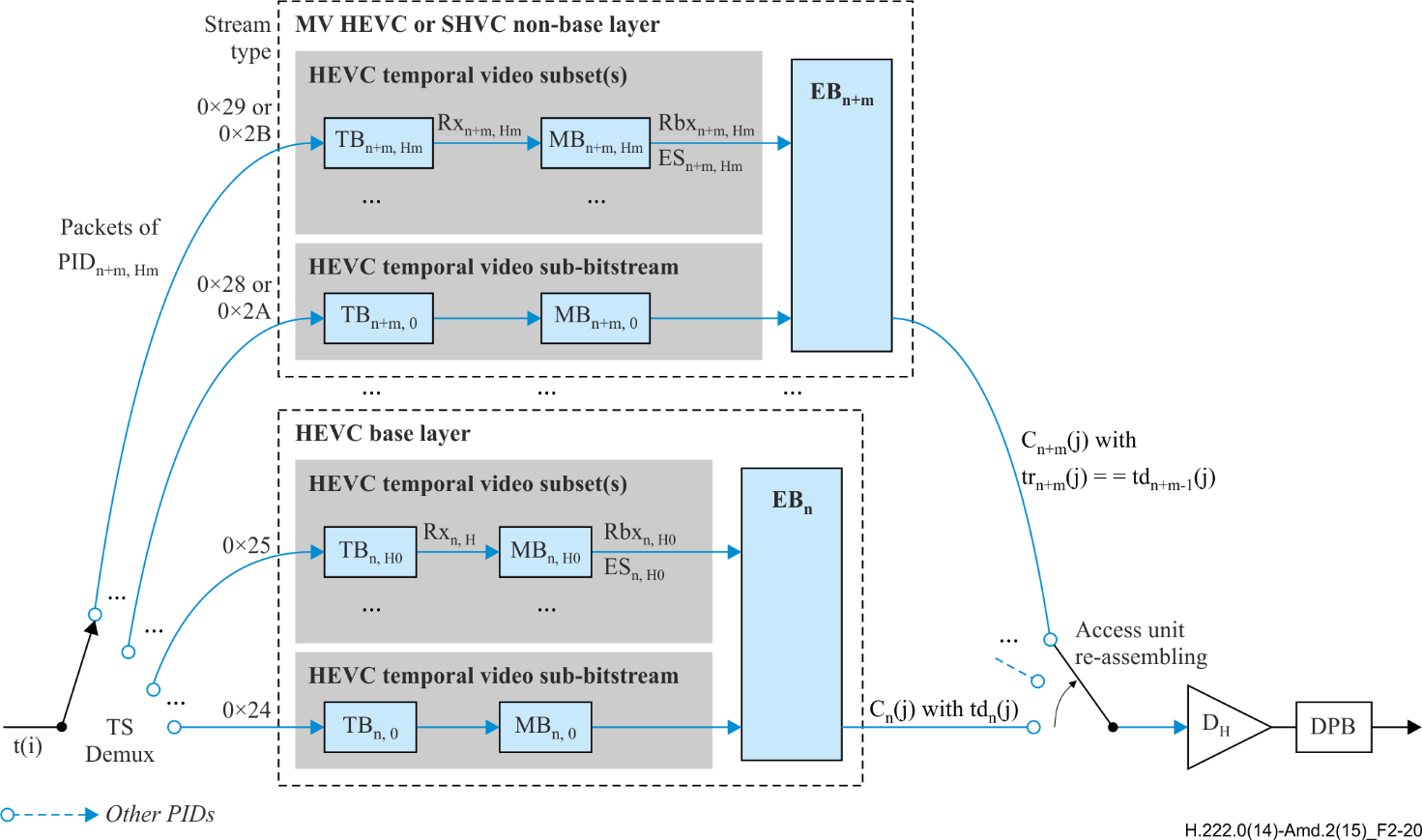


Figure 2-22 – T-STD model extensions for bitstream-partition-specific CPB operation

NOTE – The lower dashed box containing blocks that handle the HEVC base layer represents the T-STD buffer model as specified in 2.17.2, if there are no HEVC temporal video subsets, or 2.17.3, if the program contains at least one program element with stream\_type equal to 0x25.

The following additional notations are used to describe the T-STD extensions and are illustrated in Figure 2-22 above.

t(i) Indicates the time in seconds at which the i-th byte of the transport stream enters the system target decoder.

l Is an index into the received HEVC sub-partitions of stream\_type 0x28 or 0x2A (which include TemporalId 0). The order of HEVC sub-partitions is indicated by the HEVC operation point descriptor. The same index also applies to corresponding HEVC temporal enhancement sub-partitions. Here, l starts from n, which is associated with the HEVC base sub-partition, and runs up to (n+m), where m is specified below.

m Is the number of received HEVC sub-partitions of stream\_type 0x28 or 0x2A.

Hl Is the number of received HEVC corresponding temporal enhancement sub-partitions of the l-th received HEVC sub-partition of stream\_type 0x28 or 0x2A, associated by HEVC hierarchy extension descriptors with the same HEVC base sub-partition.

ESl,k Is the received elementary stream which contains the k-th HEVC corresponding temporal enhancement sub-partition of the l-th received HEVC sub-partition of stream\_type 0x28 or 0x2A, or the l-th HEVC sub-partition of stream\_type 0x28 or 0x2A if k equals 0.

ESn+m,Hm Is the received elementary stream which contains the HEVC sub-partition of the highest HEVC operation point in the set of received elementary streams.

PID n+m,Hm Is the packet identifier value which identifies ESn+m,Hm.

j Is an index to the output HEVC access units.

Cl(j) Is the j-th HEVC layer component of the l-th received HEVC sub-partition of stream\_type 0x28 or 0x2A or HEVC corresponding temporal enhancement sub-partition.

An(j) Is the j-th HEVC access unit of the HEVC complete temporal representation.

tdn(j) Is the decoding time of An(j) in the system target decoder.

trl(j) Is the value of TREF, if available in the PES header attached to Cl(j), else the decoding time of An(j) in the system target decoder.

TBl,k Is the transport buffer for elementary stream ESl,k.

TBSl,k Is the size of the transport buffer TBl,k, measured in bytes.

MBl,k Is the multiplexing buffer for elementary stream ESl,k.

MBSl,k Is the size of the multiplexing buffer MBl,k, measured in bytes.

EBl Is the elementary stream buffer for the received HEVC temporal video sub-bitstream ESl,0 and the received HEVC temporal video subsets ESl,1 to ESl,H.

NOTE 1 – Each buffer EBl contains one partition as specified in Annex F of Rec. ITU-T H.265 | ISO/IEC 23008-2.

EBSl Is the size of elementary stream buffer EBl, measured in bytes.

Rxl,k Is the transfer rate from the k-th transport buffer TBl,k to the k-th multiplex buffer MBl,k as specified below.

Rbxl,k Is the transfer rate from the k-th multiplex buffer MBl,k to the elementary stream buffer EBl as specified below.

NOTE 2 – The index n, where used, indicates that the received elementary streams and associated buffers belong to a certain HEVC base sub-partition, distinguishing these elementary streams and associated buffers from other elementary streams and buffers, maintaining consistency with the notation in Figure 2-1 and other T-STD extensions.

TBl,k, MBl,k, EBl buffer management

The following applies:

• There is one transport buffer TBl,k for each received elementary stream ESl,k, where the size TBSl,k is fixed to 512 bytes.

• There is one multiplex buffer MBl,k for each received elementary stream ESl,k, where the size MBSl,k of the multiplex buffer MBl,k is constrained as follows:

MBSl,k = BSmux + BSoh + CpbNalFactor × MaxCPB[tier, level] – cpb\_size

where

BSoh, packet overhead buffering, and BSmux, additional multiplex buffering, are as specified in 2.17.2;

MaxCPB[tier, level] and MaxBR[tier, level] are taken from the tier and level specification of the HEVC for the tier and level of ESl,k the HEVC operation point associated with ESl,k;

cpb\_size is taken from the sub-layer HRD parameters within the applicable hrd\_parameters( ), as specified in Annex F of Rec. ITU-T H.265 | ISO/IEC 23008-2, for the HEVC operation point associated with ESl,k. In the HRD parameters, cpb\_size is expressed in bits, and its value is implicitly converted into a value expressed in bytes according to Note 2 in 2.4.2.4.

• There is one elementary stream buffer EBl for the Hl + 1 elementary streams in the set of received elementary streams ESl,0 to ESlm,Hl, with a total size EBSl

EBSl = cpb\_size

where cpb\_size is taken from the sub-layer HRD parameters within the applicable hrd\_parameters( ), as specified in Annex F of Rec. ITU-T H.265 | ISO/IEC 23008-2, for the HEVC operation point associated with ESl,H. In the HRD parameters, cpb\_size is expressed in bits, and its value is implicitly converted into a value expressed in bytes according to Note 2 in 2.4.2.4.

• Transfer from TBl,k to MBl,k is applied as follows:

– When there is no data in TBl,k then Rxl,k is equal to zero.

– Otherwise, Rxl,k = bit\_rate

bit\_rate is taken from the NAL HRD parameters within the applicable hrd\_parameters( ), as specified in Annex F of Rec. ITU-T H.265 | ISO/IEC 23008-2, for the HEVC operation point associated with ESl,k, if sub-layer HRD parameters are present in the VPS and nal\_hrd\_parameters\_present\_flag is set to '1'.

Otherwise:

bit\_rate = BrNalFactor / BrVclFactor × BitRateVCL, if sub-layer HRD parameters within the applicable hrd\_parameters( ), as specified in Annex F of Rec. ITU-T H.265 | ISO/IEC 23008-2, for the HEVC operation point associated with ESl,k are present in the VPS and vcl\_hrd\_parameters\_present\_flag is set to '1'; BitRate VCL is taken from the VCL HRD parameters.

BrNalFactor and BrVclFactor are as defined in Rec. ITU-T H.265 | ISO/IEC 23008-2 for the profile, tier, and level of ESl,k the HEVC operation point associated with ESl,k.

In both cases, the index of the applicable hrd\_parameters( ) syntax structure in the active VPS of the HEVC video stream is given by: bsp\_hrd\_idx[ target\_ols ][ 0 ][ max\_temporal\_id ][ target\_schedule\_idx ][l + 1] as specified in Annex F of Rec. ITU-T H.265 | ISO/IEC 23008-2, where target\_ols and max\_temporal\_id are signalled in the HEVC operation point descriptor for the HEVC operation point associated with ESl,k, and target\_schedule\_idx is signalled in the HEVC timing and HRD descriptor.

• Transfer from MBl,k to EBl is applied as follows:

– If the HEVC\_timing\_and\_HRD\_descriptor is present with the hrd\_management\_valid\_flag set to '1' for the HEVC video sub-bitstream, then the transfer of data from MBl,k to EBl shall follow the HRD defined scheme for data arrival in the CPB of elementary stream ESl,H as defined in Annex C of Rec. ITU-T H.265 | ISO/IEC 23008-2.

– Otherwise, the leak method shall be used to transfer data from MBl,k to EBl as follows:

Rbxn,k = BrNalFactor × MaxBR[tier, level]

where MaxBR[tier, level] is as defined for the byte stream format in the tier and level specification of Rec. ITU T H.265 | ISO/IEC 23008-2 (Table A.2) for the tier and level for ESl,k signalled in the HEVC operation point descriptor for ESl,k.

If there is PES packet payload data in MBl,k, and EBl is not full, the PES packet payload is transferred from MBl,k to EBl at a rate equal to Rbxl,k. If EBl is full, data are not removed from MBl,k. When a byte of data is transferred from MBl,k to EBl, all PES packet header bytes that are in MBl,k and precede that byte are instantaneously removed and discarded. When there is no PES packet payload data present in MBl,k, no data is removed from MBl,k. All data that enters MBl,k leaves it. All PES packet payload data bytes enter EBl instantaneously upon leaving MBl,k.

Aggregation of elementary streams

The HEVC layer list for an HEVC operation point is the OperationPointESList[] associated with the HEVC operation point.

When there is no hierarchy descriptor or HEVC hierarchy extension descriptor present in the program map table associated with an Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, a value for the hierarchy\_layer\_index is implicitly assigned for each elementary stream with stream\_type 0x24, 0x25, 0x28, 0x29, 0x2A and 0x2B as described in Table 2‑161.

Table 2-161 – Implied hierarchy\_layer\_index if no hierarchy descriptors are used

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Existing stream types | Implied hierarchy\_layer\_index for program element with stream\_type value | | | | | |
| 0x24 | 0x25 | 0x28 | 0x29 | 0x2A | 0x2B |
| 0x24 | 0 |  |  |  |  |  |
| 0x24, 0x25 | 0 | 1 |  |  |  |  |
| 0x24, 0x28 | 0 |  | 1 |  |  |  |
| 0x24, 0x25, 0x28 | 0 | 1 | 2 |  |  |  |
| 0x24, 0x25, 0x28, 0x29 | 0 | 1 | 2 | 3 |  |  |
| 0x24, 0x2A | 0 |  |  |  | 1 |  |
| 0x24, 0x2A, 0x2B | 0 |  |  |  | 1 | 2 |
| 0x24, 0x25, 0x2A | 0 | 1 |  |  | 2 |  |
| 0x24, 0x25, 0x2A, 0x2B | 0 | 1 |  |  | 2 | 3 |

The HEVC operation point is aggregated from the HEVC layer components at the output of the elementary stream buffers EBl by determining the value of TREF, if available in the PES header, or else TREF is set to DTS, for the next HEVC layer component at the output of EBn+m, and gathering all HEVC layer components with a DTS equal to TREF, in the order given by the HEVC layer list as specified above, and transferring them to the HEVC decoder DH.

Carriage in PES packets

For correct re-assembling of the HEVC layer components to an HEVC access unit, if there is an HEVC dependency representation in the same HEVC access unit in more than one elementary stream, the following applies:

• Each PES packet shall contain exactly one HEVC layer component;

• The PTS and, if applicable, the DTS value shall be provided in the PES header of each HEVC layer component;

• If the DTS value of the HEVC layer component in an elementary stream is different from the DTS value of the HEVC layer component of the same HEVC access unit in an ES listed in the hierarchy descriptor or the HEVC hierarchy extension descriptors of the first elementary stream, the TREF field as defined in 2.4.3.7 shall be present in the PES header extension of the HEVC layer component of first elementary stream and the TREF field value shall be equal to the DTS value of the HEVC layer component of the second elementary stream.

STD delay

The STD delay of any Rec. ITU-T H.265 | ISO/IEC 23008-2 data other than HEVC still picture data through the System Target Decoders buffers TBl,k, MBl,k, and EBl shall be constrained by tdl(j) – t(i) ≤ 10 seconds for all l, all k, all j, and all bytes i in HEVC access unit An(j).

The delay of any HEVC still picture data through the system target decoders TBl,k, MBl,k, and EBl shall be constrained by tdl(j) – t(i) ≤ 60 seconds for all l. all k, all j, and all bytes i in HEVC access unit An(j).

Buffer management conditions

Transport streams shall be constructed so that the following conditions for buffer management are satisfied:

• Each TBl,k shall not overflow and shall be empty at least once every second;

• Each MBl,k, EBl, and DPB shall not overflow;

• EBl shall not underflow, except when VUI parameters are present for the HEVC video sequence with the low\_delay\_hrd\_flag set to '1'. Underflow of EBl occurs for HEVC access unit An(j) when one or more bytes of An(j) are not present in EBl at the decoding time tdn(j).

### 2.17.5 T-STD Extensions for the carriage of HEVC tiles

#### 2.17.5.1 Carriage of HEVC motion-constrained tile sets

There are alternative options to carry the HEVC tile substreams:

I. Each HEVC tile substream is carried in a separate ES.

II. More than one HEVC tile substreams is carried in a common ES. Within this ES, HEVC tile substream af\_descriptors are used to indicate different HEVC tile substreams.

In the first approach, the sender generates one slice per tile and all slice headers are stripped off before packetizing the slice data of each slice into a PES packet, building a separate ES per HEVC tile substream. In addition, it generates a separate ES per subregion layout that provides suitable slice headers and parameters sets that, when combined with the slice data, result in a compliant HEVC bitstream. All ESs with stream type equal to 0x31 that belong to the same program shall be referenced in the HEVC subregion descriptor that is associated to that program. If using stream type equal to 0x31, all ESs referring to the picture shall be of type equal to 0x31.

The second approach eases the allocation of receiver buffers while it allows to remove HEVC tile substreams that are not part of the desired subregion from the output of the demultiplexer before they reach the CPB of the decoder. The usage of HEVC tile substream af\_descriptors is indicated in the HEVC subregion descriptor. HEVC tile substream af\_descriptors assign a unique substream ID to each TS packet. Adaptation field padding shall be used in order to align the end of an HEVC tile substream for an access unit with the end of a TS packet.

#### 2.17.5.2 Carriage of HEVC motion-constrained tile sets as separate elementary streams

When HEVC motion-constrained tile sets are carried as separate elementary streams using the stream\_type value 0x31, the data in each SBi, where i ranges from n to (n+m) in Figure 2-23, belongs to a certain HEVC tile substream and data of multiple substream buffers can be gathered to form the encoded representation of a subregion. Which set of ESs needs to be aggregated in which order is signalled by the information conveyed in descriptors specified below, which are found in the PMT for that program. The buffer model shown in 2-23 is applicable when SubstreamMarkingFlag is set to '1'.

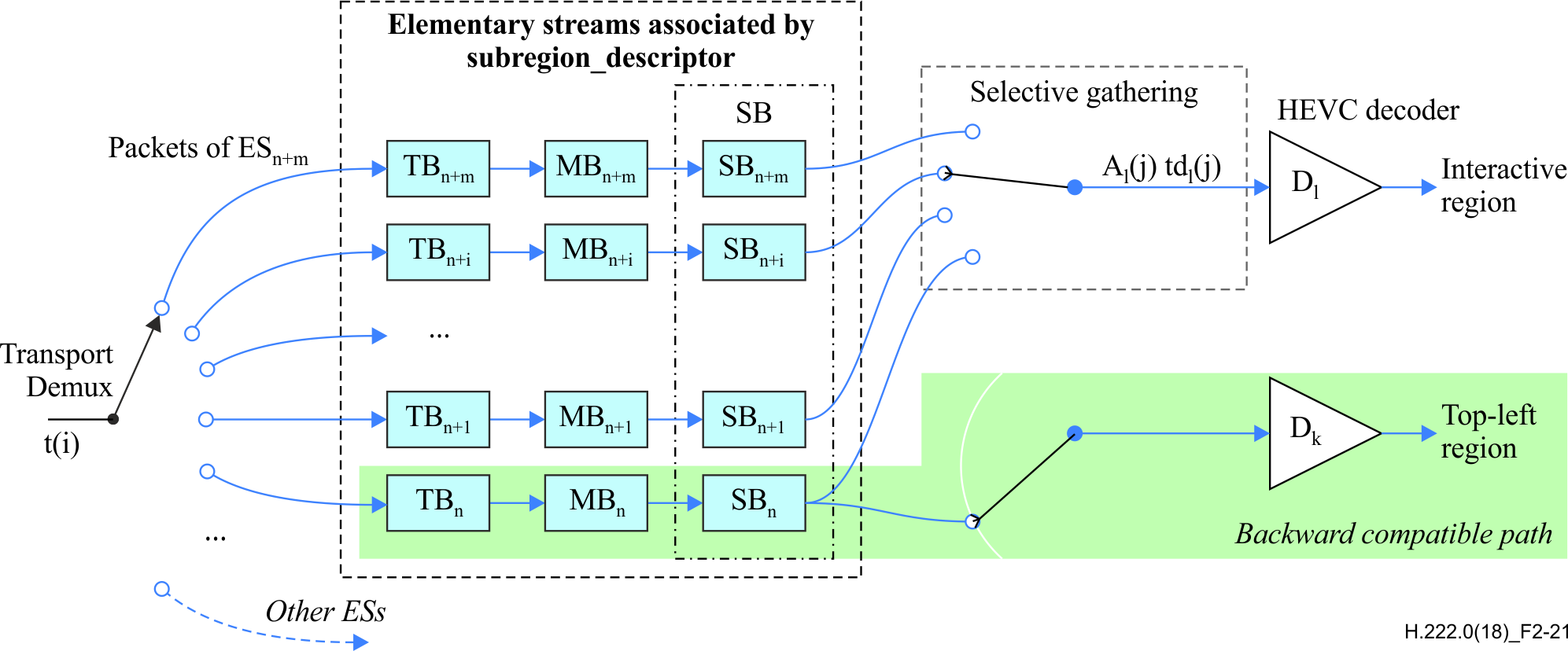


Figure 2-23 – T-STD model extensions for transport of HEVC tiles through individual ESs

For each subregion indexed by the value of i in the HEVC subregion descriptor, the parameter sets corresponding to the subregion shall be present in the HEVC tile substream with the SubstreamID equal to to PreambleSubstreamID[i] signalled in the HEVC subregion descriptor. In an HEVC tile substream with an HEVC tile substream descriptor with a PreambleFlag resp. Flag[0] value equal to '0', the parameter sets corresponding to the subregion shall also be present.

NOTE – If any HEVC tile substream descriptor of an ES that belongs to a certain program contains a PreambleFlag resp. Flag[0] equal to '0', only one subregion size can be signalled in the HEVC subregion descriptor for that program.

Process to access a subregion:

The HEVC subregion descriptor and all HEVC tile substream descriptors shall be read.

First, the receiver chooses a suitable subregion size from the HEVC subregion descriptor based on the level indication or the subregion dimension. This selection implicitly results in a value i.

Next, it selects the ES that contains the upper left HEVC tile substream of the region to be displayed based on the SubstreamID.

Next, it checks if the applicable HEVC tile substream descriptor provides a PatternReference value.

With that PatternReference, it selects the applicable SubstreamOffset values:

*SubstreamOffset[k][****PatternReference****][i] with 0 < k < SubstreamCountMinus1[i]*

If there is no explicit PatternReference value, which means that either descriptor\_length equals 1 or ReferenceFlag equals 0, the index j defaults to a value of 0.

For each PES packet of the ESx with an HEVC tile substream descriptor indicating a SubstreamID equal to SubstreamIDx, the following operations are executed, if the stated condition is met:

• If PreambleFlag resp. Flag[0] signalled in the HEVC tile substream descriptor equals '1': prepend the PES packets with the same DTS of the ES with an HEVC tile substream descriptor indicating a SubstreamID equal to PreambleSubstreamID[i] signalled in the HEVC subregion descriptor

• If SubstreamCountMinus1[i] > 0: append the PES packets with the same DTS of ESs with an HEVC tile substream descriptor indicating a SubstreamID equal to AdditionalSubstreamID[i] given in the HEVC tile substream descriptor of ESx resp. (SubstreamIDx + SubstreamOffset[k][j]), where the SubstreamOffset array is found in the subregion descriptor, with j given by the value of PatternReference in the HEVC tile substream descriptor of ESx and k ranging from 0 to SubstreamCountMinus1[i] prior to decoding. The order of the PES packets in the assembled bitstream is given by values of SubstreamID ordered with increasing values of the index k.

TBn, MBn, EBn buffer management

The following additional notations are used to describe the T-STD extensions and are illustrated in Figure 2-23.

H is the number of received HEVC tile substreams of the subregion.

ESn is the received elementary stream of the n-th HEVC tile substream of the subregion.

TBn is the transport buffer for the n-th elementary stream of the subregion.

TBSn is the size of the transport buffer TBn measured in bytes.

MBn is the multiplex buffer for the n-th elementary stream.

MBSn is the size of the multiplex buffer MBn measured in bytes.

SBn is the substream buffer for the n-th elementary stream.

SBSn is the size of the n-th substream buffer SBn measured in bytes.

SB is the substream buffer for all the received H elementary streams.

SBS is the size of elementary stream buffer EBH, measured in bytes.

The following applies:

• There is exactly one transport buffer TBn for each received elementary stream of the subregion as shown in Figure 2-23, where the size TBSn is fixed to 512 bytes.

• There is exactly one multiplex buffer MBn for each received elementary stream of the subregion as shown in Figure 2-23, where the size of the multiplex buffer MBSn is constrained as follows:

MBSn = BSmux + BSoh + CpbNalFactor × MaxCPB[tier, level] – cpb\_size (measured in bytes)

where BSoh, packet overhead buffering, is defined as:

BSoh = (1/750) seconds × max{ BrNalFactor × MaxBR[tier, level], 2 000 000 bit/s}

and BSmux, additional multiplex buffering, is defined as:

BSmux = 0.004 seconds × max{ BrNalFactor × MaxBR[tier, level], 2 000 000 bit/s}

MaxCPB[tier, level] and MaxBR[tier, level] are taken from Annex A of Rec. ITU-T H.265 | ISO/IEC 23008‑2 for the tier and level as indicated by the parameter sets set carried in the substream with the SubstreamID equal to to PreambleSubstreamID[i] signalled in the HEVC subregion descriptor. cpb\_size is taken from the HRD parameters, as specified in Annex E of Rec. ITU-T H.265 | ISO/IEC 23008‑2, present in the parameter sets carried in the substream with the SubstreamID equal to to PreambleSubstreamID[i] signalled in the HEVC subregion descriptor.

• There is exactly one subregion buffer SB for all the elementary streams in the set of received elementary streams of the subregion as shown in Figure 2-23, of which the size SBS has the following value:

SBS = cpb\_sizeH

where cpb\_sizeH is the cpb\_size taken from the HRD parameters, as specified in Annex E of Rec. ITU-T H.265 | ISO/IEC 23008‑2, present in the parameter sets carried in the substream with the SubstreamID equal to to PreambleSubstreamID[i] signalled in the HEVC subregion descriptor.

• There is exactly one substream buffer SBk for each elementary stream in the set of received elementary streams of the subregion as shown in Figure 2-23, where each substream buffer SBk in the set of received elementary streams of the subregion is allocated within SB. Even though the size of the substream buffer SBk is not constrained, the sum of the sizes of all substream buffers SBn..n+m is constrained as follows:

SBS = ∑k (SBSk) with n ≤ k ≤ n+m

NOTE – If PreambleFlag resp. Flag[0] signalled in the HEVC tile substream descriptor equals '1', the parameter sets present in the HEVC tile substream with the SubstreamID equal to to PreambleSubstreamID[i] signalled in the HEVC subregion descriptor are the same as the parameter sets encapsulated in the MCTS-EIS SEI message for the subregion. If PreambleFlag resp. Flag[0] signalled in the HEVC tile substream descriptor equals '0', the parameter sets present in that HEVC tile substream are the same as the parameter sets encapsulated in the MCTS-EIS SEI message for the subregion.

#### 2.17.5.3 Carriage of HEVC motion-constrained tile sets in a common ES using AF descriptors

The following approach allows very simple buffer allocation for the buffer that gathers the ES data for the case that SubstreamMarkingFlag is set to 0, because it allows that only a single ESn is used for the whole panorama. In this case, a third buffer is introduced after MB, called "RB" (subregion buffer) if it gathers the data that belongs to a subregion (cf. Figure 2-24), or “EB” if it gathers an elementary stream that contains the whole panorama (cf. Figure 2-25).

If SubstreamMarkingFlag is set to 0, no TS packet shall contain data of different HEVC tile substreams. The signalling using HEVC tile substream af\_descriptors as specified in U.3.15 allows removing unnecessary HEVC tile substreams before they enter the MB and RB buffers.

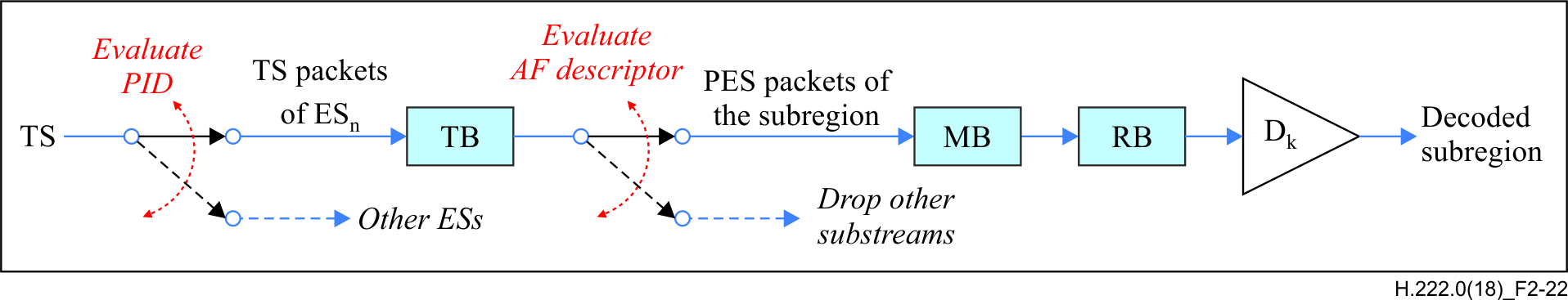


Figure 2-24 – T-STD model extensions for transport of HEVC tiles in a common ES using AF descriptors

The process to access a subregion is similar to the process described in 2.17.5.2:

First, the receiver chooses a suitable subregion size from the subregion descriptor based on the level indication or the subregion dimension. This selection implicitly results in a value l.

Next, it selects the SubstreamIDx that relates to the upper left HEVC tile substream of the region to be displayed.

All adaptation fields of TS packets that belong to ESn are read and HEVC tile substream af\_descriptors are evaluated whenever they occur. A SubstreamID that is indicated in an HEVC tile substream af\_descriptor applies to the sequence of TS packets of the same ESn starting with the TS packet that contains the HEVC tile substream af\_descriptor indicating that SubstreamID up to the TS packet preceding the TS packet that contains an HEVC tile substream af\_descriptor indicating a different SubstreamID.

NOTE 1 – Adaptation field padding is used in the last TS packet that belongs to a certain HEVC tile substream, if otherwise the data block belonging to that substream does not extend to the last payload byte of a TS packet.

The receiver sets up an empty list of SubstreamIDs that belong to the subregion and adds

• the value 0 and

• the SubstreamIDx of the upper left HEVC tile substream.

NOTE 2 – SubstreamID value 0 indicates a substream that belongs to any subregion, regardless of its position.

Next, it checks if the applicable HEVC tile substream af\_descriptor provides a PatternReference value. If there is no explicit PatternReference value, which means that either descriptor\_length equals 1 or ReferenceFlag equals 0, the index j defaults to a value of 0.

With that PatternReference, it selects the applicable SubstreamOffset values:

SubstreamOffset[k][PatternReference][i] with 0 < k < SubstreamCountMinus1[i]

For each TS packet of the ESn with an HEVC tile substream af\_descriptor indicating a SubstreamID equal to SubstreamIDx, the following operations are executed, if the stated condition is met:

• If SubstreamCountMinus1[i] > 0: add SubstreamID equal to AdditionalSubstreamID[i] to the list of SubstreamIDs that belong to the subregion, with i ranging from 0 to SubstreamCountMinus1[i]).

• Else if the SubstreamOffset array is found in the subregion descriptor (for the selected subregion layout indexed by i), add (SubstreamIDx + SubstreamOffset[k][j][i]) to the list of SubstreamIDs that belong to the subregion, with j given by the value of PatternReference in the HEVC tile substream af\_descriptor with SubstreamIDx and k ranging from 0 to SubstreamCountMinus1[i].

For each sequence of TS packets of the ESn starting with an HEVC tile substream af\_descriptor indicating a SubstreamID that is contained in the list of SubstreamIDs that belong to the subregion, the TS packets are stored in the MB.

TB, MB, RB buffer management

The following notations are used to describe the T-STD extensions and are illustrated in Figure 2-24.

H is the number of received HEVC tile substreams of the subregion.

ESn is the received elementary stream that contains all HEVC tile substreams.

TB is the transport buffer for the elementary stream that contains all HEVC tile substreams.

TBS is the size of the transport buffer TB, measured in bytes.

MB is the multiplex buffer for the elementary stream. It gathers HEVC tile substreams of the subregion.

MBS is the size of the multiplex buffer MB, measured in bytes.

RB is the subregion buffer for all the received H elementary streams of the subregion.

RBS is the size of the subregion buffer, measured in bytes.

The following applies:

• There is exactly one transport buffer TB for each received elementary stream of the subregion as shown in Figure 2‑24, where the size TBS is fixed to 512 bytes.

• There is exactly one multiplex buffer for each received elementary stream of the subregion as shown in Figure 2‑24, where the size of the multiplex buffer MBS is constrained as follows:

MBS = BSmux + BSoh + CpbNalFactor × MaxCPB[tier, level] – cpb\_size (measured in bytes)

where

BSoh, packet overhead buffering, is defined as:

BSoh = (1/750) seconds × max{ BrNalFactor × MaxBR[tier, level], 2 000 000 bit/s}

and BSmux, additional multiplex buffering, is defined as:

BSmux = 0.004 seconds × max{ BrNalFactor × MaxBR[tier, level], 2 000 000 bit/s}

MaxCPB[tier, level] and MaxBR[tier, level] are taken from Annex A of Rec. ITU-T H.265 | ISO/IEC 23008‑2 for the tier and level as indicated by the temporal motion-constrained tile sets SEI message or within the parameter sets in the motion-constrained tile sets extraction information set SEI message, as specified in D.2.30 and D.2.43 respectively of Rec. ITU-T H.265 | ISO/IEC 23008‑2, for the temporal motion-constrained tile set conformed by all tiles within the H substreams of the subregion. cpb\_size is taken from the HRD parameters, as specified in Annex E of Rec. ITU-T H.265 | ISO/IEC 23008‑2, present in the parameter sets, encapsulated in the MCTS-EIS SEI message, for the temporal motion-constrained tile set conformed by all tiles within the H HEVC tile substreams of the subregion, as defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2.

• There is exactly one subregion buffer RB for all HEVC tile substreams identified by the list of SubstreamIDs that belong to the subregion, as shown in Figure 2-24, of which the size RBS has the following value:

RBS = cpb\_sizeH

where cpb\_sizeH is the cpb\_size taken from the HRD parameters, as specified in Annex E of Rec. ITU-T H.265 | ISO/IEC 23008‑2, present in the parameter sets, encapsulated in the MCTS-EIS SEI message, for the temporal motion-constrained tile set conformed by all tiles within the H HEVC tile substreams of the subregion, as defined in Rec. ITU-T H.265 | ISO/IEC 23008‑2.

#### 2.17.5.4 Carriage of HEVC motion-constrained tile sets in a common ES ignoring AF descriptors

Alternatively, a receiver might decide to ignore the AF descriptors and select the HEVC tile substreams at bitstream level following the motion-constrained tile sets extraction information sets SEI message. In such a case, the profile and level required for decoding are the ones signalled for the subregion. The size of the buffers TB, MB, EB are specified below.

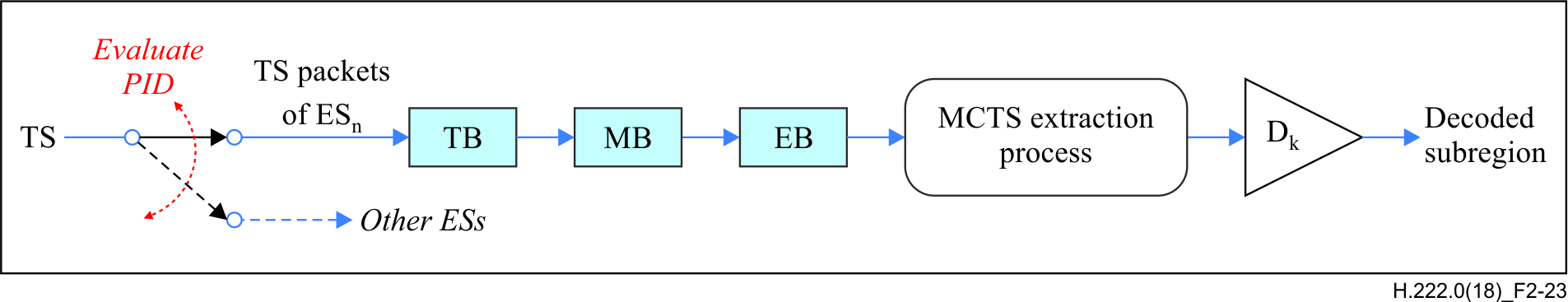


Figure 2-25 – T-STD model extension for transport of HEVC tiles in a common ES ignoring AF descriptors

TB, MB, EB buffer management

The following notations are used to describe the T-STD extensions and are illustrated in Figure 2-25 – T-STD model extension for transport of HEVC tiles in a common ES ignoring AF descriptors above.

ESn is the received elementary stream that contains all HEVC tile substreams.

TB is the transport buffer for the elementary stream that contains all HEVC tile substreams.

TBS is the size of the transport buffer TB, measured in bytes.

MB is the multiplex buffer for the elementary stream.

MBS is the size of the multiplex buffer MB, measured in bytes.

EB is the substream buffer for all the received H elementary streams of the subregion.

EBS is the size of the substream buffer, measured in bytes.

The following applies:

• There is exactly one transport buffer TB for the received elementary stream as shown in Figure 2‑25, where the size TBS is fixed to 512 bytes.

• There is exactly one multiplex buffer MB for the received elementary stream as shown in Figure 2‑25, where the size of the multiplex buffer MBS is constrained as follows:

MBS = BSmux + BSoh + CpbNalFactor × MaxCPB[tier, level] – cpb\_size (measured in bytes)

where BSoh, packet overhead buffering, is defined as:

BSoh = (1/750) seconds × max{ BrNalFactor × MaxBR[tier, level], 2 000 000 bit/s}

and BSmux, additional multiplex buffering, is defined as:

BSmux = 0.004 seconds × max{ BrNalFactor × MaxBR[tier, level], 2 000 000 bit/s}

MaxCPB[tier, level] and MaxBR[tier, level] are taken from Annex A of Rec. ITU-T H.265 | ISO/IEC 23008‑2 for the tier and level of the HEVC video stream or HEVC temporal video sub-bitstream. cpb\_size is taken from the HRD parameters, as specified in Annex E of Rec. ITU-T H.265 | ISO/IEC 23008‑2, included in the HEVC video stream or HEVC temporal video sub-bitstream.

• There is exactly one elementary stream buffer EB for all HEVC tile substreams within the elementary stream, as shown in Figure 2-25, of which the size EBS has the following value:

EBS = cpb\_size

where cpb\_size is the cpb\_size taken from the HRD parameters, as specified in Annex E of Rec. ITU-T H.265 | ISO/IEC 23008‑2, included in the HEVC video stream or the HEVC temporal video sub-bitstream.

