

TECHNICAL SPECIFICATION

**Encoding guidelines for portable multimedia CE products using MP4 file format
with AVC video codec and AAC audio codec**





THIS PUBLICATION IS COPYRIGHT PROTECTED

Copyright © 2012 IEC, Geneva, Switzerland

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either IEC or IEC's member National Committee in the country of the requester.

If you have any questions about IEC copyright or have an enquiry about obtaining additional rights to this publication, please contact the address below or your local IEC member National Committee for further information.

IEC Central Office
3, rue de Varembe
CH-1211 Geneva 20
Switzerland

Tel.: +41 22 919 02 11
Fax: +41 22 919 03 00
info@iec.ch
www.iec.ch

About the IEC

The International Electrotechnical Commission (IEC) is the leading global organization that prepares and publishes International Standards for all electrical, electronic and related technologies.

About IEC publications

The technical content of IEC publications is kept under constant review by the IEC. Please make sure that you have the latest edition, a corrigenda or an amendment might have been published.

Useful links:

IEC publications search - www.iec.ch/searchpub

The advanced search enables you to find IEC publications by a variety of criteria (reference number, text, technical committee,...).

It also gives information on projects, replaced and withdrawn publications.

IEC Just Published - webstore.iec.ch/justpublished

Stay up to date on all new IEC publications. Just Published details all new publications released. Available on-line and also once a month by email.

Electropedia - www.electropedia.org

The world's leading online dictionary of electronic and electrical terms containing more than 30 000 terms and definitions in English and French, with equivalent terms in additional languages. Also known as the International Electrotechnical Vocabulary (IEV) on-line.

Customer Service Centre - webstore.iec.ch/csc

If you wish to give us your feedback on this publication or need further assistance, please contact the Customer Service Centre: csc@iec.ch.

TECHNICAL SPECIFICATION

**Encoding guidelines for portable multimedia CE products using MP4 file format
with AVC video codec and AAC audio codec**

INTERNATIONAL
ELECTROTECHNICAL
COMMISSION

PRICE CODE

XD

ICS 33.160.01

ISBN 978-2-83220-220-3

Warning! Make sure that you obtained this publication from an authorized distributor.

CONTENTS

FOREWORD	6
0 Introduction	8
0.1 MP4 file format	8
0.2 Issues to be considered in implementation on portable CE devices	8
0.3 Encoding rules for open system and application	9
1 Scope	10
2 Normative references	10
3 Terms, definitions, abbreviations and conventions	11
3.1 Terms and definitions	11
3.2 Abbreviations	13
3.3 Conventions	13
3.3.1 Method of presenting box definition	13
3.3.2 Unit prefixes	14
3.3.3 Numeric representation	14
3.3.4 Character data	14
3.3.5 Nominal time length	14
3.3.6 Reserved fields and values	14
4 Design rules	14
4.1 General	14
4.2 File structure	15
4.2.1 General	15
4.2.2 Movie structure	16
4.2.3 Fragmented movie structure	17
4.3 Box order	18
4.4 Operational rules for boxes defined in ISO file format standards	20
4.4.1 General	20
4.4.2 Size fields	20
4.4.3 Versions and flags	20
4.4.4 File identification	20
4.4.5 File extension	21
4.4.6 Handler reference types	21
4.4.7 Movie header box	22
4.4.8 Track header box	22
4.4.9 Media header box	22
4.4.10 Sample size box	22
4.4.11 Time to sample boxes	22
4.4.12 Sample entries	22
4.4.13 Movie fragment boxes	22
4.4.14 Track fragment random access boxes – Version fields	23
4.4.15 Edit list entries	23
4.4.16 Template fields used	23
4.5 Additional definitions	24
4.5.1 Private extension box definitions	24
4.5.2 Profile configuration	30
4.5.3 Function flags	30

4.5.4	Codec specific information.....	31
4.5.5	Attribute flags	33
4.6	Operational rules for tracks	33
4.6.1	General	33
4.6.2	Main audio track	34
4.6.3	Main video track	35
4.6.4	Thumbnail picture track	36
4.7	Operational rules for media data	38
4.7.1	MPEG-4 AAC elementary stream.....	38
4.7.2	AVC video elementary stream.....	41
4.7.3	JPEG image sequence for thumbnail pictures	66
4.8	Logical structure of media data.....	67
4.8.1	General	67
4.8.2	Interleaving of main audio and video	67
4.8.3	Arrangement of data chunks	68
4.8.4	Placement of thumbnail picture.....	68
4.8.5	Media data structure in movie fragment	68
4.8.6	Exception for logical structure of media data	68
4.9	Presentation arrangement for tracks.....	68
4.9.1	Alternate group support	68
4.9.2	Edit list support	69
4.10	Metadata	70
4.10.1	Metadata types	70
4.10.2	Metadata types definition.....	71
4.11	Pixel aspect ratios	74
5	PCE AV classes	76
5.1	Class.....	76
5.2	Class PT	76
5.2.1	General	76
5.2.2	Level 1	77
5.2.3	Level 2	77
5.2.4	Level 3	77
5.3	Class MB.....	77
5.3.1	General	77
5.3.2	Level 1	78
5.3.3	Level 2	78
5.3.4	Level 3	79
5.3.5	Level 4	79
5.4	Class SD	79
5.5	Class network SD.....	80
5.6	Class HD.....	81
5.6.1	General	81
5.6.2	Level 1	82
5.6.3	Level 2	82
5.7	Class network HD.....	82
5.7.1	General	82
5.7.2	Level 1	83
5.7.3	Level 2	83
5.7.4	Level 3	84

6	Definitions for other files.....	84
Annex A (normative)	Operation rules for compliant products	85
Annex B (normative)	Capabilities required for compliant products	89
Annex C (informative)	Recommended recording modes	91
Annex D (normative)	Operation rules for rotated video file	96
Bibliography.....		98
Figure 1 – Example of a simple interchange file		15
Figure 2 – Examples of movies		17
Figure 3 – Example of a fragmented movie		18
Figure 4 – Example of AAC bitstream		38
Figure 5 – Example of a closed GOP		52
Figure 6 – Example of an open GOP		52
Figure 7 – Reference structure of a reference B-picture		60
Figure 8 – Reference structure of a non-reference B-picture		61
Figure 9 – Example of a closed GOP		61
Figure 10 – Example of an open GOP		62
Table 1 – Supported track types		16
Table 2 – Supported stream and data		16
Table 3 – Box types, structure, and cross-reference		19
Table 4 – brand identifier		21
Table 5 – handler_types		22
Table 6 – Common portion of the audio attribute flags		26
Table 7 – Common portion of the video attribute flags.....		27
Table 8 – encoding_type		29
Table 9 – Function flags.....		30
Table 10 – Data form type.....		31
Table 11 – Data form type.....		32
Table 12 – Codec specific portion of audio attribute flags		33
Table 13 – Codec specific portion of video attribute flags.....		33
Table 14 – The values of pic_struct and ct_type		54
Table 15 – Allowed values of frame_mbs_only_flag, pic_width_in_mbs_minus1 and pic_height_in_map_units_minus1		63
Table 16 – Allowed values of frame cropping related fields		63
Table 17 – data_type_ID		71
Table 18 – Presentation types.....		72
Table 19 – Track attribute flags.....		73
Table 20 – Usage of the metadata		74
Table 21 – Resolution values in visual sample entry		75
Table 22 – spacing values in pixel_aspect_ratio		75
Table C.1 – Recommended recording modes for Class PT.....		91
Table C.2 – Recommended recording modes for Class MB		92

Table C.3 – Recommended recording modes for Class SD	92
Table C.4 – Parameters for progressive video 25[fps] and 29.97[fps]	93
Table C.5 – Recommended recording modes for Class network SD	93
Table C.6 – Recommended recording modes for Class HD	94
Table C.7 – Recommended recording modes for Class network HD	95
Table D.1 – Width and height in the track header box	97

INTERNATIONAL ELECTROTECHNICAL COMMISSION

ENCODING GUIDELINES FOR PORTABLE MULTIMEDIA CE PRODUCTS USING MP4 FILE FORMAT WITH AVC VIDEO CODEC AND AAC AUDIO CODEC

FOREWORD

- 1) The International Electrotechnical Commission (IEC) is a worldwide organization for standardization comprising all national electrotechnical committees (IEC National Committees). The object of IEC is to promote international co-operation on all questions concerning standardization in the electrical and electronic fields. To this end and in addition to other activities, IEC publishes International Standards, Technical Specifications, Technical Reports, Publicly Available Specifications (PAS) and Guides (hereafter referred to as "IEC Publication(s)"). Their preparation is entrusted to technical committees; any IEC National Committee interested in the subject dealt with may participate in this preparatory work. International, governmental and non-governmental organizations liaising with the IEC also participate in this preparation. IEC collaborates closely with the International Organization for Standardization (ISO) in accordance with conditions determined by agreement between the two organizations.
- 2) The formal decisions or agreements of IEC on technical matters express, as nearly as possible, an international consensus of opinion on the relevant subjects since each technical committee has representation from all interested IEC National Committees.
- 3) IEC Publications have the form of recommendations for international use and are accepted by IEC National Committees in that sense. While all reasonable efforts are made to ensure that the technical content of IEC Publications is accurate, IEC cannot be held responsible for the way in which they are used or for any misinterpretation by any end user.
- 4) In order to promote international uniformity, IEC National Committees undertake to apply IEC Publications transparently to the maximum extent possible in their national and regional publications. Any divergence between any IEC Publication and the corresponding national or regional publication shall be clearly indicated in the latter.
- 5) IEC provides no marking procedure to indicate its approval and cannot be rendered responsible for any equipment declared to be in conformity with an IEC Publication.
- 6) All users should ensure that they have the latest edition of this publication.
- 7) No liability shall attach to IEC or its directors, employees, servants or agents including individual experts and members of its technical committees and IEC National Committees for any personal injury, property damage or other damage of any nature whatsoever, whether direct or indirect, or for costs (including legal fees) and expenses arising out of the publication, use of, or reliance upon, this IEC Publication or any other IEC Publications.
- 8) Attention is drawn to the Normative references cited in this publication. Use of the referenced publications is indispensable for the correct application of this publication.
- 9) Attention is drawn to the possibility that some of the elements of this IEC Publication may be the subject of patent rights. IEC shall not be held responsible for identifying any or all such patent rights.

