International Standard

ISO/IEC 13818-1

Information technology — Generic coding of moving pictures and associated audio information —

Part 1: Systems

Technologies de l'information — Codage générique des images animées et du son associé —

Partie 1: Systèmes

Tenth edition
2025-08

Reference number
ISO/IEC 13818-1:2025(en)

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted.

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by ITU-T (as ITU-T H.222.0) and drafted in accordance with its editorial rules, in collaboration with Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This tenth edition cancels and replaces the ninth edition (ISO/IEC 13818-1:2023), which has been technically revised.

The main changes are as follows:

— the usage of certain codec parameters and the definition of certain program element descriptors is clarified;

— in Annex T.4 the interpretation of the optional 'codecs' parameter for the audio layer is clarified;

— a field name is adjusted in the syntax and semantics of the MPEG-H 3D audio descriptor;

— in 2.6.110 the syntax of the MPEG-H 3D audio scene descriptor is corrected and field names are aligned with the names in the specification to which they refer;

— in 2.6.115 some clarification is added for the MPEG-H 3D audio multi-stream descriptor with respect to the semantics of the fields *thisStreamID* and auxiliaryStreamID.

A list of all parts in the ISO/IEC 13818 series can be found on the ISO and IEC websites.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

INTERNATIONAL STANDARD ISO/IEC 13818-1
RECOMMENDATION ITU-T H.222.0 v10

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.

This tenth edition was initially developed as Amd. 1 to the ninth edition and was eventually approved as a new edition.

The main changes are as follows:

– The usage of certain codec parameters and the definition of certain program element descriptors is clarified;
– in clause T.4 the interpretation of the optional 'codecs' parameter for the audio layer is clarified;
– a field name is adjusted in the syntax and semantics of the MPEG-H 3D audio descriptor;
– in 2.6.110 the syntax of the MPEG-H 3D audio scene descriptor is corrected and field names are aligned with the names in the specification to which they refer;
– in 2.6.115 some clarification is added for the MPEG-H 3D audio multi-stream descriptor with respect to the semantics of the fields thisStreamID and auxiliaryStreamID;