## 2.18 Carriage of green access units

### 2.18.1 Carriage of green access units in MPEG-2 sections

Green access units are carried using the MPEG-2 private section syntax with the section\_syntax\_indicator element set to '0'. See Table 2-162.

| Table 2-162 – Green access unit section syntax | | |
| --- | --- | --- |
| Syntax | Bits | Mnemonic |
| Green\_access\_unit\_section\_message(){ |  |  |
| **table\_ID** | **8** | **uimsbf** |
| **section\_syntax\_indicator**  **private\_indicator** | **1**  **1** | **bslbf**  **bslbf** |
| **reserved** | **2** | **bslbf** |
| **private\_section\_length** | **12** | **uimsbf** |
| **'0010'**  **Display\_in\_PTS [32..30]**  **marker\_bit**  **Display\_in\_PTS [29..15]**  **marker\_bit**  **Display\_in\_PTS [14..0]**  **marker\_bit** | **4**  **3**  **1**  **15**  **1**  **15**  **1** | **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf** |
| Green\_AU() |  |  |
| **CRC\_32** | **32** | **rpchof** |
| } |  |  |

### 2.18.2 Semantics of green access unit section

**table\_id** – This shall be set to 0x09.

**section\_syntax\_indicator** – This shall be set to '0'.

**Display\_in\_PTS** – This is the 33-bit PTS specified similar to that defined in the PES header and is used with the associated video access unit.

**Green\_AU()** – Defined in 2.18.3.

### 2.18.3 Green access unit

The format of the green access unit is defined in Table 2-163. Green access units contain dynamic metadata and are carried in MPEG private section format.

Table 2-163 – Green access unit

| Syntax | No. bits | Mnemonic |
| --- | --- | --- |
| Green\_AU(){ |  |  |
| **num\_quality\_levels**  **reserved**  for (k=0; k < num\_constant\_backlight\_voltage\_time\_intervals; k++) {  for (j=0; j < num\_max\_variations; j++) {  **lower\_bound**  if (lower\_bound > 0)  **upper\_bound**  **rgb\_component\_for\_infinite\_psnr**  for (i=1; i <= num\_quality\_levels; i++){  **max\_rgb\_component**  **scaled\_psnr\_rgb**  }  }  }  } | **4**  **4**  **8**  **8**  **8**  **8**  **8** | **uimsbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf** |

As explained in 6.4 of ISO/IEC 23001-11, each combination of constant\_backlight\_voltage\_time\_interval and max\_variation is associated with contrast-enhancement metadata and a set of quality levels defined in Table 2‑139.

The metadata in the Green\_AU is applicable to the presentation subsystem until the next Green\_AU containing metadata arrives.

Semantics for all the elements in Table 2-139 is defined in 6.4 of ISO/IEC 23001-11.

### 2.18.4 Timing relationship between green access unit and media access unit

The green access unit should be decoded and information should be available before the associated media access unit is decoded. Such a timing relationship guarantees that the metadata within the green access unit is made available to the display with sufficient lead time relative to the PTS of the associated media access unit. Note that the PTS of the media access unit and the PTS of the green access unit are identical. The green access unit is transmitted in the transport stream with a sufficient lead time so that the display control settings can be adjusted in advance of presentation time for correct operation. If num\_constant\_backlight\_voltage\_time\_intervals > 1, then the lead time should be equal to or larger than the largest constant\_backlight\_voltage\_time\_interval. The PMT shall not contain more than one green metadata component (stream\_type equal to 0x2C).

NOTE – Applications that use carouseling of green access unit data carouseling in a given program can do so as long as the display\_in\_PTS value is adjusted to conform to the PCR clock and T-STD buffer.

### 2.18.5 Buffer model for processing green access units

The buffer model reflects the processing required to handle green access units. The model can be used to establish constraints which can be used to verify the validity of dynamic green metadata prepared in accordance with this standard.

Figure 2-26 illustrates the buffer model for processing green access units.

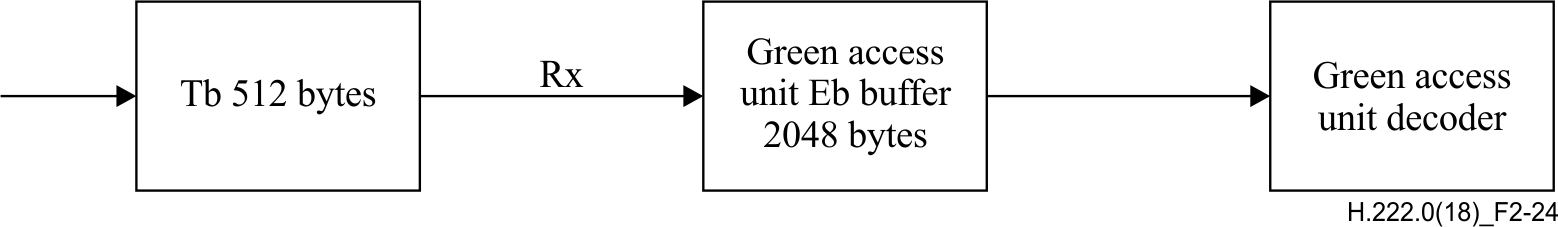


Figure 2-26 – T-STD model extension for green access units

MPEG-2 transport stream packets come into the model at the left, and are filtered by PID. Packets whose PID matches the green access unit PID flow into the 512 byte transport buffer. These buffered packets are removed at a rate of Rx = 300 kbps (kilobits/second) and stored in the green access unit Eb buffer (2048 Bytes). Green access unit table sections are removed from the Eb buffer immediately after the full access unit is available (based on section length) and these are passed onto the green access unit decoder at a rate Rbx = 300 kbps for decoding and each decoded access unit is associated with the video at time = display\_in\_PTS. The Eb buffer shall not overflow and the green access unit section shall be available in the Eb buffer at least 100 ms before display\_in\_PTS.

NOTE – In the worst-case, a green AU would contain 4488 bits. Under such conditions, the Eb buffer is large enough to hold up to three green AUs and the rate Rx is high enough to allow the removal of green AUs that are associated with video frames at 60 fps.

## 2.19 Carriage of ISO/IEC 23008-3 MPEG-H 3D audio data

### 2.19.1 Introduction

A Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream may carry ISO/IEC 23008-3 audio elementary streams. Typically, an ISO/IEC 23008-3 stream will be an element of a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, as defined by the PMT in a Transport Stream and the PSM in a Program Stream.

### 2.19.2 Carriage in PES packets

Individual ISO/IEC 23008-3 elementary streams shall be carried in PES packets as PES\_packet\_data\_bytes. For PES packetization no specific data alignment constraints apply. For synchronization PTS is encoded in the header of the PES packet that carries the ISO/IEC 23008-3 elementary stream data; for PTS encoding the same constraints apply as for ISO/IEC 13818 elementary streams.

Before PES packetization the elementary stream data shall be first encapsulated in the MHAS transport syntax defined in clause 14 of ISO/IEC 23008-3. If a PTS is present in the PES packet header, it shall refer to the first mpeghAudioStreamPacket (MHAS packet) of MHASPacketType PACTYP\_MPEGH3DAFRAME that commences in the payload of the PES packet. This MHAS packet shall occur after the first MHAS packet of MHASPacketTypes PACTYP\_SYNC and PACTYPE\_MPEGH3DACFG that commences in the payload of the PES packet, if any are present. Note that additional MHAS packets of other packet types may be present between those MHAS packets (PACTYP\_SYNC, PACTYPE\_MPEGH3DACFG and PACTYP\_MPEGH3DAFRAME), e.g., PACTYP\_SYNCGAP, PACTYP\_AUDIOSCENEINFO or PACTYP\_AUDIOTRUNCATION.

It is recommended but not mandatory that in this case the first byte of the PES\_packet\_data\_bytes is the first byte of an mpeghAudioStreamPacket with MHASPacketType set to PACTYP\_SYNC. In this case, the data\_alignment\_indicator in the PES packet header is set to '1'. According to 14.4.4 of ISO/IEC 23008-3, the mpeghAudioStreamPacket yields an overall syncword of 0xC001A5.

To provide appropriate specific information, it is strongly recommended that each MPEG-H 3D audio stream carries audio scene information data with sufficient information to ensure that the decoded and rendered MPEG-H 3D audio stream can be presented correctly by receivers.

Carriage of ISO/IEC 23008-3 elementary streams in PES packets shall be identified by appropriate stream\_id and stream\_type values, indicating the use of ISO/IEC 23008-3 3D audio. In addition, an MPEG-H\_3dAudio\_descriptor() shall be present in the descriptor loop for the respective elementary stream entry in the Program Map Table in case of a transport stream.

### 2.19.3 STD extensions for ISO/IEC 23008-3 elementary streams

The T-STD model as defined in 2.4.2.1 includes a transport buffer TBn and a multiplex buffer Bn prior to decoding of each ISO/IEC 23008-3 3D audio elementary stream n. For buffers TBn and Bn and the rate Rxn between TBn and Bn the following constraints apply for the carriage of an ISO/IEC 23008-3 3D audio stream:

The size BSn of the buffer Bn, is defined as BSn = BSmux + BSdec + BSoh with:

BSn = 3584 bytes for 1-2 encoded audio channel signals

Here, the size of the access unit decoding buffer BSdec, and the PES packet overhead buffer BSoh are constrained by: BSdec + BSoh ≤ 2848 bytes; a portion (736 bytes) of the 3584 byte buffer is allocated for buffering to allow multiplexing. The rest, 2848 bytes, are shared for access unit buffering BSdec, BSoh and additional multiplexing.

BSn = 8976 bytes for 3-8 encoded audio channel signals

BSn = 12804 bytes for 9-12 encoded audio channel signals

BSn = 51216 bytes for 13-48 encoded audio channel signals

BSn = 136576 bytes for 49-128 encoded audio channel signals

Rate Rxn:

Rxn = 2 000 000 bit/s for 1-2 encoded audio channel signals

Rxn = 5 529 600 bit/s for 3-8 encoded audio channel signals

Rxn = 8 294 400 bit/s for 9-12 encoded audio channel signals

Rxn = 33 177 600 bit/s for 13-48 encoded audio channel signals

Rxn = 88 473 600 bit/s for 49-128 encoded audio channel signals

### 2.19.4 STD extensions for multiple ISO/IEC 23008-3 elementary streams

ISO/IEC 23008-3 MPEG-H 3D audio allows encoding an audio program as several elementary streams. One single audio decoder decodes all elementary streams to one audio presentation. Each of those elementary streams carries one or more encoded audio channel signals.

For every single elementary stream the constraints for the buffer size BSn and the rate Rxn apply as described in 2.19.3 above.

As described in 14.6 of ISO/IEC 23008-3, the MHAS packets of each elementary stream are merged into one single stream before forwarding the data to the decoder. Note that based on user interactivity only a subset of the elementary streams may be merged into this single stream instead of the complete set of elementary streams.

Figure 2-27 illustrates an example of multiple elementary streams. Those streams are merged into one single intermediate stream and decoded by a single ISO/IEC 23008-3 3D audio decoder.

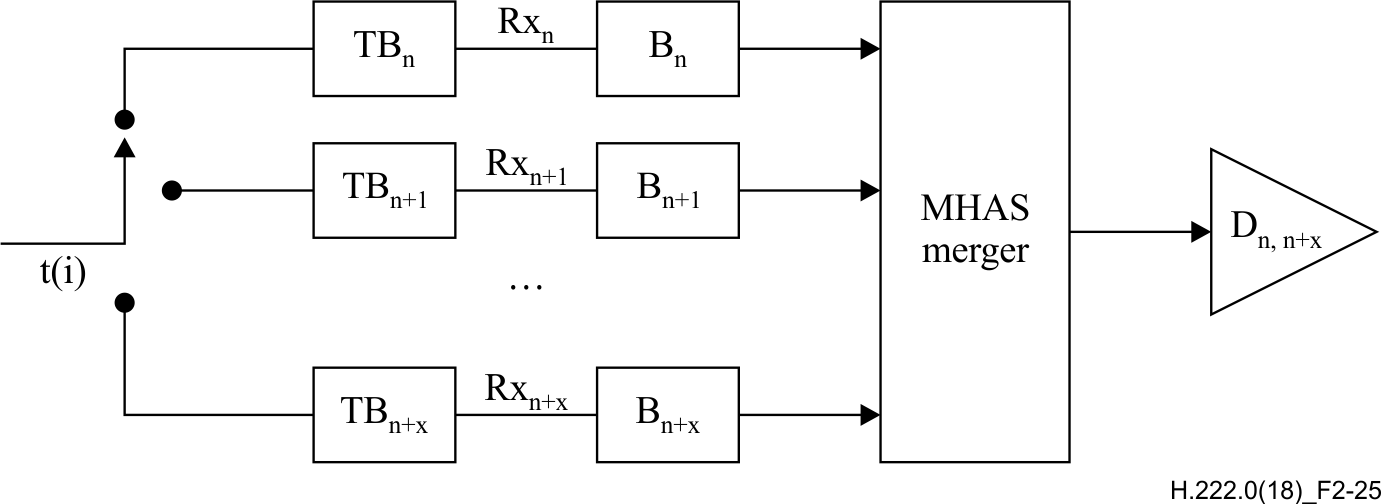


Figure 2-27 – Transport stream system target decoder for multiple audio elementary streams

In addition to the notation used and described in 2.4.2.1 to describe Figure 2-1 the following notation is used in Figure 2‑27:

Dn,n+x is the audio decoder for all x audio elementary streams n to n+x.

### 2.19.5 MPEG-2 Transport stream random access constraints and signalling

A TS packet containing the PES packet header of a 3D audio random access point (RAP) shall have an adaptation field. The payload\_unit\_start\_indicator bit shall be set to '1' in the TS packet header and the adaptation\_field\_control bits shall be set to '11' (as per 2.4.3.3). In addition, the random\_access\_indicator bit in the adaptation field of the TS packet that contains the PES packet header of the 3D audio RAP shall be set to '1' and follow the constraints specified in 2.4.3.5.

If the PES packet contains a RAP, then the MHAS packet of type PACTYP\_MPEGH3DAFRAME containing the RAP shall be the first MHAS packet of that type that commences in the PES packet. The audio data encapsulated in this MHAS packet shall follow the rules for a random access point as defined in 5.7 of ISO/IEC 23008-3.

## 2.20 Carriage of Quality Access Units in MPEG-2 sections

### 2.20.1 General description

Quality Access Units carrying dynamic quality metadata associated with one or more video frames, i.e., a given quality metric sample can be applicable for multiple video frames for a given DTS interval, until a new quality metric sample is declared. These access units are encapsulated in MPEG sections identified by stream\_type value of 0x2F.

Each Quality Access Unit contains configuration and timing information, as well as an array of quality values, corresponding one for one to the declared metrics. Each value is padded by preceding zero bytes, as needed, to the number of bytes indicated by field\_size\_bytes.

Quality\_access\_unit shall be carried in MPEG sections.

Every Quality Access Unit shall be a random access point.

NOTE – in order to make processing easier, each Quality Access Unit should be contained in a single TS packet.

The syntax for the Quality Access Units is provided in Table 2-164. The attribute definitions are aligned with those in ISO/IEC 23001-10.

Table 2-164 – Quality Access Unit

| Syntax | No. bits | Mnemonic |
| --- | --- | --- |
| Quality\_Access\_Unit() { |  |  |
| **field\_size\_bytes** | **8** | **uimsbf** |
| **metric\_count**  for ( i = 0; i < metric\_count; i++) {  **metric\_code**  **sample\_count**  for ( j = 0; j < sample\_count; j++) {  **'0010'**  **media\_DTS [32..30]**  **marker\_bit**  **media\_DTS [29..15]**  **marker\_bit**  **media\_DTS [14..0]**  **marker\_bit**  **quality\_metric\_sample**  }  } | **8**  **32**  **8**  **4**  **3**  **1**  **15**  **1**  **15**  **1**  **N** | **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf**  **uimsbf** |
| } |  |  |

**field\_size\_bytes** – Size of the quality\_metric\_sample field. As such, N=8\*field\_size\_bytes

**metric\_count** – The number of metrics for quality values in the Quality Access Unit, as defined in 4.2 of ISO/IEC 23001‑10.

**metric\_code** – The code name of the metrics in the Quality Access unit, as defined in 4.2 of ISO/IEC 23001-10.

**sample\_count** – Number of quality metric samples per metric.

**media\_DTS** – DTS of the media access unit described by the quality metric sample.

As such, quality\_metric\_sample is an array signalled using the 'i' and 'j' values of the loop.

### 2.20.2 Buffer model for processing Quality Access Units

The buffer model reflects processing required to handle Quality Access Units. The model can be used to establish constraints which can be used to verify the validity of dynamic Quality metadata prepared in accordance with this Standard.

Figure 2-28 illustrates the buffer model for processing Quality Access Units.

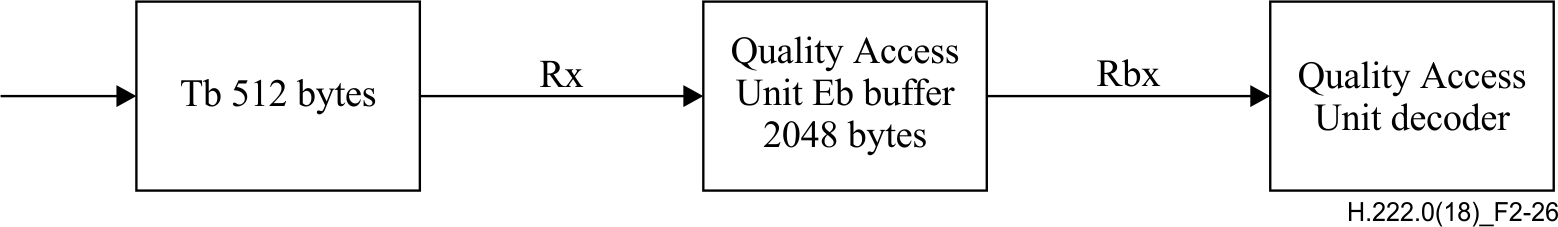


Figure 2-28 – Quality Access Unit decoder processing model

MPEG-2 Transport Stream packets come into the model at the left, and are filtered by PID. Packets whose PID matches the Quality Access Unit PID flow into the 512 byte transport buffer. These buffered packets are removed at a rate of Rx = 300 kbps (kilobits/second) and stored in the Quality Access Unit Eb buffer (2048 Bytes). Quality Access Unit table sections are removed from the Eb buffer immediately after the full access unit is available (based on section length) and these are passed onto the Quality Access Unit decoder at a rate Rbx = 300 kbps for decoding and association with the corresponding video frames. Eb buffer shall not overflow and the Quality Access Unit section shall be available in the Eb buffer to meet the requirement of DTS earlier than the DTS of the latest media access unit described in the corresponding Quality Access Unit.

## 2.21 Carriage of sample variants

### 2.21.1 Introduction

Carriage of metadata over a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream as defined in 2.12 allows for carriage of sample variants as defined in clause 11 of ISO/IEC 23001-12:2018.

A sample variants metadata stream is associated with an audiovisual content in a Rec. ITU-T H.222.0 | ISO/IEC 13818‑1 stream. There shall be only one audiovisual content stream associated with a sample variants metadata stream. There may be more than one sample variants metadata stream in a program that are associated with the same audiovisual content stream. Access to a sample variant is controlled by the protection scheme and media key and variant processing may be performed as specified in the variant processing model in clause 11.2.2 of ISO/IEC 23001-12:2018.

The decoder configuration data in the metadata descriptor of a sample variants metadata stream, as defined in clause 11.2.3 of ISO/IEC 23001-12:2018, associates the sample variants metadata stream with an audiovisual content stream.

### 2.21.2 Constraints

Sample variants shall be carried in an MPEG-2 TS metadata stream that complies with the following constraints:

1. The sample variant metadata shall be carried in a private metadata stream with stream\_id set to 0xBD (private\_stream\_id\_1).

2. The sample variant metadata stream shall be carried in a PES stream with stream\_type set to 0x15.

3. A metadata\_descriptor (specified in 2.6.60) shall be used to signal the sample variants metadata stream in the PMT, in the es\_loop for the elementary stream.

4. A four-character string, as defined in clause 11.2.3 of ISO/IEC 23001-12:2018, shall be used as the format identifier for both metadata\_format\_identifier and metadata\_application\_format\_identifier in the metadata\_descriptor to signal the sample variants metadata stream.

### 2.21.3 Sample Variants decoder configuration data

Decoding of sample variants requires the availability of decoder configuration data as defined in clause 11.2.4 of ISO/IEC 23001-12:2018. Consequently, when sample variants are carried in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream, then the metadata descriptor shall provide associated decoder configuration data in the same Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream and the coding of the decoder\_config\_flags shall be set to '001'.

## 2.22 Carriage of Media Orchestration Access Units

Media Orchestration Access Units shall adhere to the syntaxt and semantics of timed\_metadata\_access\_unit (), as specified in clause 8.2.2 of ISO/IEC 23001-13.

## 2.23 Carriage of VVC

### 2.23.1 Constraints for the transport of VVC

For VVC video streams, VVC temporal video sub-bitstreams or VVC temporal video subsets, the following constraints additionally apply:

• Each VVC access unit shall contain an access unit delimiter NAL unit;

NOTE 1 – VVC requires that an access unit delimiter NAL unit is the first NAL unit within a VVC access unit. Access unit delimiter NAL units simplify the ability to detect the boundary between VVC access units.

• A VVC video stream or VVC temporal video sub-bitstream shall be an element of a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program and the stream\_type for this elementary stream shall be equal to 0x33.

• The video parameter sets (when required), sequence parameter sets and picture parameter sets, as specified in Rec. ITU-T H.266 | ISO/IEC 23090-3, necessary for decoding a VVC video stream or VVC temporal video sub-bitstream shall be present within the elementary stream carrying that VVC video stream or VVC temporal video sub-bitstream.

NOTE 2 – A VVC video stream might not have any VPS NAL unit. In that case, the sequence parameter set refers to a video parameter set with ID equal to 0.

• For each VVC temporal video subset that is an element of the same Rec. ITU-T H.222.0 | ISO/IEC 13818‑1 program, the stream\_type for this elementary stream shall be equal to 0x34.

• When a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program includes more than one VVC temporal video subset, or more than one VVC temporal video sub-bitstream and at least one VVC temporal video subset, a hierarchy descriptor as defined in 2.6.7 shall be present for all associated elementary streams with stream type equal to 0x32 or 0x33. The hierarchy descriptors shall be used to indicate the dependencies of all VVC temporal video sub-bitstreams and all VVC temporal video subsets.

• In each elementary stream with stream\_type equal to 0x33 with a hierarchy descriptor the hierarchy\_type in the hierarchy descriptor shall be equal to 15.

• In each elementary stream with stream\_type equal to 0x34 with a hierarchy descriptor, the hierarchy\_type in the hierarchy descriptor shall be equal to 3.

• The video parameter sets (when required), sequence parameter sets and picture parameter sets, as specified in Rec. ITU-T H.266 | ISO/IEC 23090-3, necessary for decoding the VVC highest temporal sublayer representation of a VVC temporal video subset shall be present within the elementary stream carrying the VVC temporal video sub-bitstream associated by a hierarchy descriptor.

• The aggregation of the VVC temporal video sub-bitstream with associated VVC temporal video subsets according to the hierarchy descriptors, as specified in 2.23.3, shall result in a valid VVC video stream.

NOTE 3 – The resulting VVC video stream contains a set of temporal sublayers, as specified in Rec. ITU-T H.266 | ISO/IEC 23090‑3, with TemporalId values forming a contiguous range of integer numbers.

Carriage in PES packets

Rec. ITU-T H.266 | ISO/IEC 23090-3 Video is carried in PES packets as PES\_packet\_data\_bytes, using one of the 16 stream\_id values assigned to video, while signalling the Rec. ITU-T H.266 | ISO/IEC 23090-3 video stream by means of the assigned stream-type value in the PMT (see Table 2-34). The highest level that may occur in a VVC video stream as well as a profile and tier that the entire stream conforms to should be signalled using the VVC video descriptor. If a VVC video descriptor is associated with a VVC video stream, a VVC temporal video sub-bitstream, a VVC temporal video subset, then this descriptor shall be conveyed in the descriptor loop for the respective elementary stream entry in the Program Map Table. This Recommendation | International Standard does not specify presentation of Rec. ITU-T H.266 | ISO/IEC 23090-3 streams in the context of a program.

For PES packetization, no specific data alignment constraints apply. For synchronization and STD management, PTSs and, when appropriate, DTSs are encoded in the header of the PES packet that carries the Rec. ITU-T H.266 | ISO/IEC 23090-3 video elementary stream data. For PTS and DTS encoding, the constraints and semantics apply as defined in 2.4.3.7 and 2.7.

DPB buffer management

Carriage of a VVC video stream, a VVC temporal video sub-streams or a VVC temporal video subset over Rec. ITU-T H.222.0 | ISO/IEC 13818-1 does not impact the size of buffer DPB. For decoding of a VVC video stream, a VVC temporal video sub-bitstream or a VVC temporal video sub-bitstream and its associated VVC temporal video subsets in the STD, the size of DPB is as defined in Rec. ITU-T H.266 | ISO/IEC 23090-3. The DPB shall be managed as specified in Annex C of Rec. ITU-T H.266 | ISO/IEC 23090-3 (clauses C.3 and C.5). A decoded VVC access unit enters the DPB instantaneously upon decoding of the VVC access unit, hence at the CPB removal time of the VVC access unit. A decoded VVC access unit is presented at the DPB output time. If the VVC video stream, VVC temporal video sub-bitstream or VVC temporal video subset provides insufficient information to determine the CPB removal time and the DPB output time of VVC access units, then these time instants shall be determined in the STD model from PTS and DTS timestamps as follows:

1) The CPB removal time of VVC access unit n is the instant in time indicated by DTS(n) where DTS(n) is the DTS value of VVC access unit n.

2) The DPB output time of VVC access unit n is the instant in time indicated by PTS(n) where PTS(n) is the PTS value of VVC access unit n.

NOTE 4 – VVC video sequences in which the low\_delay\_hrd\_flag[ i ] in the syntax structure hrd\_parameters() is set to 1 carry sufficient information to determine the DPB output time and the CPB removal time of each VVC access unit. Hence for VVC access units for which STD underflow may occur, the CPB removal time and the DPB output time are defined by HRD parameters, and not by DTS and PTS timestamps.

NOTE 5 – A VVC video stream may carry information to determine compliance of the VVC video stream to the HRD, as specified in Annex C of Rec. ITU-T H.266 | ISO/IEC 23090-3. The presence of this information can be signalled in a transport stream using the VVC timing and HRD descriptor with the hrd\_management\_valid\_flag set to '1'. Irrespective of the presence of this information, compliance of a VVC video stream to the T-STD ensures that HRD buffer management requirements for CPB are met when each byte in the VVC video stream is delivered to and removed from CPB in the HRD at exactly the same instant in time at which the byte is delivered to and removed from EBn in the T-STD.

### 2.23.2 T-STD Extensions for single layer VVC

When there is a VVC video stream or VVC temporal video sub-bitstream in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program and there is no VVC temporal video subset associated with this elementary stream of stream\_type 0x33 in the same Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, the T-STD model as described in 2.4.2 is extended as illustrated in Figure 2-29 and as specified below.

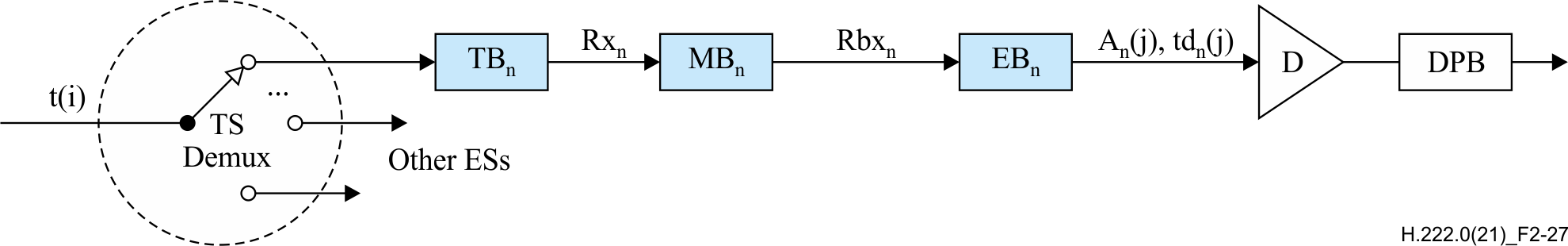


Figure 2-29 – T-STD model extensions for single layer VVC

TBn, MBn, EBn buffer management

The following additional notations are used to describe the T-STD extensions and are illustrated in Figure 2-29 above.

t(i) indicates the time in seconds at which the i-th byte of the Transport Stream enters the system target decoder

TBn is the transport buffer for elementary stream n

TBS is the size of the transport buffer TBn, measured in bytes

MBn is the multiplexing buffer for elementary stream n

MBSn is the size of the multiplexing buffer MBn, measured in bytes

EBn is the elementary stream buffer for the VVC video stream

j is an index to the VVC access unit of the VVC video stream

An(j) is the j-th access unit of the VVC video bitstream

tdn (j) is the decoding time of An(j), measured in seconds, in the system target decoder

Rxn is the transfer rate from the transport buffer TBn to the multiplex buffer MBn as specified below.

Rbxn is the transfer rate from the multiplex buffer MBn to the elementary stream buffer EBn as specified below.

The following applies:

• There is exactly one transport buffer TBn for the received VVC video stream or VVC temporal video sub-bitstream where the size TBS is fixed to 512 bytes.

• There is exactly one multiplexing buffer MBn for the VVC video stream or VVC temporal video sub-bitstream, where the size MBSn of the multiplexing buffer MBis constrained as follows:

MBSn = BSmux + BSoh + CpbNalFactor × MaxCPB[tier, level] – cpb\_size

where BSoh, packet overhead buffering, is defined as:

BSoh = (1/750) seconds × max{ BrNalFactor × MaxBR[tier, level], 2 000 000 bit/s}

and BSmux, additional multiplex buffering, is defined as:

BSmux = 0.004 seconds × max{ BrNalFactor × MaxBR[tier, level], 2 000 000 bit/s}

CpbNalFactor and BrNalFactor are taken from Annex A of Rec. ITU-T H.266 | ISO/IEC 23090-3 for the profile of the VVC video stream or VVC temporal video sub-bitstream. MaxCPB[tier, level] and MaxBR[tier, level] are taken from Annex A of Rec. ITU-T H.266 | ISO/IEC 23090-3 for the tier and level of the VVC video stream or VVC temporal video sub-bitstream. cpb\_size is CpbSize[ k ][ SchedSelIdx ] corresponding to the NAL HRD parameters of data flow into the CPB for the byte stream format as specified in clause 7 of Rec. ITU-T H.266 | ISO/IEC 23090-3, included in the VVC video stream or VVC temporal video sub-bitstream, and k is the highest temporalId in the VVC video stream or VVC temporal video sub-bitstream.

• There is exactly one elementary stream buffer EBn for all the elementary streams in the set of received elementary streams associated by hierarchy descriptors, with a total size EBSn

EBSn = cpb\_size (measured in bytes)

where cpb\_size is CpbSize[ k ][ SchedSelIdx ] corresponding to the NAL HRD parameters of data flow into the CPB for the byte stream format as specified in clause 7 of Rec. ITU-T H.266 | ISO/IEC 23090-3, included in the VVC video stream or the VVC temporal video sub-bitstream, and k is the highest temporalId in the VVC video stream or VVC temporal video sub-bitstream.

• Transfer from TBn to MBn is applied as follows:

When there is no data in TBn then Rxn is equal to zero. Otherwise:

Rxn = bit\_rate

where bit\_rate is BitRate[ k ][ SchedSelIdx ] corresponding to the NAL HRD parameters of data flow into the CPB for the byte stream format as defined in clause 7 of Rec. ITU-T H.266 | ISO/IEC 23090-3, and k is the highest temporalId in the VVC video stream or VVC temporal video sub-bitstream.

NOTE – When the NAL hrd\_parameters() is present in the SPS or VPS of the VVC video stream the bit\_rate is specified in sub\_layer\_hrd\_parameters( k ) , with k being the highest temporalId in the VVC video stream or VVC temporal video sub-bitstream. Default values for BitRate[ k ][ SchedSelIdx ] are also specified based on profile, tier and level when NAL HRD parameters are not present in the SPS.

• Transfer from MBn to EBn is applied as follows:

If the VVC\_timing\_and\_HRD\_descriptor is present with the hrd\_management\_valid\_flag set to '1' for the elementary stream, then the transfer of data from MBn to EBn shall follow the HRD defined scheme for data arrival in the CPB of elementary stream as defined in Annex C of Rec. ITU-T H.266 | ISO/IEC 23090-3.

Otherwise, the leak method shall be used to transfer data from MBn to EBn as follows:

Rbxn = BrNalFactor × MaxBR[tier, level]

where BrNalFactor is taken from Annex A of Rec. ITU-T H.266 | ISO/IEC 23090-3 for the profile of the VVC video stream or VVC temporal video sub-bitstream and MaxBR[tier, level] is taken from Annex A of Rec. ITU‑T H.266 | ISO/IEC 23090-3 for the tier and level of the VVC video stream or VVC temporal video sub-bitstream.

If there is PES packet payload data in MBn, and buffer EBn is not full, the PES packet payload is transferred from MBn to EBn at a rate equal to Rbxn. If EBn is full, data are not removed from MBn. When a byte of data is transferred from MBn to EBn, all PES packet header bytes that are in MBn and precede that byte are instantaneously removed and discarded. When there is no PES packet payload data present in MBn, no data is removed from MBn. All data that enters MBn leaves it. All PES packet payload data bytes enter EBn instantaneously upon leaving MBn.

STD delay

The STD delay of any Rec. ITU-T H.266 | ISO/IEC 23090-3 data other than VVC still picture data through the System Target Decoders buffers TBn, MBn, and EBn shall be constrained by tdn(j) – t(i) ≤ 10 seconds for all j, and all bytes i in access unit An(j).

The delay of any VVC still picture data through the System Target Decoders TBn, MBn, and EBn shall be constrained by tdn(j) – t(i) ≤ 60 seconds for all j, and all bytes i in access unit An(j).

Buffer management conditions

Transport streams shall be constructed so that the following conditions for buffer management are satisfied:

• Each TBn shall not overflow and shall be empty at least once every second.

• Each MBn, EBn, and DPB shall not overflow.

• EBn shall not underflow, except when HRD parameters are present for the VVC video sequence with the low\_delay\_hrd\_flag[ i ] set to '1'. Underflow of EBn occurs for VVC access unit An(j) when one or more bytes of An(j) are not present in EBn at the decoding time tdn(j).

### 2.23.3 T-STD Extensions for layered transport of VVC temporal video subsets

When there is a VVC video sub-bitstream and at least one associated elementary stream of type 0x34 in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, the T-STD model as described in 2.4.2 is extended as illustrated in Figure 2-30 and as specified below.

A screenshot of a computer

Description automatically generated

Figure 2-30 – T-STD model extensions for layered transport of VVC temporal video subsets

The following additional notations are used to describe the T-STD extensions and are illustrated in Figure 2-30 above.

t(i) indicates the time in seconds at which the i-th byte of the Transport Stream enters the system target decoder

H is the number of received VVC temporal video subsets, associated by hierarchy descriptors with the same VVC temporal video sub-bitstream.

k is an index identifying the H+1 received elementary streams which contain exactly one VVC temporal video sub-bitstream and H VVC temporal video subsets associated by hierarchy descriptors. The index value k equal to 0 identifies the elementary stream which contains the VVC temporal video sub-bitstream and index values k ranging from 1 up to H identify the associated VVC temporal video subsets.

ESn,k is the received elementary stream which contains the k-th VVC temporal video subset or the VVC temporal video sub-bitstream if k equals 0

ESn,H is the received elementary stream containing the highest VVC temporal video subset present in the set of received elementary streams

PIDH is the packet identifier value which identifies ESn,H

j is an index to the output access units

An(j) is the j-th access unit of the VVC complete temporal representation

tdn(j) is the decoding time of An(j) in the system target decoder

TBn,k is the transport buffer for elementary stream k

TBSn,k is the size of the transport buffer TBn,k, measured in bytes

MBn,k is the multiplexing buffer for elementary stream k

MBSn,k is the size of the multiplexing buffer MBn,k, measured in bytes

EBn is the elementary stream buffer for the received VVC temporal video sub-bitstream ESn,0 and the received VVC temporal video subsets ESn,1 to ESn,H

EBSn is the size of elementary stream buffer EBn, measured in bytes

Rxn,k is the transfer rate from the k-th transport buffer TBn,k to the k-th multiplex buffer MBn,k as specified below

Rbxn,k is the transfer rate from the k-th multiplex buffer MBn,k to the elementary stream buffer EBn as specified below

NOTE – The index n, where used, indicates that the received elementary streams and associated buffers belong to a certain VVC temporal video sub-bitstream and its associated VVC temporal video subsets, distinguishing these elementary streams and associated buffers from other elementary streams and buffers, maintaining consistency with the notation in Figure 2-29.

TBn,k, MBn,k, EBn buffer management

The following applies:

• There is one transport buffer TBn,k for each received elementary stream ESn,k, where the size TBSn,k is fixed to 512 bytes.

• There is one multiplex buffer MBn,k for each received elementary stream ESn,k, where the size MBSn,k of the multiplex buffer MBn,k is constrained as follows:

MBSn,k = BSmux + BSoh + CpbNalFactor × MaxCPB[tier, level] – cpb\_size (measured in bytes)

where

BSoh, packet overhead buffering, and BSmux, additional multiplex buffering, are as specified in 2.23.2;

CpbNalFactor and BrNalFactor are taken from Annex A of Rec. ITU-T H.266 | ISO/IEC 23090-3 for the profile of the VVC video stream. MaxCPB[tier, level] and MaxBR[tier, level] are taken from Annex A of Rec. ITU-T H.266 | ISO/IEC 23090-3 for the tier and level of the VVC highest temporal sublayer representation associated with ESn,k;

cpb\_size is CpbSize[ k ][ SchedSelIdx ] corresponding to the NAL HRD parameters of data flow into the CPB for the byte stream format as specified in Rec. ITU-T H.266 | ISO/IEC 23090-3, included in the VVC highest temporal sublayer representation associated with ESn,k and k is the highest temporalId in the VVC highest temporal sublayer representation associated with ESn,k.