The main task of IEC technical committees is to prepare International Standards. In exceptional circumstances, a technical committee may propose the publication of a Technical Specification when

- the required support cannot be obtained for the publication of an International Standard, despite repeated efforts, or
- the subject is still under technical development or where, for any other reason, there is the future but no immediate possibility of an agreement on an International Standard.

Technical Specifications are subject to review within three years of publication to decide whether they can be transformed into International Standards.

This Technical Specification cancels and replaces the first edition published in 2009. This edition constitutes a technical revision.

This edition includes the following significant technical changes with respect to the previous edition:

- a) added 23,976 fps to all classes;

- b) added new class (class network SD);
- c) added some metadata for CE content;
- d) inserted some editorial improvements and clarifications.

IEC/TS 62592, which is a Technical Specification, has been prepared by technical area 6: Storage media, data structures, equipment and systems, of IEC technical committee 100: Audio, video and multimedia systems and equipment.

The text of this Technical Specification is based on the following documents:

Enquiry draft	Report on voting
100/1926/DTS	100/1970/RVC

Full information on the voting for the approval of this Technical Specification can be found in the report on voting indicated in the above table.

This publication has been drafted in accordance with the ISO/IEC Directives, Part 2.

The committee has decided that the contents of this publication will remain unchanged until the stability date indicated on the IEC web site under "<http://webstore.iec.ch>" in the data related to the specific publication. At this date, the publication will be

- transformed into an International standard,
- reconfirmed,
- withdrawn,
- replaced by a revised edition, or
- amended.

A bilingual version of this publication may be issued at a later date.

0 Introduction

0.1 MP4 file format

ISO base media file format, ISO/IEC 14496-12, has been developed by ISO/IEC JTC 1/SC 29/WG 11 and WG1 as a common base media file format of audio, video and image applications. It provides a file format to contain timed media information for a presentation in a flexible, extensible format. The extensions to support specific codecs and systems are specified as the series of standard ISO/IEC 14496. The MP4 file format (MPEG-4 file format), ISO/IEC 14496-14, is an extension to support MPEG-4 systems in ISO base media file format. The extension to support Advanced Video Coding (AVC), ISO/IEC 14496-10 and ITU-T H.264¹, is standardized as AVC file format, ISO/IEC 14496-15. In this Technical Specification, the family of ISO base media file format is referred to as MP4 file format, which is the name widely used in the industry. The MP4 file format is designed very flexibly so that the series of the standard can be applied to various kinds of applications and can bring the maximum performances for the applications.

MP4 file format is adopted by various Consumer Electronics (CE) devices, e.g. broadcasting receivers, disc recorders / players, AV content distribution, portable AV recorders / players, and so on. In the CE audio, video and multimedia applications, MP4 file format with Advanced Video Coding (AVC), (ISO/IEC 14496-10 and ITU-T H.264) and Advanced Audio Coding (AAC), (ISO/IEC 13818-3 and ISO/IEC 14496-3), is employed most popularly.

AVC (ISO/IEC 14496-10 and ITU-T H.264) specifies profiles and levels to ensure interoperability of decoding procedure of a video elementary stream, which fits various levels of application. AVC, ISO/IEC 14496-10 and ITU-T H.264, specifies the bitstream syntax and its decoding process.

0.2 Issues to be considered in implementation on portable CE devices

To implement applications using MP4 file format into CE products, the characteristics of CE industry must be considered. The characteristics of CE industry and products are different from IT equipment especially personal computers (PCs) which have strong computational power and flexible software solution. That is:

- CE devices are designed with limited resources, processing power and memory size;
- most CE devices can not update software or hardware afterwards; and
- manufacturers have to assure the capability and the quality of the product for consumers, who are not familiar with processes inside the device;
- CE devices need to support functionalities, e.g. fast forward / backward play which are widely supported by existing CE devices. Manufacturers have to ensure such functionalities regardless of creators of the content (bitstreams).

Given these criteria, manufacturers need to check all the performance of the products because any defect in the products may generate serious complaints among consumers. Since the MP 4 file format is so flexible, the number of combination of setting parameters is very large and it takes tremendous workload and cost to check the performance for each combination of parameters. Even AVC (ISO/IEC 14496-10 and ITU-T H.264) specifies the decoding procedure and conformance point by a profile and a level. Furthermore, the constraints on bitstreams must be specified in order to ensure functionalities, which are widely supported by CE devices. Therefore, in most CE systems, the combination of encoding parameters is limited and the dedicated encoding rules are specified for the system, which enable designing and manufacturing process practical to guarantee the quality of the product. In general, these encoding rules are proprietary to the involved parties.

¹ ITU-T H.264 is equivalent to ISO/IEC 14496-10.

0.3 Encoding rules for open system and application

Currently, the content of audio, video and multimedia products are provided by commercial content providers through specific sales channels such as optical discs, CDs, DVDs and Blu-ray Discs. Additional content is provided by broadcasters and commercial content providers. However, more recently end-user generated content is increasing and placed on many Internet sites enabled by the availability of digital video cameras and Internet related technology. In this situation, portable CE devices are required to store and reproduce such content for consumer satisfaction. Accordingly, portable CE devices need to guarantee the capability to decode MP4 files with AVC (ISO/IEC 14496-10 and ITU-T H.264) and AAC (ISO/IEC 13818-3 and ISO/IEC 14496-3) which are most commonly used as the file format and codecs.

- To assure decoding of such MP4 files, CE device manufacturers have to check the performance for the files encoded by unknown parameters. However, it is impossible to achieve this with limited manpower and technical resources as mentioned above. This situation may cause a problem in decoding quality and is not beneficial for both end-users and manufacturers. CE device manufacturers also have to ensure functionalities, e.g. fast forward / backward play, regardless of the creators of the content (bitstreams).

To solve this issue, this Technical Specification specifies the encoding rules for the MP4 files targeted to be stored and reproduced by portable CE devices so that the decoding of the content encoded by these guidelines is guaranteed by the portable CE product manufacturer with reasonable cost and resources.

The encoding rules are specified considering the capability of current portable CE products. However, the rules should be revised depending on the progress of CE technologies in an appropriate time frame.

ENCODING GUIDELINES FOR PORTABLE MULTIMEDIA CE PRODUCTS USING MP4 FILE FORMAT WITH AVC VIDEO CODEC AND AAC AUDIO CODEC

1 Scope

This Technical Specification specifies encoding guidelines for portable multimedia CE products using ISO base media file format and its family (ISO/IEC 14496-12, ISO/IEC 14496-14, and ISO/IEC 14496-15) with AVC² (ISO/IEC 14496-10 and ITU-T H.264) and AAC² (ISO/IEC 13818-3 and ISO/IEC 14496-3). These guidelines may also be applicable to portable non-CE products.

This Technical Specification is applicable to the creation of MP4 files with AVC and AAC which are intended for storage, reproduction and display by portable CE products. It is applicable to both content generation software and hardware.

These guidelines are intended to encourage global interoperability between portable CE products by recommending the use of specific parameters to allow efficient interworking of devices which may have limited resources so that the decoding of content encoded in accordance with these guidelines is assured. In addition, the guidelines provide for simplified testing and verification of the interoperability of portable CE products.

2 Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

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

ISO/IEC 10646:2012, *Information technology – Universal Coded Character Set (UCS)*

ISO/IEC 10918-1, *Information technology – Digital compression and coding of continuous-tone still images – Part 1: Requirements and guidelines*

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

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 14496-3:2009, *Information technology – Coding of audio-visual objects – Part 3: Audio*

ISO/IEC 14496-10:2012, *Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding*

² The full wording for AAC and AVC can be found in 3.2.