History *

Edition	Recommendation	Approval	Study Group	Unique ID
1.0	ITU-T H.222.0	1995-07-10	15	11.1002/1000/1071
1.1	ITU-T H.222.0 (1995) Amd. 1	1996-11-11	16	11.1002/1000/3834
1.2	ITU-T H.222.0 (1995) Amd. 2	1996-11-11	16	11.1002/1000/4096
1.3	ITU-T H.222.0 (1995) Technical Cor. 1	1998-02-06	16	11.1002/1000/4532
1.4	ITU-T H.222.0 (1995) Amd. 3	1998-02-06	16	11.1002/1000/4228
1.5	ITU-T H.222.0 (1995) Amd. 4	1998-02-06	16	11.1002/1000/4229
1.6	ITU-T H.222.0 (1995) Amd. 5	1999-05-27	16	11.1002/1000/4498
1.7	ITU-T H.222.0 (1995) Amd. 6	1999-05-27	16	11.1002/1000/4671
1.8	ITU-T H.222.0 (1995) Amd. 7	2000-02-17	16	11.1002/1000/4926
2.0	ITU-T H.222.0	2000-02-17	16	11.1002/1000/5142
2.1	ITU-T H.222.0 (2000) Technical Cor. 1	2001-03-01	16	11.1002/1000/5419
2.2	ITU-T H.222.0 (2000) Technical Cor. 2	2002-03-29	16	11.1002/1000/5675
2.3	ITU-T H.222.0 (2000) Amd. 1	2002-12-14	16	11.1002/1000/6190
2.4	ITU-T H.222.0 (2000) Amd. 1 Cor. 1	2003-06-29	16	11.1002/1000/6449
2.5	ITU-T H.222.0 (2000) Amd. 2	2003-06-29	16	11.1002/1000/6363
2.6	ITU-T H.222.0 (2000) Amd. 3	2004-03-15	16	11.1002/1000/7208
2.7	ITU-T H.222.0 (2000) Technical Cor. 3	2005-01-08	16	11.1002/1000/7435
2.8	ITU-T H.222.0 (2000) Amd. 4	2005-01-08	16	11.1002/1000/7436
2.9	ITU-T H.222.0 (2000) Amd. 5	2005-01-08	16	11.1002/1000/7437
2.10	ITU-T H.222.0 (2000) Technical Cor. 4	2005-09-13	16	11.1002/1000/8560
3.0	ITU-T H.222.0	2006-05-29	16	11.1002/1000/8802
3.1	ITU-T H.222.0 (2006) Amd. 1	2007-01-13	16	11.1002/1000/9024
3.2	ITU-T H.222.0 (2006) Amd. 2	2007-08-29	16	11.1002/1000/9214
3.3	ITU-T H.222.0 (2006) Cor. 1	2008-06-13	16	11.1002/1000/9471
3.4	ITU-T H.222.0 (2006) Cor. 2	2009-03-16	16	11.1002/1000/9692
3.5	ITU-T H.222.0 (2006) Amd. 3	2009-03-16	16	11.1002/1000/9691
3.6	ITU-T H.222.0 (2006) Cor. 3	2009-12-14	16	11.1002/1000/10621
3.7	ITU-T H.222.0 (2006) Cor. 4	2009-12-14	16	11.1002/1000/10622
3.8	ITU-T H.222.0 (2006) Amd. 4	2009-12-14	16	11.1002/1000/10623
3.9	ITU-T H.222.0 (2006) Amd. 5	2011-05-14	16	11.1002/1000/11287
3.10	ITU-T H.222.0 (2006) Amd. 6	2011-05-14	16	11.1002/1000/11288
4.0	ITU-T H.222.0	2012-06-29	16	11.1002/1000/11655
4.1	ITU-T H.222.0 (2012) Amd. 1	2014-01-13	16	11.1002/1000/12054
4.2	ITU-T H.222.0 (2012) Amd. 2	2014-01-13	16	11.1002/1000/12055
4.3	ITU-T H.222.0 (2012) Amd. 3	2014-01-13	16	11.1002/1000/12056
4.4	ITU-T H.222.0 (2012) Amd. 4	2014-01-13	16	11.1002/1000/12057
4.5	ITU-T H.222.0 (2012) Amd. 5	2014-10-14	16	11.1002/1000/12306
5.0	ITU-T H.222.0	2014-10-14	16	11.1002/1000/12359
5.1	ITU-T H.222.0 (2014) Amd. 1	2015-04-29	16	11.1002/1000/12452
5.2	ITU-T H.222.0 (2014) Amd. 1 Cor. 1	2015-11-29	16	11.1002/1000/12625
5.3	ITU-T H.222.0 (2014) Amd. 2	2015-12-14	16	11.1002/1000/12632
5.4	ITU-T H.222.0 (2014) Amd. 3	2015-12-14	16	11.1002/1000/12633
5.5	ITU-T H.222.0 (2014) Amd. 1 Cor. 2	2016-07-14	16	11.1002/1000/12899
5.5	ITU-T H.222.0 (2014) Cor. 1	2016-07-14	16	11.1002/1000/12903
5.7	ITU-T H.222.0 (2014) Amd. 4	2016-07-14	16	11.1002/1000/12900
5.8	ITU-T H.222.0 (2014) Amd. 5	2016-07-14	16	11.1002/1000/12901
5.9	ITU-T H.222.0 (2014) Amd. 6	2016-07-14	16	11.1002/1000/12902

* To access the Recommendation, type the URL https://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID.