• There is exactly one elementary stream buffer EBn for the H + 1 elementary streams in the set of received elementary streams ESn,0 to ESn,H, with a total size EBSn

EBSn = cpb\_size (measured in bytes)

where cpb\_size is CpbSize[ k ][ SchedSelIdx ] corresponding to the NAL HRD parameters of data flow into the CPB for the byte stream format as specified in Rec. ITU-T H.266 | ISO/IEC 23090-3, included in the VVC highest temporal sublayer representation associated with ESn,H and k is the highest temporalId in the VVC highest temporal sublayer representation associated with ESn,H.

• Transfer from TBn,k to MBn,k is applied as follows:

When there is no data in TBn,k then Rxn,k is equal to zero. Otherwise:

Rxn,k = bit\_rate

where bit\_rate is BitRate[ k ][ SchedSelIdx ] corresponding to the NAL HRD parameters of data flow into the CPB for the byte stream format as defined in clause 7 of Rec. ITU-T H.266 | ISO/IEC 23090-3, included in the VVC highest temporal sublayer representation associated with ESn,k and k is the highest temporalId in the VVC highest temporal sublayer representation associated with ESn,k.

• Transfer from MBn,k to EBn is applied as follows:

If the VVC\_timing\_and\_HRD\_descriptor is present with the hrd\_management\_valid\_flag set to '1' for the VVC video sub-bitstream, then the transfer of data from MBn,k to EBn shall follow the HRD defined scheme for data arrival in the CPB of elementary stream ESn,H as defined in Annex C of Rec. ITU-T H.266 | ISO/IEC 23090-3.

Otherwise, the leak method shall be used to transfer data from MBn,k to EBn as follows:

Rbxn,k = BrNalFactor × MaxBR[tier, level]

where BrNalFactor is taken from Annex A of Rec. ITU-T H.266 | ISO/IEC 23090-3 for the profile of the VVC video stream and MaxBR[tier, level] is defined for the byte stream format in Annex A of Rec. ITU-T H.266 | ISO/IEC 23090-3 for the tier and level of the VVC video stream or the VVC highest temporal sublayer representation associated with ESn,H.

If there is PES packet payload data in MBn,k, and EBn is not full, the PES packet payload is transferred from MBn,k to EBn at a rate equal to Rbxn,k. If EBn is full, data are not removed from MBn,k. When a byte of data is transferred from MBn,k to EBn,, all PES packet header bytes that are in MBn,k and precede that byte are instantaneously removed and discarded. When there is no PES packet payload data present in MBn,k, no data is removed from MBn,k. All data that enters MBn,k leaves it. All PES packet payload data bytes enter EBn instantaneously upon leaving MBn,k.

At the output of the elementary stream buffer EBn, the elementary streams are aggregated by removing all VVC access units in ascending DTS order and transferring them to the VVC decoder DH, irrespective of which elementary stream ESn,k each VVC access unit belongs to.

STD delay

The STD delay of any Rec. ITU-T H.266 | ISO/IEC 23090-3 data other than VVC still picture data through the System Target Decoders buffers TBn,k, MBn,k, and EBn shall be constrained by tdn(j) – t(i) ≤ 10 seconds for all k, all j, and all bytes i in access unit An(j).

The delay of any VVC still picture data through the System Target Decoders TBn,k, MBn,k, and EBn shall be constrained by tdn(j) – t(i) ≤ 60 seconds for all k, all j, and all bytes i in access unit An(j).

Buffer management conditions

Transport streams shall be constructed so that the following conditions for buffer management are satisfied:

• Each TBn,k shall not overflow and shall be empty at least once every second.

• Each MBn,k, EBn, and DPB shall not overflow.

• EBn shall not underflow, except when HRD parameters are present for the VVC video sequence with the low\_delay\_hrd\_flag[ i ] set to '1'. Underflow of EBn occurs for VVC access unit An(j) when one or more bytes of An(j) are not present in EBn at the decoding time tdn(j).

## 2.24 Carriage of EVC

### 2.24.1 Constraints for the transport of EVC

For EVC video streams or EVC temporal video sub-bitstreams, the following constraints additionally apply:

• Each EVC access unit shall contain an access unit delimiter NAL unit;

NOTE 1 – EVC requires that an access unit delimiter NAL unit is the first NAL unit within an EVC access unit. Access unit delimiter NAL units simplify the ability to detect the boundary between EVC access units.

• An EVC video stream or EVC temporal video sub-bitstream shall be an element of a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program and the stream\_type for this elementary stream shall be equal to 0x35.

• The sequence parameter sets, picture parameter sets and adaptation parameter sets, as specified in ISO/IEC 23094-1, necessary for decoding an EVC video stream or EVC temporal video sub-bitstream shall be present within the elementary stream carrying that EVC video stream or EVC temporal video sub-bitstream.

• In each elementary stream with stream\_type equal to 0x35 with a hierarchy descriptor the hierarchy\_type in the hierarchy descriptor shall be equal to 16.

Carriage in PES packets

ISO/IEC 23094-1 Video is carried in PES packets as PES\_packet\_data\_bytes, using one of the 16 stream\_id values assigned to video, while signalling the ISO/IEC 23094-1 video stream by means of the assigned stream-type value in the PMT (see Table 2-34). The highest level that may occur in an EVC video stream as well as profile and toolset identifications that the entire stream conforms to should be signalled using the EVC video descriptor. If an EVC video descriptor is associated with an EVC video stream or an EVC temporal video sub-bitstream, then this descriptor shall be conveyed in the descriptor loop for the respective elementary stream entry in the Program Map Table. This Recommendation | International Standard does not specify presentation of ISO/IEC 23094-1 streams in the context of a program.

Every EVC access unit shall be carried on exactly one PES packet. For PES packetization, no specific data alignment constraints apply. For synchronization and STD management, PTSs and, when appropriate, DTSs are encoded in the header of the PES packet that carries the ISO/IEC 23094-1 video elementary stream data. For PTS and DTS encoding, the constraints and semantics apply as defined in 2.4.3.7 and 2.7.

NOTE 2 – EVC video streams can emulate PES start codes (in contrast to other video codecs that mandate start code emulation prevention for the elementary stream).

DPB buffer management

Carriage of a EVC video stream or a EVC temporal video sub-streams over Rec. ITU-T H.222.0 | ISO/IEC 13818-1 does not impact the size of buffer DPB. The DPB shall be managed as specified in Annex C of ISO/IEC 23094-1 (clauses C.3 and C.5). A decoded EVC access unit enters the DPB instantaneously upon decoding of the EVC access unit, hence at the CPB removal time of the EVC access unit. A decoded EVC access unit is presented at the DPB output time. If the EVC video stream or an EVC temporal video sub-bitstream provides insufficient information to determine the CPB removal time and the DPB output time of EVC access units, then these time instants shall be determined in the STD model from PTS and DTS timestamps as follows:

1) The CPB removal time of EVC access unit n is the instant in time indicated by DTS(n) where DTS(n) is the DTS value of EVC access unit n.

2) The DPB output time of EVC access unit n is the instant in time indicated by PTS(n) where PTS(n) is the PTS value of EVC access unit n.

NOTE 3 – EVC video sequences in which the low\_delay\_hrd\_flag in the syntax structure hrd\_parameters() is set to 1 carry sufficient information to determine the DPB output time and the CPB removal time of each EVC access unit. Hence for EVC access units for which STD underflow may occur, the CPB removal time and the DPB output time are defined by HRD parameters, and not by DTS and PTS timestamps.

NOTE 4 – An EVC video stream may carry information to determine compliance of the EVC video stream to the HRD, as specified in Annex C of ISO/IEC 23094-1. The presence of this information can be signalled in a transport stream using the EVC timing and HRD descriptor with the hrd\_management\_valid\_flag set to '1'. Irrespective of the presence of this information, compliance of a EVC video stream to the T-STD ensures that HRD buffer management requirements for CPB are met when each byte in the EVC video stream is delivered to and removed from CPB in the HRD at exactly the same instant in time at which the byte is delivered to and removed from EBn in the T-STD.

### 2.24.2 T-STD Extensions for single layer EVC

When there is a EVC video stream or EVC temporal video sub-bitstream in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, the T-STD model as described in 2.4.2 is extended as illustrated in Figure 2-31 and as specified below.

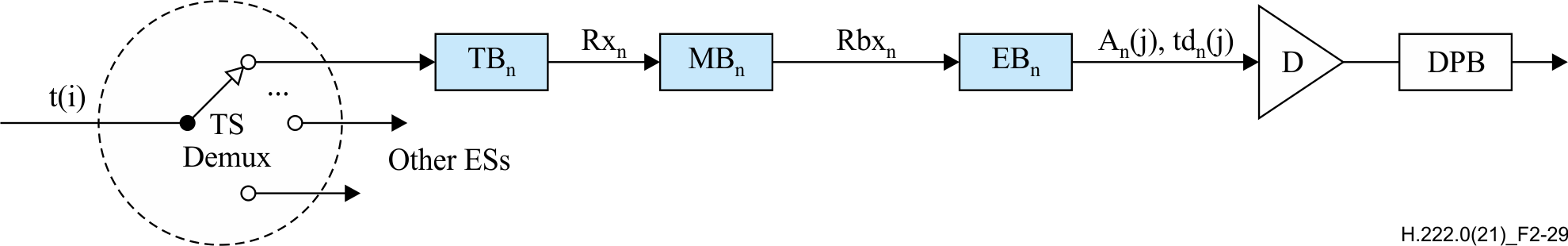


Figure 2-31 – T-STD model extensions for single layer EVC

TBn, MBn, EBn buffer management

The following additional notations are used to describe the T-STD extensions and are illustrated in Figure 2-31.

t(i) indicates the time in seconds at which the i-th byte of the Transport Stream enters the system target decoder

TBn is the transport buffer for elementary stream n

TBS is the size of the transport buffer TBn, measured in bytes

MBn is the multiplexing buffer for elementary stream n

MBSn is the size of the multiplexing buffer MBn, measured in bytes

EBn is the elementary stream buffer for the EVC video stream

j is an index to the EVC access unit of the EVC video stream

An(j) is the j-th access unit of the EVC video bitstream

tdn (j) is the decoding time of An(j), measured in seconds, in the system target decoder

Rxn is the transfer rate from the transport buffer TBn to the multiplex buffer MBn as specified below.

Rbxn is the transfer rate from the multiplex buffer MBn to the elementary stream buffer EBn as specified below.

The following applies:

• There is exactly one transport buffer TBn for the received EVC video stream or EVC temporal video sub-bitstream where the size TBS is fixed to 512 bytes.

• There is exactly one multiplexing buffer MBn for the EVC video stream or EVC temporal video sub-bitstream, where the size MBSn of the multiplexing buffer MB is constrained as follows:

MBSn = BSmux + BSoh + CpbBrNalFactor × MaxCPB[level] – cpb\_size

where BSoh, packet overhead buffering, is defined as:

BSoh = (1/750) seconds × max{ CpbBrNalFactor × MaxBR[level], 2 000 000 bit/s}

and BSmux, additional multiplex buffering, is defined as:

BSmux = 0.004 seconds × max{ CpbBrNalFactor × MaxBR[level], 2 000 000 bit/s}

MaxCPB [level] and MaxBR[level] are taken from Annex A of ISO/IEC 23094-1 for the level of the EVC video stream or EVC temporal video sub-bitstream. cpb\_size is taken from the HRD parameters cpb\_size\_value\_minus1[ SchedSelIdx ], as specified in Annex E of ISO/IEC 23094-1, included in the EVC video stream or EVC temporal video sub-bitstream.

• There is exactly one elementary stream buffer EBn for all the elementary streams in the set of received elementary streams associated by hierarchy descriptors, with a total size EBSn

EBSn = cpb\_size (measured in bytes)

where cpb\_size is taken from the HRD parameters, as specified in clause 7 of ISO/IEC 23094-1, included in the EVC video stream or the EVC temporal video sub-bitstream.

• Transfer from TBn to MBn is applied as follows:

When there is no data in TBn then Rxn is equal to zero. Otherwise:

Rxn = bit\_rate

where bit\_rate is bit\_rate\_value\_minus1[ SchedSelIdx ] +1 corresponding to the NAL HRD parameters of data flow into the CPB for the byte stream format and bit\_rate\_value\_minus1 [ SchedSelIdx ] is as defined in Annex E of ISO/IEC 23094-1.

NOTE – When the NAL hrd\_parameters() is present in the SPS of the EVC video stream the bit\_rate is specified in sub\_layer\_hrd\_parameters( i ). Default values for bit\_rate\_value\_minus1 [ SchedSelIdx] are also specified based on profile, tier and level when NAL HRD parameters are not present in the SPS. .

• Transfer from MBn to EBn is applied as follows:

If the EVC\_timing\_and\_HRD\_descriptor is present with the hrd\_management\_valid\_flag set to '1' for the elementary stream, then the transfer of data from MBn to EBn shall follow the HRD defined scheme for data arrival in the CPB of elementary stream as defined in Annex C of ISO/IEC 23094-1.

Otherwise, the leak method shall be used to transfer data from MBn to EBn as follows:

Rbxn = CpbBrNalFactor × MaxBR[level]

where MaxBR[level] is taken from Annex A of ISO/IEC 23094-1 for the tier and level of the EVC video stream or EVC temporal video sub-bitstream.

If there is PES packet payload data in MBn, and buffer EBn is not full, the PES packet payload is transferred from MBn to EBn at a rate equal to Rbxn. If EBn is full, data are not removed from MBn. When a byte of data is transferred from MBn to EBn, all PES packet header bytes that are in MBn and precede that byte are instantaneously removed and discarded. When there is no PES packet payload data present in MBn, no data is removed from MBn. All data that enters MBn leaves it. All PES packet payload data bytes enter EBn instantaneously upon leaving MBn.

STD delay

The STD delay of any ISO/IEC 23094-1 data other than EVC still picture data through the System Target Decoders buffers TBn, MBn, and EBn shall be constrained by tdn(j) – t(i) ≤ 10 seconds for all j, and all bytes i in access unit An(j).

The delay of any EVC still picture data through the System Target Decoders TBn, MBn, and EBn shall be constrained by tdn(j) – t(i) ≤ 60 seconds for all j, and all bytes i in access unit An(j).

Buffer management conditions

Transport streams shall be constructed so that the following conditions for buffer management are satisfied:

• Each TBn shall not overflow and shall be empty at least once every second.

• Each MBn, EBn, and DPB shall not overflow.

• EBn shall not underflow, except when HRD parameters are present for the EVC video sequence with the low\_delay\_hrd\_flag set to '1'. Underflow of EBn occurs for EVC access unit An(j) when one or more bytes of An(j) are not present in EBn at the decoding time tdn(j).

## 2.25 Carriage of LCEVC

### 2.25.1 Constraints for the transport of LCEVC

For LCEVC video streams, the following constraints additionally apply:

• An LCEVC video stream shall be an element of an Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program and the stream\_type for this elementary stream shall be equal to the value defined in Table 2-34.

• A Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program may include more than one LCEVC video stream.

• Each LCEVC video stream in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program shall have an LCEVC video descriptor present in the descriptor loop of the program map section for the LCEVC video stream.

• It is possible that the same video elementary stream constitutes a base stream for more than one LCEVC video stream.

• For each LCEVC video stream in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, a video elementary stream in the same Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program shall exist with an LCEVC linkage descriptor and the array of lcevc\_stream\_tags of this LCEVC linkage descriptor shall include one with the same value of lcevc\_stream\_tag of the LCEVC video descriptor of this LCEVC video stream.

• An LCEVC video stream enhances the base video elementary stream, element of the same Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program, that carries the same lcevc\_stream\_tag, in its LCEVC linkage descriptor, as the lcevc\_stream\_tag indicated in the LCEVC video descriptor.

• The NAL units, as specified in ISO/IEC 23094-2, necessary for decoding an LCEVC video stream, shall be present within the elementary stream carrying that LCEVC video stream.

### 2.25.2 Carriage in PES packets

ISO/IEC 23094-2 video is carried in PES packets as PES\_packet\_data\_bytes, using one of the 16 stream\_id values assigned to video, while signalling the ISO/IEC 23094-2 video stream by means of the assigned stream-type value in the PMT (see Table 2-34). The highest level that may occur in an LCEVC video stream as well as a profile and tier that the entire stream conforms to should be signalled using the LCEVC video descriptor. If an LCEVC video descriptor is associated with an LCEVC video stream, then this descriptor shall be conveyed in the descriptor loop for the respective elementary stream entry in the program map table.

This Recommendation | International Standard does not specify the presentation of ISO/IEC 23094-2 streams in the context of a program.

For PES packetization, the following data alignment constraints apply:

• Only one access unit shall be carried within each PES packet.

• PES\_packet\_length of value 0 is allowed and shall indicate to the decoder that it should detect PES boundaries by means of the payload\_unit\_start\_indicator.

For synchronization and STD management, PTSs shall be encoded in the header of each PES packet that carries the ISO/IEC 23094-2 video elementary stream data. For PTS encoding, the constraints and semantics apply as defined in 2.4.3.7 and 2.7.

### 2.25.3 DPB buffer management

Carriage of a LCEVC video stream over Rec. ITU-T H.222.0 | ISO/IEC 13818-1 does not impact the size of buffer DPB. For decoding of a LCEVC video stream in the STD, the size of DPB is as defined in ISO/IEC 23094-2. The DPB shall be managed as specified in Annex C of ISO/IEC 23094-2 (clauses C.3 and C.5). A decoded LCEVC access unit enters the DPB instantaneously upon decoding of the LCEVC access unit, hence at the CPB removal time of the LCEVC access unit. A decoded LCEVC access unit is presented at the DPB output time. If the base video stream provides insufficient information to determine the CPB removal time and the DPB output time of LCEVC access units, then these time instants shall be determined in the STD model from PTS and DTS timestamps as follows:

• The CPB removal time of LCEVC access unit n is the instant in time indicated by DTS(n), where DTS(n) is the DTS value of base access unit n.

• The DPB output time of LCEVC access unit n is the instant in time indicated by PTS(n), where PTS(n) is the PTS value of base access unit n.

NOTE 1 – LCEVC video sequences in which the low\_delay\_hrd\_flag in the syntax structure hrd\_parameters() on the base stream is set to 1 carry sufficient information to determine the DPB output time and the CPB removal time of each LCEVC access unit. Hence, for base access units for which STD underflow can occur, the CPB removal time and the DPB output time are defined by HRD parameters, and not by DTS and PTS timestamps.

NOTE 2 – An LCEVC video stream may carry information to determine compliance of the LCEVC video stream to the HRD, as specified in Annex C of ISO/IEC 23094-2. The presence of this information can be signalled in a transport stream using the base timing and HRD descriptor with the hrd\_management\_valid\_flag set to '1'. Irrespective of the presence of this information, compliance of a base video stream to the T-STD ensures that HRD buffer management requirements for CPB are met when each byte in the LCEVC video stream is delivered to and removed from CPB in the HRD at exactly the same instant in time at which the byte is delivered to and removed from EBn in the T-STD.

Annex A  
  
CRC decoder model

(This annex forms an integral part of this Recommendation | International Standard.)

## A.1 CRC decoder model

The 32-bit CRC decoder model is specified in Figure A.1.



Figure A.1 – 32-bit CRC decoder model

The 32-bit CRC decoder operates at bit level and consists of 14 adders '+' and 32 delay elements z(i). The input of the CRC decoder is added to the output of z(31), and the result is provided to the input z(0) and to one of the inputs of each remaining adder. The other input of each remaining adder is the output of z(i), while the output of each remaining adder is connected to the input of z(i + 1), with i = 0, 1, 3, 4, 6, 7, 9, 10, 11, 15, 21, 22, and 25. Refer to Figure A.1.

This is the CRC calculated with the polynomial:

x32 + x26 + x23 + x22 + x16 + x12 + x11 + x10 + x8 + x7 + x5 + x4 + x2 + x + 1 (A-1)

Bytes are received at the input of the CRC decoder. Each byte is shifted into the CRC decoder one bit at a time, with the left most bit (msb) first. For example, if the input is byte 0x01 the seven '0's enter the CRC decoder first, followed by the one '1'. Before the CRC processing of the data of a section the output of each delay element z(i) is set to its initial value '1'. After this initialization, each byte of the section is provided to the input of the CRC decoder, including the four CRC\_32 bytes. After shifting the last bit of the last CRC\_32 byte into the decoder, i.e., into z(0) after the addition with the output of z(31), the output of all delay elements z(i) is read. In the case where there are no errors, each of the outputs of z(i) shall be zero. At the CRC encoder the CRC\_32 field is encoded with a value such that this is ensured.

Annex B  
  
Digital storage medium command and control (DSM-CC)

(This annex does not form an integral part of this Recommendation | International Standard.)

## B.1 Introduction

The DSM-CC protocol is a specific application protocol intended to provide the basic control functions and operations specific to managing a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream on digital storage media. This DSM-CC is a low-level protocol above network/OS layers and below application layers.

The DSM-CC shall be transparent in the following sense:

• It is independent of the DSM used;

• it is independent of whether the DSM is located at a local or remote site;

• it is independent of the network protocol with which the DSM-CC is interfaced;

• it is independent of the various operating systems on which the DSM is operated.

### B.1.1 Purpose

Many applications of Rec. ITU-T H.222.0 | ISO/IEC 13818-1 DSM Control Commands require access to a Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 bitstream stored on a variety of digital storage media at a local or remote site. Different DSM have their own specific control commands and thus, a user would need to know different sets of specific DSM control commands in order to access Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstreams from different DSM. This brings many difficulties to the interface design of a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 or ISO/IEC 11172-1 application system. To overcome this difficulty, a set of common DSM control commands, which is independent of the specific DSM used, is suggested in this annex. This annex is informative only. ISO/IEC 13818-6 defines DSM-CC extension with a broader scope.

### B.1.2 Future applications

Beyond the immediate applications supported by the current DSM control commands, future applications based on extensions of DSM command control could include the following:

Video on demand

Video programs are provided as requested by a customer through various communication channels. The customer could select a video program from a list of programs available from a video server. Such applications could be used by hotels, cable TV, educational institutions, hospitals, etc.

Interactive video services

In these applications, the user provides frequent feedback controlling the manipulation of stored video and audio. These services can include video-based games, user-controlled video tours, electronic shopping, etc.

Video networks

Various applications may wish to exchange stored audio and video data through some type of computer network. Users could route AV information through the video network to their terminals. Electronic publishing and multimedia applications are examples of this kind of application.

### B.1.3 Benefits

Specifying the DSM control commands independent of the DSM, end-users can perform Rec. ITU-T H.222.0 | ISO/IEC 13818-1 decoding without having to fully understand the detailed operation of the specific DSM used.

The DSM control commands are codes to give end users the assurance that the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstreams can be played and stored with the same semantics, independent of the DSM and user interface. They are fundamental commands for the control of DSM operation.

### B.1.4 Basic functions

#### B.1.4.1 Stream selection

The DSM-CC provides the means to select a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream upon which to perform the succeeding operations. Such operations include creation of a new bitstream. Parameters of this function include:

• index of the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream (the mapping between this index and a name meaningful to an application is outside the scope of the current DSM-CC);

• mode (retrieval/storage).

#### B.1.4.2 Retrieval

The DSM-CC provides the means to:

• play an identified Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream;

• play from a given presentation time;

• set the playback speed (normal or fast);

• set the playback duration (until a specified presentation time, the end of the bitstream in forward play or the beginning in reverse play or the issuance of a stop command);

• set the direction (forward or reverse);

• pause;

• resume;

• change the access point in the bitstream;

• stop.

#### B.1.4.3 Storage

The DSM-CC provides the means to:

• cause storage of a valid bitstream for a specified duration;

• cause storage to stop.

DSM-CC provides a useful but limited subset of functionality that may be required in DSM based Rec. ITU-T H.222.0 | ISO/IEC 13818-1 applications. It is fully expected that significant additional capabilities will be added through subsequent extensions.

## B.2 General elements

### B.2.1 Scope

The scope of this work consists of the development of a Recommendation | International Standard to specify a useful set of commands for control of digital storage media on which a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream is stored. The commands can perform remote control of a digital storage media in a general way independently of the specific DSM and apply to any Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream stored on a DSM.

### B.2.2 Overview of the DSM-CC application

The current DSM-CC syntax and semantics cover the single user to DSM application. The user's system is capable of retrieving a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream and is also (optionally) capable of generating a Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 bitstream. The control channel over which the DSM commands and acknowledgements are sent is shown in Figure B.1 as an out-of-band channel. This can also be accomplished by inserting the DSM-CC commands and acknowledgements into the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstreams if an out-of-band channel is not available.



Figure B.1 – Configuration of DSM-CC application

### B.2.3 The transmission of DSM-CC commands and acknowledgements

The DSM-CC is encoded into a DSM-CC bitstream according to the syntax and semantics defined in B.3.2 through B.3.9. The DSM-CC bitstream can be transmitted both as a stand-alone bitstream and in a Rec. ITU-T H.222.0 | ISO/IEC 13818‑1 Systems bitstream.

When the DSM-CC bitstream is transmitted in stand-alone mode, its relationship to the Systems bitstream and the decoding process is illustrated in Figure B.2. In this case, the DSM-CC bitstream is not embedded in the Systems bitstream. This transmission mode can be used in the applications when the DSM is connected directly with the Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 decoder. It can also be used in the applications where the DSM-CC bitstream could be controlled and transmitted by other types of network multiplexors.



Figure B.2 – BSM-CC bitstream decoded as a stand-alone bitstream

For some applications, it is desirable to transmit the DSM-CC in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 systems bitstream so that some features of the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 systems bitstream could be applied to the DSM-CC bitstream as well. In this case, the DSM-CC bitstream is embedded in the systems bitstream by the systems multiplexor.

The DSM-CC bitstream is encoded by the systems encoder in the following process. First, the DSM-CC bitstream is packetized into a packetized elementary stream (PES) according to the syntax described in 2.4.3.6. The PES packet is then multiplexed into either a program stream (PS) or a transport stream (TS) according to the requirement of the transmission media. The decoding procedures are the inverse of the encoding procedures and are illustrated in the block diagram of the Systems decoder depicted in Figure B.3.

In Figure B.3, the output of the Systems decoder is a video bitstream, audio bitstream and/or DSM-CC bitstream. The DSM-CC bitstream is identified by the stream\_id, value '1111 0010' as defined by the stream\_id Table 2-22. Once the DSM-CC bitstream is identified, it follows the rules as specified by T-STD or P-STD.



Figure B.3 – DSM-CC bitstream decoded as part of the system bitstream

## B.3 Technical elements

### B.3.1 Definitions

For the purposes of this Recommendation | International Standard, the following definitions apply:

**B.3.1.1 DSM-CC**: Digital Storage Media Command and Control Commands that are specified by Rec. ITU‑T H.222.0 | ISO/IEC 13818-1 for the control of digital storage media at a local or remote site containing a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream.

**B.3.1.2 DSM ACK**: The acknowledgement from the DSM-CC command receiver to the command initiator.

**B.3.1.3 MPEG bitstream**: An ISO/IEC 11172-1 Systems stream, Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program stream or Rec. ITU-T H.222.0 | ISO/IEC 13818-1 transport stream.

**B.3.1.4 DSM-CC server**: A system, either local or remote, used to store and/or retrieve a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream.

**B.3.1.5 point of random access**: A point in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream with the property that for at least one elementary stream within the bitstream, the next access unit, 'N', completely contained in the bitstream can be decoded without reference to previous access units, and for every elementary stream in the bitstream all access units with the same or later presentation times are completely contained subsequently in the bitstream and can be completely decoded by a system target decoder without access to information prior to the point of random access. The bitstream as stored on the DSM may have certain points of random access; the output of the DSM may include additional points of random access manufactured by the DSM's own manipulation of the stored material (e.g., storing quantization matrices so that a sequence header can be generated whenever necessary). A point of random access has an associated PTS, namely the actual or implied PTS of access unit 'N'.

**B.3.1.6 current operational PTS value**: The actual or implied PTS associated with the last point of random access preceding the last access unit provided from the DSM from the currently selected Rec. ITU-T H.222.0 | ISO/IEC 13818‑1 bitstream. If no access unit has been provided from this Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream, the DSM is incapable of providing random access into the current bitstream, then the current operational PTS value is the first point of random access in the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream.

**B.3.1.7 DSM-CC bitstream**: A sequence of bits satisfying the syntax of B.3.2.

### B.3.2 Specification of DSM-CC syntax

• Every DSM control command shall commence with a start\_code, as specified in Table B.1.

• Every DSM control command shall have a packet\_length to specify the number of byte in a DSM‑CC packet.

• When the DSM-CC bitstream is transmitted as a PES packet as defined in 2.4.3.6, the fields up to the packet\_length field are identical to those specified in 2.4.3.6. In other words, if the DSM-CC packet is encapsulated in a PES packet, the PES packet start code is the only start code at the beginning of the packet.

• The actual control command or acknowledgement shall follow the last byte of the packet\_length field.

• An acknowledgement stream shall be provided by the DSM control bitstream receiver after the requested operation is started or is completed, depending on the command received.

• At all times the DSM is responsible for providing a normative Rec. ITU-T H.222.0 | ISO/IEC 13818-1 stream. This may include manipulating the trick mode bits defined in 2.4.3.6.

Table B.1 – DSM-CC syntax

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| DSM\_CC() { |  |  |
| **packet\_start\_code\_prefix** | **24** | **bslbf** |
| **stream\_id** | **8** | **uimsbf** |
| **packet\_length** | **16** | **uimsbf** |
| **command\_id** | **8** | **uimsbf** |
| If (command\_id = = '01') { |  |  |
| control() |  |  |
| } else if (command\_id = = '02') { |  |  |
| ack() |  |  |
| } |  |  |
| } |  |  |

### B.3.3 Semantics of fields in specification of DSM-CC syntax

**packet\_start\_code\_prefix** – This is a 24-bit code. Together with the stream\_id that follows it constitutes a DSM-CC packet start code that identifies the beginning of a DSM-CC packet bitstream. The packet\_start\_code\_prefix is the bit string '0000 0000 0000 0000 0000 0001' (0x000001).

**stream\_id** – This 8-bit field specifies the bitstream type and shall have a value '1111 0010' for the DSM-CC bitstream. Refer to Table 2-22.

**packet\_length** – This 16-bit field specifies the number of bytes in the DSM-CC packet immediately following the last byte of this field.

**command\_id** – This 8-bit unsigned integer identifies the bitstream is a control command or an acknowledgement stream. The values are defined in Table B.2.

Table B.2 – Command\_id assigned values

|  |  |
| --- | --- |
| Value | Command\_id |
| 0x00 | Forbidden |
| 0x01 | Control |
| 0x02 | Ack |
| 0x03 .. 0xFF | Reserved |

### B.3.4 Control layer

Constraints on setting flags in DSM-CC control

• At most one of the flags for select, playback and storage shall be set to '1' for each DSM control command. If none of these bits are set, then this command shall be ignored.

• At most one of pause\_mode, resume\_mode, stop\_mode, play\_flag, and jump\_flag shall be set for each retrieval command. If none of these bits are set, then this command shall be ignored.

• At most one of record\_flag and stop\_mode shall be selected for each storage command. If none of these bits are set, then this command shall be ignored.

See Table B.3.

Table B.3 – DSM-CC control

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| control() { |  |  |
| **select\_flag** | **1** | **bslbf** |
| **retrieval\_flag** | **1** | **bslbf** |
| **storage\_flag** | **1** | **bslbf** |
| **reserved** | **12** | **bslbf** |
| **marker\_bit** | **1** | **bslbf** |
| If (select\_flag = = '1') { |  |  |
| **bitstream\_id [31..17]** | **15** | **bslbf** |
| **marker\_bit** | **1** | **bslbf** |
| **bitstream\_id [16..2]** | **15** | **bslbf** |
| **marker\_bit** | **1** | **bslbf** |
| **bitstream\_id [1..0]** | **2** | **bslbf** |
| **select\_mode** | **5** | **bslbf** |
| **marker\_bit** | **1** | **bslbf** |
| } |  |  |
| if ( retrieve\_flag = = '1') { |  |  |
| **jump\_flag** | **1** | **bslbf** |
| **play\_flag** | **1** | **bslbf** |
| **pause\_mode** | **1** | **bslbf** |
| **resume\_mode** | **1** | **bslbf** |
| **stop\_mode** | **1** | **bslbf** |
| **reserved** | **10** | **bslbf** |
| **marker\_bit** | **1** | **bslbf** |
| if (jump\_flag = ='1') { |  |  |
| **reserved** | **7** | **bslbf** |
| **direction\_indicator** | **1** | **bslbf** |
| time\_code() |  |  |
| } |  |  |
| if (play\_flag = ='1'){ |  |  |
| **speed\_mode** | **1** | **bslbf** |
| **direction\_indicator** | **1** | **bslbf** |
| **reserved** | **6** | **bslbf** |
| time\_code() |  |  |
| } |  |  |
| } |  |  |
| if (storage\_flag = ='1') { |  |  |
| **reserved** | **6** | **bslbf** |
| **record\_flag** | **1** | **bslbf** |
| **stop\_mode** | **1** | **bslbf** |
| if (record\_flag = ='1') { |  |  |
| time\_code() |  |  |
| } |  |  |
| } |  |  |
| } |  |  |

### B.3.5 Semantics of fields in control layer

**marker\_bit** – This is a 1-bit marker that is always set to '1' to avoid start code emulation.

**reserved\_bits** – This 12-bit field is reserved for future use by this Recommendation | International Standard for DSM control commands. Until otherwise specified by ITU-T | ISO/IEC it shall have the value '0000 0000 0000'.

**select\_flag** – This 1-bit flag when set to '1' specifies a bitstream selection operation. When it is set to '0' no bitstream selection operation shall occur.

**retrieval\_flag** – This 1-bit flag when set to '1' specifies that a specific retrieval (playback) action will occur. The operation starts from the current operational PTS value.

**storage\_flag** – This 1-bit flag when set to '1' specifies that a storage operation is to be executed.

**bitstream\_ID** – This 32-bit field is coded in three parts. The parts are combined to form an unsigned integer specifying which Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream is to be selected. It is the DSM server's responsibility to map the names of the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstreams stored on its DSM uniquely to a series of numbers which could be represented by the bitstream\_ID.

**select\_mode** – This 5-bit unsigned integer specifies which mode of bitstream operation is requested. Table B.4 specifies the defined modes.

Table B.4 – Select mode assigned values

|  |  |
| --- | --- |
| Code | Mode |
| 0x00 | Forbidden |
| 0x01 | Storage |
| 0x02 | Retrieval |
| 0x03 .. 0x1F | Reserved |

**jump\_flag** – This 1-bit flag when set to '1' specifies a jump in the playback pointer to a new access unit. The new PTS is specified by a relative time\_code with respect to the current operational PTS value. This function is only valid when the current Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream is in the "stop" mode.

**play\_flag** – This 1-bit flag when set to '1' specifies to play a bitstream for a certain time period. The speed, direction, and play duration are additional parameters in the bit stream. The play starts from the current operational PTS value.

**pause\_mode** – This is a one-bit code specifying to pause the playback action and keep the playback pointer at the current operational PTS value.

**resume\_mode** – This is a one-bit code specifying to continue the playback action from the current operational PTS value. Resume only has meaning if the current bitstream is in the "pause" state, and the bitstream will be set to the forward play state at normal speed.

**stop\_mode** – This is a one-bit code specifying to stop a bitstream transmission.

**direction\_indicator** – This is a one-bit code to indicate the playback direction. If this bit is set to '1', it stands for a forward play. Otherwise it stands for a backward play.

**speed\_mode** – This is a 1-bit code to specify the speed scale. If this bit is set to '1', it specifies that the speed is normal play. If this bit is set to '0', it specifies that the speed is fast play (i.e., fast forward or fast reverse).

**record\_flag** – This is one-bit flag to specify the request of recording the bitstream from an end user to a DSM for a specified duration or until the reception of a stop command, whichever comes first.

### B.3.6 Acknowledgement layer

Constraints on setting flags in DSM-CC control

Only one of the acks bits specified below can be set to '1' for each DSM ack bitstream (see Table B.5).

Table B.5 – DSM-CC Acknowledgement

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| ack() { |  |  |
| **select\_ack** | **1** | **bslbf** |
| **retrieval\_ack** | **1** | **bslbf** |
| **storage\_ack** | **1** | **bslbf** |
| **error\_ack** | **1** | **bslbf** |
| **reserved** | **10** | **bslbf** |
| **marker\_bit** | **1** | **bslbf** |
| **cmd\_status** | **1** | **bslbf** |
| If (cmd\_status = = '1' &&  (retrieval\_ack = ='1' || storage\_ack = = '1')) { |  |  |
| time\_code() |  |  |
| } |  |  |
| } |  |  |

### B.3.7 Semantics of fields in acknowledgement layer

**select\_ack** – This 1-bit field when it is set to '1' indicates that the ack() command is to acknowledge a select command.

**retrieval\_ack** – This 1-bit field when set to '1' indicates that the ack() command is to acknowledge a retrieval command.

**storage\_ack** – This 1-bit field when set to '1' indicates that the ack() command is to acknowledge a storage command.

**error\_ack** – This 1-bit field when set to '1' indicates a DSM error. The defined errors are EOF (end of file on forward play or start of file on reverse play) on a stream being retrieved and Disk Full on a stream being stored. If this bit is set to '1', cmd\_status is undefined. The current bitstream is still selected.

**cmd\_status** – This 1-bit flag set to '1' indicates that the command is accepted. When set to '0' it indicates the command is rejected. The semantics vary according to the command received as follows:

• If select\_ack is set and cmd\_status is set to '1', it specifies that the Rec. ITU-T H.222.0 | ISO/IEC 13818‑1 bitstream is selected and the server is ready to provide the selected mode of operation. The current operational PTS value is set to the first point of random access of the newly selected Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream. If cmd\_status is set to '0', the operation has failed and no bitstream is selected.

• If retrieval\_ack is set and cmd\_status is set to '1', it specifies that the retrieval operation is initiated for all retrieval commands. The position of the current operational PTS pointer is reported by the succeeding time\_code.

• For the play\_flag command with infinite\_time\_flag != '1', a second acknowledgement will be sent. This will acknowledge that the play operation has ended by reaching the duration defined by the play\_flag command.

• If the cmd\_status is set to '0' in a retrieval acknowledgement, the operation has failed. Possible reasons for this failure include an invalid bitstream\_ID, jumping beyond the end of a file, or a function not supported such as reverse play in standard speed.