ISO/IEC 14496-12:2008, *Information technology – Coding of audio-visual objects – Part 12: ISO base media file format*

ISO/IEC 14496-14:2003, *Information technology – Coding of audio-visual objects – Part 14: MP4 file format*

ISO/IEC 14496-15:2010, *Information technology – Coding of audio-visual objects – Part 15: Advanced Video Coding (AVC) file format*

ISO/IEC 15444-12, *Information technology – JPEG 2000 image coding system – Part 12: ISO base media file format*

3 Terms, definitions, abbreviations and conventions

3.1 Terms and definitions

For the purposes of this document, the terms and definitions specified in ISO/IEC 14496-1, ISO/IEC 14496-2, ISO/IEC 14496-3, ISO/IEC 14496-10, ISO/IEC 14496-12 and ISO/IEC 14496-15, as well as the following apply.

3.1.1

B-field

field consisting only of B slices

3.1.2

B-frame

frame consisting only of B slices

3.1.3

bitrate

rate at which the coded bitstream is delivered to the input of a decoder

Note 1 to entry: Bitrate is an index to indicate how many bits are transmitted per second.

Note 2 to entry: The unit is bps (bits per second).

3.1.4

B-picture

frame or a complementary field pair consisting only of B slices

3.1.5

class

video and audio data formats and constraints on the values of the syntax

3.1.6

fragmented movie

movie that consists of a single Movie Box ('moov'), Movie Fragment Boxes ('moof'), and Media Data Boxes ('mdat') that contain fragmented media data

Note 1 to entry: The Movie Box of a fragmented movie is referred to as an initial Movie Box to distinguish it from that of an ordinary movie in this Technical Specification. An initial 'moov' box contains a Movie Extends Box ('mvex') that indicates that the movie is possibly fragmented, which means 'moof's should be found and used in the file.

3.1.7

frame rate

rate at which frames are to be output from the composition process

Note 1 to entry: Frame rate is an index to indicate how many frames can be drawn per second in video playback.

Note 2 to entry: The unit is fps (frames per second). When the frame rate is 30 fps, it means that 30 frames are drawn per second.

3.1.8

I-field

field consisting only of I slices

3.1.9

I-frame

frame consisting only of I slices

3.1.10

I-picture

frame or a complementary field pair consisting only of I slices

3.1.11

main audio track

audio track that contains an audio stream as main presentation

3.1.12

main video track

video track that contains a video stream as main presentation

3.1.13

P-field

field consisting only of P slices

3.1.14

P-frame

frame consisting only of P slices

3.1.15

player

device or software with a function to play back the PCE AV files

3.1.16

PCE AV file

file that complies with the file format described in this Technical Specification

3.1.17

P-picture

frame or a complementary field pair consisting only of P slices

3.1.18

product

generic term of players and recorders

3.1.19

PCE AV Classes

collective term for Class PT, Class MB, Class SD, Class network SD, Class HD and Class network HD

Note 1 to entry: PCE AV Classes are described in Clause 5.

3.1.20

recorder

device or software with a function to record/delete the PCE AV files and directories, compliant with this Technical Specification

Note 1 to entry: A playback function is not necessarily required. Products capable of deleting files/directories or updating existing files are also included even if they cannot record new files.

3.1.21

thumbnail picture track

video track that contains one or more still images for thumbnail picture use

3.2 Abbreviations

AAC	Advanced Audio Coding
AAC LC	AAC Low Complexity Profile
ADIF	Audio Data Interchange Format
ADTS	Audio Data Transport Stream
AVC	Advanced Video Coding
bslbf	bit string, left bit first
CE	Consumer Electronics
DCF	Design rule for Camera File system
Exif	Exchangeable image file format for digital still cameras
GOP	Group Of Pictures
HDTV	High Definition Television
JPEG	Joint Photographic Experts Group
MPEG	Moving Picture Experts Group
PNG	Portable Network Graphics
PPS	Picture Parameter Set
QVGA	Quarter Video Graphics Array
SDTV	Standard Definition Television
SPS	Sequence Parameter Set
URL	Uniform Resource Locator
UTC	Coordinated Universal Time
VGA	Video Graphics Array

3.3 Conventions

3.3.1 Method of presenting box definition

The term “Mandatory: Yes” or “Mandatory: No” is placed at the beginning of each definition of a box, as described in ISO/IEC 14496-12, ISO/IEC 14496-14 and ISO/IEC 14496-15.

“Mandatory: Yes” means that this box is mandatory in its container.

NOTE “Mandatory: Yes/No” is for containers, not for entire structures.

The syntax of a box is documented in the syntactic description language (SDL) defined in ISO/IEC 14496-1.

3.3.2 Unit prefixes

In this Technical Specification, the following four unit prefixes are used:

- 1 KB = 1 024 B
- 1 MB = 1 024 KB
- 1 kbps = 1 000 bps
- 1 Mbps = 1 000 kbps

3.3.3 Numeric representation

A decimal number consists of digits “0” to “9”.

A hexadecimal number consists of digits “0” to “9”, upper case letters “A” to “F”, and suffix “h”.

A binary number consists of digits “0” and “1”, and suffix “b”.

3.3.4 Character data

Unless otherwise stated, character data contain a set of characters except for control codes (0000h to 001Fh, and 007Fh to 009Fh) in Unicode. Note that the code 0000h (null) and 0009h (horizontal tabulation, referred to as TAB hereinafter) may be used for specific purposes: the code 0000h (null) may be used only as the termination code for character strings in some fields that require explicit termination; the code 0009h (TAB) may be used only as the separator code for character strings.

3.3.5 Nominal time length

In this Technical Specification, second* is defined as nominal time length and it indicates the length of time calculated by the following:

$$1 \text{ second}^* = 1 \text{ nsample} \times \text{sample duration} / \text{media timescale second}$$

where nsample is the value calculated by rounding up the number of samples per second.

Note that in interlaced video, the number of samples is twice as much as the frame rate (number of frames per second).

For example, 1 second* for progressive video of 29,97 fps is calculated as follows:

$$1 \text{ second}^* = 30 \times 1\,001 / 30\,000 = 1,001 \text{ second}$$

1 second* for audio data of sampling frequency 48 kHz is calculated as follows:

$$1 \text{ second}^* = 47 \times 1\,024 / 48\,000 = 1,002\,7 \text{ second}$$

3.3.6 Reserved fields and values

Reserved fields are provided for future expansion. Unless otherwise stated, reserved fields shall be set to ‘0’ when data is written and shall not be interpreted when data is read. Any values already set in the reserved fields shall be kept unchanged under any circumstances.

4 Design rules

4.1 General

The file format specified in this Technical Specification is based on MP4 file format (ISO/IEC 14496-14) for a fundamental structure and AVC file format (ISO/IEC 14496-15) for a track of AVC video elementary stream, both of which are instances of ISO Base Media file format (ISO/IEC 14496-12 and ISO/IEC 15444-12). The general nature of MP4 file format and AVC file format is partly exercised, and there are some extensions and restrictions.

The following elements are defined in this Technical Specification.

a) Operational rules for MP4 file format

To achieve better interoperability and to make implementation of the format easier, the subset of MP4 file format is defined as well as operational rules for the use of MP4 file format such as the setting of boxes / fields and box order.

b) Extensions to MP4 file format

To provide more convenience and to improve usability, Technical Specification specific extensions to the MP4 File format are defined. The structures for improved file identification and those for handling metadata are provided in 4.5.

c) Operational rules for media data and track structure

The operational rules for encoding audio and video media data and the combinations of audio and video are defined to indicate the required capability of decoders. Also, the operational rules for the structures of tracks that store media data are defined.

d) Other operational rules for interoperability

Other operational rules such as required capabilities of decoders, the rules for handling of information contained in a compliant file and recommended recording modes are defined.

4.2 File structure

4.2.1 General

Of the usage modes documented in ISO/IEC 14496-12, only a simple interchange file and a content creation file are supported in this Technical Specification.

Figure 1 shows an example of a simple interchange file.