5.10	ITU-T H.222.0 (2014) Amd. 3 Cor. 1	2017-03-01	16	11.1002/1000/13184
5.10	ITU-T H.222.0 (2014) Cor. 2	2017-03-01	16	11.1002/1000/13188
5.12	ITU-T H.222.0 (2014) Amd. 7	2017-03-01	16	11.1002/1000/13186
5.13	ITU-T H.222.0 (2014) Amd. 8	2017-03-01	16	11.1002/1000/13187
6.0	ITU-T H.222.0	2017-03-01	16	11.1002/1000/13269
6.1	ITU-T H.222.0 (2017) Amd. 1	2017-12-14	16	11.1002/1000/13431
7.0	ITU-T H.222.0	2018-08-29	16	11.1002/1000/13664
7.1	ITU-T H.222.0 (2018) Cor. 1	2019-11-29	16	11.1002/1000/14106
7.2	ITU-T H.222.0 (2018) Amd. 1	2019-11-29	16	11.1002/1000/14105
8.0	ITU-T H.222.0 v8	2021-06-13	16	11.1002/1000/14658
8.1	ITU-T H.222.0 v8 (2021) Cor. 1	2022-12-14	16	11.1002/1000/15172
8.2	ITU-T H.222.0 v8 (2021) Amd. 1	2022-12-14	16	11.1002/1000/15171
9.0	ITU-T H.222.0 v9	2023-08-29	16	11.1002/1000/15655
10.0	ITU-T H.222.0 v10	2025-04-13	16	11.1002/1000/16266

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, and 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 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 https://www.itu.int/ITU-T/ipr/.

CONTENTS

List of Tables

Page

List of Figures

Page

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.

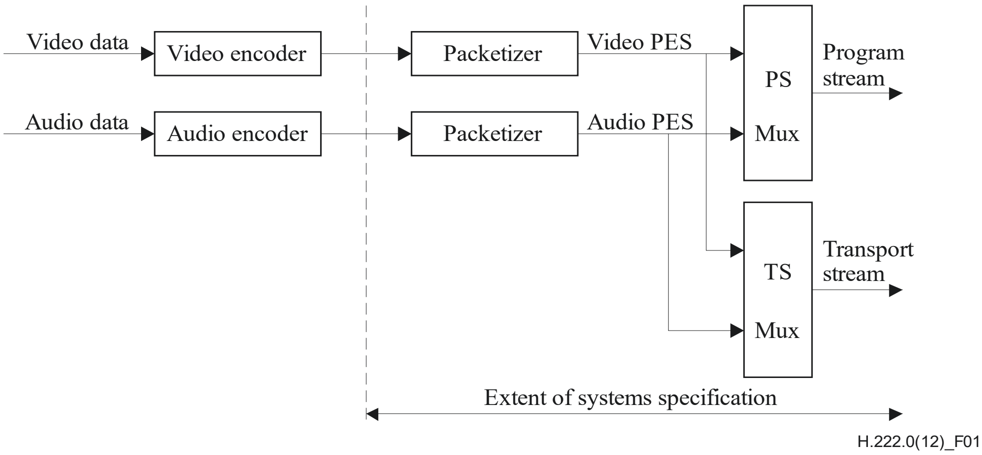

Figure Intro. 1 – Simplified overview of the scope of this Recommendation | International Standard

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 time base. 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.

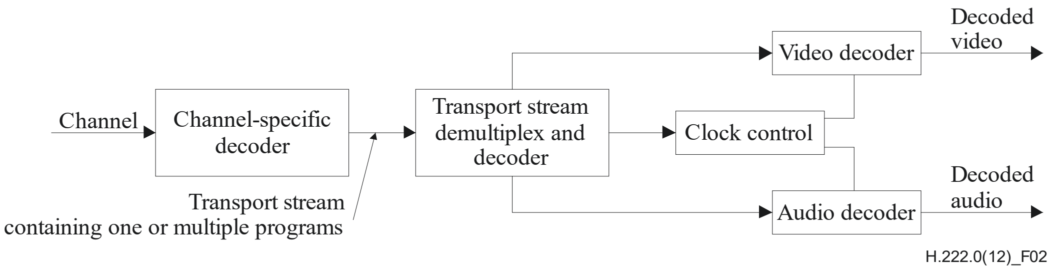

Figure Intro. 2 – Prototypical transport demultiplexing and decoding example

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.