• If storage\_ack is set, it specifies that the storage operation is being started for the record\_flag command or is completed by the stop\_mode command. The PTS of the last complete access unit stored is reported by the succeeding time\_code.

• If the recording operation is ended by reaching the duration defined by the storage\_flag command, another acknowledgement shall be sent and the current operational PTS value after the recording shall be reported.

• If the cmd\_status is set to '0' in a storage acknowledgement, the operation has failed. Possible reasons for this failure include an invalid bitstream\_ID, or the inability of the DSM to store data.

### B.3.8 Time code

Constraints on time code

• A forward operation of specified duration given by a time\_code terminates after the actual or implied PTS of an access unit is observed such that PTS minus the current operational PTS value at the start of the operation modulo 233 exceeds the duration.

• A backward operation of specified duration given by a time\_code terminates after the actual or implied PTS of an access unit is observed such that current operational PTS value at the start of the operation minus that PTS modulo 233 exceeds the duration.

• For all the commands in the control() layer, the time\_code is specified as a relative duration with respect to the current operational PTS value.

• For all the commands in the ack() layer, the time\_code is specified by the current operational PTS value.

See Table B.6.

Table B.6 – Time code

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| time\_code() { |  |  |
| **reserved** | **7** | **bslbf** |
| **infinite\_time\_flag** | **1** | **bslbf** |
| if (infinite\_time\_flag = = '0') { |  |  |
| **reserved** | **4** | **bslbf** |
| **PTS [32..30]** | **3** | **bslbf** |
| **marker** | **1** | **bslbf** |
| **PTS [29..15]** | **15** | **bslbf** |
| **marker\_bit** | **1** | **bslbf** |
| **PTS [14..0]** | **15** | **bslbf** |
| **marker\_bit** | **1** | **bslbf** |
| } |  |  |
| } |  |  |

### B.3.9 Semantics of fields in time code

**infinite\_time\_flag** – This 1-bit flag when set to '1' indicates an infinite time period. This flag is set to '1' in applications where a time period for a specific operation could not be defined in advance.

**PTS [32..0]** – The presentation timestamp of the access unit of the bitstream. Depending upon the function, this can be an absolute value or a relative time delay in cycles of the 90-kHz system clock.

Annex C  
  
Program-specific information

(This annex does not form an integral part of this Recommendation | International Standard.)

## C.1 Explanation of program-specific information in transport streams

Clause 2.4.4 contains the normative syntax, semantics and text concerning program-specific information. In all cases, compliance with the constraints of 2.4.4 is required. This annex provides explanatory information on how to use the PSI functions, and considers examples of how it may be used in practice.

## C.2 Introduction

This Recommendation | International Standard provides a method for describing the contents of transport stream packets for the purpose of the demultiplexing and presentation of programs. The coding specification accommodates this function through the program-specific information (PSI). This annex discusses the use of PSI.

The PSI may be thought of as belonging to six tables:

1) Program association table (PAT);

2) TS program map table (PMT);

3) Network information table (NIT);

4) Conditional access table (CAT).

5) Transport stream description table (TSDT); and

6) IPMP control information table.

The contents of the PAT, PMT, CAT and TSDT are specified in this Recommendation | International Standard. ICIT is defined in ISO/IEC 13818-11 (MPEG-2 IPMP).

The NIT is a private table, and the PID value of the transport stream packets which carry it is specified in the PAT. Both the NIT and ICIT must follow the structure defined in this Recommendation | International Standard.

## C.3 Functional mechanism

The tables listed above are conceptual in that they need never be regenerated in a specified form within a decoder. While these structures may be thought of as simple tables, they may be partitioned before they are sent in transport stream packets. The syntax supports this operation by allowing the tables to be partitioned into sections and by providing a normative mapping method into transport stream packet payloads. A method is also provided to carry private data in a similar format. This is advantageous as the same basic processing in the decoder can then be used for both the PSI data and the private data helping to keep cost down. For advice on the optimum placing of PSI in the transport stream, see Annex D.

Each section is uniquely identified by the combination of the following elements:

i) **table\_id**

The 8-bit table\_id identifies to which table the section belongs.

• Sections with table\_id 0x00 belong to the Program Association Table.

• Sections with table\_id 0x01 belong to the Conditional Access Table.

• Sections with table\_id 0x02 belong to the TS Program Map Table.

• Sections with table\_id 0x03 belong to the TS\_description\_section.

• Sections with table\_id 0x04 belong to the ISO\_IEC\_14496\_scene\_description\_section.

• Sections with table\_id 0x05 belong to the ISO\_IEC\_14496\_object\_descriptor\_section.

• Sections with table\_id 0x06 belong to the metadata\_section.

• Sections with table\_id 0x07 belong to the IPMP\_Control\_Information\_section.

Other values of the table\_id can be allocated by the user for private purposes.

It is possible to set up filters looking at the table\_id field to identify whether a new section belongs to a table of interest or not.

ii) **table\_id\_extension**

This 16-bit field exists in the long version of a section. In the Program Association Table it is used to identify the transport\_stream\_id of the stream – effectively a user-defined label which allows one transport stream to be distinguished from another within a network or across networks. In the conditional access table this field currently has no meaning and is therefore marked as "reserved" meaning that it shall be coded as 0xFFFF, but that a meaning may be defined by ITU-T | ISO/IEC in a subsequent revision of this Recommendation | International Standard. In a TS Program Map section the field contains the program\_number, and thereby identifies the program to which the data in the section refers. The table\_id\_extension can also be used as a filter point in certain cases.

iii) **section\_number**

The section\_number field allows the sections of a particular table to be reassembled in their original order by the decoder. There is no obligation within this Recommendation | International Standard that sections must be transmitted in numerical order, but this is recommended, unless it is desired to transmit some sections of the table more frequently than others, e.g., due to random access considerations.

iv) **version\_number**

When the characteristics of the transport stream described in the PSI change (e.g., extra programs added, different composition of elementary streams for a given program), then new PSI data has to be sent with the updated information as the most recently transmitted version of the sections marked as "current" must always be valid. Decoders need to be able to identify whether the most recently received section is identical with the section they have already processed/stored (in which case the section can be discarded), or whether it is different, and may therefore signify a configuration change. This is achieved by sending a section with the same table\_id, table\_id\_extension, and section\_number as the previous section containing the relevant data, but with the next value version\_number.

v) **current\_next\_indicator**

It is important to know at what point in the bitstream the PSI is valid. Each section can therefore be numbered as valid "now" (current), or as valid in the immediate future (next). This allows the transmission of a future configuration in advance of the change, giving the decoder the opportunity to prepare for the change. There is however no obligation to transmit the next version of a section in advance, but if it is transmitted, then it shall be the next correct version of that section.

## C.4 The mapping of sections into transport stream packets

Sections are mapped directly into transport stream packets, that is to say, without a prior mapping into PES packets. Sections do not have to start at the beginning of transport stream packets, (although they may), because the start of the first section in the payload of a transport stream packet is pointed to by the pointer\_field. The presence of the pointer\_field is signalled by the payload\_unit\_start\_indicator being set to a value of '1' in PSI packets. (In non‑PSI packets, the indicator signals that a PES packet starts in the transport stream packet.) The pointer\_field points to the start of the first section in the transport stream packet. There is never more than one pointer\_field in a transport stream packet, as the start of any other section can be identified by counting the length of the first and any subsequent sections, since no gaps between sections within a transport stream packet are allowed by the syntax.

It is important to note that within transport stream packets of any single PID value, one section must be finished before the next one is allowed to be started, or else it is not possible to identify to which section header the data belongs. If a section finishes before the end of a transport stream packet, but it is not convenient to open another section, a stuffing mechanism is provided to fill up the space. Stuffing is performed by filling each remaining byte of the packet with the value 0xFF. Consequently the table\_id value 0xFF is forbidden, or else this would be confused with stuffing. Once a 0xFF byte has occurred at the end of a section, then the rest of the transport stream packet must be stuffed with 0xFF bytes, allowing a decoder to discard the rest of the transport stream packet. Stuffing can also be performed using the normal adaptation\_field mechanism.

## C.5 Repetition rates and random access

In systems where random access is a consideration, it is recommended to re-transmit PSI sections several times, even when changes do not occur in the configuration, as in the general case, a decoder needs the PSI data to identify the contents of the transport stream, to be able to start decoding. This Recommendation | International Standard does not place any requirements on the repetition or occurrence rate of PSI sections. Clearly though, repeating sections frequently helps random access applications, whilst causing an increase in the amount of bitrate used by PSI data. If program mappings are static or quasi-static, they may be stored in the decoder to allow faster access to the data than having to wait for it to be re-transmitted. The trade-off between the amount of storage required and the desired impact on channel acquisition time may be made by the decoder manufacturer.

## C.6 What is a program?

The concept of a program has a precise definition within this Recommendation | International Standard [refer to 2.1.123 program (system)]. For a transport stream the time base is defined by the PCR. This effectively creates a virtual channel within the transport stream.

Note that this is not the same definition as is commonly used in broadcasting, where a "program" is a collection of elementary streams not only with a common timebase, but also with a common start and end time. A series of "broadcaster programs" (referred to in this annex as events) can be transmitted sequentially in a transport stream using the same program\_number to create a "broadcasting conventional" TV-channel (sometimes called a service).

Event descriptions could be transmitted in private\_sections().

A program is denoted by a program\_number which has significance only within a transport stream. The program\_number is a 16-bit unsigned integer and thus permits 65535 unique programs to exist within a transport stream (program\_number 0 is reserved for identification of the NIT). Where several transport streams are available to the decoder (e.g., in a cable network), in order to successfully demultiplex a program, the decoder must be notified of both the transport\_stream\_id (to find the right multiplex) and the program\_number of the service (to find the right program within the multiplex).

The transport stream mapping may be accomplished via the optional network information table. Note that the network information table may be stored in decoder non-volatile memory to reduce channel acquisition time. In this case, it needs to be transmitted only often enough to support timely decoder initialization set-up operations. The contents of the NIT are private, but shall take at least the minimum section structure.

## C.7 Allocation of program\_number

It may not be convenient in all cases to group together all the program elements which share a common clock reference as one program. It is conceivable to have a multi-service transport stream with only one set of PCRs, common to all. In general, though, a broadcaster may prefer to logically split up the transport stream into several programs, where the PCR\_PID (location of the clock reference) is always the same. This method of splitting the program elements into pseudo-independent programs can have several uses. Two examples follow:

i)multilingual *transmissions into separate markets*

One video stream may be accompanied by several audio streams in different languages. It is advisable to include an example of the ISO\_639\_language\_descriptor associated with each audio stream to enable the selection of the correct program and audio. It is reasonable to have several program definitions with different program\_numbers, where all the programs reference the same video stream and PCR\_PID, but have different audio PIDs. It is, however, also reasonable and possible to list the video stream and all the audio streams as one program, where this does not exceed the section size limit of 1024 bytes.

ii)*Very large program definitions*

There is a maximum limit on the length of a section of 1024 bytes (including section header and CRC\_32). This means that no single program definition may exceed this length. For the great majority of cases, even with each program element having several descriptors, this size is adequate. However, one may envisage cases in very high bitrate systems, which could exceed this limit. It is then in general possible to identify methods of splitting the references of the streams, so that they do not all have to be listed together. Some program elements could be referenced under more than one program, and some under only one or the other, but not both.

## C.8 Usage of PSI in a typical system

A communications system, especially in broadcast applications, may consist of many individual transport streams. Each one of the four PSI data structures may appear in each and every transport stream in a system. There must always be a complete version of the program association table listing all programs within the transport stream and a complete TS program map table, containing complete program definitions for all programs within the transport stream. If any streams are scrambled, then there must also be a conditional access table present listing the relevant Entitlement Management Messages (EMM) streams. The presence of a NIT is fully optional.

The PSI tables are mapped into transport stream packets via the section structure described above. Each section has a table\_id field in its header, allowing sections from PSI tables and private data in private\_sections to be mixed in transport stream packets of the same PID value or even in the same transport stream packet. Note, however, that within packets of the same PID, a complete section must be transmitted before the next section can be started. This is only possible for packets labelled as containing TS Program Map Table section or NIT packets however, since private sections may not be mapped into PAT or CAT packets.

It is required that all PAT sections be mapped into transport stream packets with PID = 0x0000 and all CA sections be mapped into packets with PID = 0x0001. PMT sections may be mapped into packets of user-selected PID value, listed as the PMT\_PID for each program in the Program Association Table. Likewise, the PID for the NIT-bearing transport stream packets is user-selected, but must be pointed to by the entry "program\_number = = 0x00" in the PAT, if the NIT exists.

The contents of any CA parameter streams are entirely private, but EMMs and ECMs must also be sent in transport stream packets to be compliant with this Recommendation | International Standard.

Private data tables may be sent using the private\_section() syntax. Such tables could be used for example in a broadcasting environment to describe a service, an upcoming event, broadcast schedules and related information.

## C.9 The relationships of PSI structures

Figure C.1 shows an example of the relationship between the four PSI structures and the transport stream. Other examples are possible, but the figure shows the primary connections.

In the following subclauses, each PSI table is described.



Figure C.1 – Program and network mapping relationships

### C.9.1 Program Association Table

Every transport stream must contain a complete valid Program Association Table. The Program Association Table gives the correspondence between a program\_number and the PID of the transport stream packets that carry the definition of that program (the PMT\_PID). The PAT may be partitioned into up to 255 sections before it is mapped into transport stream packets. Each section carries a part of the overall PAT. This partitioning may be desirable to minimize data loss in error conditions. That is, packet loss or bit errors may be localized to smaller sections of the PAT, thus allowing other sections to still be received and correctly decoded. If all PAT information is put into one section, an error causing a changed bit in the table\_id, for example, would cause the loss of the entire PAT. However, this is still permitted as long as the section does not extend beyond the 1024-byte maximum length limit.

Program 0 (zero) is reserved and is used to specify the Network PID. This is a pointer to the transport stream packets which carry the network information table.

The program association table is always transmitted without encryption.

### C.9.2 Program map table

The program map table (MPT) provides the mapping between a program number and the program elements that comprise it. This table is present in transport stream packets having one or more privately-selected PID values. These transport stream packets may contain other private structures as defined by the table\_id field. It is possible to have TS PMT sections referring to different programs carried in transport stream packets having a common PID value.

This Recommendation | International Standard requires a minimum of program identification: program number, PCR PID, stream types and program elements PIDs. Additional information for either programs or elementary streams may be conveyed by use of the descriptor() construct. Refer to C.9.6.

Private data may also be sent in transport stream packets denoted as carrying TS program map table sections. This is accomplished by the use of the private\_section(). In a private\_section() the application decides whether version\_number and current\_next\_indicator represent the values of these fields for a single section or whether they are applicable to many sections as parts of a larger private table.

NOTE 1 – Transport stream packets containing the Program Map Table are transmitted unencrypted.

NOTE 2 – It is possible to transmit information on events in private descriptors carried within the TS\_program\_map\_section()s.

### C.9.3 Conditional access table

The conditional access table (CAT) gives the association between one or more conditional access (CA) systems, their EMM streams and any special parameters associated with them.

NOTE – The (private) contents of the transport stream packets containing EMM and CA parameters if present will, in general, be encrypted (scrambled).

### C.9.4 Network information table

The contents of the network information table (NIT) are private and not specified by this Recommendation | International Standard. In general, it will contain mappings of user-selected services with transport\_stream\_ids, channel frequencies, satellite transponder numbers, modulation characteristics, etc.

### C.9.5 Private\_section()

Private\_sections() can occur in two basic forms, the short version (where only the fields up to and including section\_length are included) or the long version (where all the fields up to and including last\_section\_number are present, and after the private data bytes the CRC\_32 field is present).

Private\_section()s can occur in PIDs which are labelled as PMT\_PIDs or in transport stream packets with other PID values which contain exclusively private\_sections(), including the PID allocated to the NIT. If the transport stream packets of the PID carrying the private\_section()s are identified as a PID carrying private\_sections (stream\_type assignment value 0x05), then only private\_sections may occur in transport stream packets of that PID value. The sections may be either of the short or long type.

### C.9.6 Descriptors

There are several normative descriptors defined in this Recommendation | International Standard. Many more private descriptors may also be defined. All descriptors have a common format: {tag, length, data}. Any privately defined descriptors must adhere to this format. The data portion of these private descriptors are privately defined.

One descriptor (the CA\_descriptor()), is used to indicate the location (PID value of transport packets) of ECM data associated with program elements when it is found in a TS PMT section. When found in a CA section it refers to EMMs.

In order to extend the number of private\_descriptors available, the following mechanism could be used: A private descriptor\_tag could be privately defined to be constructed as a composite descriptor. This entails privately defining a further sub\_descriptor as the first field of the private data bytes of the private descriptor. The described structure is as indicated in Tables C.1 and C.2.

Table C.1 – Composite\_descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| Composite\_descriptor(){ |  |  |
| **descriptor\_tag(privately defined)** | **8** | **uimsbf** |
| **descriptor\_length** | **8** | **uimsbf** |
| for (i = 0; i < N; i++){ |  |  |
| sub\_descriptor() |  |  |
| } |  |  |
| } |  |  |

Table C.2 – Sub-descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| sub\_descriptor() { |  |  |
| **sub\_descriptor\_tag** | **8** | **uimsbf** |
| **sub\_descriptor\_length** | **8** | **uimsbf** |
| for (i = 0; i < N; i++) { |  |  |
| **private\_data\_byte** | **8** | **uimsbf** |
| } |  |  |
| } |  |  |

## C.10 Bandwidth utilization and signal acquisition time

Any implementation of a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 bitstream must make reasonable bandwidth demands for PSI information and, in applications where random access is a consideration, should promote fast signal acquisition. This subclause analyses this issue and gives some broadcast application examples.

The packet-based nature of the transport stream allows for the interspersing of PSI information with fine granularity in the multiplexed data. This provides significant flexibility in the construction and transmission of PSI.

Signal acquisition time in a real decoder is dependent on many factors, including: FDM tuning slew time, demultiplexing time, sequence headers, I-frame occurrence rate and scrambling key retrieval and processing.

This subclause examines both the bitrate and signal acquisition time impacts of the PSI syntax in 2.4.4.4 and 2.4.4.9. It is assumed that the conditional access table does not need to be received dynamically at every program change. This assumption is also made of the private EMM streams. This is because these streams do not contain the quickly-varying ECM components used for program element scrambling (encryption).

Also, in the discussion below, the time to acquire and process ECMs has been neglected.

Tables C.3 and C.4 provide bandwidth usage values for a range of transport stream conditions. One axis of the table is the number of programs contained in a single transport stream. The other axis is the frequency with which the PSI information is transmitted in the transport stream.

Table C.3 – Program association table bandwidth usage (bit/s)  
Number of programs per transport stream

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | 1 | 5 | 10 | 32 | 128 |
|  | 1 | 1504 | 1504 | 1504 | 1504 | 4512 |
| Frequency of PA table | 10 | 15040 | 15040 | 15040 | 15040 | 45120 |
| Information | 25 | 37600 | 37600 | 37600 | 37600 | 112800 |
| (s–1) | 50 | 75200 | 75200 | 75200 | 75200 | 225600 |
|  | 100 | 150400 | 150400 | 150400 | 150400 | 451200 |
| NOTE – Since 46 **program\_association\_sections** fit into one transport packet, the numbers in the table do not change until the last column. | | | | | | |

Table C.4 – Program map table bandwidth usage (bit/s)  
Number of programs per transport stream

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | 1 | 5 | 10 | 32 | 128 |
|  | 1 | 1504 | 1504 | 3008 | 7520 | 28576 |
| Frequency of PM table | 10 | 15040 | 15040 | 30080 | 75200 | 285760 |
| Information | 25 | 37600 | 37600 | 75200 | 188000 | 714400 |
| (s–1) | 50 | 75200 | 75200 | 150400 | 376000 | 1428800 |
|  | 100 | 150400 | 150400 | 300800 | 601600 | 2857600 |

This frequency will be a key determinant of the component of signal acquisition time due to PSI structures.

Both bandwidth usage tables assume that only the minimum program mapping information is provided. This means that the PID values and stream types are provided with no additional descriptors. All programs in the example are composed of two elementary streams. Program associations are 2 bytes long, while the minimal program map is 26 bytes long. There is additional overhead associated with version numbers, section lengths, etc. This will be on the order of 1-3% of the total PSI bitrate usage in sections of moderate to maximum length (a few hundred bytes to 1024 bytes) and will thus be ignored here.

The above assumptions allow forty-six (46) program associations to map into one Program Association Table transport stream packet (if no adaptation field is present). Similarly, seven (7) TS\_program\_map\_sections fit into a single transport stream packet. It may be noted that to facilitate easy "drop/add" it is possible to transmit only one (1) TS\_program\_map\_section per PMT\_PID. This may cause an undesirable increase in PSI bitrate usage, however.

Using a frequency of 25 Hz for the two PSI Tables, yields a worst-case contribution to the signal acquisition time of approximately 80 ms. This would only occur when the required PAT data was "just missed" and then, once the PAT was acquired and decoded, the required PMT data was also "just missed". This doubling of the worst case acquisition time is one disadvantage of the extra level of indirection introduced by the PAT structure. This effect could be reduced by coordinated transmission of related PAT and PMT packets. Presumably, the advantage that this approach offers for "drop/add" re-multiplexing operations is compensatory.

With the 25-Hz PSI frequency, the following examples may be constructed (all examples leave ample allowance for various datalink, FEC, CA and routing overheads):

6-MHz CATV channel

• five 5.2-Mbit/s programs: 26.5 Mbit/s (includes transport overhead)

• total PSI bandwidth: 5.2 kbit/s

• CA bandwidth: 500 kbit/s

*total Rec. ITU-T H.222.0 | ISO/IEC 13818-1 transport bandwidth: 27.1 Mbit/s*

• PSI Overhead: 0.28%

OC-3 fibre channel (155 Mbit/s)

• 32 3.9-Mbit/s programs: 127.5 Mbit/s (includes transport overhead)

• total PSI bandwidth: 225.6 kbit/s

• CA bandwidth: 500 kbit/s

*total Rec. ITU-T H.222.0 | ISO/IEC 13818-1 transport bandwidth: 128.2 Mbit/s*

• PSI Overhead: 0.18%

C-band satellite transponder

• 128 256-kbit/s audio programs: 33.5 Mbit/s (includes transport overhead)

• total PSI bandwidth: 826.4 kbit/s

• CA bandwidth: 500 kbit/s

*total Rec. ITU-T H.222.0 | ISO/IEC 13818-1 transport bandwidth: 34.7 Mbit/s*

• PSI overhead: 2.4%(actually would be lower if only one PID used per program)

As expected, the percent overhead increases for lower-rate services since many more services are possible per transport stream. However, the overhead is not excessive in all cases. Higher transmission rates (than 25 Hz) for the PSI data may be used to decrease the impact on channel acquisition time with only modest bitrate demand increases.

Annex D  
  
Systems timing model and application implications  
of this Recommendation | International Standard

(This annex does not form an integral part of this Recommendation | International Standard.)

## D.1 Introduction

The Rec. ITU-T H.222.0 | ISO/IEC 13818-1 Systems specification includes a specific timing model for the sampling, encoding, encoder buffering, transmission, reception, decoder buffering, decoding, and presentation of digital audio and video in combination. This model is embodied directly in the specification of the syntax and semantic requirements of compliant Rec. ITU-T H.222.0 | ISO/IEC 13818-1 data streams. Given that a decoding system receives a compliant bit stream that is delivered correctly in accordance with the timing model it is straightforward to implement the decoder such that it produces as output high quality audio and video which are properly synchronized. There is no normative requirement, however, that decoders be implemented in such a way as to provide such high quality presentation output. In applications where the data are not delivered to the decoder with correct timing, it may be possible to produce the desired presentation output; however, such capabilities are not in general guaranteed. This informative annex describes the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 Systems timing model in detail, and gives some suggestions for implementing decoder systems to suit some typical applications.

### D.1.1 Timing model

Rec. ITU-T H.222.0 | ISO/IEC 13818-1 Systems embodies a timing model in which all digitized pictures and audio samples that enter the encoder are presented exactly once each, after a constant end to end delay, at the output of the decoder. As such, the sample rates, i.e., the video frame rate and the audio sample rate, are precisely the same at the decoder as they are at the encoder. This timing model is diagrammed in Figure D.1:



Figure D.1 – Constant delay model

As indicated in Figure D.1, the delay from the input to the encoder to the output or presentation from the decoder is constant in this model[[2]](#footnote-3)1), while the delay through each of the encoder and decoder buffers is variable. Not only is the delay through each of these buffers variable within the path of one elementary stream, the individual buffer delays in the video and audio paths differ as well. Therefore the relative location of coded bits representing audio or video in the combined stream does not indicate synchronization information. The relative location of coded audio and video is constrained only by the System Target Decoder (STD) model such that the decoder buffers must behave properly; therefore coded audio and video that represent sound and pictures that are to be presented simultaneously may be separated in time within the coded bit stream by as much as one second, which is the maximum decoder buffer delay that is allowed in the STD model.

The audio and video sample rates at the encoder are significantly different from one another, and may or may not have an exact and fixed relationship to one another, depending on whether the combined stream is a program stream or a transport stream, and on whether the System\_audio\_locked and System\_video\_locked flags are set in the program stream. The duration of a block of audio samples (an audio presentation unit) is generally not the same as the duration of a video picture.

There is a single, common system clock in the encoder, and this clock is used to create timestamps that indicate the correct presentation and decoding timing of audio and video, as well as to create timestamps that indicate the instantaneous values of the system clock itself at sampled intervals. The timestamps that indicate the presentation time of audio and video are called Presentation Time Stamps (PTS). Those that indicate the decoding time are called Decoding Timestamps (DTS), and those that indicate the value of the system clock are called the System Clock Reference (SCR) in program streams and the Program Clock Reference (PCR) in transport streams. It is the presence of this common system clock in the encoder, the timestamps that are created from it, and the recreation of the clock in the decoder and the correct use of the timestamps that provide the facility to synchronize properly the operation of the decoder.

Encoder implementations may not follow this model exactly; however, the data stream which results from the actual encoder, storage system, network, and one or more multiplexor must follow the model precisely. (Delivery of the data may deviate somewhat, depending on the application). Therefore in this annex, the term "encoder system clock" is used to mean either the actual common system clock as described in this model or the equivalent function, however it may be implemented.

Since the end-to-end delay through the entire system is constant, the audio and video presentations are precisely synchronized. The construction of System bit streams is constrained such that when they are decoded by a decoder that follows this model with the appropriately sized decoder buffers, those buffers are guaranteed never to overflow nor underflow, with specific exceptions allowing intentional underflow.

In order for the decoder system to incur the precise amount of delay that causes the entire end-to-end delay to be constant, it is necessary for the decoder to have a system clock whose frequency of operation and absolute instantaneous value match those of the encoder. The information necessary to convey the encoder's system clock is encoded in the SCR or PCR; this function is explained below.

Decoders which are implemented in accordance with this timing model such that they present audio samples and video pictures exactly once (with specific intentionally coded exceptions), at a constant rate, and such that decoder buffers behave as in the model, are referred to in this annex as precisely timed decoders, or those that produce precisely timed output. Decoder implementations are not required by this International Standard to present audio and video in accordance with this model; it is possible to construct decoders that do not have constant delay, or equivalently do not present each picture or audio sample exactly once. In such implementations, however, the synchronization between presented audio and video may not be precise, and the behaviour of the decoder buffers may not follow the reference decoder model. It is important to avoid overflow at the decoder buffers, as overflow causes a loss of data that may have significant effects on the resulting decoding process. This annex covers primarily the operation of such precisely timed decoders and some of the options that are available in implementing these decoders.

### D.1.2 Audio and video presentation synchronization

Within the coding of this Recommendation | International Standard Systems data are timestamps concerning the presentation and decoding of video pictures and blocks of audio samples. The pictures and blocks are called "Presentation Units", abbreviated PU. The sets of coded bits which represent the PUs and which are included within the Rec. ITU-T H.222.0 | ISO/IEC 13818‑1 bit stream are called "Access Units", abbreviated AU. An audio access unit is abbreviated AAU, and a video access unit is abbreviated VAU. In ISO/IEC 13818-3 audio the term "audio frame" has the same meaning as AAU or APU (audio presentation unit) depending on the context. A video presentation unit (VPU) is a picture, and a VAU is a coded picture.

Some, but not necessarily all, AAUs and VAUs have associated with them PTSs. A PTS indicates the time that the PU which results from decoding the AU which is associated with the PTS should be presented to the user. The audio PTSs and video PTSs are both samples from a common time clock, which is referred to as the System Time Clock or STC. With the correct values of audio and video PTSs included in the data stream, and with the presentation of the audio and video PUs occurring at the time indicated by the appropriate PTSs in terms of the common STC, precise synchronization of the presented audio and video is achieved at the decoding system. While the STC is not part of the normative content of this Recommendation | International Standard, and the equivalent information is conveyed in this Recommendation | International Standard via such terms as the system\_clock\_frequency, the STC is an important and convenient element for explaining the timing model, and it is generally practical to implement encoders and decoders which include an STC in some form.

PTSs are required for the conveyance of accurate relative timing between audio and video, since the audio and video PUs generally have significantly different and essentially unrelated duration. For example, audio PUs of 1152 samples each at a sample rate of 44 100 samples per second have a duration of approximately 26.12 ms, and video PUs at a frame rate of 29.97 Hz have a duration of approximately 33.76 ms. In general the temporal boundaries of APUs and VPUs rarely, if ever, coincide. Separate PTSs for audio and video provide the information that indicates the precise temporal relation of audio and video PUs without requiring any specific relationship between the duration and interval of audio and video PUs.

The values of the PTS fields are defined in terms of the System Target Decoder or STD, which is a fundamental normative constraint on all System bit streams. The STD is a mathematical model of an idealized decoder which specifies precisely the movement of all bits into and out of the decoder's buffers, and the basic semantic constraint imposed on the bit stream is that the buffers within the STD must never overflow nor underflow, with specific exceptions provided for underflow in special cases. In the STD model the virtual decoder is always exactly synchronized with the data source, and audio and video decoding and presentation are exactly synchronized. While exact and consistent, the STD is somewhat simplified with respect to physical implementations of decoders in order to clarify its specification and to facilitate its broad application to a variety of decoder implementations. In particular, in the STD model each of the operations performed on the bit stream in the decoder is performed instantaneously, with the obvious exception of the time that bits spend in the decoder buffers. In a real decoder system the individual audio and video decoders do not perform instantaneously, and their delays must be taken into account in the design of the implementation. For example, if video pictures are decoded in exactly one picture presentation interval 1/P, where P is the frame rate, and compressed video data are arriving at the decoder at bit rate R, the completion of removing bits associated with each picture is delayed from the time indicated in the PTS and DTS fields by 1/P, and the video decoder buffer must be larger than that specified in the STD model by R/P. The video presentation is likewise delayed with respect to the STD, and the PTS should be handled accordingly. Since the video is delayed, the audio decoding and presentation should be delayed by a similar amount in order to provide correct synchronization. Delaying decoding and presentation of audio and video in a decoder may be implemented for example by adding a constant to the PTS values when they are used within the decoder.

Another difference between the STD and precise practical decoder implementation is that in the STD model the explicit assumption is made that the final audio and video output is presented to the user instantaneously and without further delay. This may not be the case in practice, particularly with cathode-ray tube displays, and this additional delay should also be taken into account in the design. Encoders are required to encode audio and video such that the correct synchronization is achieved when the data is decoded with the STD. Delays in the input and sampling of audio and video, such as video camera optical charge integration, must be taken into account in the encoder.

In the STD model proper synchronization is assumed and the timestamps and buffer behaviour are tested against this assumption as a condition of bit stream validity. Of course in a physical decoder precise synchronization is not automatically the case, particularly upon start-up and in the presence of timing jitter. Precise decoder timing is a goal to be targeted by decoder designs. Inaccuracy in decoder timing affects the behaviour of the decoder buffers. These topics are covered in more detail in later subclauses of this annex.

The STD includes Decoding Time Stamps (DTS) as well as PTS fields. The DTS refers to the time that an AU is to be extracted from the decoder buffer and decoded in the STD model. Since the audio and video elementary stream decoders are instantaneous in the STD, the decoding time and presentation time are identical in most cases; the only exception occurs with video pictures which have undergone re-ordering within the coded bit stream, i.e., I- and P‑pictures in the case of non-low-delay video sequences. In cases where re-ordering exists, a temporary delay buffer in the video decoder is used to store the appropriate decoded I- or P-picture until it should be presented. In all cases where the decoding and presentation times are identical in the STD, i.e., all AAUs, B-picture VAUs, and I- and P-picture VAUs within low-delay video sequences, the DTS is not coded, as it would have the same value as the PTS. Where the values differ, both are coded if either is coded. For all AUs where only the PTS is coded, this field may be interpreted as being both the PTS and the DTS.

Since PTS and DTS values are not required for every AAU and VAU, the decoder may choose to interpolate values which are not coded. PTS values are required with intervals not exceeding 700 ms in each elementary audio and video stream. These time intervals are measured in presentation time, that is, in the same context as the values of the fields, not in terms of the times that the fields are transmitted and received. In cases of data streams where the system, video and audio clocks are locked, as defined in the normative part of this Recommendation | International Standard, each AU following one for which a DTS or PTS is explicitly coded has an effective decoding time of the sum of that for the previous AU plus a fixed and specified difference in value of the STC. For example, in video coded at 29.97 Hz each picture has a difference in time of 3003 cycles of the 90-kHz portion of the STC from the previous picture when the video and system clocks are locked. The same time relationship exists for decoding successive AUs, although re‑ordering delay in the decoder affects the relationship between decoder AUs and presented PUs. When the data stream is coded such that the video or audio clock is not locked to the system clock the time difference between decoding successive AUs may be estimated using the same values as indicated above; however, these time differences are not exact due to the fact that relationships between the frame rate, audio sample rate, and system clock frequency were not exact at the encoder.

Note that the PTS and DTS fields do not, by themselves, indicate the correct fullness of the decoder buffers at start up nor at any other time, and equivalently, they do not indicate the amount of time delay that should elapse upon receiving the initial bits of a data stream before decoding should start. This information is retrieved by combining the functions of the PTS and DTS fields and correct clock recovery, which is covered below. In the STD model, and therefore in decoders which are modelled after it, the decoder buffer behaviour is determined completely by the SCR (or PCR) values, the times that they are received, and the PTS and DTS values, assuming that data is delivered in accordance with the timing model. This information specifies the time that coded data spends in the decoder buffers. The amount of data that is in the coded data buffers is not explicitly specified, and this information is not necessary, since the timing is fully specified. Note also that the fullness of the data buffers may vary considerably with time in a fashion that is not predictable by the decoder, except through the proper use of the timestamps.

In order for the audio and video PTSs to refer correctly to a common STC, a correctly timed common clock must be made available within the decoder system. This is subject of the next subclause.

### D.1.3 System time clock recovery in the decoder

Within the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 Systems data stream there are, in addition to the PTS and DTS fields, clock reference timestamps. These references are samples of the system time clock, which are applicable both to a decoder and to an encoder. They have a resolution of one part in 27 000 000 per second, and occur at intervals up to 100 ms in transport streams, or up to 700 ms in program streams. As such, they can be utilized to implement clock reconstruction control loops in decoders with sufficient accuracy for all identified applications.

In the program stream, the clock reference field is called the System Clock Reference or SCR. In the transport stream, the clock reference field is called the Program Clock Reference or PCR. In general the SCR and PCR definitions may be considered to be equivalent, although there are distinctions. The remainder of this subclause uses the term SCR for clarity; the same statements apply to the PCR except where otherwise noted. The PCR in transport streams provides the clock reference for one program, where a program is a set of elementary streams that have a common time base and are intended for synchronized decoding and presentation. There may be multiple programs in one transport stream, and each may have an independent time base and a separate set of PCRs.

The SCR field indicates the correct value of the STC when the SCR is received at the decoder. Since the SCR occupies more than one byte of data, and System data streams are defined as streams of bytes, the SCR is defined to arrive at the decoder when the last byte of the system\_clock\_reference\_base field is received at the decoder. Alternatively the SCR can be interpreted as the time that the SCR field should arrive at the decoder, assuming that the STC is already known to be correct. Which interpretation is used depends on the structure of the application system. In applications where the data source can be controlled by the decoder, such as a locally attached DSM, it is possible for the decoder to have an autonomous STC frequency, and so the STC need not be recovered. In many important applications, however, this assumption cannot be made correctly. For example, consider the case where a data stream is delivered simultaneously to multiple decoders. If each decoder has its own autonomous STC with its own independent clock frequency, the SCRs cannot be assured to arrive at the correct time at all decoders; one decoder will in general require the SCRs sooner than the source is delivering them, while another requires them later. This difference cannot be made up with a finite size data buffer over an unbounded length of time of data reception. Therefore the following addresses primarily the case where the STC must slave its timing to the received SCRs (or PCRs).

In a correctly constructed and delivered Rec. ITU-T H.222.0 | ISO/IEC 13818-1 data stream, each SCR arrives at the decoder at precisely the time indicated by the value of that SCR. In this context, "time" means correct value of the STC. In concept, this STC value is the same value that the encoder's STC had when the SCR was stored or transmitted. However, the encoding may have been performed not in real time or the data stream may have been modified since it was originally encoded, and in general the encoder or data source may be implemented in a variety of ways such that the encoder's STC may be a theoretical quantity.

If the decoder's clock frequency matches exactly that of the encoder, then the decoding and presentation of video and audio will automatically have the same rate as those at the encoder, and the end-to-end delay will be constant. With matched encoder and decoder clock frequencies, any correct SCR value can be used to set the instantaneous value of the decoder's STC, and from that time on the decoder's STC will match that of the encoder without the need for further adjustment. This condition remains true until there is a discontinuity of timing, such as the end of a program stream or the presence of a discontinuity indicator in a transport stream.

In practice a decoder's free-running system clock frequency will not match the encoder's system clock frequency which is sampled and indicated in the SCR values. The decoder's STC can be made to slave its timing to the encoder using the received SCRs. The prototypical method of slaving the decoder's clock to the received data stream is via a phase-locked loop (PLL). Variations of a basic PLL, or other methods, may be appropriate, depending on the specific application requirements.

A straight forward PLL which recovers the STC in a decoder is diagrammed and described here.

Figure D.2 shows a classic PLL, except that the reference and feedback terms are numbers (STC and SCR or PCR values) instead of signal events such as edges.