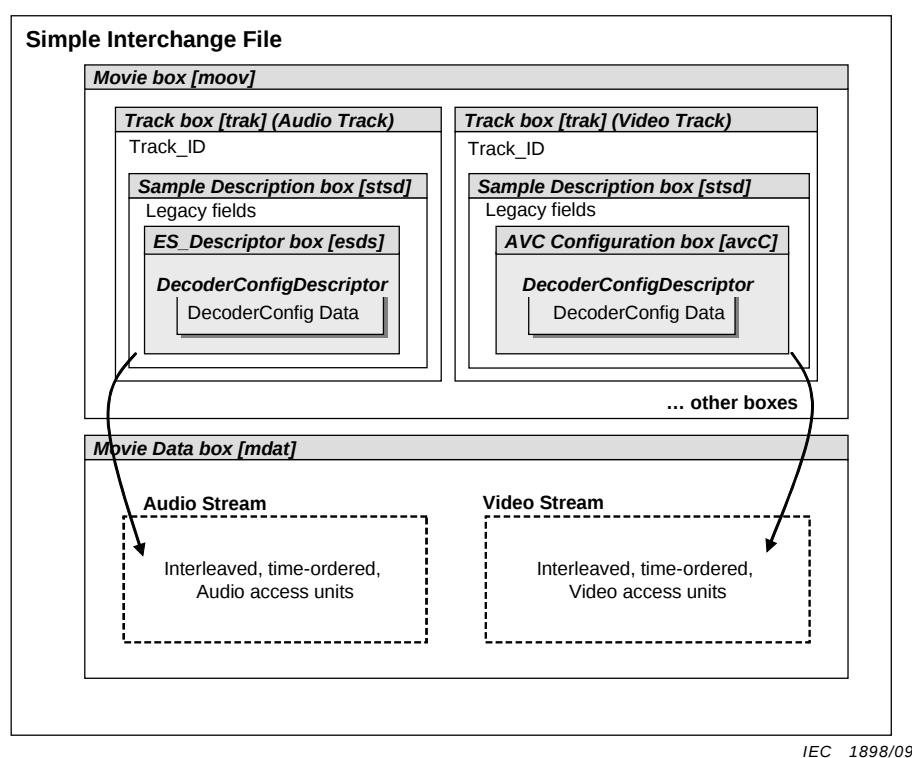


Figure 1 – Example of a simple interchange file

A PCE AV file supports the structures of both a (regular) movie and a fragmented movie.

Note that the Movie Box of a fragmented movie is referred to as an initial Movie Box or an initial 'moov' to distinguish it from that of a movie without fragmented structure. These two types of Movie Boxes, however, are simply different in that the initial Movie Box further contains a Movie Extends Box. The following subclauses describe both structures, and 4.8 defines some restrictions for the logical structure of media data in both movies.

A PCE AV file shall consist of at least one main video track. It may also contain one or more main audio, thumbnail picture and other tracks (in the future) all together. Table 1 and Table 2 show track types supported in this Technical Specification.

See 4.6 for details on operational rules for such tracks, and 4.10.2.2 for the presentation types.

Table 1 – Supported track types

Track types	Stream / data	Handler types	Media header	Presentation types
Main audio track	MPEG-4 AAC elementary stream	'soun'	'smhd'	00000001h
Main video track	AVC video elementary stream	'vide'	'vmhd'	00000001h
Thumbnail picture track	JPEG image sequence	'vide'	'vmhd'	00000002h

Table 2 – Supported stream and data

Stream / data	Format types	ES descriptor	
		Object types	Audio / visual object types
MPEG-4 AAC elementary stream	'mp4a' (MP4AudioSampleEntry)	40h (ISO/IEC 14496-3)	02h (AAC LC)
AVC video elementary stream	'avc1' (AVCSampleEntry)	–	–
JPEG image sequence	'mp4v' (MP4VisualSampleEntry)	6Ch (ISO/IEC 10918-1)	–

4.2.2 Movie structure

This subclause defines restrictions on a (regular) movie. Subclause 4.8 describes the logical structure of a media data, and Figure 2 (Figure 2a and Figure 2b) shows examples of a movie structure.

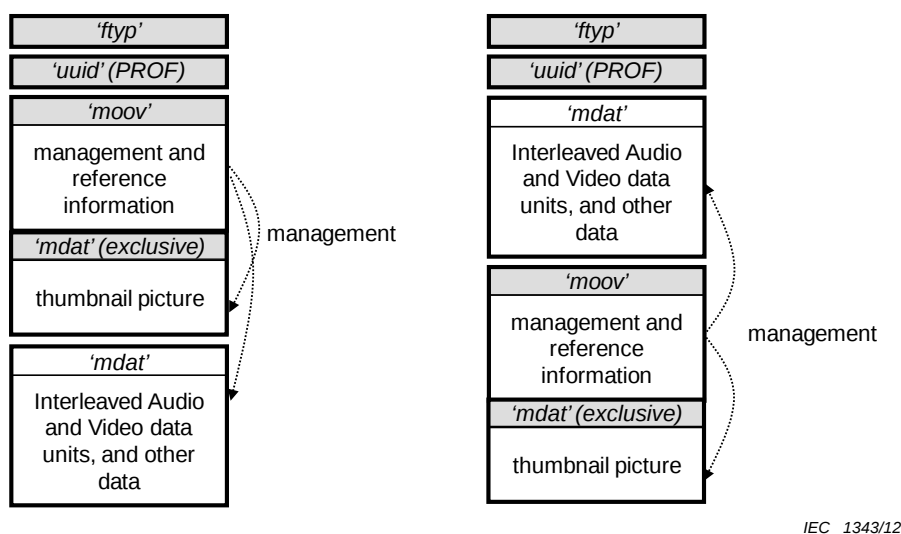


Figure 2a – 'mdat' placed after 'moov'

Figure 2b – 'mdat' placed before 'moov'

Figure 2 – Examples of movies

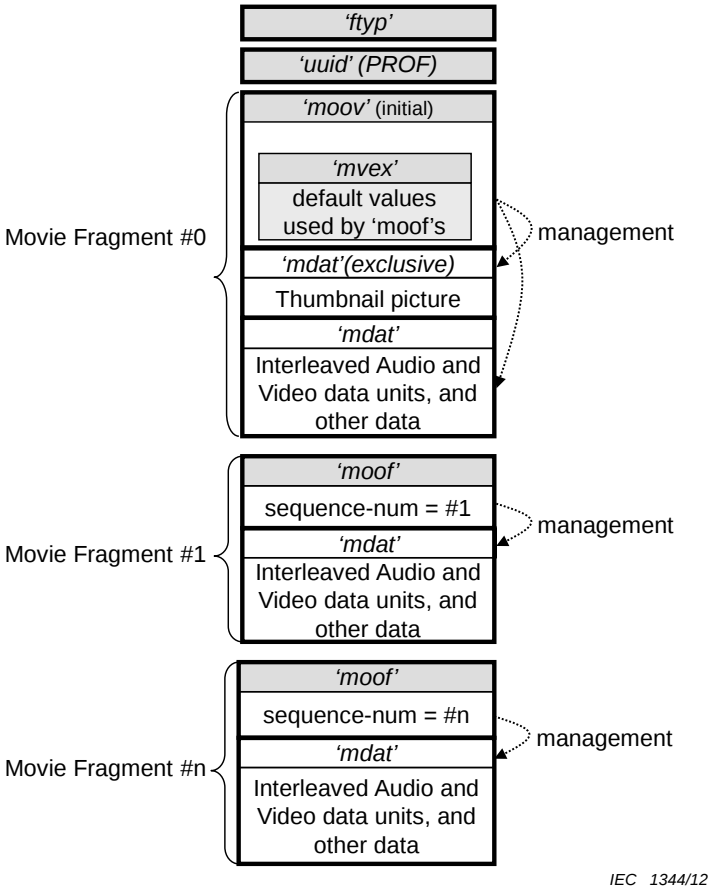
The size of a PCE AV file in the form of a regular movie shall not exceed 4 gigabytes. If the size of a PCE AV file exceeds 4 gigabytes, it shall be in the form of a fragmented movie.

4.2.3 Fragmented movie structure

Some additional or overridden restrictions are defined in this subclause for the fragmented movie. The basic technique to construct fragmented movies shall conform to Annex A of ISO/IEC 14496-12. Subclause 4.8 describes the logical structure of a movie fragment, and Figure 3 shows an example of a fragmented movie structure. When the file size exceeds 4 gigabytes, Fragment movie is required.

A PCE AV file as a fragmented movie shall also satisfy the following additional or overridden rules.

- a) A "movie fragment" consists of the following items when at least one track is self-contained:
 - fragmented media data (Media Data Box(es) ('mdat') that contains samples);
 - single initial 'moov' or 'moof' box that contains management and reference information about those samples.
- b) A "movie fragment" consists of the following items when all tracks reference to media data in other file(s):
 - single initial 'moov' or 'moof' box that contains management and reference information about the samples in other file(s).
- c) In each movie fragment, a 'moof' box has only one Track Fragment Box ('traf') per existing track in the movie fragment.
- d) A duration-is-empty flag in a 'traf' box is prohibited for use, which means the PCE AV file shall not contain any empty-duration fragments.



IEC 1344/12

Figure 3 – Example of a fragmented movie

4.3 Box order

An overall view of the normal encapsulation structure is provided in the following table.

Table 3 shows the boxes that are used in this Technical Specification that are subset of the boxes defined in ISO/IEC 14496-12, ISO/IEC 14496-14 and ISO/IEC 14496-15 as well as additionally defined boxes in this Technical Specification. In other words, boxes that are defined in those file formats, but not described in Table 3 shall not be used in all files conforming to a specific media file format that uses this Technical Specification as a core. Note that other boxes may be additionally defined in a specific media file format that uses this Technical Specification. Refer to ISO/IEC 14496-12 for details on boxes with no reference (no number or mark appears in “Reference” column of the table).

See the description of the individual boxes in ISO/IEC 14496-12, ISO/IEC 14496-14 and ISO/IEC 14496-15 for a discussion of what shall be assumed if the optional boxes are not present.