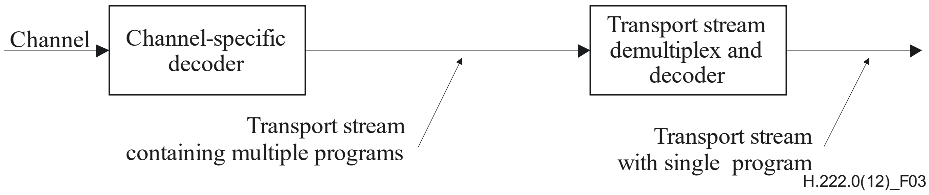

Figure Intro. 3 – Prototypical transport multiplexing example

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.

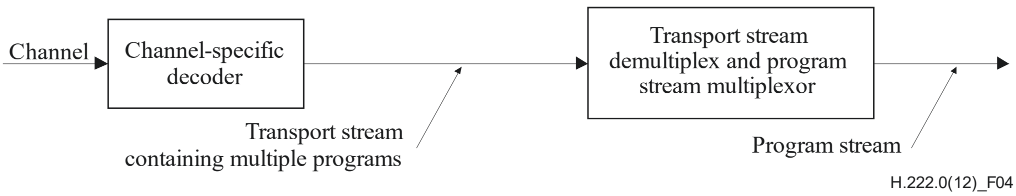

Figure Intro. 4 – Prototypical transport stream to program stream conversion

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.

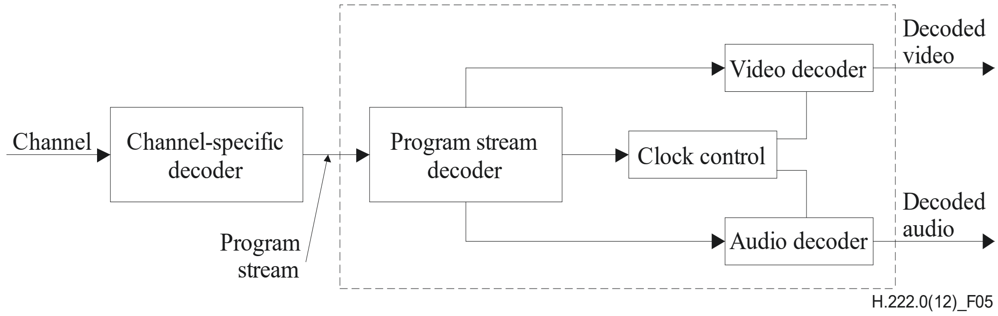

Figure Intro. 5 – Prototypical decoder for program streams

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 (2024) | ISO/IEC 15444-1:2024, *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 (2024), *Advanced video coding for generic audiovisual services*.

ISO/IEC 14496-10:2025, *Information technology – Coding of audio-visual objects – Part 10: Advanced video coding*.

– Recommendation ITU-T H.265 (2024), *High efficiency video coding.*

ISO/IEC 23008-2:2025, *Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 2: High efficiency video coding*.

– Recommendation. ITU-T H.273 (2024), *Coding-independent code points for video signal type identification*.

ISO/IEC 23091-2:2025, *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 (2023), *Versatile video coding*.

ISO/IEC 23090-3:2024, *Information technology – Coded Representation of Immersive Media – Part 3: Versatile video coding*.

– Recommendation ITU-T H.274 (2023), *Versatile supplemental enhancement information messages for coded video bitstreams*.

ISO/IEC 23002-7:2024 – *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.2100-3 (2025), *Image parameter values for high dynamic range television for use in production and international programme exchange.*

– ISO 639:2023, *Code for individual languages and individual groups*.

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

– ISO 15706-1:2023, *Information and documentation – International Standard Audiovisual Number (ISAN) – Part 1: Audiovisual work identifier*.

– ISO 15706-2:2023, *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 21122-1:2024, *JPEG XS low-latency lightweight image coding system – Part 1: Core coding system*.

- ISO/IEC 21122-2:2024, *JPEG XS low-latency lightweight image coding system – Part 2: Profiles and buffer models*.
- ISO/IEC 21122-3:2024, *JPEG XS low-latency lightweight image coding system – Part 3: Transport and container formats*.
- 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 23003-4:2020, *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 23091-3:2018, *Information technology – Coding-independent code points – Part 3: Audio*.
- 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.*
- ANSI/SCTE 35 (2019), *Digital Program Insertion Cueing Message for Cable*.
- 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*.