Figure D.2 – STC recovery using PLL

Upon initial acquisition of a new time base, i.e., a new program, the STC is set to the current value encoded in the SCRs. Typically the first SCR is loaded directly into the STC counter, and the PLL is subsequently operated as a closed loop. Variations on this method may be appropriate, i.e., if the values of the SCRs are suspect due to jitter or errors.

The closed-loop action of the PLL is as follows. At the moment that each SCR (or PCR) arrives at the decoder, that value is compared with the current value of the STC. The difference is a number, which has one part in units of 90 kHz and one part in terms of 300 times this frequency, i.e., 27 MHz. The difference value is linearized to be in a single number space, typically units of 27 MHz, and is called "**e**", the error term in the loop. The sequence of **e** terms is input to the low-pass filter and gain stage, which are designed according to the requirements of the application. The output of this stage is a control signal "**f**" which controls the instantaneous frequency of the Voltage Controlled Oscillator (VCO). The output of the VCO is an oscillator signal with a nominal frequency of 27-MHz; this signal is used as the system clock frequency within the decoder. The 27-MHz clock is input to a counter which produces the current STC values, which consist of both a 27-MHz extension, produced by dividing by 300, and a 90-kHz base value which is derived by counting the 90-kHz results in a 33-bit counter. The 33-bit, 90-kHz portion of the STC output is used as needed for comparison with PTS and DTS values. The complete STC is also the feedback input to the subtractor.

The bounded maximum interval between successive SCRs (700 ms) or PCRs (100 ms) allows the design and construction of PLLs which are known to be stable. The bandwidth of the PLLs has an upper bound imposed by this interval. As shown below, in many applications the PLL required has a very low bandwidth, and so this bound typically does not impose a significant limitation on the decoder design and performance.

If the free-running or initial frequency of the VCO is close enough to the correct encoder's system clock frequency, the decoder may be able to operate satisfactorily as soon as the STC is initialized correctly, before the PLL has reached a defined locked state. For a given decoder STC frequency which differs by a bounded amount from the frequency encoded in the SCRs and which is within the absolute frequency bounds required by the decoder application, the effect of the mismatch between the encoder's and the decoder's STC frequencies if there were not PLL is the gradual and unavoidable increase or decrease of the fullness of the decoder's buffers, such that overflow or underflow would occur eventually with any finite size of decoder buffers. Therefore the amount of time allowable before the decoder's STC frequency is locked to that of the encoder is determined by the allowable amount of additional decoder buffer size and delay.

If the SCRs are received by the decoder with values and timing that reflect instantaneously correct samples of a constant frequency STC in the encoder, then the error term **e** converges to an essentially constant value after the loop has reached the locked state. This condition of correct SCR values is synonymous with either constant-delay storage and transmission of the data from the encoder to the decoder, or if this delay is not constant, the effective equivalent of constant delay storage and transmission with the SCR values having been corrected to reflect the variations in delay. With the values of **e** converging to a constant, variations in the instantaneous VCO frequency become essentially zero after the loop is locked; the VCO is said to have very little jitter or frequency slew. While the loop is in the process of locking, the rate of change of the VCO frequency, the frequency slew rate, can be controlled strictly by the design of the low pass filter and gain stage. In general the VCO slew rate can be designed to meet application requirements, subject to constraints of decoder buffer size and delay.

### D.1.4 SCR and PCR jitter

If a network or a transport stream re-multiplexor varies the delay in delivering the data stream from the encoder or storage system to the decoder, such variations tend to cause a difference between the values of the SCRs (or PCRs) and the values that they should have when they are actually received. This is referred to as SCR or PCR jitter. For example, if the delay in delivering one SCR is greater than the delay experienced by other similar fields in the same program, that SCR is late. Similarly, if the delay is less than for other clock reference fields in the program, the field is early.

Timing jitter at the input to a decoder is reflected in the combination of the values of the SCRs and the times when they are received. Assuming a clock recovery structure as illustrated in Figure D.2, any such timing jitter will be reflected in the values of the error term **e**; and non-zero values of **e** induce variations in the values of **f**, resulting in variations in the frequency of the 27-MHz system clock. Variations in the frequency of the recovered clock may or may not be acceptable within decoder systems, depending on the specific application requirements. For example, in precisely timed decoders that produce composite video output, the recovered clock frequency is typically used to generate the composite video sample clock and the chroma sub-carrier; the applicable specifications for sub-carrier frequency stability may permit only very slow adjustment of the system clock frequency. In applications where a significant amount of SCR or PCR jitter is present at the decoder input and there are tight constraints on the frequency slew rate of the STC, the constraints of reasonable additional decoder buffer size and delay may not allow proper operation.

The presence of SCR or PCR jitter may be caused for example by network transmission which incorporates packet or cell multiplexing or variable delay of packets through the network, as may be caused by queuing delays or by variable network access time in shared-media systems.

Multiplexing or re-multiplexing of transport or program streams changes the order and relative temporal location of data packets and therefore also of SCRs or PCRs. The change in temporal location of SCRs causes the value of previously correct SCRs to become incorrect, since in general the time at which they are delivered via a constant delay network is not correctly represented by their values. Similarly, a program stream or transport stream with correct SCRs or PCRs may be delivered over a network which imposes a variable delay on the data stream, without correcting the SCR or PCR values. The effect is once again SCR or PCR jitter, with attendant effects on the decoder design and performance. The worst case amount of jitter which is imposed by a network on the SCRs or PCRs received at a decoder depends on a number of factors which are beyond the scope of this Recommendation | International Standard, including the depth of queues implemented in each of the network switches and the total number of network switches or re-multiplexing operations which operate in cascade on the data stream.

In the case of a transport stream, correction of PCRs is necessary in a re-multiplex operation, creating a new transport stream from one or more transport streams. This correction is accomplished by adding a correction term to the PCR; this term can be computed as:

ΔPCR = delact – delconst

where delact is the actual delay experienced by the PCR, and delconst is a constant which is used for all PCRs of that program. The value which should be used for delconst will depend on the strategy used by the original encoder/multiplexor. This strategy could be, for instance, to schedule packets as early as possible, in order to allow later transmission links to delay them. In Table D.1, three different multiplex strategies are shown together with the appropriate value for delconst.

Table D.1 – Re-multiplexing strategy

|  |  |
| --- | --- |
| Strategy | delconst |
| Early | delmin |
| Late | delmax |
| Middle | delavg |

When designing a system, private agreements may be needed as to what strategy should be used by the encoder/multiplexors, since this will have an effect on the ability to perform any additional re-multiplexing.

The amount of multiplex jitter allowed is not normatively bounded in this Recommendation | International Standard. However, 4 ms is intended to be the maximum amount of jitter in a well-behaved system.

In systems which include re-multiplexors special care might be necessary to ensure that the information in the transport stream is consistent. In particular, this applies to PSI and to discontinuity points. Changes in PSI tables might need to be inserted into a transport stream in such a way that subsequent re-multiplexor steps never move them so far that information becomes incorrect. For instance, a new version of PMT section in some cases should not be sent within 4 ms of the data affected by the change.

Similarly, it may be necessary for an encoder/mux to avoid inserting PTS or DTS in a ±4-ms window around a discontinuity point.

### D.1.5 Clock recovery in the presence of network jitter

In applications in which there is any significant amount of jitter present in the received clock reference timestamps, there are several choices available for decoder designs; how the decoder is designed depends in large part on the requirements for the decoder's output signal characteristics as well as the characteristics of the input data and jitter.

Decoders in various applications may have differing requirements for the accuracy and stability of the recovered system clock, and the degree of this stability and accuracy that is required may be considered to fall along a single axis. One extreme of this axis may be considered to be those applications where the reconstructed system clock is used directly to synthesize a chroma sub-carrier for use in composite video. This requirement generally exists where the presented video is of the precisely timed type, as described above, such that each coded picture is presented exactly once, and where the output is composite video in compliance with the applicable specifications. In that case the chroma sub-carrier, the pixel clock, and the frame rate all have exactly specified ratios, and all of these have a defined relationship to the system clock. The composite video sub-carrier must have at least sufficient accuracy and stability that any normal television receiver's chroma sub-carrier PLL can lock to the sub-carrier, and the chroma signals which are demodulated using the recovered sub-carrier do not show visible chrominance phase artefacts. The requirement in some applications is to use the system clock to generate a sub-carrier that is in full compliance with the NTSC, PAL, or SECAM specifications, which are typically even more stringent than those imposed by typical television receivers. For example, the SMPTE specification for NTSC requires a sub-carrier accuracy of 3 ppm, with a maximum short term jitter of 1 ns per horizontal line time and a maximum long term drift of 0.1 Hz per second.

In applications where the recovered system clock is not used to generate a chroma sub-carrier, it may still be used to generate a pixel clock for video and it may be used to generate a sample clock for audio. These clocks have their own stability requirements that depend on the assumptions made about the receiving display monitor and on the acceptable amount of audio frequency drift, or "wow and flutter", at the decoder's output.

In applications where each picture and each audio sample are not presented exactly once, i.e., picture and audio sample "slipping" is allowed, the system clock may have relatively loose accuracy and stability requirements. This type of decoder may not have precise audio-video presentation synchronization, and the resulting audio and video presentation may not have the same quality as for precisely timed decoders.

The choice of requirements for the accuracy and stability of the recovered system clock is application dependent. The following focuses on the most stringent requirement which is identified above, i.e., where the system clock is to be used to generate a chroma sub-carrier.

### D.1.6 System clock used for chroma sub-carrier generation

The decoder design requirements can be determined from the requirements on the resulting sub-carrier and the maximum amount of network jitter that must be accepted. Similarly, if the system clock performance requirements and the decoder design's capabilities are known, the tolerable maximum network jitter can be determined. While it is beyond the scope of this Recommendation | International Standard to state such requirements, the numbers which are needed to specify the design are identified in order to clarify the statement of the problem and to illustrate a representative design approach.

With a clock recovery PLL circuit as illustrated in Figure D.2, the recovered system clock must meet the requirements of a worst case frequency deviation from the nominal, measured in units of ppm (parts per million), and a worst case frequency slew rate, measured in ppm/s (ppm per second). The peak-to-peak uncorrected network timing jitter has a value that may be specified in milliseconds. In such a PLL the network timing jitter appears as the error term **e** in the diagram, and since the PLL acts as a low-pass filter on jitter at its input, the worst case effect on the 27-MHz output frequency occurs when there is a maximum amplitude step function of PCR timing at the input. The value **e** then has a maximum amplitude equal to the peak-to-peak jitter, which is represented numerically as the jitter times 2\*\*33 in the base portion of the SCR or PCR encoding. The maximum rate of change of the output of the low pass filter (LPF), **f**, with this maximum value of **e** at its input, directly determines the maximum frequency slew rate of the 27-MHz output. For any given maximum value of **e** and maximum rate of change of **f** a LPF can be specified. However, as the gain or cut-off frequency of the LPF is reduced, the time required for the PLL to lock to the frequency represented by the SCRs or PCRs is increased. Implementation of PLLs with very long time constants can be achieved through the use of digital LPF techniques, and possibly analogue filter techniques. With digital LPF implementations, when the frequency term **f** is the input to an analogue VCO, **f** is quantized by a digital to analogue converter, whose step size should be considered when calculating the maximum slew rate of the output frequency.

In order to ensure that **e** converges to a value that approaches zero, the open loop gain of the PLL must be very high, such as might be implemented in an integrator function in the low-pass filter in the PLL.

With a given accuracy requirement, it may be reasonable to construct the PLL such that the initial operating frequency of the PLL meets the accuracy requirement. In this case the initial 27-MHz frequency before the PLL is locked is sufficiently accurate to meet the stated output frequency requirement. If it were not for the fact that the decoder's buffers would eventually overflow or underflow, this initial system clock frequency would be sufficient for long term operation. However, from the time the decoder begins to receive and decode data until the system clock is locked to the time and clock frequency that is represented by the received SCRs or PCRs, data is arriving at the buffers at a different rate than it is being extracted, or equivalently the decoder is extracting access units at times that differ from those of the System Target Decoder (STD) model. The decoder buffers will continue to become more or less full than those of the STD according to the trajectory of recovered system clock frequency with respect to the encoder's clock frequency. Depending on the relative initial VCO frequency and encoder system clock frequency, decoder buffer fullness is either increasing or decreasing. Assuming this relationship is not known, the decoder needs additional data buffering to allow for either case. The decoder should be constructed to delay all decoding operations by an amount of time that is at least equal to the amount of time that is represented by the additional buffering that is allocated for the case of the initial VCO frequency being greater than the encoder's clock frequency, in order to prevent buffer underflow. If the initial VCO frequency is not sufficiently accurate to meet the stated accuracy requirements, then the PLL must reach the locked state before decoding may begin, and there is a different set of considerations regarding the PLL behaviour during this time and the amount of additional buffering and static delay which is appropriate.

A step function in the input timing jitter which produces a step function in the error term **e** of the PLL in Figure D.2 must produce an output frequency term **f** such that when it is multiplied by the VCO gain the maximum rate of change is less than the specified frequency slew rate. The gain of the VCO is stated in terms of the amount of the change in output frequency with respect to a change in control input. An additional constraint on the LPF in the PLL is that the static value of **e** when the loop is locked must be bounded in order to bound the amount of additional buffering and static decoding delay that must be implemented. This term is minimized when the LPF has very high DC gain.

Clock recovery circuits which differ somewhat from that shown in Figure D.2 may be practical. For example, it may be possible to implement a control loop with a Numerically Controlled Oscillator (NCO) instead of a VCO, wherein the NCO uses a fixed frequency oscillator and clock cycles are inserted or deleted from normally periodic events at the output in order to adjust the decoding and presentation timing. There may be some difficulties with this type of approach when used with composite video, as there is a tendency to cause either problematic phase shifts of the sub‑carrier or jitter in the horizontal or vertical scan timing. One possible approach is to adjust the period of horizontal scans at the start of vertical blanking, while maintaining the phase of the chroma sub-carrier.

In summary, depending on the values specified for the requirements, it may or may not be practical to construct a decoder which reconstructs the system clock with sufficient accuracy and stability, while maintaining desired decoder buffer sizes and added decoding delay.

### D.1.7 Component video and audio reconstruction

If component video is produced at the decoder output, the requirements for timing accuracy and stability are generally less stringent than is the case for composite video. Typically the frequency tolerance is that which the display deflection circuitry can accept, and the stability tolerance is determined by the need to avoid visible image displacement on the display.

The same principles as illustrated above apply; however, the specific requirements are generally easier to meet.

Audio sample rate reconstruction again follows the same principles; however, the stability requirement is determined by the amount of acceptable long and short term sample rate variation. Using a PLL approach as illustrated in the previous subclause, short term deviation can be made to be very small, and longer term frequency variation is manifested as variation in perceived pitch. Again, once specified bounds on this variation are set specific design requirements can be determined.

### D.1.8 Frame slipping

In some applications where precise decoder timing is not required, the decoder's system time clock may not adjust its operating frequency to match the frequency represented by received SCRs (or PCRs); it may have a free-running 27 MHz clock instead, while still slaving the decoder's STC to the received data. In this case the STC value must be updated as needed to match the received SCRs. Updating the STC upon receipt of SCRs causes discontinuities in the STC value. The magnitude of these discontinuities depends upon the difference between the decoder's 27-MHz frequency and the encoder's 27-MHz, i.e., that which is represented by the received SCRs, and upon the time interval between successive received SCRs or PCRs. Since the decoder's 27-MHz system clock frequency is not locked to that of the received data, it cannot be used to generate the video or audio sample clocks while maintaining the precise timing assumptions of presenting each video and audio presentation unit exactly once and of maintaining the same picture and audio presentation rate at the decoder and the encoder, with precise audio and video synchronization. There are multiple possibilities for implementing decoding and presentation systems using this structure.

In one type of implementation, the pictures and audio samples are decoded at the time indicated by the decoder's STC, while they are presented at slightly different times, according to the locally produced sample clocks. Depending on the relationships of the decoder's sample clocks to the encoder's system clock, pictures and audio samples may on occasion be presented more than one each or not at all; this is referred to as "frame slipping" or "sample slipping", in the case of audio. There may be perceptible artefacts introduced by this mechanism. The audio-video synchronization will in general not be precise, due to the units of time over which pictures, and perhaps audio presentation units, are repeated or deleted. Depending on the specific implementation, additional buffering in the decoder is generally needed for coded data or decoded presentation data. Decoding may be performed immediately before presentation, and not quite at the time indicated in the decoder's STC, or decoded presentation units may be stored for delayed and possibly repeated presentation. If decoding is performed at the time of presentation, a mechanism is required to support deleting the presentation of pictures and audio samples without causing problems in the decoding of predictively coded data.

### D.1.9 Smoothing of network jitter

In some applications it may be possible to introduce a mechanism between a network and a decoder in order to reduce the degree of jitter which is introduced by a network. Whether such an approach is feasible depends on the type of streams received and the amount and type of jitter which is expected.

Both the transport stream and the program stream indicate within their syntax the rate at which the stream is intended to be input to a decoder. These indicated rates are not precise, and cannot be used to reconstruct data stream timing exactly. They may, however, be useful as part of a smoothing mechanism.

For example, a transport stream may be received from a network such that the data is delivered in bursts. It is possible to buffer the received data and to transmit data from the buffer to the decoder at an approximately constant rate such that the buffer remains approximately one-half full.

However, a variable rate stream should not be delivered at constant rate, and with variable rate streams the smoothing buffer should not always be one-half full. A constant average delay through the buffer requires a buffer fullness that varies with the data rate. The rate that data should be extracted from the buffer and input to the decoder can be approximated using the rate information present in the data stream. In transport streams the intended rate is determined by the values of the PCR fields and the number of transport stream bytes between them. In program streams the intended rate is explicitly specified as the Program\_mux\_rate, although as specified in this Recommendation | International Standard the rate may drop to zero at SCR locations, i.e., if the SCR arrives before the time expected when the data is delivered at the indicated rate.

In the case of variable rate streams, the correct fullness of the smoothing buffer varies with time, and may not be determined exactly from the rate information. In an alternative approach, the SCRs or PCRs may be used to measure the time when data enter the buffer and to control the time when data leave the buffer. A control loop can be designed to provide constant average delay through the buffer. It may be observed that such a design is similar to the control loop illustrated in Figure D.2. The performance obtainable from inserting such a smoothing mechanism before a decoder can also be achieved by cascading multiple clock recovery PLLs. The rejection of jitter from the received timing will benefit from the combined low pass filter effect of the cascaded PLLs.

Annex E  
  
Data transmission applications

(This annex does not form an integral part of this Recommendation | International Standard.)

## E.1 General considerations

• Rec. ITU-T H.222.0 | ISO/IEC 13818-1 transport multiplex will be used to transmit data as well as video and audio.

• Data elementary streams are not continuous as may appear video and audio streams in broadcast applications.

• While it is already possible to identify the beginning of a PES packet, it is not always possible to identify the end of a PES packet by the beginning of the next PES packet, because it is possible for one or more Transport packet carrying PES packets to be lost.

## E.2 Suggestion

A suitable solution is to transmit the following PES packet just after an associated PES packet. A PES packet without payload may be sent when there are no further PES packets to send.

Table E.1 is an example of such a PES packet.

Table E.1 – PES packet header example

|  |  |
| --- | --- |
| PES packet header fields | Values |
| packet\_start\_code\_prefix | 0x000001 |
| stream\_id | assigned |
| PES\_packet\_length | 0x0003 |
| '10' | '10' |
| PES\_scrambling\_control | '00' |
| PES\_priority | '0' |
| data\_alignment\_indicator | '0' |
| copyright | '0' |
| original\_or\_copy | '0' |
| PTS\_DTS\_flags | '00' |
| ESCR\_flag | '0' |
| ES\_rate\_flag | '0' |
| DSM\_trick\_mode\_flag | '0' |
| additional\_copy\_info\_flag | '0' |
| PES\_CRC\_flag | '0' |
| PES\_extension\_flag | '0' |
| PES\_header\_data\_length | 0x00 |

Annex F  
  
Graphics of syntax for this Recommendation | International Standard

(This annex does not form an integral part of this Recommendation | International Standard.)

## F.1 Introduction

This annex is an informative annex presenting graphically the transport stream and program stream syntax. This annex in no way replaces any normative clause(s).

In order to produce clear drawings, not all fields have been fully described or represented. Reserved fields may be omitted or indicated by areas with no detail. Fields lengths are indicated in bits.

### F.1.1 Transport stream syntax

See Figure F.1.



Figure F.1 – Transport stream syntax diagram

### F.1.2 PES packet

See Figure F.2.



Figure F.2 – PES packet syntax diagram

### F.1.3 Program Association section

See Figure F.3.



Figure F.3 – Program association section diagram

### F.1.4 CA section

See Figure F.4.



Figure F.4 – Conditional access section diagram

### F.1.5 TS program map section

See Figure F.5.



Figure F.5 – TS program map section diagram

### F.1.6 Private section

See Figure F.6.



Figure F.6 – Private section diagram

### F.1.7 Program stream

See Figure F.7.



Figure F.7 – Program stream diagram

### F.1.8 Program stream map

See Figure F.8.



Figure F.8 – Program stream map diagram

Annex G  
  
General information

(This annex does not form an integral part of this Recommendation | International Standard.)

## G.1 General information

### G.1.1 Sync byte emulation

In the choice of PID values it is recommended that the periodic emulation of sync bytes be avoided. Such emulation may potentially occur within the PID field or as a combination of the PID field and adjacent flag settings. It is recommended that emulation of the sync byte be permitted to occur in the same position of the packet header for a maximum of 4‑consecutive transport packets.

### G.1.2 Skipped picture status and decoding process

Assume that the sequence being displayed contains only I- and P-frames. Denote the next picture to be decoded by picture\_next, and the picture currently being displayed by picture\_current. Because of the fact that the video encoder may skip pictures, it is possible that not all of the bits of picture\_next are present in the STD buffers EBn or Bn when the time arrives to remove those bits for instantaneous decoding and display. When this case arises, no bits are removed from the buffer and picture\_current is displayed again. When the next picture display time arrives, if the remainder of the bits corresponding to picture\_next are now in buffer EBn or Bn, all the bits of picture\_next are removed and picture\_next is displayed. If all the bits of picture\_next are not in the buffer EBn or Bn, the above process of re‑displaying picture\_current is repeated. This process is repeated until picture\_next can be displayed. Note that if a PTS preceded picture\_next in the bitstream, it will be incorrect by some multiple of the picture display interval, which itself may depend on some parameters, and must be ignored.

Whenever the skipped picture situation described above occurs, the encoder is required to insert a PTS before the picture to be decoded after picture\_next. This allows the decoder to immediately verify that it has correctly displayed the received picture sequence.

### G.1.3 Selection of PID values

Applications are encouraged to use low numbered PID values (avoiding reserved values as specified in Table 2-4) and group values together as much as possible.

### G.1.4 PES start\_code emulation

Three consecutive bytes having the value of a packet\_start\_code\_prefix (0x000001), which when concatenated with a fourth byte, may emulate the four bytes of a PES\_packet\_header at a unintended place in the stream.

Such, so called, start code emulation is not possible in video elementary streams other than EVC video streams. It is possible in audio and data elementary streams. It is also possible at the boundary of a PES\_packet\_header and a PES\_packet payload, even if the PES\_packet payload is video.

Annex H  
  
Private data

(This annex does not form an integral part of this Recommendation | International Standard.)

## H.1 Private data

Private data is any user data which is not coded according to a standard specified by ITU-T | ISO/IEC and referred to in this Specification. The contents of this data is not and shall not be specified within this Recommendation | International Standard in the future. The STD defined in this Specification does not cover private data other than the demultiplex process. A private party may define each STD for private streams.

Private data may be carried in the following locations within the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 syntax.

1) *Transport stream packet Table 2-2*

The data bytes of the transport\_packet() syntax may contain private data. Private data carried in this format is referred to as user private within the stream\_type Table 2-34. It is permitted for transport stream packets containing private data to also include adaptation\_field()s.

2) *Transport stream adaptation field Table 2-6*

The presence of any optional private\_data\_bytes in the adaptation\_field() is signalled by the transport\_private\_data\_flag. The number of the private\_data\_bytes is inherently restricted by the semantic of the adaptation\_field\_length field, where the value of the adaptation\_field\_length shall not exceed 183 bytes.

3) *PES packet Table 2-21*

There are two possibilities for carrying private data within PES packets. The first possibility is within the PES\_packet\_header, within the optional 16 bytes of PES\_private\_data. The presence of this field is signalled by the PES\_private\_data\_flag. The presence of the PES\_private\_data\_flag is signalled by the PES\_extension\_flag. If present, these bytes, when considered with the adjacent fields, shall not emulate the packet\_start\_code\_prefix.

The second possibility is within the PES\_packet\_data\_byte field. This may be referred to as private data within PES packets under the stream\_type Table 2-34. This category of private data can be split in two: private\_stream\_1 refers to private data within PES packets which follow the PES\_packet() syntax such that all fields up to and including, but not limited to, PES\_header\_data\_length are present. private\_stream\_2 refers to private data within PES packets where only the first three fields shall be present followed by the PES\_packet\_data\_bytes containing private data.

Note that PES packets exist within both program streams and transport streams therefore private\_stream\_1 and private\_stream\_2 exist within both program streams and transport streams.

4) *Descriptors*

Descriptors exist within program streams and transport streams. A range of private descriptors may be defined by the user. These descriptors shall commence with descriptor\_tag and descriptor\_length fields. For private descriptors, the value of descriptor\_tag may take the values 64-255 as identified in Table 2-45. These descriptors may be placed within a program\_stream\_map() Table 2-34, a CA\_section() Table 2‑32, a TS\_program\_map\_section(), Table 2-33 and in any private\_section(), Table 2-35.

Specifically private\_data\_bytes also appear in the CA\_descriptor().

5) *Private Section*

The private\_section Table 2-35 provides a further means to carry private data also in two forms. This type of elementary stream may be identified under stream\_type Table 2-34 as private\_data in PSI sections. One type of private\_section() includes only the first five defined fields, and is followed by private data. For this structure the section\_syntax\_indicator shall be set to a value of '0'. For the other type, the section\_syntax\_indicator shall be set to a value of '1' and the full syntax up to and including last\_section\_number shall be present, followed by private\_data\_bytes and ending with the CRC\_32.

Annex I  
  
Systems conformance and real-time interface

(This annex does not form an integral part of this Recommendation | International Standard.)

## I.1 Systems conformance and real-time interface

Conformance for Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program streams and transport streams is specified in terms of the normative specifications in this Recommendation | International Standard. These specifications include, among other requirements, a System Target Decoder (T-STD and P-STD) which specifies the behaviour of an idealized decoder when the stream is the input to such a decoder. This model, and the associated verification, do not include information concerning the real-time delivery performance of the stream, except for the accuracy of the system clock frequency which is represented by the transport stream and the program stream. All transport streams and program streams must comply with this Recommendation | International Standard.

In addition, there is a real-time interface specification for input of transport streams and program streams to a decoder. This Recommendation | International Standard allows standardization of the interface between MPEG decoders and adapters to networks, channels, or storage media. The timing effects of channels, and the inability of practical adapters to eliminate completely these effects, causes deviations from the idealized byte delivery schedule to occur. While it is not necessary for all MPEG decoders to implement this interface, implementations which include the interface shall adhere to the specifications. This Recommendation | International Standard covers the real-time delivery behaviour of transport streams and Program streams to decoders, such that the coded data buffers in decoders are guaranteed not to overflow nor underflow, and decoders are guaranteed to be able to perform clock recovery with the performance required by their applications.

The MPEG real-time interface specifies the maximum allowable amount of deviation from the idealized byte delivery schedule which is indicated by the Program Clock Reference (PCR) and System Clock Reference (SCR) fields encoded in the stream.

Annex J  
  
Interfacing jitter-inducing networks to MPEG-2 decoders

(This annex does not form an integral part of this Recommendation | International Standard.)

## J.1 Introduction

In this annex the expression "system stream" will be used to refer to both Rec. ITU-T H.222.0 | ISO/IEC 13818-1 transport streams and Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program streams. When the term "STD" is used, it is understood to mean the P-STD (Program System Target Decoder) for program streams and the T-STD (Transport System Target Decoder) for transport streams.

The intended byte delivery schedule of a system stream can be deduced by analysing the stream. A system stream is compliant if it can be decoded by the STD, which is a mathematical model of an idealized decoder. If a compliant system stream is transmitted over a jitter-inducing network, the true byte delivery schedule may differ significantly from the intended byte delivery schedule. In such cases it may not be possible to decode the system stream on such an idealized decoder, because jitter may cause buffer overflows or underflows and may make it difficult to recover the time base. An important example of such a jitter-inducing network is ATM.

The purpose of this annex is to provide guidance and insight to entities concerned with sending system streams over jitter-inducing networks. Network-specific compliance models for transporting system streams are likely to be developed for several types of networks, including ATM. The STD plus a real-time interface definition can play an integral role in defining such models. A framework for developing network compliance models is presented in J.2.

Three examples of network encoding to enable the building of jitter-smoothing network adapters are discussed in J.3. In the first example, a constant bitrate system stream is assumed and a FIFO is used for jitter smoothing. In the second example, the network adaptation layer includes timestamps to facilitate jitter smoothing. In the final example, a common network clock is assumed to be available end-to-end, and is exploited to achieve jitter smoothing.

Clause J.4 presents two examples of decoder implementations in which network-induced jitter can be accommodated. In the first example, a jitter-smoothing network adapter is inserted between a network's output and an MPEG-2 decoder. The MPEG-2 decoder is assumed to conform to a real-time MPEG-2 interface specification. This interface requires an MPEG-2 decoder with more jitter tolerance than the idealized decoder of the STD. The network adapter processes the incoming jittered bitstream and outputs a system stream whose true byte delivery schedule conforms to the real-time specification. Example one is discussed in J.4.1. For some applications the network adapter approach will be too costly because it requires two stages of processing. Therefore, in the second example the dejittering and MPEG-2 decoding functions are integrated. The intermediate processing of the jitter-removal device is bypassed, so only a single stage of clock recovery is required. Decoders that perform integrated dejittering and decoding are referred to in this annex as integrated network-specific decoders, or simply integrated decoders. Integrated decoders are discussed in J.4.2.

In order to build either network adapters or integrated decoders a maximum value for the peak-to-peak network jitter must be assumed. In order to promote interoperability, a peak-to-peak jitter bound must be specified for each relevant network type.

## J.2 Network compliance models

One way to model the transmission of a system stream across a jitter-inducing network is shown in Figure J.1.

The system stream is input to a network-specific encoding device that converts the system stream into a network‑specific format. Information to assist in jitter removal at the network output may be part of this format. The network decoder comprises a network-specific decoder and a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 decoder. The Rec. ITU-T H.222.0 | ISO/IEC 13818-1 decoder is assumed to conform to a real-time interface specification, and could have the same architecture as the STD with appropriate buffers made larger to provide more jitter tolerance. The network-specific decoder removes the non-Rec. ITU-T H.222.0 | ISO/IEC 13818-1 data added by the network-specific encoder and dejitters the network's output. The output of the network-specific decoder is a system stream that conforms to the real-time specification.

A network target decoder (NTD) can be defined based on the above architecture. A compliant network bitstream would be one that was able to be decoded by the NTD. A network decoder would be compliant provided it could decode any network bitstream able to be decoded by the NTD. A real network decoder might or might not have the architecture of the NTD.



Figure J.1 – Sending system streams over a jitter-inducing network

## J.3 Network specification for jitter smoothing

In the case of constant bit rate system streams, jitter smoothing can be accomplished with a FIFO. Additional data that provides specific support for dejittering is not required in the network adaptation layer. After the bytes added by the network encoding are removed, the system stream data is placed in a FIFO. A PLL keeps the buffer approximately half full by adjusting the output rate in response to changes in buffer fullness. In this example the amount of jitter-smoothing achieved will depend on the size of the FIFO and the characteristics of the PLL.

Figure J.2 illustrates a second way to accomplish jitter smoothing. In this example timestamp support from a network adaptation layer is assumed. Using this technique, both constant bit rate and variable bit rate system streams can be dejittered.



Figure J.2 – Jitter-smoothing using network-layer timestamps

Assume the network adapter is designed to compensate for a peak-to-peak jitter of J seconds. The intended byte delivery schedule is reconstructed using Clock Reference (CRs) samples taken from a Time Clock (TC). The CRs and the TC are analogous to PCRs and the STC. The Network Data Packet (NDP) encode converts each system stream packet into a Network Data Packet (NDP). The network data packets contain a field for carrying CR values, and the current value of the TC is inserted into this field as the NDP leaves the NDP encoder. The Network Transport Packetization (NXP) function encapsulates the NDPs into network transport packets. After transmission across the network, the CRs are extracted by the NDP decoder as the NDPs enter the NDP decoder. The CRs are used to reconstruct the TC, for example by using a PLL. The first MPEG-2 packet is removed from the dejittering buffer when the delayed TC (TCd) is equal to the first MPEG-2 packet's CR. Subsequent MPEG-2 packets are removed when their CR values equal the value of the TCd.

Ignoring implementation details such as the speed of the TC clock recovery loop and the spectral purity of the TC, the size of the dejittering buffer depends only on the maximum peak-to-peak jitter to be smoothed and the largest transport rate that occurs in the system stream. The dejittering buffer size, Bdj, is given by

Bdj = JRmax

where Rmax is the maximum data rate of the system stream in bits per second. When packets traversing the network experience the nominal delay, the buffer is half full. When they experience a delay of J/2 seconds, the buffer is empty, and when they experience a delay (advance) of –J/2 seconds the buffer is full.

As a final example, in some cases a common network clock will be available end-to-end, and it may be feasible to lock the system clock frequency to the common clock. The network adapter can smooth jitter with a FIFO. The adapter uses PCRs or SCRs to reconstruct the original byte delivery schedule.

## J.4 Example decoder implementations

### J.4.1 Network adapter followed by an MPEG-2 decoder

In this implementation a network adapter conforming to the network compliance specification is connected to an MPEG‑2 decoder conforming to the real-time interface specification.

### J.4.2 Integrated decoder

The example presented in J.4.1 requires two stages of processing. The first stage is necessary to dejitter the network's output. The second stage, recovering the STC by processing PCRs or SCRs, is required for STD decoding. The example presented in this subclause is a decoder that integrates the dejittering and decoding functions in a single system. The STC clock is recovered directly using the jittered PCR or SCR values. For presenting this example, an MPEG-2 transport stream will be assumed.

Figure J.3 illustrates the operation of the integrated decoder. The stream of network packets input to the decoder is assumed to be the same as the one shown in Figure J.2.



Figure J.3 – Integrated dejittering and MPEG-2 decoding

The incoming network packets are reassembled into MPEG-2 transport stream data by the NXP and NDP decode functions. The jittered Rec. ITU-T H.222.0 | ISO/IEC 13818-1 transport stream packets are then filtered to extract packets with the desired PID. For the case illustrated, the PID being decoded is also carrying the PCRs. The PCR values are sent to a PLL to recover the STC. Entire packets for the selected PID are placed in the integrated buffer. A positive value of J/2 s is subtracted from the STC to obtain the delayed STC, STCd. Again, J is the peak-to-peak jitter the network-savvy decoder can accommodate. The delay is introduced to guarantee that all the data required for an access unit has arrived in the buffer when the PTS/DTS of the access unit equals the current value of the STCd.

Ignoring implementation details such as the speed of the STC clock recovery loop and the spectral purity of the STC:

*Bsize* = *Bdec* + *Bmux* + *BOH* + 512 + *Bj*  
 = *Bn* + 512 + *Bj*

where *Bj* = *Rmax* *J* and *Rmax* is the maximum rate at which data is input to the PID filter. Depending on the implementation, the integrated memory could be broken into two components as in the transport STD.

Annex K  
  
Splicing transport streams

(This annex does not form an integral part of this Recommendation | International Standard.)

## K.1 Introduction

For the purposes of this annex, the term 'splicing' refers to the concatenation performed on the Transport level of two different elementary streams, the resulting transport stream conforming totally to this Recommendation | International Standard. The two elementary streams may have been generated at different locations and/or at different times, and were not necessarily intended to be spliced together when they were generated. In the following we will call the 'old' stream a continuous elementary stream (video or audio), which has been superseded by another stream (the 'new' one) from a certain point on. This point is called the splice. It is the boundary between data belonging to the 'old' stream and data belonging to the 'new' one.

A splice can be seamless or non-seamless:

• A seamless splice is a splice inducing no decoding discontinuity (refer to 2.7.6). This means that the decoding time of the first access unit of the 'new' stream is consistent with respect to the decoding time of the access unit of the 'old' stream preceding the splice, i.e., it is equal to the one that the next access unit would have had if the 'old' stream had continued. In the following, we will call this decoding time the 'seamless decoding time'.

• A non-seamless splice is a splice which results in a decoding discontinuity, i.e., the decoding time of the first access unit of the 'new' stream is greater than the seamless decoding time.

NOTE – A decoding time lower than the seamless decoding time is forbidden.

Splicing is allowed to be performed at any transport stream packet boundary, since the resulting stream is legal. But in a general case, if nothing is known about the location of PES packet starts and access unit starts, this constraint imposes that not only the Transport layer is parsed, but also the PES layer and the Elementary Stream layer, and may in some cases, make some processing on the payload of transport stream packets necessary. If such complex operations are wished to be avoided, splicing should be performed at locations where the transport stream has favourable properties, these properties being indicated by the presence of a splicing point.

The presence of a splicing point is indicated by the splice\_flag and splice\_countdown fields (refer to 2.4.3.4 for the semantics of these fields). In the following, the transport stream packet in which the splice\_countdown field value reaches zero will be called 'splicing packet'. The splicing point is located immediately after the last byte of the splicing packet.

## K.2 The different types of splicing point

A splicing point can be either an ordinary splicing point or a seamless splicing point.

### K.2.1 Ordinary splicing points

If the seamless\_splice\_flag field is not present, or if its value is zero, the splicing point is ordinary. The presence of an ordinary splicing point only signals alignment properties of the Elementary Stream: the splicing packet ends on the last byte of an Access Unit, and the payload of the next transport stream packet of the same PID will start with the header of a PES packet, the payload of which will start with an Elementary Stream Access Point (or with a sequence\_end\_code() immediately followed by an Elementary Stream Access Point, in the case of video). These properties allow 'Cut and Paste' operations to be performed easily on the Transport level, while respecting syntactical constraints and ensuring bit stream consistency. However, it does not provide any information concerning timing or buffer properties. As a consequence, with such splicing points, seamless splicing can only be done with the help of private arrangements, or by analysing the payload of the transport stream packets and tracking buffer status and timestamp values.