In order to improve interoperability and utility of the files, the rules and guidelines that are described in the subclause “Box Order” of ISO/IEC 14496-12 shall be followed for the order of boxes. Some additional rules are given below.

- a) The Profile Box (Private Extension) shall occur immediately after the File Type Compatibility Box, before the Movie Box and the Media Data Box.
- b) The order of the Movie Box and the Media Data Box that contains audio, video and image sample data is not specified as long as they come after the File Type Box and The Profile Box (Technical Specification specific extension), except for ‘exclusive’ Media Data Box

that is placed immediately after the Movie Box. See 4.8.4 for details on the ‘exclusive’ Media Data Box.

- c) The boxes within the Sample Table Box should be in the following order: Sample Description, Decoding Time to Sample, Composition Time to Sample, Sample to Chunk, Sample Size, Chunk Offset, and Sync Sample.
- d) The Handler Reference Box shall precede the Media Information Box.
- e) The Meta Type Definition Box should precede the Metadata Box in User Specific Metadata Box (Technical Specification specific extension).

In case of fragmented movie files, the following rules shall also be satisfied.

- f) An initial Movie Box (‘moov’) is placed before any Movie Fragment Box (‘moof’).
- g) In the first movie fragment, an exclusive ‘mdat’ box is placed immediately after the initial ‘moov’ box. See 4.8.4 for details on the ‘exclusive’ Media Data Box.
- h) In each movie fragment, an initial ‘moov’ or ‘moof’ box is placed before any ‘mdat’ box when at least one track is self-contained.
- i) The movie fragment random access box (‘mfra’) shall exist in a file.

Table 3 – Box types, structure, and cross-reference

Box types							Reference	Description
ftyp							a	file type and compatibility
uuid							a 4.5.1.2	UUID extension (‘PROF’); file profile
moov							a	container for all the movie resources
	mvhd						a	movie header, overall declarations
	trak						a	container for an individual track or stream
		tkhd					a	track header, overall information about the track
		tref						track reference container
			sync					synchronization indication
		edts					a	edit list container
			elst				a	an edit list
		mdia					a	container for the media information in a track
			mdhd				a	media header, overall information about the media
			hdlr				a	media handler, the media (handler) type
			minf				a	media information container
				vmhd			c	video media header (video track only)
				smhd			c	sound media header (audio track only)
				dinf			a	data information, container
					dref		a	data reference, declares source(s) of media in track
						url	a	URL reference entry
				stbl			a	sample table, container for the time/space map
				stsd			a	sample descriptions (codec types, initialization etc.)
				stts			a	(decoding) time to sample
				ctts				composition time to sample
				stsc			a	sample to chunk, partial data-offset information
				stsz			a	sample sizes (framing)

Box types						Reference	Description
				stco		a	chunk offset, partial data-offset information
				stss			sync (key frame) sample table (random access points)
		uuid				a 4.5.1.3	UUID extension ('USMT'); user specific metadata
			MTDT			a 4.5.1.4	metadata list
	mvex						movie extends, container
		mehd					movie extends header, overall duration of fragments
		trex				b	track extends defaults
	uuid					4.5.1.3	private extension ('USMT'); user specific metadata
		MTDT				4.5.1.4	metadata list
moof							movie fragment, container
	mfhd					b	movie fragment header
	traf						track fragment, container
		tfhd				b	track fragment header
		trun					track fragment run
mfra							movie fragment random access points
uuid							user private extension d
free							free space
mdat							Media data container
a Mandatory boxes in this Technical Specification. b Mandatory box when the container that contains it is present. c Only one of these boxes shall be present in the 'minf' container. d It is allowed to contain one or more user private extension uuid boxes that are defined outside of this Technical Specification.							

4.4 Operational rules for boxes defined in ISO file format standards

4.4.1 General

This subclause describes operational rules for boxes defined in ISO/IEC 14496-12, ISO/IEC 14496-14 and ISO/IEC 14496-15.

4.4.2 Size fields

The size field of all boxes shall not be set to 1. In other words, the largesize field is prohibited for use.

4.4.3 Versions and flags

The version and flags of all boxes shall be set to '0', unless this document states otherwise.

4.4.4 File identification

4.4.4.1 General

The File Type Box identifies which specification is the 'best use' (major brand) of a file, a minor version of that specification, and a set of other specifications (compatible brands) with which the file complies. In this Technical Specification, operational rules for the File Type Box are defined for better file identification.

4.4.4.2 Brand identifier

The brand identifier precisely identifies the format that files conform to. Only brand identifiers defined in Table 4 are currently used in this Technical Specification.

The brand identifier 'MSNV' is defined as the major brand identifying this Technical Specification. This brand shall appear in the major_brand field and the compatible_brands list of the File Type box as defined in ISO/IEC 14496-12, in all files conforming to this Technical Specification.

The brand identifier 'mp42' is defined as identifying the latest version of ISO/IEC 14496-14, which shall appear in the compatible_brands list of the File Type box.

The brand identifier 'isom' is defined as identifying ISO/IEC 14496-12, which shall appear in the compatible_brands list of the File Type box.

Note that each brand identifier shall appear in the compatible_brands list only once.

The major brand will be changed when modification or updating that does not maintain backward compatibility with this version of the Technical Specification is carried out.

Table 4 – brand identifier

Values	Description
'MSNV'	This Technical Specification
'mp42 '	MP4 file format
'isom'	ISO Base Media file format

4.4.4.3 Minor-version definition

This subclause discusses in detail the minor_version value. The following structure is used for any specific media file format, but the values of these fields are specified in the format.

The minor_version has the following structure:

```
unsigned int(4) reserved = Fh;
unsigned int(12) version_number;
unsigned int(16) reserved = 1FFFh;
```

The semantics of the "version_number" field is defined as follows:

version_number A 12-bit integer that specifies the version number of this Technical Specification. The upper 4 bits indicate the major version number, and the lower 8 bits indicate the minor version number. Set 101h for this technical specification.

4.4.5 File extension

The file extension for the files compliant with this Technical Specification should be ".MP4".

4.4.6 Handler reference types

Only the following values for handler_type, which is in the Handler Reference Box ('hdlr') as defined in ISO/IEC 14496-12, are used in this Technical Specification (see Table 5).

Table 5 – handler_types

Values	Description
'vide'	Video Stream, and Image
'soun'	Sound Stream

4.4.7 Movie header box

The creation_time field should be set to the date and time when the file is opened or closed when creating a file, and the modification_time field should be set to the date and time when the file is opened or closed when modifying a file.

4.4.8 Track header box

The track_ID field in a track header box shall be set as follows:

Track_ID (corresponds to ES_ID): 00000001h – 0000007Fh (1 – 127) or,
00000400h – 0000F7FFh (1 024 – 63 487);
other values are reserved.

The creation_time field should be set to the date and time when the file is opened or closed when creating a file, and the modification_time field should be set to the date and time when the file is opened or closed when modifying a file.

4.4.9 Media header box

The creation_time field should be set to the date and time when the file is opened or closed when creating a file, and the modification_time field should be set to the date and time when the file is opened or closed when modifying a file.

The code of “undetermined” ('und') can be set to the language field in the Media Header Box for the track that is not dependent on any language.

4.4.10 Sample size box

A dummy sample that has no data is prohibited, and therefore, entry_size fields in the Sample Size Box shall not be set to '0'.

4.4.11 Time to sample boxes

In the Time to Sample Boxes (Decoding Time to Sample Box and Composition Time to Sample Box), futility time information shall not be held, and therefore, entry_count fields in the Time to Sample Boxes shall not be set to '0'.

4.4.12 Sample entries

It is recommended to keep one of them if some Sample Entries for a track can be put together in accordance with the definitions in ISO/IEC 14496-12, ISO/IEC 14496-14 and ISO/IEC 14496-15, for example when their contents are identical.

4.4.13 Movie fragment boxes

4.4.13.1 General

This subclause defines restrictions when the movie is fragmented.

4.4.13.2 Sample size fields

A dummy sample that has no data is prohibited. Therefore, a `sample_size` field in the Track Fragment Run Box and a `default_sample_size` field in a Track Fragment Header Box shall not be set to '0' if present.

A `default_sample_size` field in a Track Extends Box shall not be set to '0' if the field is used.

4.4.13.3 Sample flags fields

A field or flag in `sample_flags` fields corresponding to unsupported function or box in this Technical Specification such as sample padding and degradation control shall not be used and such fields or flags shall be set to '0'.

4.4.14 Track fragment random access boxes – Version fields

The version field in the Track Fragment Random Access Box may be set to 1.

4.4.15 Edit list entries

When the Edit List Box is used, each edit list entry in the box except for 'empty' edit entry shall satisfy the following rule:

$$\begin{aligned} &\text{media_time} + \text{segment_duration (converted into the media timescale)} \\ &\leq \text{the sum of the sample durations in the track} + \text{ct_offset} \end{aligned}$$

where 'ct_offset' is the composition time of the first sample to be presented when decoding delay is a non-zero value.