### K.2.2 Seamless splicing points

If the seamless\_splice\_flag field is present and its value is one, information is given by the splicing point, indicating some properties of the 'old' stream. This information is not aimed at decoders. Its primary goal is to facilitate seamless splicing. Such a splicing point is called a seamless splicing point. The available information is:

• The seamless decoding time, which is encoded as a DTS value in the DTS\_next\_AU field. This DTS value is expressed in the time base which is valid in the splicing packet.

• In the case of a video elementary stream, the constraints that have been applied to the 'old' stream when it was generated, aiming at facilitating seamless splicing. These conditions are given by the value of the splice\_type field, in the table corresponding to the profile and level of the video stream.

Note that a seamless splicing point can be used as an ordinary splicing point, by discarding this additional information. This information may also be used if judged helpful to perform non-seamless splicing, or for purposes other than splicing.

## K.3 Decoder behaviour on splices

### K.3.1 On non-seamless splices

As described above, a non-seamless splice is a splice which results in a decoding discontinuity.

It shall be noted that with such a splice, the constraints related to the decoding discontinuity (see 2.7.6) shall be fulfilled. In particular:

• a PTS shall be encoded for the first access unit of the 'new' stream (except during trick mode operation or when low\_delay = '1');

• the decoding time derived from this PTS (or from the associated DTS) shall not be earlier than the seamless decoding time;

• in the case of a video elementary stream, if the splicing packet does not end on a sequence\_end\_code(), the 'new' stream shall begin with a sequence\_end\_code() immediately followed by a sequence\_header().

In theory, since they introduce decoding discontinuities, such splices result in a non-continuous presentation of presentation units (i.e., a variable length dead time between the display of two consecutive pictures, or between two consecutive audio frames). In practice, the result will depend on how the decoder is implemented, especially in video. With some video decoders, the freezing of one or more pictures may be the preferred solution. See Part 4 of ISO/IEC 13818.

### K.3.2 On seamless splices

The aim of having no decoding discontinuity is to allow having no presentation discontinuity. In the case of audio, this can always be ensured. But it has to be noted that in the case of video, presentation continuity is in theory not possible in cases 1) and 2) below:

1) The 'old' stream ends on the end of a low-delay sequence, and the 'new' stream begins with the start of a non-low-delay sequence.

2) The 'new' stream ends on the end of a non-low-delay sequence, and the 'new' stream begins with the start of a low-delay sequence.

The effects induced by such situations is implementation dependent. For instance, in case 1, a picture may have to be presented during two frame periods, and in case 2, a picture may have to be skipped. However, it is technically possible that some implementations support such situations without any undesirable effect.

In addition, referring to 6.1.1.6 of Rec. ITU-T H.262 | ISO/IEC 13818-2, a sequence\_end\_code() shall be present before the first sequence\_header() of the 'new' stream, if at least one sequence parameter (i.e., a parameter defined in the sequence header or in a sequence header extension) has a different value in both streams, with the only exception of those defining the quantization matrix. As an example, if the bit rate field has not the same value in the 'new' stream as in the 'old' one, a sequence\_end\_code() shall be present. Thus, if the splicing packet does not end on a sequence\_end\_code, the 'new' stream shall begin with a sequence\_end\_code followed by a sequence\_header.

According to the previous paragraph, a sequence\_end\_code will be mandatory in most splices, even seamless ones. It has to be noted that Rec. ITU-T H.262 | ISO/IEC 13818-2 specifies the decoding process of video sequences (i.e., data comprised between a sequence\_header() and a sequence\_end\_code()), and nothing is specified about how to handle a sequence change. Thus, for the behaviour of the decoders when such splices are encountered, refer to Part 4 of ISO/IEC 13818.

### K.3.3 Buffer overflow

Even if both elementary streams obey the T-STD model before being spliced, it is not necessarily ensured that the STD buffers do not overflow with the spliced stream in the time interval during which bits of both streams are in these buffers.

In the case of constant bit rate video, if no particular conditions have been applied to the 'old' stream, and if no particular precautions have been taken during splicing, this overflow is possible in the case where the video bit rate of the 'new' stream is greater than the video bit rate of the 'old' one. Indeed, it is certainly true that the buffers MBn and EBn of the T‑STD do not overflow if bits are delivered to the T-STD at the 'old' rate. But if the delivery rate is switched to a higher value at the input of TBn before 'old' bits are completely removed from the T-STD, the fullness of the STD buffers will become higher than if the 'old' stream had continued without splicing, and may cause overflow of EBn and/or MBn. In the case of variable bit rate video, the same problem can occur if the delivery rate of the 'new' stream is higher than the one for which provision was made during the creation of the 'old' stream. Such a situation is forbidden.

However, it is possible for the encoder generating the 'old' stream to add conditions in the VBV buffer management in the neighbourhood of splicing points, so that provision is made for any 'new' video bit rate lower than a chosen value. For instance, in the case of a seamless splicing point, such additional conditions can be indicated by a 'splice\_type' value to which entries correspond in Table 2-7 through Table 2-20 for 'splice\_decoding\_delay' and 'max\_splice\_rate'. In that case, if the video bit rate of the 'new' stream is lower than 'max\_splice\_rate', it is ensured that the spliced stream will not lead to overflow during the time interval during which bits of both streams are in the T-STD buffer.

In the case where no such constraints have been applied, this problem can be avoided by introducing a dead time in the delivery of bits between the 'old' stream and the 'new' one, in order to let the T-STD buffers get sufficiently empty before the bits of the 'new' stream are delivered. If we call tin the time at which the last byte of the last access unit of the 'old' stream enters the STD, and tout the time at which it exits the STD, it is sufficient to ensure that no more bits enter the T‑TD the time interval [tin, tout] with the spliced stream than if the 'old' stream had continued without splicing. As an example, in the case where the 'old' stream has a constant bit rate Rold, and the 'new' one a constant bit rate Rnew, it is sufficient to introduce a dead time Td satisfying the following relations to avoid this risk of overflow:

Td ≥ 0 and Td ≥ (tout – tin) × (1 – Rold/Rnew)

Annex L  
  
Registration procedure (see 2.9)

(This annex does not form an integral part of this Recommendation | International Standard.)

## L.1 Procedure for the request of a Registered Identifier (RID)

Requesters of a RID shall apply to the Registration Authority. Registration forms shall be available from the Registration Authority. Information which the requester shall provide is given in L.3. Companies and organizations are eligible to apply.

## L.2 Responsibilities of the Registration Authority

The primary responsibilities of the Registration Authority administrating the registration of copyright\_identifiers is outlined in this subclause; certain other responsibilities may be found in the JTC 1 Directives. The Registration Authority shall:

a) implement a registration procedure for application for a unique RID in accordance with Annex H/JTC 1 Directives;

b) receive and process the applications for allocation of the work type code identifier from Copyright Registration Authority;

c) ascertain which applications received are in accordance with this registration procedure, and to inform the requester within 30 days of receipt of the application of their assigned RID;

d) inform application providers whose request is denied in writing within 30 days of receipt of the application, and also inform the requesting party of the appeals process;

e) maintain an accurate register of the allocated RID. Revisions to the contact information and technical specifications shall be accepted and maintained by the Registration Authority;

f) make the contents of this register available upon request to any interested party;

g) maintain a database of RID request forms, granted and denied. Parties seeking technical information on the format of private data which has a copyright\_identifier shall have access to such information which is part of the database maintained by the Registration Authority;

h) report its activities to JTC 1, the ITTF and the JTC 1/SC 29 Secretariat, or their respective assignees, annually on a schedule mutually agreed upon.

### L.2.1 Contact information of the Registration Authority

The contact information for the Registration Authority concerning provisions of Rec. ITU-T H.222.0 | ISO/IEC 13818-1 can be found at <https://www.iso.org/maintenance_agencies.html> or <https://handle.itu.int/11.1002/plink/0724859361>.

## L.3 Responsibilities of parties requesting an RID

The party requesting an RID for the purpose of copyright identification shall:

a) apply using the form and procedures supplied by the Registration Authority;

b) provide contact information describing how a complete description of the copyright organization can be obtained on a non-discriminatory basis;

c) include technical details of the syntax and semantics of the data format used to describe the audiovisual works or other copyrighted works within the **additional\_copyright\_info** field. Once registered, the syntax used for the additional copyright information shall not change;

d) agree to institute the intended use of the granted copyright\_identifier within a reasonable time-frame;

e) maintain a permanent record of the application form and the notification received from the Registration Authority of each granted copyright\_identifier.

## L.4 Appeal procedure for denied applications

The registration management group is formed to have jurisdiction over appeals relating to a denied request for an RID. The RMG shall have a membership who are nominated by P and L members of the ISO technical body responsible for this Recommendation | International Standard. It shall have a convenor and secretariat nominated from its members. The Registration Authority is entitled to nominate one non-voting observing member.

The responsibilities of the RMG shall be:

a) to review and act on all appeals within a reasonable time-frame;

b) to inform, in writing, organizations which make an appeal for reconsideration of its petition of the RMG's disposition of the matter;

c) to review the annual report of the Registration Authority summary of activities;

d) to supply ISO member bodies with information concerning the scope of operation of the Registration Authority.

Annex M  
  
Registration application form (see 2.9)

(This annex does not form an integral part of this Recommendation | International Standard.)

## M.1 Contact information of organization requesting a Registered Identifier (RID)

Organization Name:

Address:

Telephone:

Fax:

email:

## M.2 Statement of an intention to apply the assigned RID

RID application domain: using guidelines to be provided by the Registration Authority.

## M.3 Date of intended implementation of the RID

## M.4 Authorized representative

Name:

Title:

Address:

Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## M.5 For official use only of the Registration Authority

|  |
| --- |
| Registration rejected: \_\_\_\_\_\_\_  Reason for rejection of the application:  Registration granted: \_\_\_\_\_\_\_\_ Registration value: \_\_\_\_\_\_\_ |

Attachment 1 – Attachment of technical details of the registered data format.

Attachment 2 – Attachment of notification of appeal procedure for rejected applications.

Annex N  
  
Registration Authority  
Diagram of administration structure (see 2.9)

(This annex does not form an integral part of this Recommendation | International Standard.)



Examples



All the copyright\_identifiers are registered by the Registration Authority,   
uniquely for copyright\_numbers standardized by ISO. Each organization   
which allocates copyright\_numbers, requests a specific copyright\_identifier  
from the Registration Authority, e.g., Staatsbibliothek Preussischer Kulturbesitz,   
designated by ISO to manage I.S.B.N., asks for a specific copyright\_identifier   
from the R.A. for book numbering.

Annex O  
  
Registration procedure (see 2.10)

(This annex does not form an integral part of this Recommendation | International Standard.)

## O.1 Procedure for the request of an RID

Requesters of an RID shall apply to the Registration Authority. Registration forms shall be available from the Registration Authority. The requester shall provide the information specified in O.4. Companies and organizations are eligible to apply.

## O.2 Responsibilities of the Registration Authority

The primary responsibilities of the Registration Authority administrating the registration of private data format\_identifiers is outlined in this annex; certain other responsibilities may be found in the JTC 1 Directives. The Registration Authority shall:

a) implement a registration procedure for application for a unique RID in accordance with the JTC 1 Directives;

b) receive and process the applications for allocation of an identifier from application providers;

c) ascertain which applications received are in accordance with this registration procedure, and to inform the requester within 30 days of receipt of the application of their assigned RID;

d) inform application providers whose request is denied in writing within 30 days of receipt of the application, and to consider resubmission of the application in a timely manner;

e) maintain an accurate register of the allocated identifiers. Revisions to format specifications shall be accepted and maintained by the Registration Authority;

f) make the contents of this register available upon request to National Bodies of JTC 1 that are members of ISO or IEC, to liaison organizations of ISO or IEC and to any interested party;

g) maintain a database of RID request forms, granted and denied. Parties seeking technical information on the format of private data which has an RID shall have access to such information which is part of the database maintained by the Registration Authority;

h) report its activities to JTC 1, the ITTF, and the SC 29 Secretariat, or their respective designees, annually;

i) accommodate the use of existing RIDs whenever possible.

## O.3 Contact information for the Registration Authority

The contact information for the Registration Authority concerning provisions of Rec. ITU-T H.222.0 | ISO/IEC 13818-1 can be found at <https://www.iso.org/maintenance_agencies.html> or <https://handle.itu.int/11.1002/plink/0724859361>.

## O.4 Responsibilities of parties requesting an RID

The party requesting a format\_identifier shall:

a) apply, using the form and procedures supplied by the Registration Authority;

b) include a description of the purpose of the registered bit stream, and the required technical details as specified in the application form;

c) provide contact information describing how a complete description can be obtained on a non‑discriminatory basis;

d) agree to institute the intended use of the granted RID within a reasonable time-frame;

e) to maintain a permanent record of the application form and the notification received from the Registration Authority of a granted RID.

## O.5 Appeal procedure for denied applications

The Registration Management Group is formed to have jurisdiction over appeals to denied requests for an RID. The RMG shall have a membership who is nominated by P- and L-members of the ISO technical committee responsible for this Specification. It shall have a convenor and secretariat nominated from its members. The Registration Authority is entitled to nominate one non-voting observing member.

The responsibilities of the RMG shall be:

a) to review and act on all appeals within a reasonable time-frame;

b) to inform, in writing, organizations which make an appeal for reconsideration of its petition of the RMG's disposition of the matter;

c) to review the annual report of the Registration Authorities summary of activities;

d) to supply member bodies of ISO and National Committees of IEC with information concerning the scope of operation of the Registration Authority.

Annex P  
  
Registration application form

(This annex does not form an integral part of this Recommendation | International Standard.)

## P.1 Contact information of organization requesting an RID

Organization Name:

Address:

Telephone:

Fax:

email:

Telex:

## P.2 Request for a specific RID

NOTE – If the system has already been implemented and is in use, fill in this item and also the item P.3 and then skip to P.6; otherwise leave this space blank and skip to P.4.

## P.3 Short description of RID that is in use and date system that was implemented

## P.4 Statement of an intention to apply the assigned RID

## P.5 Date of intended implementation of the RID

## P.6 Authorized representative

Name:

Title:

Address:

Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## P.7 For official use of the Registration Authority

|  |
| --- |
| Registration rejected: \_\_\_\_\_\_\_  Reason for rejection of the application:  Registration granted: \_\_\_\_\_\_\_\_ Registration value: \_\_\_\_\_\_\_ |

Attachment 1 – Attachment of technical details of the registered data format.

Attachment 2 – Attachment of notification of appeal procedure for rejected applications.

Annex Q  
  
T-STD and P-STD buffer models for ISO/IEC 13818-7 ADTS

(This annex does not form an integral part of this Recommendation | International Standard.)

## Q.1 Introduction

The transport stream system target decoder model for audio streams is defined in 2.4.2. In this annex, the buffer model for ISO/IEC 13818-7 ADTS is described.

ISO/IEC 13818-7 ADTS audio streams can be recognized in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 multiplex through the presence of stream\_id='110y yyyy' ('y' = "do not care") and stream\_type=0x0F as defined in Tables 2-22 and 2-34.

## Q.2 Leak rate from transport buffer

For audio except ISO/IEC 13818-7 ADTS, the leak rate from Transport Buffer is 2 Mbit/s. This rate is, however, lower than the maximum rate of ISO/IEC 13818-7 ADTS. Therefore, the leak rate for ISO/IEC 13818-7 ADTS streams is set to a different value from ISO/IEC 11172-3 and ISO/IEC 13818-3 audio streams.

ISO/IEC 13818-7 ADTS elementary stream consists of one or more channels. The maximum rate of each channel is 576 kbit/s where the sampling frequency is 96 kHz. Therefore, the leak rate for ISO/IEC 13818-7 ADTS is calculated as in the following equation.

Rxn = 1.2 × Rmax × N bits per second

where:

Rmax is a constant 576 kbit/s as defined in 3.2.2 of ISO/IEC 13818-7. It is an upper bound of the bit rate per channel of AAC ADTS stream corresponding to the maximum value of sampling frequency (i.e., Fs = 96 kHz),

and where:

N is the number of audio channels that require their own decoder buffer in this elementary stream (i.e., individual channel streams in a single channel element or channel pair element and independently switched coupling channel elements).

## Q.3 Buffer size

For audio except ISO/IEC 13818-7 ADTS, the main buffer size is 3584 bytes. This size is, however, smaller than the maximum decoder input buffer size of ISO/IEC 13818-7 ADTS. Therefore, the main buffer size for the ISO/IEC 13818‑7 ADTS stream is set to a different value from ISO/IEC 11172-3 and ISO/IEC 13818-3 audio streams.

The main buffer size for ISO/IEC 13818-7 ADTS is calculated as follows:

BSn = BSmux + BSdec + BSoh

where BSoh, PES packet overhead buffering is defined as:

BSoh = 528 bytes

and BSmux, additional multiplexing buffering is defined as:

BSmux = 0.004 seconds × Rmax × N

and BSdec, access unit buffering is defined as:

BSdec = 6144 bits × N

where:

Rmax is a constant 576 kbit/s as defined in 3.2.2 of ISO/IEC 13818-7. It is an upper bound of the bit rate per channel of AAC ADTS stream corresponding to the maximum value of sampling frequency (i.e., Fs = 96 kHz),

and where:

N is the number of audio channels that require their own decoder buffer in this elementary stream (i.e., individual channel streams in a single channel element or channel pair element and independently switched coupling channel elements).

### Q.3.1 TBSn: same as other audio

In term of the smoothing buffer, there is no difference in TBn between ISO/IEC 13818-7 ADTS and other audio streams. Consequently, it is not necessary to change TBSn, which is size of TBn.

### Q.3.2 BSmux: different from other audio

BSmux, additional multiplexing buffering, shall be changed to accept up to 4 ms of delay jitter. This is similar to the approach taken for other streams in Rec. ITU-T H.222.0 | ISO/IEC 13818-1.

### Q.3.3 BSdec: different from other audio

BSdec, access unit buffering is based on the decoder input buffer size of the elementary stream. As defined in 3.2.2 of ISO/IEC 13818-7, total decoder input buffer size is 6144 bits multiplied by the number of channels which require their each decoder input buffer.

### Q.3.4 BSoh: different from other audio

BSoh corresponds to the PES packet header overhead.

In 2.4.2.7,

*The delay of any data through the System Target Decoders buffers shall be less than or equal to one second except for still picture video data.*

Besides, in 2.7.4,

*The program stream and transport stream shall be constructed so that the maximum difference between coded presentation time stamp referring to each elementary video or audio stream is 0.7 seconds.*

BSoh shall be set to the appropriate size corresponding to the PES packet header overhead when AAC stream is packetized with the above rules. The maximum size of PES packet header is 264 bytes. Therefore, BSoh = 528 bytes, i.e., twice the maximum size of PES packet header, assures that at least two PES packet headers can enter the main buffer regardless of the size of PES packet header. It means that PES packet header with PTS can be inserted at less than 0.7 seconds intervals even when the data of one second will be in the main buffer.

Example: sampling frequency is 48 kHz

The size of PES packet header without any optional fields except PTS is 18 bytes. The number of Access Units of one second is about 47. When the data of one second is in the main buffer (i.e., the worst case), PES packet header overhead can fit the BSoh with packetizing more than or equal to two Access Units into one packet.

number\_of\_AU = 48 kHz/1024 = 46 875 per second

(number\_of\_AU/2) × 18[byte] = 421 875 bytes < BSoh

More frequent PES packet headers can fit BSoh, if the delay of any data through the main buffer is shorter than one second.

## Q.4 Conclusion

The decoder buffer model should cover the maximum size of buffer; however, AAC can handle up to 48 channels and very high bitrate. Therefore the 3 levels of number of channels, 2, 8 and 48, are used to define the leak rate and the main buffer size. In a case of 2, the same leak rate and main buffer size as the conventional values are used to keep the compatibility. In other cases (8 and 48), the proposed formulas are applied.

*T-STD leak rate for ISO/IEC 13818-7 ADTS audio,*

|  |  |
| --- | --- |
| Number of channels | Rxn [bit/s] |
| 1 .. 2 | 2 000 000 |
| 3 .. 8 | 5 529 600 |
| 9 .. 12 | 8 294 400 |
| 13 .. 48 | 33 177 600 |

Channels: The number of full-bandwidth audio output channels plus the number of independently switched coupling channel elements within the same elementary audio stream. For example, in the typical case that there are no independently switched coupling channel elements, mono is 1 channel, stereo is 2 channels and 5.1 channel surround is 5 channels (the LFE channel is not counted).

*T-STD main buffer size or ISO/IEC 13818-7 ADTS audio*

|  |  |
| --- | --- |
| Number of channels | BSn [bytes] |
| 1 .. 2 | 3 584 |
| 3 .. 8 | 8 976 |
| 9 .. 12 | 12 804 |
| 13 .. 48 | 51 216 |

Channels: The number of full-bandwidth audio output channels plus the number of independently switched coupling channel elements within the same elementary audio stream. For example, in the typical case that there are no independently switched coupling channel elements, mono is 1 channel, stereo is 2 channels and 5.1 channel surround is 5 channels (the LFE channel is not counted).

For program stream, the above main buffer size should be set in the P-STD\_buffer\_scale and P-STD\_buffer\_size as follows.

|  |  |  |
| --- | --- | --- |
| Number of Channels | P-STD\_buffer\_scale | P-STD\_buffer\_size |
| 1 .. 2 | 0 | 28 |
| 3 .. 8 | 0 | 71 |
| 9 .. 48 | 0 | 401 |

Annex R  
  
Carriage of ISO/IEC 14496 scenes in Rec. ITU-T H.222.0 | ISO/IEC 13818-1

(This annex does not form an integral part of this Recommendation | International Standard.)

## R.1 Content access procedure for ISO/IEC 14496 program components within a program stream

The following provides a reference receiver acquisition procedure for accessing ISO/IEC 14496 program elements in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 program stream. Here, it is assumed that the program stream includes a Program Stream Map (conveyed in a PES packet having stream\_id equal to 0xBC):

• Acquire the Program Stream Map.

• Identify the IOD descriptor in the first descriptor loop.

• Identify the ES\_IDs of the object descriptor, scene description and other streams described within the initial object descriptor.

• Acquire the SL descriptor and FMC descriptor in the second descriptor loop for elementary\_stream\_ids 0xFA and 0xFB, as applicable.

• Generate from these descriptors a stream map table between ES\_IDs and associated elementary\_stream\_id plus M4Mux channel, if applicable.

• Locate the object descriptor stream using its ES\_ID and the stream map table.

• Locate other streams described in the initial object descriptor using their ES\_ID and the stream map table.

• Continuously monitor the object descriptor stream and identify ES\_IDs of additional streams.

• Locate the additional streams using their ES\_ID and the stream map table.

Figure R.1 gives an example of ISO/IEC 14496 content in a program stream, consisting of object descriptor stream, scene description stream (BIFS-Command), BIFS-Anim stream and IPMP stream. All ISO/IEC 14496 streams are multiplexed in a single M4Mux stream.



Figure R.1 – Example of ISO/IEC 14496 content in a program stream

## R.2 Content access procedure for ISO/IEC 14496 program components within a transport stream

The following provides a reference receiver acquisition procedure for accessing ISO/IEC 14496 program elements in a Rec. ITU-T H.222.0 | ISO/IEC 13818-1 transport stream:

• Acquire the Program Map Table for the desired program.

• Identify the IOD descriptor in the first descriptor loop.

• Identify the ES\_IDs of the object descriptor, scene description and other streams described within the initial object descriptor.

• Acquire the set of all SL descriptors and FMC descriptors present in the second descriptor loop for any of the elementary\_PIDs.

• Generate from these descriptors a stream map table between ES\_IDs and associated elementary\_PID plus M4Mux channel, if applicable.

• Locate the object descriptor stream using its ES\_ID and the stream map table.

• Locate other streams described in the initial object descriptor using their ES\_ID and the stream map table.

• Continuously monitor the object descriptor stream and identify ES\_IDs of additional streams.

• Locate the additional streams using their ES\_ID and the stream map table.

• If ODUpdate\_descriptors are present in the first descriptor loop, process the ObjectDescriptor\_Update as defined in 2.6.92.

Figure R.2 gives an example of ISO/IEC 14496 program elements in a transport stream, consisting of object descriptor stream, scene description stream (BIFS-Command), BIFS-Anim and IPMP elementary streams. BIFS-Command and OD stream are conveyed by means of ISO\_IEC\_14496\_sections, while BIFS-Anim and IPMP elementary streams are conveyed in PES packets referenced by two distinct elementary\_PID values, without the use of the ISO/IEC 14496-1 M4Mux tool.



Figure R.2 – Example of ISO/IEC 14496 content in a transport stream

Figure R.3 gives an example of ISO/IEC 14496 program elements in a transport stream, consisting of a scene description stream (BIFS-Command), and audio and video streams natively carried over PES (no SL packetization or M4Mux). BIFS-Command stream are conveyed by means of ISO\_IEC\_14496\_sections, and the BIFS scene directly refers to the audio and video streams in the transport stream through "pid://" URLs in the BIFS media nodes.



Figure R.3 – Usage of MPEG-4 in a transport stream with BIFS scene referring to native PES

Figure R.4 gives an example of ISO/IEC 14496 program elements in a transport stream, consisting of a scene description stream (BIFS-Command) and an image stream. The BIFS-Command stream is conveyed by means of ISO\_IEC\_14496\_sections, the image stream is conveyed in PES packets using SL packetization. The ObjectDescriptor associated with the image stream is conveyed by means of an ODUpdate\_Descriptor in the first descriptor loop of the PMT.



Figure R.4 – Usage of MPEG-4 in a transport stream with an ODUpdate\_descriptor   
carrying an image ObjectDescriptor in the PMT

Annex S  
  
Carriage of JPEG 2000 part 1 video over MPEG-2 transport streams

(This annex forms an integral part of this Recommendation | International Standard.)

## S.1 Introduction

This annex specifies normative constraints for the carriage of JPEG 2000 video in an MPEG-2 transport stream. The parameters specified include mapping of J2K video streams into MPEG-2 transport packets, signalling of J2K video streams as well as T-STD parameters for various profiles. Transport of J2K video shall be limited to transport stream only. Program stream support may be added in the future based on application requirements.

## S.2 J2K video access unit, J2K video elementary stream, J2K video sequence and J2K still picture

The J2K video access unit contains the elementary stream (elsm) header created as defined in S.3 concatenated with self-contained Rec. ITU-T T.800 | ISO/IEC 15444-1 codestream(s). The (elsm) header contains all video-related parameters necessary to display the decoded codestream(s). Multiple codestreams may comprise an access unit in the two following situations:

1. When the access unit is an interlaced (portion of) frame.

2. When the J2K stripe mode is enabled, as defined in S.4.

The J2K video elementary stream is a progression of J2K access units and the J2K video sequence is a subset of J2K video elementary stream where all the J2K access units have the same parameters in the (elsm) header.

The J2K still picture (system) consists of a J2K video sequence which contains exactly one J2K access unit. This still picture has an associated PTS and the presentation time of succeeding pictures, if any, is later than that of the still picture by at least two picture periods. The J2K still picture (system) mode is used to support transmission of J2K video access units at a rate much lower than the display frame rate (determined by the difference in PTS values between successive J2K access units). J2K still picture can be used in applications such as 'slide show' and 'stills with Music'.

## S.3 Optional J2K block mode for high resolution support

By default, each video frame is encoded into one self-contained Rec. ITU-T T.800 | ISO/IEC 15444-1 codestream for progressive content, and into two such codestreams for interlaced content (one per field). The codestream(s) are then encapsulated in a J2K access unit, which therefore always contains an entire video frame.

For 4k resolution (3840x2160) and beyond, the J2K block mode can be used. This mode allows to divide each video frame horizontally and vertically in the spatial domain, into multiple video blocks, each block being processed the same way entire frames were processed in the default mode. The main use case for this mechanism is to be able to stream different quadrants or blocks of a video separately. As an example, a 3840x2160p/59.94 video can be divided into four 1920x1080p/59.94 video blocks.

For this mode to be correctly enabled, the video descriptor shall have *extended\_capability\_flag* and *block\_flag* both set to '1' and shall contain the block dimensions (stored in *horizontal\_size* and *vertical\_size* fields), the frame dimensions (stored in *full\_horizontal\_size* and *full\_vertical\_size* fields), the maximum values of the horizontal and vertical block index (*max\_blk\_idx\_h*, *max\_blk\_idx\_v*), and the horizontal and vertical index value for the block being processed by the video descriptor (*blk\_idx\_h*, *blk\_idx\_v*). See Table 2-100 for further details on the J2K video descriptor. The blocks shall be constructed so that they have the same horizontal and vertical size, with the exception of the ones located on the last row and the ones located on the last column, which may have a different size. The following method shall be used to derive the width and height of a block:

if (blk\_idx\_h == max\_blk\_idx\_h):

width = full\_horizontal\_size – ( horizontal\_size \* max\_blk\_idx\_h )

otherwise:

width = horizontal\_size

if (blk\_idx\_v == max\_blk\_idx\_v):

height = full\_vertical\_size – ( vertical\_size \* max\_blk\_idx\_v )

otherwise:

height = vertical\_size

When this mode is enabled, each J2K video access unit contains the Rec. ITU-T T.800 | ISO/IEC 15444-1 codestream(s) related to a specific video block only. Corresponding J2K video blocks across video frames compose an independent J2K video elementary stream (with its own J2K video descriptor) that shall have its own PID.

## S.4 Optional J2K stripe mode for Ultra-Low Latency

The J2K stripe mode has been defined to allow carriage of JPEG 2000 video in an MPEG-2 transport stream with a sub-frame end-to-end latency. This mode is implemented through an horizontal striping mechanism: the picture embedded in a J2K access unit (whether it be an entire progressive video frame, the two fields of an interlaced video frame, a progressive video block or the two fields of an interlaced video block as defined in S.3) is divided horizontally into an integer number of stripes, each stripe being encoded as a self-contained Rec. ITU-T T.800 | ISO/IEC 15444-1 codestream.

To produce a compliant Ultra-Low Latency (ULL) J2K video stream, the J2K video descriptor (see Table 2-100) shall have the following information set accordingly

• The *extended\_capability\_flag* and the *stripe\_flag* both set to '1'.

• The fields *strp\_max\_idx* and *strp\_height* respectively indicating the number of stripes per frame minus one, and the stripe height.

In the JPEG 2000 Elementary Stream header, they shall ensure the Stripe Coding box (*j2k\_strp*) is present, and they shall ensure the Time Coding box (*j2k\_tcod*) is not present. Stripes shall be included in raster order, from top to bottom. All stripes shall have the same size, with the exception of the very last (bottom) one, which may have a different height. The stripe height is given by the following formula:

if last stripe in the frame/field/block:

*height = vertical\_size – (strp\_height \* strp\_max\_idx)*

otherwise:

*height = strp\_height*

When this mode is enabled, each J2K access unit contains *N* Rec. ITU-T T.800 | ISO/IEC 15444-1 codestreams where:

*N* = (strp*\_max\_idx* + 1) for progressive content

*N* = 2\*(*strp\_max\_idx* + 1) for interlaced content

## S.5 Elementary stream header (elsm) and mapping to PES packets

Each J2K access unit from a J2K video elementary stream shall start with an elementary stream header (elsm header) as detailed in Table S.1.

Table S.1 – J2K Access unit elementary stream header

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| J2K\_elsm\_header() {  **elsm\_box\_code** '0x656c 736d'  // j2k\_frat  **frat\_box\_code** '0x6672 6174'  **frat\_denominator**  **frat\_numerator**  // j2k\_brat  **brat\_box\_code** '0x6272 6174'  **brat\_max\_br**  **brat\_auf1**  If (interlaced\_video == 1) {  **brat\_auf2**  }  // j2k\_fiel  If (interlaced\_video == 1) {  **fiel\_box\_code** '0x6669 656c'  **fiel\_fic**  **fiel\_fio**  }  If (extended\_capability\_flag == 0 || stripe\_flag == 0) {  // j2k\_tcod  **tcod\_box\_code** '0x7463 6f64'  **tcod\_hh** (0-23)  **tcod\_mm** (0-59)  **tcod\_ss** (0-59)  **tcod\_ff** (1-60)  } else { // extended\_capability\_flag == 1 && stripe\_flag == 1  // j2k\_strp  **strp\_box\_code** '0x7374 7270'  **strp\_max\_idx**  **frame\_vertical\_size**  **reserved**  }  // j2k\_bcol  If (extended\_capability\_flag == 0) {  **bcol\_box\_code** '0x6263 6f6c'  **bcol\_colcr**  **reserved**  } else { // extended\_capability\_flag == 1  **color\_primaries**  **transfer\_characteristics**  **matrix\_coefficients**  **video\_full\_range\_flag**  **reserved**  }  } | 32  32  16  16  32  32  32  32  32  8  8  32  8  8  8  8  32  8  16  8  32  8  8  8  8  8  1  23 | bslbf  bslbf  uimsbf  uimsbf  bslbf  uimsbf  uimsbf  uimsbf  bslbf  uimsbf  uimsbf  bslbf  uimsbf  uimsbf  uimsbf  uimsbf  bslbf  uimsbf  uimsbf  bsfbf  bslbf  uimsbf  bsfbf  uimsbf  uimsbf  uimsbf  bsfbf  bsfbf |

**Semantics**

**j2k\_frat** – The two fields *frat\_denominator* and *frat\_numerator* are used to derive the frame rate in frames per second. If either is zero, the frame rate is variable or undefined. Otherwise, the frame rate is expressed by a rational number of the form numerator/denominator. If the frame rate is an integer, the denominator shall be equal to 1. In case of interlaced content (two fields per frame), then the field rate is twice the frame rate.

**frat\_denominator** – This field is defined as a 2-byte big-endian unsigned integer. It specifies the denominator used to calculate the frame rate.

**frat\_numerator** – The field is defined as a 2-byte big-endian unsigned integer. It specifies the numerator used to calculate the frame rate.

**j2k\_brat** – The three fields *brat\_max\_br, brat\_auf1* and *brat\_auf2* specify the maximum bit rate of the elementary stream and the actual access unit size.

**brat\_max\_br** – This field is defined as a 4-byte big-endian unsigned integer which specifies a maximum instantaneous bit rate that is not to be exceeded, expressed in bits per second for the elementary stream at the frame rate specified by *j2k\_frat* fields. The maximum bit rate shall not exceed the bit rate specified for a given profile and level.

**brat\_auf1** – This field is defined as a 4-byte big-endian unsigned integer which specifies the size of the contiguous codestream corresponding to the frame in case of progressive content and to the first field in case of interlaced content. This field shall be set to zero if J2K stripe mode is enabled (*stripe\_flag* == 1).

**brat\_auf2** – This field is defined as a 4-byte big-endian unsigned integer which specifies the size of the contiguous codestream corresponding to the second field in case of interlaced content. This field shall be omitted in case of progressive content. This field shall be set to zero in case of interlaced content and J2K stripe mode enabled.

**j2k\_fiel** – The fields *fiel\_fic* and *fiel\_fio* specify the field order in case of interlaced content. These two fields, together with the field *fiel\_box\_code* shall only be present in case of interlaced content (*interlaced\_video* flag in the *j2k\_video\_descriptor* set to '1').

**fiel\_fic** – This field is defined as a 1-byte unsigned integer which specifies the number of fields in the access unit. As this field is only present in case of interlaced content, it shall always be set to 2, and is only present for backward compatibility reasons.

**fiel\_fio** – This field is defined as a 1-byte unsigned integer which describes the order of the two fields:

0 field coding unknown;

1 field with the topmost line is stored first in the access unit; fields are in temporal order;

6 field with the topmost line is stored second in the access unit; fields are in temporal order.

**j2k\_tcod** – The fields *tcod\_hh*, *tcod\_mm*, *tcod\_ss*, and *tcod\_ff* specify the time code of the access unit. These fields, together with *tcod\_box\_code* shall only be present if the J2K stripe mode is disabled (*stripe\_flag* in the *j2k\_video\_descriptor* set to '0').

**tcod\_hh** – This field is a 1-byte unsigned integer which specifies the hour. The value ranges from 0 to 23, inclusively.

**tcod\_mm** – This field is a 1-byte unsigned integer which specifies the minutes. The value ranges from 0 to 59, inclusively.

**tcod\_ss** – This field is a 1-byte unsigned integer which specifies the seconds. The value ranges from 0 to 59, inclusively.

**tcod\_ff** – This field is a 1-byte unsigned integer which specifies the frame count. The value ranges from 1 to 60, inclusively.

**j2k\_strp** – The fields *strp\_max\_idx* and *frame\_vertical\_size* specify the settings for the striping mechanism. These fields, together with *strp\_box\_code* and the *reserved* field shall only be present if the J2K stripe mode is enabled (*extended\_capability\_flag* and *stripe\_flag* in the *j2k\_video\_descriptor* both set to '1').

**strp\_max\_idx** – This field is a 1-byte unsigned integer which specifies the number of J2K stripes per frame (for progressive content) or per field (for interlaced content) in this access unit, minus one. This value shall be equal to the *strp\_max\_idx* value found in the *j2k\_video\_descriptor.*

**frame\_vertical\_size** – This field is a 2-byte unsigned integer which specifies the vertical size of the frame (for progressive) or field (for interlaced) comprised in this J2K access unit. This value shall be equal to the *vertical\_size* value found in the *j2k\_video\_descriptor.*

**j2k\_bcol** – The j2k\_bcol fields described hereunder specify the color parameters of the access unit. They define one method by which an application can interpret the colourspace of the decompressed access unit. This colour specification is to be applied to the access unit after it has been decompressed and after any reverse decorrelating component transform has been applied to the image data.

**bcol\_colcr** – This field is a 1-byte integer which specifies the corresponding colour standard as defined in Table M.2 of Rec. ITU-T T.800 | ISO/IEC 15444-1. It shall only be present if, in the J2K video descriptor, *extended\_capability\_flag* is set to '0'.

NOTE – In Table S.1, it should be noted that the *bcol\_box\_code* is only present when the *bcol\_colcr* field is present. It is indeed only present for a purpose of backward compatibility with previous version of this Recommendation | International Standard.

**color\_primaries, transfer\_characteristics, matrix\_coefficients, video\_full\_range\_flag** – These four fields (three 1-byte integers and one 1-bit flag) shall be coded according to the semantics with the same name defined in Rec. ITU-T H.273 | ISO/IEC 23001-8. They shall only be present if, in the J2K video descriptor *extended\_capability\_flag* is set to '1'.

Figure S.1 shows the structure and mapping of J2K video access unit into PES packets. JPEG 2000 represents each frame or block (if J2K block mode is enabled) as *N* Part 1 (Rec. ITU-T T.800 | ISO/IEC 15444-1) contiguous codestreams, where

*N* = 1, for progressive content, J2K stripe mode disabled

2, for interlaced content, J2K stripe mode disabled

*strp\_max\_idx* + 1, for progressive content, J2K stripe mode enabled

2\*(*strp\_max\_idx* + 1), for interlaced content, J2K stripe mode enabled

The codestream main header, included within each contiguous codestream, contains all information to decode its image, including the image size and the profile indicator, called a SIZ marker in Rec. ITU-T T.800 | ISO/IEC 15444-1. Preceding the contiguous codestreams, an elementary stream (elsm) header containing video-related information, as shown in Table S.1, shall be present.

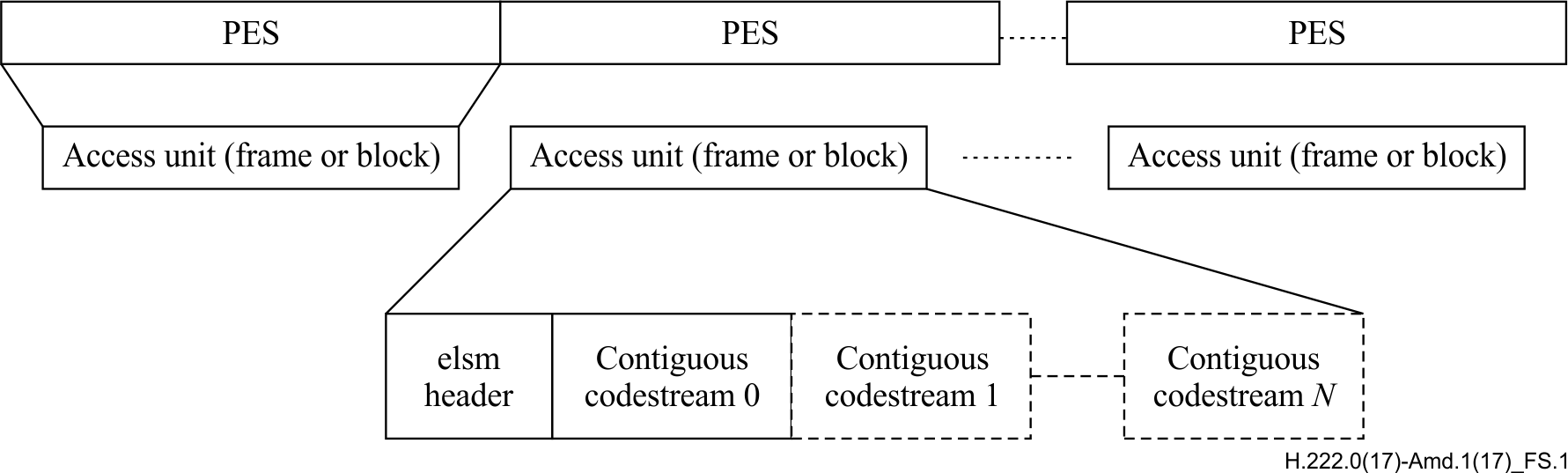


Figure S.1 – Structure and order of JPEG 2000 access units

## S.6 J2K transport constraints

When a J2K video elementary stream conforming to one or more profiles as defined in Rec. ITU-T T.800 | ISO/IEC 15444-1 is transported using MPEG-2 systems, the following constraints apply:

1) Each J2K access unit shall contain an elementary stream header (elsm) defined in Table S.1 followed by *N* Rec. ITU-T T.800 | ISO/IEC 15444-1 codestream(s) (see S.5).

2) Each J2K codestream main header contains a SIZ marker segment that includes a RSIZ capability parameter. Both the SIZ marker segment and RSIZ capability parameter are defined in Annex A of Rec. ITU-T T.800 | ISO/IEC 15444-1. The 15 less significant bits of the RSIZ parameter are equal to the *profile\_and\_level*parameter of the *j2k\_video\_descriptor* (see 2.6.81).

3) The J2K video access units shall be ordered in the J2K video elementary stream in a monotonic display order.

4) Each J2K video access unit shall include a PES header with PTS and each PES packet shall contain exactly one J2K video access unit.

5) For successive J2K video access units, the increments to PTS shall be consistent with increments to corresponding J2K\_tcod parameters in the elsm header.

6) The following constraints apply to the coding of syntax elements in the adaptation header for transport of J2K video elementary stream:

a) Both random\_access\_indicator and elementary\_stream\_priority\_indicator flags can be set to '1' for each J2K video access unit contained in the transport packet. Applications may limit the signalling of random access based on their use cases. Any J2K video access unit is also a J2K video 'access point' required for random access.

b) All other flags should be set appropriately.

7) The following constraints apply to the coding of syntax elements in the PES header for transport of J2K video elementary stream:

a) stream\_id shall be set to 0xBD (same as Private\_stream\_1).

b) PES\_packet\_length shall be set to 0x0000.

c) data\_alignment\_indicator shall be set to '1'. Also see S.5 for alignment constraints.

d) PTS\_DTS\_flags shall be set to '10'. This precludes display re-ordering for J2K video access units.

e) All other flags should be set appropriately.

## S.7 Interpretation of flags in adaptation and PES headers for J2K video elementary streams

The interpretation, extensions, use and constraints for the following syntax elements in the adaptation header (2.4.3.4 and 2.4.3.5) and PES header (2.4.3.6 and 2.4.3.7) for JPEG 2000 Part 1 video are defined in this clause:

• In the semantics for discontinuity\_indicator (see 2.4.3.5), a J2K video elementary stream access point is defined as the first byte of the J2K video access unit.

• The elementary\_stream\_priority\_indicator in adaptation header (see 2.4.3.4 and 2.4.3.5) may be set to '1' if the transport packet payload contains the start of a J2K video access unit.

• For J2K video elementary streams, when the data\_alignment\_indicator (see 2.4.3.6) is set to '1', the PES packet header shall be immediately followed by the first byte of J2K video access unit. The data\_stream\_alignment\_descriptor is optional and, if included for J2K video, the alignment\_type shall be set to 0x00.

For J2K video elementary streams conforming to one or more profiles defined in Rec. ITU-T T.800 | ISO/IEC 15444‑1, if a PTS is present in the PES packet header (see 2.4.3.7), it shall refer to the first byte of J2K video access unit that commences in this PES packet.

## S.8 T-STD extension for J2K video elementary streams

### S.8.1 General

The T-STD model includes a transport buffer TBn and J2K video elementary stream buffer EBn for decoding of each J2K video elementary stream n. See Figure S.2.



Figure S.2 – T-STD model extensions for J2K Video

The following notation is used to describe the transport stream system target decoder and is partially illustrated in Figure 2-1.

i index to bytes in the transport stream. The first byte has index 0.

j is an index to access units in the elementary streams.

k index to presentation units in the elementary streams.

n is an index to the elementary streams.

p is an index to transport stream packets in the transport stream.

t(i) indicates the time in seconds at which the i-th byte of the transport stream enters the system target decoder. The value t(0) is an arbitrary constant.

PCR(i) is the time encoded in the PCR field measured in units of the period of the 27 MHz system clock where i is the byte index of the final byte of the program\_clock\_reference\_base field.

An(j) is the j-th access unit in elementary stream n. An(j) is indexed in decoding order.

tdn(j) is the decoding time, measured in seconds, in the system target decoder of the j-th access unit in elementary stream n.

Pn(k) is the k-th presentation unit in elementary stream n. Pn(k) results from decoding An(j). Pn(k) is indexed in presentation order.

tpn(k) is the presentation time, measured in seconds, in the system target decoder of the k-th presentation unit in elementary stream n.

t is time measured in seconds.

Fn(t) is the fullness, measured in bytes, of the system target decoder input buffer for elementary stream n at time t.

Bn is the main buffer for elementary stream n. It is present only for audio elementary streams.

BSn is the size of buffer, Bn, measured in bytes.

EBn is the elementary stream buffer for elementary stream n. It is present only for video elementary streams.

EBSn is the size of the elementary stream buffer EBn, measured in bytes.

TBn is the transport buffer for elementary stream n.

TBSn is the size of TBn, measured in bytes.

Rxn is the rate at which data are removed from TBn.

**TBn and EBn buffer management**

The input to buffer TBn and its size TBSn are specified in 2.4.2.4. The rate Rxn between TBn and buffer EBn and the following constraints apply for the carriage of a J2K video elementary stream:

Size EBSn of buffer EBn: See Table S.2 for the maximum buffer size based on profile and level of J2K video stream.

Rate Rxn:

when there is no data in TBn then Rxn is equal to zero,

otherwise: Rxn = bit\_rate,

where bit\_rate is the bit rate as specified for the profile level in Table S.2. All J2K video payload data bytes enter EBn instantaneously upon leaving TBn.

***Removal of J2K access units from EBn***

Each J2K video access unit An(j) that is present in EBn is removed instantaneously at time tdn(j). The decoding time tdn(j) is specified by the PTS (as PTS = DTS for J2K video elementary streams) in the PES header for the J2K video access unit.

**STD delay**

The total delay of any J2K video elementary stream data other than J2K still picture data through the system target decoders buffers TBn and EBn shall be constrained by tdn(j) – t(i) ≤ 1 second for all j, and all bytes i in J2K video access unit An(j).

The delay of any J2K still picture data through the system target decoders buffers TBn and EBn shall be constrained by tdn(j)  t(i) ≤ 60 seconds for all j, and all bytes i in J2K video access unit An(j).

**Buffer management conditions**

Transport streams shall be constructed so that the following conditions for buffer management are satisfied:

• TBn shall not overflow and shall be empty at least once every second.

• EBn shall not overflow or underflow.

NOTE – EBn shall not underflow as J2K video elementary streams do not support low delay mode.

### S.8.2 J2K video elementary stream buffer size

A J2K video elementary stream supplies parameters suitable for determining a standard decoder model buffer size. This CODEC uses variable rate compression. Profiles from Rec. ITU-T T.800 (2015) | ISO/IEC 15444-1:2016 specify operating levels that limit the maximum compressed bit rate of the Rec. ITU-T T.800 | ISO/IEC 15444-1 codestream. Moreover, within access units an elementary stream header contains the present maximum bit rate an encoder must support to decode video without overflow. A decoder may scale the buffer size by reading the maximum bit rate expected by a particular sequence of frames from the elsm header. This header information shall not exceed the limit set by the operating level specified in Rec. ITU-T T.800 (2015) | ISO/IEC 15444-1:2016, Table A.49.

Table S.2 – Operating levels and maximum buffer size for JPEG 2000 broadcast profiles  
(from Table A.49 in Rec. ITU-T T.800 (2015) | ISO/IEC 15444-1:2016)

|  |  |  |
| --- | --- | --- |
| Levels  (Note) | Max. compressed bit rate (Mbit/s) | Maximum buffer size for a 60.0 Hz frame rate, progressive video (MBytes) |
| Level 0 | Unspecified | Unspecified |
| Level 1 | 200 | 1.25 |
| Level 2 | 200 | 1.25 |
| Level 3 | 200 | 1.25 |
| Level 4 | 400 | 2.5 |
| Level 5 | 800 | 5 |
| Level 6 | 1600 | 10 |
| Level 7 | 3200 | 20 |
| Level 8 | 6400 | 40 |
| Level 9 | 12800 | 80 |
| Level 10 | 25600 | 160 |
| Level 11 | 51200 | 320 |
| NOTE – Levels are specified in the JPEG 2000 codestream main header SIZ marker segment, RSIZ capability parameter. Refer to Rec. ITU-T T.800 (2015) | ISO/IEC 15444-1:2016, Table A.10. | | |

Annex T  
  
MIME type for MPEG-2 transport streams

(This annex does not form an integral part of this Recommendation | International Standard.)

## T.1 Introduction

This annex provides the formal MIME type registration for MPEG-2 transport streams. It is referenced from the registry at http://www.iana.org/.

## T.2 MIME type and subtype

MIME media type name:  
 video

MIME subtype name:  
 mp2t

Required parameters:  
 none

Optional parameters:  
 The 'profiles' parameter as documented in T.2.1

The 'codecs' parameter as document in T.2.2

Encoding considerations:

This type is defined for general use; for transfer via RTP see IETF RFC 3550.

Security considerations:  
 see T.3

Interoperability considerations:

The specification defines a platform-independent expression of a presentation, and it is intended that wide interoperability can be achieved.

Published specification:  
 ITU-T H.222.0 | ISO/IEC 13818-1, Information technology – Generic coding of moving pictures and associated audio information: Systems

Applications that use this media type:  
 various, including video streaming and video broadcasting applications

Additional information:  
 File extension(s):

.ts

Intended usage:

COMMON

Other information/General comments:

none

Person to contact for further information:

Name:

David Singer

e-mail:

[Singer@apple.com](mailto:Singer@apple.com)

Change controller:

ISO/IEC JTC1/SC29 (MPEG)

## T.3 Security considerations

It is possible to inject non-compliant streams (audio, video and systems) in the transport stream to overload the receiver/decoder's buffers. This might compromise the functionality of the receiver or even crash it.

An MPEG-2 transport stream is an extensible container format, and hence might carry streams that have active aspects (e.g., contain script snippets). If those subsystems are not properly defined or implemented, it may be possible to crash the receiver or temporarily make it unavailable.

## T.4 Parameters

### T.4.1 The profiles parameter

Parameter Name: profiles

Parameter Value: The 'profiles' parameter is an optional parameter that indicates one or more profiles  
 to which the stream claims conformance. The contents of this attribute shall conform  
 to either the pro-simple or pro-fancy productions of IETF RFC 6381, Section 4.5.   
 The profile identifiers reported in the MIME type parameter takes as value the  
 transport\_profile, coded as a decimal integer, e.g., profiles="1" for streams  
 conforming to the 'complete' profile.

Example: video/mp2t;profiles="1"

### T.4.2 The codecs parameter

Parameter Name: codecs

Parameter Value:

The 'codecs' parameter is an optional parameter that indicates one or more codecs which are used for the elementary streams in the MPEG-2 TS. The contents of this attribute shall conform to either the pro-simple or pro-fancy productions of IETF RFC 6381, Section 3.2. IETF RFC 6381 defines 'codecs' parameter as a single value, or a comma-separated list of values identifying the codec(s), where each value consists of one or more dot-separated elements. The first element of such a value can be derived from the value of stream\_type in the program\_map\_section. If the stream\_type value appears in Table T.1, the value in the first element column shall be used as the first element. If the stream\_type value is between 0x80 and 0xFF, the first element may be derived in ways not specified by this annex (e.g., by examining the contents of registration\_descriptor, if used).

Table T.1 – 'codecs' parameter values for some specific   
stream\_type values

|  |  |
| --- | --- |
| stream\_type | first element |
| 0x01 | mp1v |
| 0x02 | mp2v |
| 0x03 | mp1a |
| 0x04 | mp2a |
| 0x0F | mp2a |
| 0x10 | mp4v |
| 0x11 | mp4a |
| 0x1B | avc1 |
| 0x1C | mp4a |
| 0x1D | tx3g |
| 0x1F | svc1 |
| 0x20 | mvc1 |
| 0x21 | mjp2 |
| 0x2F | vqme |

Wherever a definition for additional elements exists in IETF RFC 6381 for a given first element, the definition in IETF RFC 6381 shall be followed.

When the first element of a value is 'mp1a', 'mp1v','mp2a' or 'mp2v', the second element of the codecs parameter value is the hexadecimal representation of the MP4 Registration Authority ObjectTypeIndication (OTI) for the appropriate specification and profile.

When the first element of a value is 'mp1a.6B' (ISO/IEC 11172-3), or 'mp2a.069' (i.e., ISO/IEC 13818-3), the third element of the codecs parameter value is the hexadecimal representation of the 2-bit layer, as defined in 2.6.4.

Examples:

ISO/IEC 13818-2 Main Profile

video/mp2ts;codecs="mp2v.61"

ISO/IEC 11172-3 layer 3 is represented

video/mp2ts;codecs="mp2a.6B.03"

ISO/IEC 13818-3 layer 2 is represented

video/mp2ts;codecs="mp2a.69"

ISO/IEC 13818-7 Low Complexity Profile

video/mp2ts;codecs="mp2a.67"

Dolby AC-3 audio (per ATSC A/52, AC-3 audio has stream\_type 0x81 and format\_identifier "AC-3" in the registration\_descriptor )

video/mp2ts;codecs="ac-3"

ISO/IEC 13818-2 Main Profile Video together with ISO/IEC 13818-7 audio

video/mp2ts;codecs="mp2v.61,mp2a.67"

Annex U  
  
Carriage of timeline and external media information over MPEG-2 transport streams

(This annex forms an integral part of this Recommendation | International Standard.)

## U.1 Introduction

### U.1.1 General

This annex specifies a format for carriage of timeline and location of external media resource that may be used as a synchronized enhancement of an MPEG-2 transport stream. The possible resolving, consumption and rendering of external media indicated in the stream are outside the scope of this Recommendation | International Standard.

The format specifies the mapping of the transport stream program clock to an embedded timeline, the signalling of associated external resources, hereafter called add-on(s), and the signalling of prefetching events. The format is designed to be compact in order to fit within one TS packet for common use cases. The mapping of the embedded timeline indicated in the PES packet payload or in the adaptation field descriptor with the PTS value of the PES header of the PES packet provides a stable timeline for media streams in the program, regardless of PCR discontinuities or other timestamps rewriting that may happen in the network.

In the context of this annex, the "timeline and external media information" is called TEMI.

TEMI data may be carried directly in the adaptation field of a media component or carried in a dedicated PES, called TEMI stream.

The TEMI stream describes external data and associated timing for the program in the MPEG-2 Transport Stream with which the TEMI stream is associated through the Program Map Table.

TEMI data carried in adaptation field of a media component describes external data and associated timing for the program in the MPEG-2 Transport Stream carrying this media component.

This annex also specifies a format for carriage of boundary and labelling descriptors that may be used to indicate a boundary type for seamless content splicing or switching in the applications of Ad insertion, cloud DVR recording and segmentation of adaptive bit rate streaming. The possible resolving and consumption of the boundary descriptor and labelling descriptor indicated in the stream are outside the scope of this Recommendation | International Standard.

### U.1.2 Notation

This annex makes extensive use of variable-length, where field length is specified prior to the field itself. An additional short-hand notation is used to improve readability in these cases: length field names are referenced within the "number of bits" column of syntax tables. The alias name for the length field is provided in parenthesis in non-bold font at the same line as the length field, and the number of bits is given as a function of that field.

In the example in Table U.1, SFL is an alias for the some\_field\_length\_qwords field. As the latter can have values of 0 .. 3, some\_field can have lengths of 0, 64, 128 and 192 bits. Stating that some\_field is a 0-bit field implies that some\_field is not present (in the example below this would result in some\_structure() being a 1-byte structure).

Table U.1 – Variable field length notation example

| Syntax | No. bits | Mnemonic |
| --- | --- | --- |
| **some\_structure** {  **some\_field\_length\_qwords** (SFL)  **reserved**  **some\_field**  } | **2**  **6**  **SFL\*64** | **uimsbf**  **bslbf**  **uismbf** |

The full notation of the same structure is given in Table U.1*bis*.

Table U.1*bis* – Table U.1 in equivalent full notation

| Syntax | No. bits | Mnemonic |
| --- | --- | --- |
| **some\_structure** {  **some\_field\_length\_qwords**  **reserved**  if ( some\_field\_length\_qwords > 0 ) {  if (some\_field\_length\_qwords == 1 ) {  **some\_field**  }  if (some\_field\_length\_qwords == 2 ) {  **some\_field**  }  if (some\_field\_length\_qwords == 3 ) {  **some\_field**  }  }  } | **2**  **6**  **64**  **128**  **192** | **uimsbf**  **bslbf**  **uismbf**  **uismbf**  **uismbf** |

### U.1.3 Annex references

– ANSI/SCTE 35 2016, Digital Program Insertion Cueing Message for Cable

## U.2 TEMI access unit and TEMI elementary stream

The format of the TEMI access unit is defined in Table U.2. TEMI access units shall be carried as PES packets using private\_stream\_1 streamID and identified in the Program Map Table by the stream type 0x27. There shall be at most one TEMI elementary stream declared in the Program Map Table.

The payload of a TEMI PES packet is a single complete TEMI\_AU, i.e., there shall be one and only one complete TEMI Access Unit in a TEMI PES packet.

The TEMI PES Packet header shall contain a PTS timestamp, whose value is used to match the current System Time Clock with the timeline value embedded in the TEMI packet payload, as defined in Table U.2.

TEMI data is made of one or several AF descriptors. These AF descriptors may be sent in different access units and at different rates. A TEMI\_AU contains one or more AF descriptors. Since AF descriptors are independently decodable, all TEMI access units are random access points.

NOTE 1 – In order to avoid interpolation issues when frame-accurate synchronization is required, the indicated PTS should be the same as the PTS of the associated video or audio stream for which frame accurate sync is needed.

NOTE 2 – It is possible to perform timeline interpolation in-between TEMI access units, for example if multiple audio frames are packed in a single PES packet, or when the TEMI AU frequency is less than the media AU frequency. However, receivers detecting PCR discontinuities in-between TEMI AUs should be careful when performing interpolation.

Table U.2 – TEMI access unit

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| TEMI\_AU { |  |  |
| **CRC\_flag**  **reserved**  for (i=0; i<N; i++) {  af\_descriptor()  }  if (CRC\_flag) {  **CRC\_32**  }  } | **1**  **7**  **32** | **bslbf**  **bslbf**  **rpchof** |

**CRC\_flag** – A 1-bit flag, which when set to '1' indicates that a CRC field is present in the packet.

**CRC\_32** – This is a 32-bit field that contains the CRC value that gives a zero output of the registers in the decoder defined in Annex A after processing the entire payload of the TEMI access unit.

## U.3 AF descriptors

### U.3.1 Introduction

AF descriptors are structures used to carry various features of the timeline or other information. All AF descriptors have a format that begins with an 8-bit tag value. The tag value is followed by an 8-bit AF descriptor length and data fields. The following semantics apply to the descriptors defined throughout Annex U.

**af\_descr\_tag** – The af\_descr\_tag is an 8-bit field that identifies each AF descriptor.

Table U.3 provides the Rec. ITU-T H.222.0 | ISO/IEC 13818-1 defined, Rec. ITU-T H.222.0 | ISO/IEC 13818-1 reserved, and user available AF descriptor tag values.

**af\_descr\_length** – The af\_descr\_length is an 8-bit field specifying the number of bytes of the AF descriptor immediately following af\_descr\_length field.

Table U.3 – AF descriptor tags

| AF descriptor tag | Identification |
| --- | --- |
| 0x00 .. 0x03 | Rec. ITU-T H.222.0 | ISO/IEC 13818-1 reserved |
| 0x04 | Timeline descriptor |
| 0x05 | Location descriptor |
| 0x06 | BaseURL descriptor |
| 0x07 | Cets\_byte\_range\_descriptor (see Note ) |
| 0x08 | AF\_MPEG-H\_3dAudio\_extStreamID\_descriptor |
| 0x09 | AF\_MPEG-H\_3dAudio\_multi-stream\_descriptor |
| 0x0A | AF\_MPEG-H\_3dAudio\_command\_descriptor |
| 0x0B | Boundary Descriptor |
| 0x0C | Labeling Descriptor |
| 0x0D | HEVC\_tile\_substream\_ af\_descriptor |
| 0x0E .. 0x7F | Rec. ITU-T H.222.0 | ISO/IEC 13818-1 reserved |
| 0x80 .. 0xFF | User Private |
| NOTE – See 6.4 in ISO/IEC 23001-9 (Common encryption of MPEG-2 transport streams) for description and usage. | |

AF descriptors may be carried in the adaptation field of TS packets of a media elementary stream, as defined in 2.4.3.5.

### U.3.2 Location descriptor

The location descriptor is used to signal the location of external data that can be synchronized with the program. It conveys several locations and their type (optionally including MIME types), along with the ability to signal upcoming external data association though a countdown until activation of the external data. It is possible to signal splicing of external data, by signalling that the newly associated data is temporary and the previous association will be re-used later on.

Table U.4 – TEMI location descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| temi\_location\_descriptor {  **af\_descr\_tag**  **af\_descr\_length**  **force\_reload**  **is\_announcement**  **splicing\_flag**  **use\_base\_temi\_url**  **reserved**  **timeline\_id**  if (is\_announcement = = '1') {  **timescale**  **time\_before\_activation**  }  if (use\_base\_temi\_url = = '0') {  **url\_scheme**  **url\_path\_length**  for (i=0; i < url\_path\_length; i++) {  **url\_path**  }  }  **nb\_addons**  for (i=0; i < nb\_addons; i++) {  **service\_type**  if (service\_type = = 0) {  **mime\_length**  for (j=0; j < mime\_length; j++) {  **mime\_type**  }  }  **url\_subpath\_len**  for (j=0; j < url\_subpath\_len; j++) {  **addon\_location**  }  }  } | **8**  **8**  **1**  **1**  **1**  **1**  **5**  **7**  **32**  **32**  **8**  **8**  **8**  **8**  **8**  **8**  **8**  **8**  **8** | **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **uimsbf**  **bslbf** |

### U.3.3 Semantic definition of fields in location descriptor

**force\_reload** – when set to '1', indicates that the add-on descriptions received or fetched prior to receiving this descriptor may be obsolete and shall be reloaded before attempting to map media times or locate media components. Reloading may typically happen for manifest-based add-on such as MPEG-DASH or MPEG-MMT.

**is\_announcement** – when set to '1', indicates that the add-on described by this descriptor is not yet active.

**splicing\_flag** – when set to '1', indicates that the new add-on indicated by this descriptor temporarily interrupts the last defined add-on for which splicing\_flag was not set. It is possible to have a sequence of add-ons with splicing\_flag set. This allows terminal to optimize loading of the add-on when splicing period ends. There shall not be two temi\_location\_descriptor pointing to the same add-on with different values for splicing\_flag, unless another temi\_location\_descriptor pointing to different add-ons is sent in-between with a splicing\_flag set to '0'.

**url\_scheme** – indicates the URL scheme to use for the URL. If url\_scheme is not equal to 0, the scheme identified shall be prepended to the url\_path, according to Table U.5.

Table U.5 – TEMI URL scheme types

| TEMI URL scheme type | Scheme value |
| --- | --- |
| 0 | Scheme URL is included in url\_path |
| 1 | "http://" |
| 2 | "https://" |
| 3 .. 0x7F | Rec. ITU-T H.222.0 | ISO/IEC 13818-1 reserved |
| 0x80 .. 0xFF | User private |

**timeline\_id** – a unique identifier for this content location. If force\_reload is set to '0' and another temi\_location\_descriptor with the same timeline\_id and splicing\_flag has already been received, the associated descriptions of the two descriptors shall be the same. If the splicing\_flag differs for the same timeline\_id, the timeline\_id is reassigned to the new URL defined in the descriptor (i.e., redefinition of timeline\_id).

**timescale** – indicates the timescale used to express the time\_before\_activation field in this message.

**use\_base\_temi\_url** – when set to '1', indicates that the URL defined in the last received temi\_base\_url\_descriptor shall be used as a base URL; when set to '0', a base URL is provided in the payload of this descriptor for the location described in this descriptor and only this descriptor.

**time\_before\_activation** – indicates the time in timescale units until the resource identified by addon\_location becomes active; the ratio ( time\_before\_activation / timescale ) indicates a duration in seconds. An implementation may use this information to start prefetching content.

**url\_path\_length** – indicates the length in bytes of the base URL path; when set to 0, indicates an empty URL path.

**url\_path** – base URL common to the different add-ons, if any; it shall be encoded without trailing zero character. This URL shall be a valid URL, as defined in clause 3 of RFC 3986, and may contain a Fragment and or a Query part.

**nb\_addons** – indicates the number of add-ons that share this timeline. If set to 0, only one add-on is present at the location indicated by url\_path, if this string is not empty. If url\_path is empty and nb\_addons is 0, this means that no service is associated with the current broadcast. If url\_path is empty and nb\_addons is not 0, url\_subpath\_len must be greater than 0.

**service\_type** – indicates the type of add-on present at the given URL, as described in Table U-6. An implementation can decide to fetch or not the add-on based on this service type indication.

Table U.6 – TEMI service types

| TEMI service type | Add-on type |
| --- | --- |
| 0 | Specified with MimeType |
| 1 | MPEG-DASH |
| 2 | ISO/IEC 14496-12 file |
| 3 | Rec. ITU-T H.222.0 | ISO/IEC 13818-1 Transport Stream |
| 0x04 .. 0x7E | Rec. ITU-T H.222.0 | ISO/IEC 13818-1 Reserved |
| 0x7F | Unknown service type |
| 0x80 .. 0xFF | User Private |

**mime\_type** – indicates the mime type of the add-on available at the indicated location, as defined in RFC 2046. An implementation can decide to fetch or not the add-on based on this mime type indication.

**url\_subpath\_length** – indicates the length in bytes of the URL sub path; when set to 0, indicates an empty URL subpath.

**url\_subpath** – indicates the URL sub path, without trailing zero character; this URL shall be a valid URL, as defined in clause 3 of RFC 3986. The URL for this add-on is obtained by merging url\_subpath with the base URL path, as defined in clause 5 of RFC 3986.

### U.3.4 Base URL descriptor

The base URL descriptor is used to assign a default base URL to all location descriptors (see Table U.7).

Table U.7 – TEMI base URL descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| temi\_base\_url\_descriptor {  **af\_descr\_tag**  **af\_descr\_length**  **url\_scheme**  for (i=0; i < N; i++) {  **base\_url\_path**  }  } | **8**  **8**  **8**  **8** | **uimsbf**  **uimsbf**  **uimsbf**  **bslbf** |

### U.3.5 Semantic definition of fields in Base URL descriptor

**url\_scheme** – indicates the URL scheme to use for the URL. If url\_scheme is not equal to 0, the scheme identified shall be prepended to the base\_url\_path, according to Table U.5.

**base\_url\_path** – base URL common to all following location descriptors, if any; it shall be encoded without trailing zero character. This URL shall be a valid path, as defined in clause 3 of RFC 3986, and may contain a fragment and or a query part.

### U.3.6 Timeline descriptor

The timeline descriptor is used to carry timing information that can be used to synchronize external data (see Table U.8). When the descriptor is carried within a TEMI access unit, the included timing information is given for the PTS value of the TEMI access unit carrying the descriptor. When the descriptor is carried in the adaptation field of a media component, the included timing information is given for the PTS found in the PES header starting in the payload of this transport stream packet or in the first subsequent transport stream packet with payload\_unit\_start\_indicator set to '1' on this component (same PID). This PES header shall have a PTS declared. For a given PES, there shall be at most one temi\_timeline\_descriptor with a timeline\_id in the range [0,0x7F], a paused flag set to '0' and for which the last temi\_location\_descriptor received had an is\_announcement flag set to '0'. A temi\_timeline\_descriptor with a timeline\_id in the range 0 .. 0x7F, for which the last temi\_location\_descriptor received had an is\_announcement flag set to '1', indicates the media time at which the timeline will start upon activation. This Recommendation | International Standard does not define any restrictions on the number of temi\_timeline\_descriptor using timeline\_id values in the range 0x80 .. 0xFF associated with a given PES.

In this section, this media PES packet is called the associated PES packet and the media PTS value is called the associated PTS.

Table U.8 – TEMI timeline descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| temi\_timeline\_descriptor {  **af\_descr\_tag**  **af\_descr\_length**  **has\_timestamp**  **has\_ntp**  **has\_ptp**  **has\_timecode**  **force\_reload**  **paused**  **discontinuity**  **reserved**  **timeline\_id**  if (has\_timestamp != 0) {  **timescale**  if (has\_timestamp = = 1) {  **media\_timestamp**  } else if (has\_timestamp == 2) {  **media\_timestamp**  }  }  if (has\_ntp = = '1') {  **ntp\_timestamp**  }  if (has\_ptp = = '1') {  **ptp\_timestamp**  }  if (has\_timecode != 0) {  **drop**  **frames\_per\_tc\_seconds**  **duration**  if (has\_timecode = = 1) {  **short\_time\_code**  } else if (has\_timecode = = 2) {  **long\_time\_code**  }  }  } | **8**  **8**  **2**  **1**  **1**  **2**  **1**  **1**  **1**  **7**  **8**  **32**  **32**  **64**  **64**  **80**  **1**  **15**  **16**  **24**  **64** | **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **uimsbf**  **bslbf**  **bslbf**  **bslbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf** |

### U.3.7 Semantic definition of fields in the timeline descriptor

**has\_timestamp** – indicates a media timestamp will be carried in this descriptor, and indicates its type. Value 0 means no media timestamp is present, value 1 means a 32 bit media timestamp is present, value 2 means a 64 bit media timestamp is present, value 3 is reserved.

**has\_ntp** – when set to '1', indicates that a NTP timestamp will be carried in this descriptor.

**has\_ptp** – when set to '1', indicates that a PTP timestamp will be carried in this descriptor.

**has\_timecode** – indicates that a frame timecode will be carried in this descriptor, and indicate its type. Value 0 means no frame timecode is present, value 1 means a compact frame timecode is present, value 2 means a full frame timecode is present, value 3 is reserved.

**force\_reload** – when set to '1', indicates that the add-on descriptions received or fetched prior to receiving this descriptor may be obsolete and shall be reloaded before attempting to map media times or locate media components. Reloading typically happens for manifest-based add-on such as MPEG-DASH or MPEG-MMT.

**paused** – when set to '1', indicates that the timeline identified by timeline\_id is currently paused; this typically happens when a timeline has to be paused but no splicing timeline is to be inserted during the pause. When a timeline whose timeline\_id is in the range [0,0x7F] is running, all other timelines defined whose timeline\_id are in the range [0, 0x7F] are implicitly in pause mode. For timelines whose timeline\_id are in the range [0x80, 0xFF], this flag only indicates the running status of the timeline.

**discontinuity** – when set to '1', indicates that a discontinuity has happened in the timeline. If set to '0', no discontinuity happened since the last received temi\_timeline\_descriptor with the same value of timeline\_id and same value of splicing\_flag, if defined.

NOTE – An implementation may use this information to optimize playback of add-on content.

**timeline\_id** – indicates the active timeline. timeline\_id values in the range [0, 0x7F] are identified in a temi\_location\_descriptor; for such values of timeline\_id, the content of this temi\_timeline\_descriptor shall be ignored if no temi\_location\_descriptor with the same timeline\_id has been received. timeline\_id values in the range [0x80, 0xFF] identify timelines defined by means beyond the scope of this Specification.

**timescale** – indicates the timescale used to express the media\_timestamp field in this message; it indicates the amount by which the media time, as indicated by the media\_timestamp field, increases within a 1 second interval when the timeline is not paused.

**media\_timestamp** – indicates the media time in timescale units corresponding to the PES PTS value of this packet for the timeline identified by timeline\_id. The value in this field is the media time modulo 2N where N is the number of bits used to represent this field. The timeline may be interpolated between two temi\_timeline\_descriptor: let PTS0 be the associated PTS of the temi\_timeline\_descriptor carrying media time MT0; until a new temi\_timeline\_descriptor with the same timeline\_id is received, the PTS of subsequent PES packets of other PIDs in this program is mapped to this TEMI timeline as follows:

MTi = timescale \* (PTSi - PTS0) / 90000 + MT0

**ntp\_timestamp** – A time value in 64-bit NTP timestamp format as defined in clause 6 of RFC 5905. The timeline may be interpolated between two temi\_timeline\_descriptor: let PTS0 be the associated PTS of the temi\_timeline\_descriptor carrying NTP timestamp NTP0; until a new temi\_timeline\_descriptor with the same timeline\_id is received, the PTS of subsequent PES packets of other PIDs in this program is mapped to the NTP time NTPi as follows (where NTP0 and NTPi are in units of seconds and fractions of a second):

NTPi = (PTSi - PTS0) / 90000.0 + NTP0

**ptp\_timestamp** – A full 80 bits PTP timestamp as defined in IEEE 1588v2. The timeline may be interpolated between two temi\_timeline\_descriptor: let PTS0 be the associated PTS of the temi\_timeline\_descriptor carrying PTP timestamp PTP0; until a new temi\_timeline\_descriptor with the same timeline\_id is received, the PTS of subsequent PES packets of other PIDs in this program is mapped to the PTP time PTPi as follows (where PTP0 and PTPi are in units of seconds and fractions of a second):

PTPi = (PTSi - PTS0) / 90000.0 + PTP0

**drop** – drop-frame indication, as defined in clause 5 of RFC 5484.

**frames\_per\_tc\_second** – the number of those frames that make a time-code second, as defined in clause 5 of RFC 5484.

**duration** – the duration in ticks of a frame expressed in the timescale of 90000 ticks per seconds, as defined in clause 5 of RFC5 484.

**short\_time\_code** – A compact 24-bit time code as defined clause 6.1 of RFC 5484.

**long\_time\_code** – A full 64-bit time code as defined clause 6.2 of RFC 5484.

**short\_time\_code** and **long\_time\_code** indicate the media time of the first access unit starting in the payload of the associated PES. Using the information of drop, duration and frames\_per\_tc\_seconds, it is possible to interpolate the timing between two temi\_timeline\_descriptor.

### U.3.8 MPEG-H 3dAudio extStreamID descriptor

The MPEG-H\_3dAudio extStreamID descriptor (see Table U.9) is used to provide unique identification of streams. This descriptor shall be directly followed by either a location descriptor or a base URL descriptor which indicates the location of the external auxiliary stream.

The used auxiliaryStreamID in this descriptor value shall match the value of auxiliaryStreamID of the MPEG-H 3D audio multi-stream descriptor.

Table U.9 – TEMI MPEG-H\_3dAudio\_extStreamID descriptor

| Syntax | No. of bits | Mnemonic |
| --- | --- | --- |
| AF\_MPEG-H\_3dAudio\_extStreamID\_descriptor {  **af\_descr\_tag**  **af\_descr\_length**  **reserved**  **auxiliaryStreamID**  **}** | **8**  **8**  **1**  **7** | **uimsbf**  **uimsbf**  **bslbf**  **uimsbf** |

### U.3.9 Semantic definition of fields in AF\_MPEG-H\_3dAudio extStreamID descriptor

**auxiliaryStreamID:** provides a unique identifier for an external data stream that can be synchronized with the program.