4.4.16 Template fields used

In ISO/IEC 14496-12, the concept of "template" fields is defined. This Technical Specification uses the following template fields.

a) Track Header Box

- `layer`;
refer to ISO/IEC 14496-12 for the semantics of this field.
- `alternate_group`;
refer to ISO/IEC 14496-12 for the semantics of this field.
- `volume`;
refer to ISO/IEC 14496-12 for the semantics of this field.
- `matrix`;
refer to ISO/IEC 14496-12 for the semantics of this field.

b) Video Media Header Box

- `graphicsmode`;
refer to ISO/IEC 14496-12 for the semantics of this field.
- `opcolor`;
refer to ISO/IEC 14496-12 for the semantics of this field.

c) Sound Media Header Box

- `Balance`;
refer to ISO/IEC 14496-12 for the semantics of this field.

d) Visual Sample Entry

- `horizresolution`;
refer to ISO/IEC 14496-12 for the semantics of this field. The minimum value of these fields is 72,0. See 4.11 for the values defined.

- `vertresolution`;
refer to ISO/IEC 14496-12 for the semantics of this field. The minimum value of these fields is 72,0. See 4.11 for the values defined.

e) Audio Sample Entry

- `channelcount`;
refer to ISO/IEC 14496-12 for the semantics of this field.
- `samplesize`;
refer to ISO/IEC 14496-12 for the semantics of this field. This field gives the number of bits for the sound sample before compressing.

These declared fields may have non-default values as required. When a file is created, other “template” fields that are not declared above shall be set to their default values. When a file is read, the values in such non-declared “template” fields shall be ignored.

4.5 Additional definitions

4.5.1 Private extension box definitions

4.5.1.1 Private extension boxes

Private Extension Boxes are described in ISO/IEC 14496-12.

Private Extension Boxes use the escape type ‘uuid’ as boxtype, and the UUID as extended_type that is composed of a four-byte type value with a twelve-byte format reserved value. The type value is the printable four-character code specified for each box and the format reserved value is XXXXXXXX-21D2-4FCE-BB88-695CFAC9C740h; the type value defined as follows replaces the XXXXXXXXh in the preceding number.

4.5.1.2 Profile box

4.5.1.2.1 Definition

Box Type : ‘uuid’
Type Value : ‘PROF’
Container : File
Mandatory : Yes
Quantity : Exactly one

This box specifies information about the profile with which the file is in compliance, which is achieved through the private extension with the ‘uuid’ type. This box shall be present after the File Type Box, before Movie Data Box and Movie Box, and also all files compliant with this Technical Specification shall contain one and only one Profile Box.

The profile is reference information for devices to facilitate finding the required capability to decode the content in the file.

A profile entry with the same track_ID for each sample entry shall be in the Profile Box if there are multiple sample entries with at least one different value the for following parameters in a track:

- a codec type;
- the parameter(s) that compose(s) a codec_specific_information;
- the parameters that compose an audio_attribute_flags or video_attribute_flags;
- a pixel aspect ratio for video profile.

A profile entry with track_ID set to ‘0’ is a representative profile entry applied to multiple tracks of the same media that have the same profile information. The representative profile entry is permitted only if all values for above parameters and following one are the same in the tracks:

- a sampling rate for audio profile;
- an average and a maximum bitrate;
- an average and a maximum frame rate.

This makes it possible to avoid that the profile entry with the same value increases immoderately.

4.5.1.2.2 Syntax

// Profile entries

```
aligned(8) class FileGlobalProfileEntry extends FullBox('FPRF', version=0, 0){
    unsigned int(32)    function_flags;
    unsigned int(32)    reserved = 0;
}
```

```
aligned(8) class AudioProfileEntry extends FullBox('APRF', version=0, 0){
    unsigned int(32)    track_ID;
    unsigned int(32)    codec_type;
    unsigned int(32)    codec_specific_information;
    unsigned int(32)    audio_attribute_flags;
    unsigned int(32)    average_bitrate;
    unsigned int(32)    max_bitrate;
    unsigned int(32)    sampling_rate;
    unsigned int(32)    audio_channel_number;
}
```

```
aligned(8) class VideoProfileEntry extends FullBox('VPRF', version=0, 0){
    unsigned int(32)    track_ID;
    unsigned int(32)    codec_type;
    unsigned int(32)    codec_specific_information;
    unsigned int(32)    video_attribute_flags;
    unsigned int(32)    average_bitrate;
    unsigned int(32)    max_bitrate;
    unsigned int(32)    average_frame_rate;
    unsigned int(32)    max_frame_rate;
    unsigned int(32)    visual_size;
    unsigned int(32)    pixel_aspect_ratio;
}
```

// Profile Box

```
aligned(8) class ProfileBox extends Box('uuid', 'PROF'){
    unsigned int(8)    version = 0;
    bit(24)            flags;
    unsigned int(32)    profile_entry_count;
    FileGlobalProfileEntry global_profile;
    // profile entry
    for (i=1; i < profile_entry_count; i++) {
        ProfileEntry()    profile;
    }
}
```

4.5.1.2.3 Semantics

version An 8-bit integer that specifies the version of this box.

flags A 24-bit space for flags; set this field to '0'

profile_entry_count A 32-bit integer that counts the profile entries. The number starts from 1.

global_profile A File Global Profile Entry that contains profile information applied globally to the file.

profile An Audio Profile or Video Profile Entry that contains profile information applied to the audio or video track.

function_flags A 32-bit flag space that indicates which optional functions are applied to this file. The syntax and value are defined in 4.5.3.

track_ID A 32-bit integer that indicates the track identifier to apply this profile information. The value 0 shall be set for a representative profile entry that contains profile information applied to some tracks.

codec_type A 32-bit integer that indicates the codec type of the track(s). The value is a four-character-code copied from format type of sample entry. For example, the value is set to 'avc1' for MPEG-4 AVC, or 'mp4a' for MPEG-4 audio. Note that some codec may have specific information in codec_specific_information field, which is specified according to the codec type.

codec_specific_information A 32-bit integer that specifies more specific information or value that depends on codec. The syntax and value are defined in 4.5.4.

audio_attribute_flags A 32-bit flag space that indicates attribute of audio stream in the track(s). The upper 16 bits contains codec specific attributes specified according to the codec type indicated in the codec_type field and lower 16 bits contains attributes for common use. The flags specified according to the codec type are defined in 4.5.5. The following flags are defined for common use (see Table 6).

Table 6 – Common portion of the audio attribute flags

Values	Description
0001h	Reserved
0002h	Deprecated field. Ignore this field regardless of whatever value it may hold
0004h	dual_mono_audio: indicates that the stream is encoded as a dual monaural audio
others	reserved for future use in this Technical Specification

video_attribute_flags A 32-bit flag space that indicates attribute of video stream in the track(s). The upper 16 bits contain codec specific attributes specified according to the codec type indicated in the codec_type field and lower 16 bits contains attributes for common use. The flags specified according to the codec type are defined in 4.5.5. The following flags are defined for common use (see Table 7).

Table 7 – Common portion of the video attribute flags

Values	Description
0001h	Reserved
0002h	Deprecated field. Ignore this field regardless of whatever value it may hold.
0004h	variable_frame_rate: indicates that the stream has a variable frame rate
0008h	interlaced_video: indicates that the video frames are encoded as interlaced video in a presentation. Note that AVC progressive video may be contained even if this flag is set (see Clause C.4).
others	reserved for future use in this Technical Specification

average_bitrate A 32-bit integer that indicates the average bitrate of the data stream in the track(s), in kilobits per second (kbps). The value is the overall average of bitrate calculated from the Sample Table Box (not considering any edit list). Note that it is permitted to set the average value set at the encoder for initialization when the file is created. The value should be rounded up to the nearest integer if it has a fractional portion. For example, the value is set to 000003E8h (1 000) for 1 Mbps data stream.

max_bitrate A 32-bit integer that indicates the maximum bitrate of the data stream in the track(s), in kilobits per second (kbps). The value is the maximum of the sliding averages of bitrate over 1 s calculated from the Sample Table Box (not considering any edit list). Note that it is permitted to set the maximum value set at the encoder for initialization when the file is created. The value should be rounded up to the nearest integer if it has a fractional portion. For example, the value field is set to 000003E8h (1 000) for 1 Mbps data stream.

sampling_rate A 32-bit integer that indicates the sampling rate of audio stream in the track(s), in hertz (Hz). The value should be rounded up to the nearest integer if it has a fractional portion. For example, the value field is set to 0000BB80h (48 000) for 48 kHz sampling.

audio_channel_number A 32-bit integer that indicates the channel number of audio stream in the track(s). The value should be the total number of audio channels. If there are different channel numbers in the track(s) that correspond(s) to a profile entry, the value should be the largest one among those. For example, the value field is set to 00000006h (6) for the case of 5,1 channels. If there are different channel numbers in the track(s) that correspond(s) to a profile entry, the value should be the largest one among those.

average_frame_rate A 32-bit fixed-point 16,16 number that indicates the average frame rate of the data stream in the track(s), in frames per second (fps). The value is the overall average of frame rate (not considering any edit list) at the time of displaying, not encoding. Note that it is permitted to set the average value set at the encoder for initialization when the file is created. The value should be rounded up to close approximation of the rate. For example, the value should be set to a value between 001DF852h and 001DF854h for 29,97 (30 000 / 1001) fps.