### U.3.10 MPEG-H\_3dAudio multi-stream and command descriptors

The following MPEG-H\_3dAudio descriptors may be present in the AF descriptor field:

**AF\_MPEG-H\_3dAudio\_multi-stream\_descriptor()** – The syntax and semantics of this descriptor is identical to the MPEG‑H\_3dAudio\_multi-stream\_descriptor() as defined in 2.6.114 and 2.6.115.

**AF\_MPEG-H\_3dAudio\_command\_descriptor()** – The syntax and semantics of this descriptor is identical to the MPEG‑H\_3dAudio\_command\_descriptor() as defined in 2.6.118.

Note – The descriptors MPEG-H\_3dAudio\_multi-stream\_descriptor() as defined in 2.6.114 and MPEG‑H\_3dAudio\_command\_descriptor() as defined in 2.6.118 are optionally part of the PMT. In case of dynamic updates of those descriptors an update of the PMT would be required. In case PMT updates are not desirable, the corresponding AF\_descriptors AF\_MPEG-H\_3dAudio\_multi-stream\_descriptor() respectively AF\_MPEG-H\_3dAudio\_command\_descriptor() can be distributed in the AF descriptor field.

### U.3.11 Boundary descriptor

The boundary descriptor (see Table U.10) is used to indicate segment boundary information that can be used to support content partitioning requirements in different types of applications. When the descriptor is carried explicitly within an elementary stream, it indicates the frame accurate location of that boundary. Information carried in the boundary descriptor describes the boundary type, e.g., SAP type, partition and an optional sequence number of the partitioned virtual segment. When it is used together with the TEMI time descriptor, it may also be used to indicate a wall clock timestamp on the boundary point.

Partitions that are indicated by the boundary descriptor are a set of continuous segments divided by boundary descriptors within a media stream. A stream can be partitioned in several ways. For example, as shown in Figure U.1, Partition A corresponds to 2-second segment, while Partition B corresponds to 5-second segment, both share a common boundary point every 10 seconds.

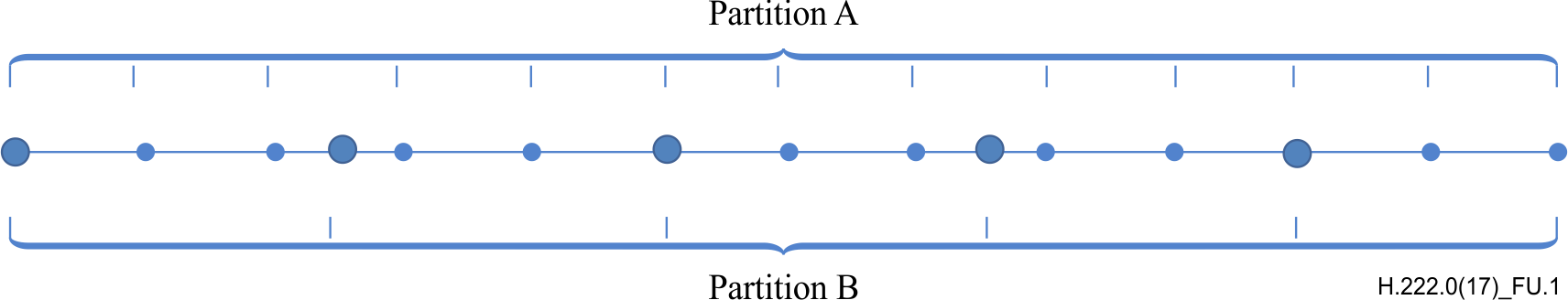


Figure U.1 – Stream partitioning into 2 and 5 second segments

Table U.10 – Boundary descriptor

| Syntax | No. of bits | Format |
| --- | --- | --- |
| boundary\_descriptor() { |  |  |
| **af\_descr\_tag** | **8** | **uimbsf** |
| **af\_descr\_length**  **num\_partitions\_minus\_1**  **SAP**  **concealment\_flag**  **reserved**  for ( i=0; i <= num\_partitions\_minus\_1; i++) {  **partition\_id**  **partition\_info\_flag**  **reserved**    if ( partition\_info\_flag == 1 ) {  **sequence\_number\_length\_code (SNL)**  **reserved**  **sequence\_number**  }  }  } | **8**  **3**  **3**  **1**  **1**  **3**  **1**  **4**  **2**  **6**  **see SNL** | **uimbsf**  **uimbsf**  **uimbsf**  **bslbf**  **"0"**  **uimbsf**  **bslbf**  **bslbf**  **bslbf**  **uimbsf**  **uismbf** |

### U.3.12 Semantic definition of fields in boundary descriptor

**SAP** – SAP type, as defined ISO/IEC 14496-12.

**num\_partitions\_minus\_1** – Number of partitions to which this boundary belongs. By default (i.e., if this field is zero), each descriptor corresponds to a single partition.

**partition\_id** – partition identifier

**partition\_info\_flag** – if set to 1and SNL is not equal to zero, thenextended partition information is available. If partition\_info\_flag is set to 1 and SNL is equal to zero, then this is reserved for future extension.

**concealment\_flag** – Indicates if the source stream may have lost a frame that was at a boundary point. If flag is set to 1, indicates this is a repair boundary, inserted due to a previous lost boundary point frame. If set to 0, indicates no repair was done to the stream."Repair" is meant to create a boundary point where it may have been skipped due to a lost boundary point frame.

**sequence\_number\_length\_code** – Specifies length of the sequence\_number field. The semantics for the width values are given in Table U.11.

Table U.11 – sequence\_number\_length\_code interpretation

|  |  |
| --- | --- |
| sequence\_number\_length\_code | length, bits |
| 0 | 0 |
| 1 | 16 |
| 2 | 32 |
| 3 | 64 |

**sequence\_number** – Field providing a unique identifier for a segment within the context of this partition and multiplex. Length of this field is defined in sequence\_number\_length\_code. If the sequence\_number\_length\_code is 0, sequence\_number field does not appear.

### U.3.13 Labelling Descriptor

The Labelling Descriptor (see Table U.12) is used to carry one or more labels. When the descriptor is carried explicitly within an elementary system, it indicates a frame accurate location for the label or set of labels as a notation to the virtual segment. When used together with a boundary descriptor, it can note the start of an ad, program, chapter, or multi-period asset and carry identifiers (e.g., EIDR, or user defined) in the stream.

Table U.12 – Labelling Descriptor

| Syntax | No. of bits | Format |
| --- | --- | --- |
| labeling\_descriptor() { |  |  |
| **af\_descr\_tag** | **8** | **uimbsf** |
| **af\_descr\_length**  l = 0  while (l < af\_descr\_length) {  **label\_length\_code**  **label\_type**  l += 2  if ( label\_length\_code == 7 ) {  **label\_length**  l++  }  **label\_bytes**  l += N  }  } | **8**  **3**  **13**  **8**  **N\*8** | **uimbsf**  **uimsbf**  **uimsbf**  **uimbsf**  **bslbf** |

NOTE – In Table U.12, the value of N is derived from label\_length\_code and label\_length (if present), as described in U.3.14. For example, N would be 20 when label\_length\_code value is 6, and 7 when label\_length\_code and label\_length both have value of 7.

### U.3.14 Semantic definition of fields in labelling descriptor

**label\_length\_code** – Length of the label field in bytes, as provided in Table U.13.

Table U.13 – label\_length\_code interpretation

|  |  |
| --- | --- |
| Value | Length, bytes |
| 0 | 0 (not present) |
| 1 | 2 |
| 2 | 4 |
| 3 | 8 |
| 4 | 12 |
| 5 | 16 |
| 6 | 20 |
| 7 | Explicit value provided in ***label\_length*** |

**label\_type** – Type of the label field, as provided in Table U.14.

Table U.14 – label\_type values

|  |  |
| --- | --- |
| Range | Definition |
| 0x00 .. 0xFF | Reserved for MPEG standardization |
| 0x100 .. 0x1FF | Value defined in ANSI/SCTE 35 Table 8-8 (segmentation\_type\_id) + 0x100 |
| 0x200 .. 0x2FF | Value defined in ANSI/SCTE 35 Table 8-7 (upid\_type\_id) + 0x200 |
| 0x300 .. 0xFFF | Reserved for MPEG standardization |
| 0x1000 .. 0x1FFF | User Private types |

**label\_length** – Length, in bytes, of the label field (i.e., number of label\_bytes).

**label\_bytes** – Bytes carrying the label. Length N is defined by label\_length\_code or label\_length. Interpretation of these bytes depends on label types.

NOTE – If label\_length\_code is zero, this allows for label\_type to exist with no label\_bytes. This can serve as a marker type to the indicated PES frame.

### U.3.15 HEVC tile substream af\_descriptor

The HEVC tile substream af\_descriptor may be present in the AF descriptor field, as defined in 2.4.3.5. The syntax of the HEVC tile substream af\_descriptor is given in Table U.15.

Table U.15 – HEVC tile substream af\_descriptor

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| HEVC\_tile\_substream\_ af\_descriptor() {  **af\_descr\_tag**  **af\_descr\_length**  HEVC\_tile\_substream\_descriptor()  } | **8**  **8** | **uimsbf**  **uimsbf** |

The structure HEVC\_tile\_substream\_descriptor() is specified in Table 2-129, semantics of its fields are defined in 2.6.123.

Annex V  
  
Transport of HEVC tiles

(This annex does not form an integral part of this Recommendation | International Standard.)

## V.1 Introduction

Region of Interest (RoI) streaming aka. Interactive Panorama Streaming allows navigating within a very wide angle and high-resolution video, showing at the receiver side only a part of the whole video, i.e. the RoI. In particular, the panorama video may be encoded at a very high resolution and should not be sent as a whole to a receiver since it cannot be decoded by deployed hardware.

HEVC bitstreams can be generated using the “tile” concept, where the whole video is structured in a rectangular pattern of N x M tiles. In this case, the encoder can be configured in a way that in-picture prediction dependencies (including entropy decoding dependencies) are broken on tile borders such that no information is shared among different tiles. The data generated by an encoder for each such tile can then be handled separately, e.g. can be processed by one processor/core.

For certain use cases, like the presentation of a smaller window (aka. RoI) taken from a large panorama, only a subset of the tiles needs to be decoded as shown in Figure V.1. However, the HEVC bitstream has to be encoded in such a way that inter-prediction is constrained in such a way that tiles of a picture are not predicted from different tiles of previous pictures. In the following text, a part of the bitstream that allows decoding a tile or a subset of tiles is called HEVC tile substream. It would be beneficial to split the bitstream into separate HEVC tile substreams of which the client can select a subset needed to decode the RoI. In this contribution, such set of HEVC tile substreams is called a subregion.

Different subregion sizes can be used, i.e., a set of subregions can consist of a certain number of rows and columns, while another set of subregions can consist of a different number of rows and columns. The following text refers to them as different subregion layouts. In the example shown in Figure V.1, it is shown how for the selected RoI (marked with the red rectangle in the figure), a tile subset is extracted (Tile Sets in terms of the HEVC specification) consisting of three by three tiles.

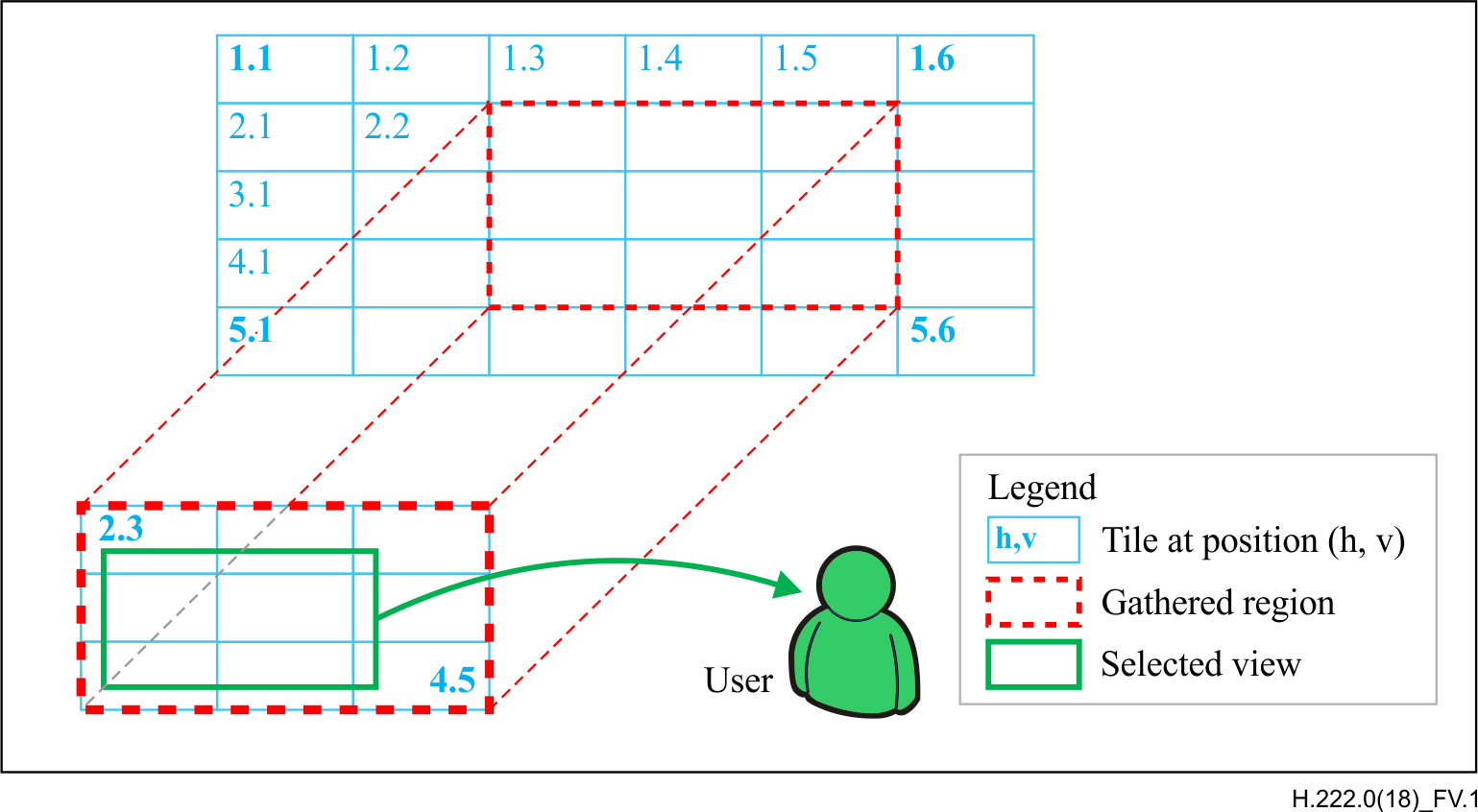


Figure V.1 – Illustration of HEVC tiled encoding of panoramic content beyond UHD

## V.2 HEVC tile substream identification example

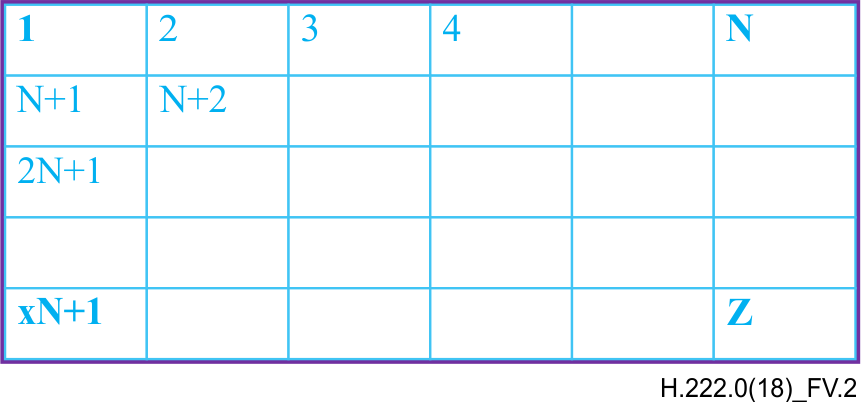


Figure V.2 – Example of HEVC tile substream identification

For the example shown in Figure V.2, the number of columns N is indicated as follows:

***N*** is signalled in the HEVC subregion descriptor in the field *SubstreamIDsPerLine*.

The number of lines can be calculated by Z / N, where Z is indicated as follows:

***Z*** is signalled in the HEVC subregion descriptor in the field *TotalSubstreamIDs*.

## V.3 Subregion layout example

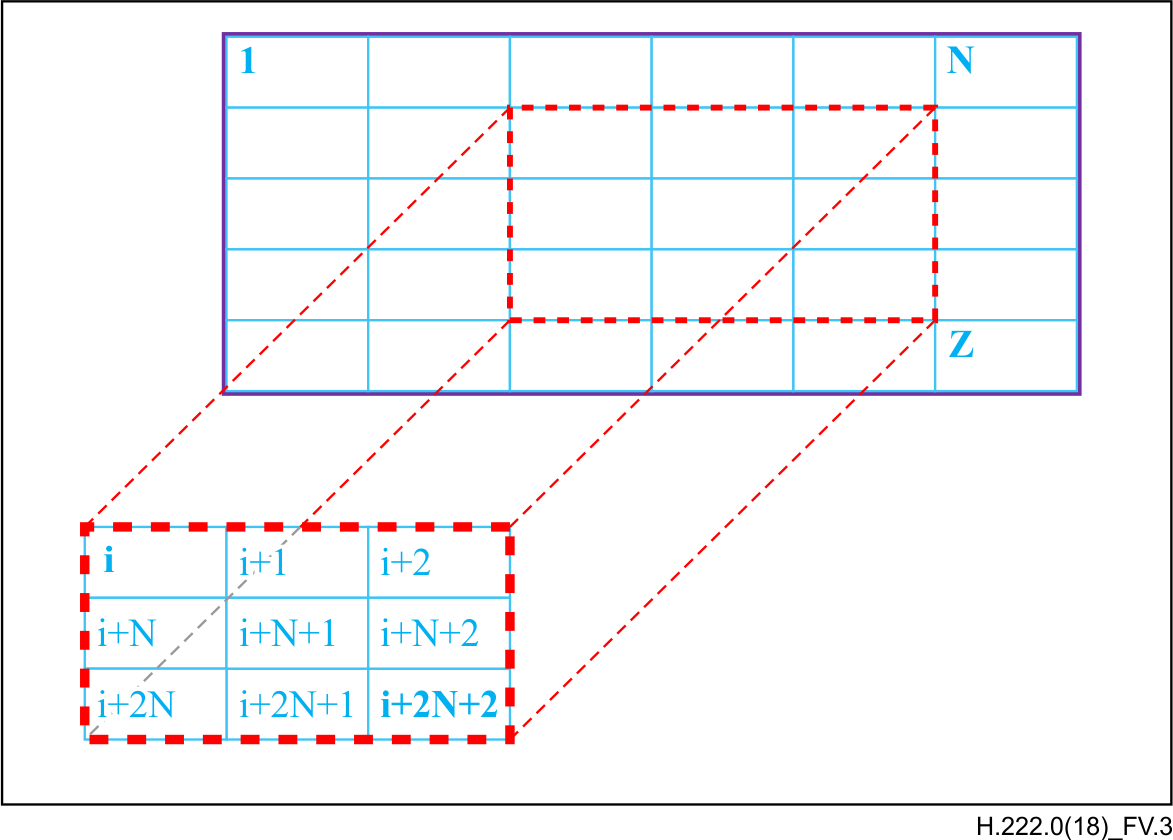


Figure V.3 – Example of subregion layout for a 3 x 3 RoI

For the example shown in Figure V.3, an offset pattern for a subregion of 3x3 HEVC tile substreams can be indicated by the array:

|  |  |
| --- | --- |
| SubstreamOffset[0][0][i]: | 1 |
| SubstreamOffset[1][0][i]: | 2 |
| SubstreamOffset[2][0][i]: | N |
| SubstreamOffset[3][0][i]: | N+1 |
| SubstreamOffset[4][0][i]: | N+2 |
| SubstreamOffset[5][0][i]: | 2N |
| SubstreamOffset[6][0][i]: | 2N+1 |
| SubstreamOffset[7][0][i]: | 2N+2 |

An offset pattern for a subregion of 2x2 HEVC tile substreams would be indicated by the array:

|  |  |
| --- | --- |
| SubstreamOffset[0][0][i]: | 1 |
| SubstreamOffset[1][0][i]: | N |
| SubstreamOffset[2][0][i]: | N+1 |

More than one pattern might be necessary, e.g. if a 360 cylinder is considered: at the right edge of the picture, a region could be taken that “wraps-around” to the left edge of the picture. In such a case, for a 3x3 subregion having the rightmost tile as the top-left tile:

|  |  |
| --- | --- |
| SubstreamOffset[0][0][i]: | -N+1 |
| SubstreamOffset[1][0][i]: | -N+2 |
| SubstreamOffset[2][0][i]: | N |
| SubstreamOffset[3][0][i]: | 1 |
| SubstreamOffset[4][0][i]: | 2 |
| SubstreamOffset[5][0][i]: | 2N |
| SubstreamOffset[6][0][i]: | N+1 |
| SubstreamOffset[7][0][i]: | N+2 |

Annex W  
  
Carriage of JPEG XS part 1 video over MPEG-2 Transport Streams

(This annex forms an integral part of this Recommendation | International Standard.)

## W.1 Introduction

This annex specifies normative constraints for the carriage of JPEG XS video in an MPEG-2 transport stream. The parameters specified include mapping of JPEG XS video streams into MPEG-2 transport packets, signalling of JPEG XS video streams as well as T-STD parameters depending on selected profile, level and sublevel. Transport of JPEG XS video shall be limited to transport stream only. Program stream support may be added in the future based on application requirements.

## W.2 JPEG XS video access unit, JPEG XS video elementary stream, JPEG XS video sequence and JPEG XS still picture

The JPEG XS video access unit contains the elementary stream (jxes) header created as defined in W.3 concatenated with self-contained ISO/IEC 21122-1 codestream(s). The (jxes) header contains all video-related parameters necessary to display the decoded codestream(s). An access unit shall contain exactly one codestream for the progressive frame mode, or shall contain two codestreams for the interlaced frame mode (one codestream per interlace field).

The JPEG XS video elementary stream is a progression of JPEG XS access units and the JPEG XS video sequence is a subset of JPEG XS video elementary stream where all the JPEG XS access units have the same parameters in the (jxes) header.

The JPEG XS still picture (system) consists of a JPEG XS video sequence which contains exactly one JPEG XS access unit. This still picture has an associated PTS, and the presentation time of succeeding pictures, if any, is later than that of the still picture by at least one picture period for progressive frame mode, but at least two picture periods for interlaced frame mode. The JPEG XS still picture (system) mode is used to support transmission of JPEG XS video access units at a rate much lower than the display frame rate (determined by the difference in PTS values between successive JPEG XS access units). JPEG XS still picture can be used in applications such as 'slide show' and 'stills with Music'.

## W.3 Elementary stream header (jxes) and mapping to PES packets

Each JPEG XS access unit from a JPEG XS video elementary stream shall start with an elementary stream header (jxes header) as detailed in Table W.1.

Table W.1 – JPEG XS Access unit elementary stream header (jxes header)

|  |  |  |
| --- | --- | --- |
| Syntax | No. of bits | Mnemonic |
| jxes\_header() {  **jxes\_length**  **jxes\_box\_code** '0x6a786573'  **brat**  **frat**  **schar**  **Ppih**  **Plev**  **color\_primaries**  **transfer\_characteristics**  **matrix\_coefficients**  **video\_full\_range\_flag**  **reserved**  **tcod**  } | **32**  **32**  **32**  **32**  **16**  **16**  **16**  **8**  **8**  **8**  **1**  **7**  **32** | **uimsbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **bslbf**  **bslbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf**  **uimsbf** |

**Semantics**

Except for the **tcod** field, all fields from the jxes header are also found in the JXS video descriptor defined in 2.6.127 and 2.6.128. This is done in purpose to allow for a greater flexibility of the implementations, at a cost of a negligible overhead. Values found in the JXS video descriptor and in the jxes headers shall be consistent. If inconsistent values are found, values from the jxes header shall take precedence.

**jxes\_length –** This field is used to indicate the length (in bytes) of the jxes\_header, which is 30 as fixed value.

**brat** – The brat parameter specifies the maximum bitrate of the elementary stream in Mbit per second. A detailed definition can be found in ISO/IEC 21122-3.

**frat** – The frat parameter specifies the frame rate of the elementary stream and also contains a flag indicating if the stream is interlaced or progressive. A detailed definition can be found in ISO/IEC 21122-3.

**schar** – The schar parameter specifies the image sample characteristics and the sampling structure. A detailed definition can be found in ISO/IEC 21122-3.

**Ppih** – This field specifies the Profile of the elementary stream. It shall be coded the same as the Ppih parameter defined in ISO/IEC 21122-1 using the values defined in ISO/IEC 21122-2.

**Plev** – This field specifies the Level and Sublevel of the elementary stream. It shall be coded the same as the Plev parameter defined in ISO/IEC 21122-1 using the values defined in ISO/IEC 21122-2.

**color\_primaries, transfer\_characteristics, matrix\_coefficients, video\_full\_range\_flag** – These four fields (three 1‑byte integers and one 1-bit flag) shall be coded according to the semantics with the same name defined in ISO/IEC 23091-2.

**tcod** – The tcod parameter specifies the timecode of an access unit. A detailed definition can be found in ISO/IEC 21122‑3.

Figure W.1 shows the structure and mapping of JPEG XS video access unit into PES packets. JPEG XS represents each frame as one or two JPEG XS codestreams (depending if the stream is interlaced or not), as defined in ISO/IEC 21122-1.The Codestream\_Header(), included within each codestream, contains all information to decode its image, including the image size and the profile, level and sublevel indicators, as defined in ISO/IEC 21122-1 and ISO/IEC 21122-2. Preceding the codestreams, a JPEG XS elementary stream (jxes) header containing video-related information, as shown in Table W.1, shall be present. Each PES packet shall contain one single JPEG XS video access unit.

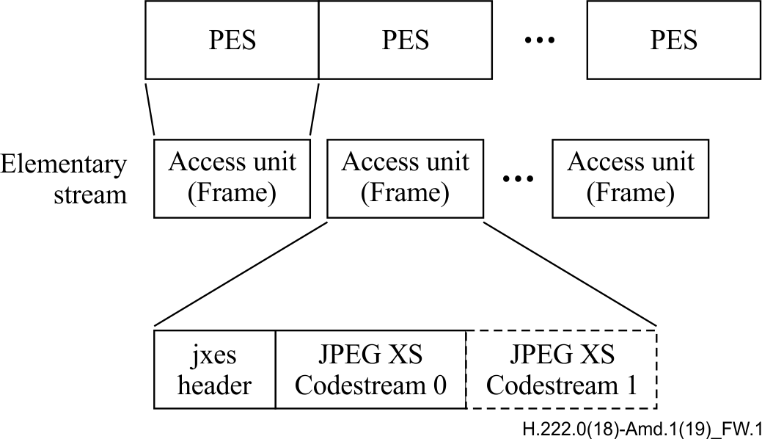


Figure W.1 – Structure and order of JPEG XS access units

## W.4 JPEG XS transport constraints

When a JPEG XS video elementary stream conforming to one or more profiles as defined in ISO/IEC 21122-1 and ISO/IEC 21122-2 is transported using MPEG-2 systems, the following constraints apply:

1. Each JPEG XS video access unit shall contain an elementary stream header (jxes header) defined in Table W.1 followed by one or two ISO/IEC 21122-1 codestream(s).

2. Each JPEG XS codestream shall contain a picture header with Ppih and Plev fields equating to corresponding values found in the JXS video descriptor detailed in 2.6.127 and in the JPEG XS elementary stream (jxes) header detailed in W.3.

3. The JPEG XS video access units shall be ordered in the JPEG XS video elementary stream in a monotonic display order.

4. Each PES packet shall contain exactly one JPEG XS video access unit.

5. Each PES packet shall include a PES header with PTS.

6. For successive JPEG XS video access units, the increments to PTS shall be consistent with increments to corresponding tcod parameters in the jxes header.

7. The following constraints apply to the coding of syntax elements in the adaptation header for transport of JPEG XS video elementary stream:

a. Both random\_access\_indicator and elementary\_stream\_priority\_indicator flags may be set to '1' for each JPEG XS video access unit contained in the transport packet. Applications may limit the signalling of random access based on their use cases. Any JPEG XS video access unit is also a JPEG XS video 'access point' required for random access.

b. All other flags should be set appropriately.

8. The following constraints apply to the coding of syntax elements in the PES header for transport of JPEG XS video elementary stream:

a. stream\_id shall be set to 0xBD (same as Private\_stream\_1).

b. PES\_packet\_length shall be set to 0x0000.

c. data\_alignment\_indicator shall be set to '1'. Also see W.5 for alignment constraints.

d. PTS\_DTS\_flags shall be set to '10'. This precludes display re-ordering for JPEG XS video access units.

e. All other flags should be set appropriately.

NOTE – As there is no emulation prevention byte in JPEG XS elementary stream, there can be misdetection of PES start code. To avoid this, and as the PES\_packet\_length is set to 'undefined' (0x0000), the start of a PES packet can be signalled using the Payload Unit Start Indicator in the transport packet header.

9. The number of horizontal and vertical blanking periods **Nb,x** and**Nb,y**, as defined in ISO/IEC 21122-2, shall be zero.

## W.5 Interpretation of flags in adaptation field and PES packet for JPEG XS video elementary streams

The interpretation, extensions, use and constraints for the following syntax elements in the adaptation field (2.4.3.4 and 2.4.3.5) and PES packet (2.4.3.6 and 2.4.3.7) for JPEG XS Part 1 video are defined in this clause:

– In the semantics for discontinuity\_indicator (see 2.4.3.5), a JPEG XS video elementary stream access point is defined as the first byte of the JPEG XS video access unit.

– The elementary\_stream\_priority\_indicator in adaptation field (see 2.4.3.4 and 2.4.3.5) may be set to '1' if the transport packet payload contains the start of a JPEG XS video access unit.

– For JPEG XS video elementary streams, when the data\_alignment\_indicator (see 2.4.3.6) is set to '1', the PES packet header shall be immediately followed by the first byte of the JPEG XS video access unit. The data\_stream\_alignment\_descriptor is optional and, if included for JPEG XS video, the alignment\_type shall be set to 0x00.

For JPEG XS video elementary streams conforming to one or more profiles defined in ISO/IEC 21122-2, if a PTS is present in the PES packet header (see 2.4.3.7), it shall refer to the first byte of the JPEG XS video access unit that commences in this PES packet.

## W.6 T-STD extension for JPEG XS video elementary streams

### W.6.1 General

The T-STD model includes a transport buffer TBn and a JPEG XS video elementary stream buffer EBn for decoding of each JPEG XS video elementary stream n. See Figure W.2.

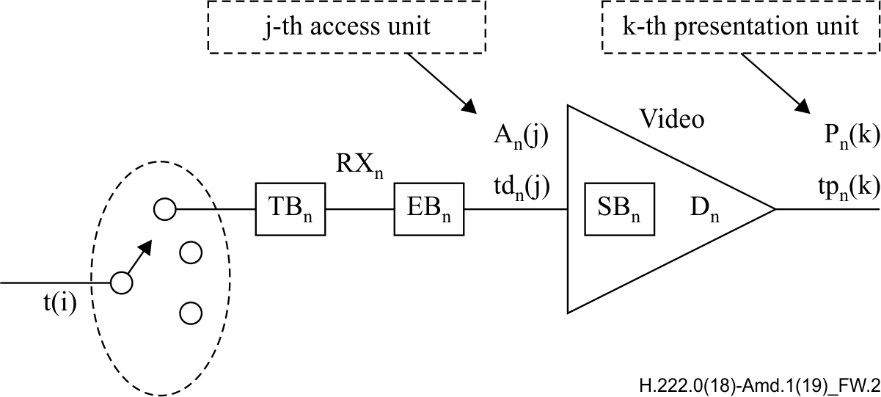


Figure W.2 – T-STD model extensions for JPEG XS Video

The following notation is used to describe the transport stream system target decoder and is partially illustrated in Figure W.2.

i index to bytes in the transport stream. The first byte has index 0.

j is an index to access units in the elementary streams.

k index to presentation units in the elementary streams.

n is an index to the elementary streams.

p is an index to transport stream packets in the transport stream.

t(i) indicates the time in seconds at which the i-th byte of the transport stream enters the system target decoder. The value t(0) is an arbitrary constant.

PCR(i) is the time encoded in the PCR field measured in units of the period of the 27 MHz system clock where i is the byte index of the final byte of the program\_clock\_reference\_base field.

An(j) is the j-th access unit in elementary stream n. An(j) is indexed in decoding order.

tdn(j) is the decoding time, measured in seconds, in the system target decoder of the j-th access unit in elementary stream n.

Pn(k) is the k-th presentation unit in elementary stream n. Pn(k) results from decoding An(j). Pn(k) is indexed in presentation order.

tpn(k) is the presentation time, measured in seconds, in the system target decoder of the k-th presentation unit in elementary stream n.

t is time measured in seconds.

Fn(t) is the fullness, measured in bytes, of the system target decoder input buffer for elementary stream n at time t.

EBn is the elementary stream buffer for elementary stream n. It is present only for video elementary streams.

EBSn is the size of the elementary stream buffer EBn, measured in bytes.

TBn is the transport buffer for elementary stream n.

TBSn is the size of TBn, measured in bytes.

Rxn is the rate at which data are removed from TBn.

Dn is the decoder unit for elementary stream n.

SBn is the smoothing buffer within decoder Dn for elementary stream n.

TBn and EBn buffer management

The input to buffer TBn and its size TBSn are specified in 2.4.2.4. The rate Rxn between TBn and buffer EBn and the following constraints apply for the carriage of a JPEG XS video elementary stream:

Size EBSn of buffer EBn: maximum buffer size based on level and sublevel of JPEG XS video stream (see max\_buffer\_size field in JXS\_video descriptor, 2.6.128).

Rate Rxn:

o when there is no data in TBn then Rxn is equal to zero,

o otherwise: Rxn = bit\_rate,

where bit\_rate is the bit rate whose maximum value is the "maximum encoded rate" specified in ISO/IEC 21122-2 for each combination of level and sublevel. All JPEG XS video payload data bytes enter EBn instantaneously upon leaving TBn.

Removal of JPEG XS access units from EBn: each JPEG XS access unit An(j) that is present in EBn is removed instantaneously at time tdn(j). The decoding time tdn(j) is specified by the PTS (as PTS = DTS for JPEG XS video elementary streams) in the PES header for the JPEG XS video access unit.

STD delay

The total delay of any JPEG XS video elementary stream data other than JPEG XS still picture data through the system target decoders buffers TBn and EBn shall be constrained by tdn(j) – t(i) ≤ 1 second for all j, and all bytes i in JPEG XS video access unit An(j).

The delay of any JPEG XS still picture data through the system target decoders buffers TBn and EBn shall be constrained by tdn(j) – t(i) ≤ 60 seconds for all j, and all bytes i in JPEG XS video access unit An(j).

Buffer management conditions

Transport streams shall be constructed so that the following conditions for buffer management are satisfied:

• TBn shall not overflow and shall be empty at least once every second.

• EBn shall not overflow or underflow.

### W.6.2 JPEG XS video elementary stream buffer size

A JPEG XS video elementary stream supplies parameters suitable for determining a standard decoder model buffer size. This CODEC uses variable rate compression. Levels and sublevels from ISO/IEC 21122-2 specify operating levels that limit the maximum compressed bit rate of the ISO/IEC 21122-1 codestream. Moreover, within access units an elementary stream header contains the present maximum bit rate an encoder must support to decode video without overflow. A decoder may scale the buffer size by reading the maximum bit rate expected by a particular sequence of frames from the jxes header. This header information shall not exceed the limit set by the operating level specified in ISO/IEC 21122-2.

Bibliography

– ISO 639-3:2007, *Codes for the representation of names of languages – Part 3: Alpha-3 code for comprehensive coverage of languages*.

– Supplement 19 to ITU-T H-series Recommendations (2021), *Usage of video signal type code points*.

– Recommendation ITU-R BT.601-7 (2011), *Studio encoding parameters of digital television for standard 4:3 and wide-screen 16.9 aspect ratios*.

– ISO/IEC 23009-1:2019 – *Information technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats*.

– ISO/IEC TR 23091-4:2021 – *Information technology* – *Coding-independent code points* – *Part 4: Usage of video signal type code points*.

– ST 2067-2:2020, *SMPTE Standard – Interoperable Master Format – Core Constraints*.

|  |  |
| --- | --- |
| **SERIES OF ITU-T RECOMMENDATIONS** | |
| Series A | Organization of the work of ITU-T |
| Series D | Tariff and accounting principles and international telecommunication/ICT economic and policy issues |
| Series E | Overall network operation, telephone service, service operation and human factors |
| Series F | Non-telephone telecommunication services |
| Series G | Transmission systems and media, digital systems and networks |
| **Series H** | **Audiovisual and multimedia systems** |
| Series I | Integrated services digital network |
| Series J | Cable networks and transmission of television, sound programme and other multimedia signals |
| Series K | Protection against interference |
| Series L | Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant |
| Series M | Telecommunication management, including TMN and network maintenance |
| Series N | Maintenance: international sound programme and television transmission circuits |
| Series O | Specifications of measuring equipment |
| Series P | Telephone transmission quality, telephone installations, local line networks |
| Series Q | Switching and signalling, and associated measurements and tests |
| Series R | Telegraph transmission |
| Series S | Telegraph services terminal equipment |
| Series T | Terminals for telematic services |
| Series U | Telegraph switching |
| Series V | Data communication over the telephone network |
| Series X | Data networks, open system communications and security |
| Series Y | Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities |
| Series Z | Languages and general software aspects for telecommunication systems |
|  |  |

1. \* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, <http://handle.itu.int/11.1002/1000/11830-en>. [↑](#footnote-ref-2)
2. 1) Constant delay as indicated for the entire system is required for correct synchronization, however some deviations are possible. Network delay is discussed as being constant. Slight deviations may be tolerated, and network adaptation may allow greater variations of network delay. Both of these are discussed later. [↑](#footnote-ref-3)