max_frame_rate A 32-bit fixed-point 16,16 number that indicates the maximum frame rate of the data stream in the track(s), in frames per second (fps). The value is the inverse of the smallest sample duration of all samples (not considering any edit list). Note that it is permitted to set the maximum value set at the encoder for initialization when the file is created. The value should be rounded up to close approximation of the rate. For example, the value should be set to a value between 001DF852h and 001DF854h for 29,97 (30 000 / 1 001) fps.

visual_size A 32-bit integer that indicates the maximum visual width and height of the data stream in the track(s), in pixels. The value is the width and height copied from

visual sample entry. The upper 16 bits indicate the width, and the lower 16 bits indicate the height. For example, the value field is set to 028001E0h for 640 pixels of width and 480 pixels of height. If there are different visual sizes in the track(s) that correspond(s) to a profile entry, the value should be the width and height of the largest area among those. For example, the value should be 640x480 when the two visual sizes exist, 720 × 352 and 640 × 480.

pixel_aspect_ratio A 32-bit integer that indicates the pixel aspect ratio of the video frame data in the track(s). The value consists of the horizontal spacing of luma sampling instants in the upper 16-bit portion and the vertical spacing of picture lines in the lower 16-bit portion. The ratio of horizontal spacing to vertical spacing is the same as the aspect ratio given by the horizresolution and vertresolution in a Visual Sample Entry. See 4.11 for the defined values.

4.5.1.3 User specific metadata box

4.5.1.3.1 Definition

Box Type : 'uuid'
 Type Value : 'USMT'
 Container : Movie Box ('moov') or Track Box ('trak')
 Mandatory : No
 Quantity : Zero or one

This box contains objects that declare user information about the containing box and its data (presentation or track), which is achieved through the private extension with the 'uuid' type.

This box may contain the Meta Type Definition Box and the Metadata Box. The Meta Type Definition Box is a container for the definitions of user defined metadata types, and the Metadata Box is a container for informative metadata. The metadata is formatted and listed as a set of units with more specific data types in the Metadata Box.

This box is always required in the Track Box.

4.5.1.3.2 Syntax

```
aligned(8) class UserSpecificMetadataBox extends Box('uuid', 'USMT'){
}
```

4.5.1.4 Metadata box

4.5.1.4.1 Definition

Box Type : 'MTDT'
 Container : User Specific Metadata Box ('uuid' with the type value of 'USMT')
 Mandatory : No
 Quantity : Zero or one

This box contains a list of user specific metadata. All metadata units have associated language codes from ISO 639-2, T and character encoding types.

Multiple metadata units that have the same combination of a data_type_ID, a language, and an encoding_type shall not exist in this box.

This box is always required in the User Specific Metadata Box that is the immediate child of the Track Box.

4.5.1.4.2 Syntax

```
aligned(8) class MetadataBox extends Box('MTDT'){
    unsigned int(16)      number_of_units;
    // metadata units
    for (i=1; i ≤ number_of_units; i++) {
        unsigned int(16)      data_unit_size;
        unsigned int(32)      data_type_ID
        bit(1)      read_only_flag;
        unsigned int(5)[3]  language;
        unsigned int(16)      encoding_type;
        bit(8)      metadata[];
    }
}
```

4.5.1.4.3 Semantics

number_of_unit A 16-bit integer that specifies the number of metadata units.

data_unit_size A 16-bit integer that specifies the number of bytes of a metadata unit that contains this field.

data_type_ID A 32-bit integer that identifies the type of the meta_data. This field contains format reserved data_type_ID. See 4.10.1 for the details of data_type_ID and the format reserved data_type_ID.

read_only_flag A 1-bit flag that indicates the creator of this file (or metadata) requires that this metadata be not modified or deleted unintentionally by a device or software with a function to modify or delete metadata. When editing the metadata for which this flag is set to '1', warning should be displayed to notify users that the metadata is intended to be read only.

language A 16-bit integer that declares the language code from ISO 639-2, T for the metadata. Refer to ISO 639-2/T for the set of three character codes. Each character is packed as the difference between its ASCII value and 60h. Since the code is confined to being three lower-case letters, these values are strictly positive. The code of "undetermined" (und) shall be set for the metadata of binary data and can be used for the metadata that is not dependent on any specific language.

encoding_type A 16-bit integer that specifies the text encoding type for the metadata. The value 0 is used only if the language field is set to "undetermined" (und). The following values are defined (see Table 8).

Table 8 – encoding_type

Values	Description
0000h	non-character data (binary data)
0001h	UTF-16BE (ISO/IEC 10646)
0002h	UTF-8 (ISO/IEC 10646) Optional
others	reserved for future use in this Technical Specification

metadata A field of metadata. If the encoding_type field is set to '0', this field contains binary data. In all other cases, this field contains human-readable information of presentation or track in a null-terminated string. The form of this field may be specified according to each data_type_ID.

4.5.2 Profile configuration

The profile of a file is described in the Profile Box (private extension) defined in 4.5.1.2. The box may have some profile entries that contain a set of profile information.

For PCE AV files, File Global Profile Entry and Profile Entries about main audio and video tracks shall be specified in the Profile Box, but the profile entry for the other tracks may be omitted. Multiple profile entries for a single sample entry shall not exist.

All fields in the Profile Entries shall be set to appropriate values.

Use of the representative profile entry (the entry with track_ID set to '0') is permitted for only main audio tracks and main video tracks. In the case where the representative profile entry is used, one such entry is permitted for the main audio and / or video tracks, and the individual profile entries for the tracks to which the representative profile entry is applied shall not exist.

In a Video Profile Entry for a track containing an AVC video elementary stream, bitrate fields shall be set as follows:

average_bitrate = the average bitrate of NAL CPB, which is set at the encoder for initialization in accordance with regulations for recording modes defined in Annex C;

max_bitrate = the maximum bitrate of NAL CPB in accordance with 4.5.1.2.

4.5.3 Function flags

The profile box (private extension) shall have a FileGlobalEntry that contains profile information applied globally to the file. The FileGlobalEntry contains a function_flags field that is coded as a 32-bit specific value. Players can understand which optional functions are applied to the file using this field.

The following flags are currently defined. The flags not defined in the Table 9 should not be ignored and they should be treated as optional functions as well.

Table 9 – Function flags

Values	Description
10000000h	movie_is_fragmented: indicates that the movie has a Movie Box that contains a Movie Extends Box, which indicates that the movie is possibly fragmented.
20000000h	additional_track_exists: indicates that the presence of the track(s) for presentation with the track_enabled flag set to ON (= 1) in their Track Header Box other than main audio and main video tracks or more than one main audio and/or video track with the track_enabled flag set to ON in their Track Header Box in a file.
40000000h	main_AV_track_is_edited: indicates that the main audio or video track is edited when any of the following applies: a) an edit list exists, except for the case where it consists of one 'invariable' edit entry or one 'initial-offset' edit entry followed by an 'invariable' edit entry; b) either the main audio or the video track is not self-contained; c) the stream does not follow the restrictions for a single stream, as defined by the codec in 4.6.

4.5.4 Codec specific information

4.5.4.1 General

The profile box (private extension) may have some profile entries that contain a set of profile information applied to each track. 4.5.5 discusses codec specific information specified by each codec type supported in this Technical Specification.

4.5.4.2 Specific information for format type 'mp4a'

4.5.4.2.1 General

When codec_type field in Audio Profile Entry is set to the format type 'mp4a', the codec_specific_information field contains an actual audio codec or stream type, and some information.

This codec_specific_information has the following structure in this case:

```

bit(8)      data_form_type;
unsigned int(24) specific_data.

```

The semantics of these fields are defined as follows:

data_form_type An 8-bit code that identifies the data form type. The type specifies actual audio codec and stream type. The following values are currently defined (see Table 10).

Table 10 – Data form type

Values	Description
00h	ISO/IEC 14496-3 (MPEG-4 audio stream)
others	reserved for future use in this Technical Specification

specific_data A 24-bit integer that contains some information specified by data_form_type.

4.5.4.2.2 Specific data for MPEG-4 AAC elementary stream

When the data_form_type field in the codec_specific_information for format type 'mp4a' is set to 00h, the specific_data field has the following structure:

```

bit(11)    reserved = 0;
bit(5)     audioObjectType;
bit(8)     audioProfileLevelIndication.

```

The semantics of these fields are defined as follows.

audioObjectType A 5-bit code that identifies the audio object type. The audio object type identifies the actual bitstream syntax of the MPEG-4 audio stream. This field contains the value of audioObjectType as found in the AudioSpecificConfig (DecoderSpecificInfo), as defined in ISO/IEC 14496-3, contained in the ESDescriptor of the MP4AudioSampleEntry. Refer to ISO/IEC 14496-3 for the values of audio object type.

audioProfileLevelIndication An 8-bit code that identifies the profile and level the MPEG-4 audio stream conforms to. This field contains appropriate value of audioProfileLevelIndication, as defined in ISO/IEC 14496-1.

As the audioProfileLevelIndication, the value for AAC Profile should be set.

For example, the `codec_specific_information` field is set to 00000229h for AAC LC in AAC Profile / Level 2.

4.5.4.3 Specific information for format type 'avc1'

4.5.4.3.1 General

When `codec_type` field in Video Profile Entry is set to the format type 'avc1', the `codec_specific_information` field contains an actual video codec and stream type, and some information.

This `codec_specific_information` has the following structure in this case:

bit(4) `data_form_type`;
unsigned int(28) `specific_data`.

The semantics of these fields are defined as follows.

data_form_type A 4-bit code that identifies the data form type. The type specifies actual video codec and stream type. The following values are currently defined (see Table 11).

Table 11 – Data form type

Values	Description
0h	ISO/IEC 14496-10 (AVC video elementary stream)
others	reserved for future use in this Technical Specification

specific_data A 28-bit integer that contains some information specified by `data_form_type`.

4.5.4.3.2 Specific data for AVC video elementary stream

When the `data_form_type` field in the `codec_specific_information` for format type 'avc1' is set to 0h, the `specific_data` has the following structure:

bit(3) reserved = 0;
bit(1) `entropy_coding_mode_flag`;
unsigned int(8) `AVCProfileIndication`;
unsigned int(8) `profile_compatibility`;
unsigned int(8) `AVCLevelIndication`.

The semantics of these fields are defined as follows:

entropy_coding_mode_flag A 1-bit flag that indicates the entropy decoding mode to be applied. The value shall be the same as `entropy_coding_mode_flag` in Picture Parameter Set (PPS), as defined in ISO/IEC 14496-10, contained in the AVC Configuration Box of the `AVCSampleEntry`.

AVCProfileIndication An 8-bit integer that identifies the profile to process the AVC video elementary stream. This field contains the value of `AVCProfileIndication` as found in the `AVCDecoderConfigurationRecord`, as defined in ISO/IEC 14496-15, contained in the `AVCSampleEntry`. The value is the same as the profile code defined in ISO/IEC 14496-10.

profile_compatibility An 8-bit integer that indicates the constraint set(s) the AVC video elementary stream obeys. This field contains the value of `profile_compatibility` as found in the `AVCDecoderConfigurationRecord`, as defined in ISO/IEC 14496-15, contained in the `AVCSampleEntry`. The value is the same as the byte that occurs between the `profile_idc` and `level_idc` in a sequence parameter set defined in ISO/IEC 14496-10.

AVCLevelIndication An 8-bit integer that identifies the level to process the AVC video elementary stream. This field contains the value of AVCLevelIndication as found in the AVCDecoderConfigurationRecord, as defined in ISO/IEC 14496-15, contained in the AVCSampleEntry. The value is the same as the level code defined in ISO/IEC 14496-10.

For example, the codec_specific_information field is set to 014D401Eh for Main Profile / Level 3.0, and the entropy decoding method to be applied is CABAC, and the AVC video elementary stream obeys all constraints specified in A.2.2. of ISO/IEC 14496-10:2012.

4.5.5 Attribute flags

4.5.5.1 General

The Profile Box (private extension) may have some profile entries that contain a set of profile information applied to each track. This subclause discusses the codec specific portion (upper 16 bits) in video and audio attribute flags, which is specified according to each codec type supported in this Technical Specification.

4.5.5.2 Specific attributes for format type 'mp4a'

When the codec_type field in Audio Profile Entry is set to the format type 'mp4a', the following flags are defined for codec specific portion of the audio_attribute_flags field (see Table 12).

Table 12 – Codec specific portion of audio attribute flags

Values	Description
0001h	SBR_data_present: indicates that the stream may contain SBR data. Currently, this flag is operated only in Class MB.
others	reserved for future use in this Technical Specification.

4.5.5.3 Specific attributes for format type 'avc1'

When the codec_type field in Video Profile Entry is set to the format type 'avc1', the following flags are defined for codec specific portion in the video_attribute_flags field (see Table 13).

Table 13 – Codec specific portion of video attribute flags

Values	Description
0001h	IDR_intervals_more_than_n_seconds*: indicates that one or more intervals between IDR pictures in a stream may be more than n seconds*, where n is the intervals specified for each Class defined in 4.6.3.2. When this flag is set to 0, every interval between IDR pictures is less than or equal to n seconds*. Currently, this flag is operated only in Class SD, Class network SD, Class HD and Class network HD.
0002h	recovery_point_sei_attached: indicates that a Recovery Point SEI message is attached to an access unit in the stream. Currently, this flag is operated only in Class SD, Class network SD, Class HD and Class network HD.
others	reserved for future use in this technical specification.

4.6 Operational rules for tracks

4.6.1 General

Subclause 4.6 describes operational rules for tracks contained in PCE AV files. The tracks are basically composed in conformity to ISO/IEC 14496-12, ISO/IEC 14496-14 and ISO/IEC 14496-15.

4.6.2 Main audio track

4.6.2.1 General

A PCE AV file may contain one or more main audio tracks. The main audio track is an audio track that contains streams described in 4.6.2. The general nature of ISO/IEC 14496-14 is partly exercised by this format for a main audio track structure. It therefore uses the following:

- a) a handler_type of 'soun' in the HandlerBox;
- b) a sound media header 'smhd';
- c) a format type of 'mp4a' in the SampleDescriptionBox;
- d) the MP4AudioSampleEntry as defined in MP4 file format for 'mp4a' and
- e) a presentation_type of 00000001h(original/main) in the track property of metadata box.

The syntax and values for the Track Box and its sub-boxes shall conform to ISO/IEC 14496-12, and the following fields of each box shall be set to the following specified values. There are some "template" fields declared to use; refer to 4.4.16.

- Track header box

flags = 000007h, except for the case where the track belongs to an alternate group;
 layer = 0;
 volume = 0100h;
 matrix = {00010000h,0,0,0, 00010000h,0,0,0, 40000000h};
 width = 0;
 height = 0.

- Handler reference box

name = "Sound Media Handler".

- Sound media header box

balance = 0.

4.6.2.2 MPEG-4 AAC elementary stream

4.6.2.2.1 General

An MPEG-4 AAC elementary stream shall be stored in the track in accordance with ISO/IEC 14496-14 basically. The following are limitations for using MPEG-4 AAC elementary streams.

- The parameter values of DecoderConfigDescriptor, program_config_element, and Sample Entry shall be consistent.
- Only one AU shall be handled as a sample. All AUs are a random access point (sync sample) and therefore, the Sync Sample Box shall not be used.

4.6.2.2.2 Sample entry

The syntax and values for sample entry shall conform to MP4AudioSampleEntry ('mp4a') defined in ISO/IEC 14496-14, and the following fields shall be set to the following specified values. There are some "template" fields declared to use; refer to 4.4.16.

The actual format type and specific parameters for each are specified in an ESD Box ('esds'), as described below.

channelcount = 1 (for single mono), 2 (for stereo or dual mono) , or 6 (for 5,1 channels);
 sampleRate = 48 000 (when SBR is not used), or 24 000 (when SBR is used);
 ES = ESD Box, see 4.7.1.2.

4.6.3 Main video track

4.6.3.1 General

A PCE AV file shall contain one or more main video tracks. The main video track is a video track that contains streams described in this subclause. The general nature of ISO/IEC 14496-14 and ISO/IEC 14496-15 is partly exercised by this format for a main video track structure. It therefore uses the following:

For AVC video elementary stream:

- a) a handler_type of 'vide' in the HandlerBox;
- b) a video media header 'vmhd';
- c) a format type of 'avc1' in the SampleDescriptionBox;
- d) the AVCSampleEntry as defined in AVC file format for 'avc1' and
- e) a presentation_type of 00000001h(original/main) in the Track Property of Metadata Box.

The syntax and values for the Track Box and its sub-boxes shall conform to ISO/IEC 14496-12, and the following fields of each box shall be set to the following specified values. There are some "template" fields declared to use; refer to 4.4.16.

- Track header box

flags = 000007h, except for the case where the track belongs to an alternate group;

volume = 0;

matrix = {00010000h,0,0,0, 00010000h,0,0,0, 40000000h}, if other tracks for visual presentation with track_enabled flags set to ON (=1) in their Track Header Box do not exist; Note that if the file contains rotated video, this field's settings shall conform to the rules in Clause D.4;

width = equal to width in visual sample entry, if other tracks for visual presentation with track_enabled flags set to ON (=1) in their Track Header Box do not exist; Note that if the file contains rotated video, this field's settings shall conform to the rules in Clause D.4;

height = equal to height in visual sample entry, if other tracks for visual presentation with track_enabled flags set to ON (=1) in their Track Header Box do not exist. Note that if the file contains rotated video, this field's settings shall conform to the rules in Clause D.4.

NOTE In the case of Class PT, Class MB, Class HD and Class network HD, the width and height in the Track Header box may be specified to form a square grid.

- Handler reference box

name = "Video Media Handler".

- Video media header box

graphicsmode = 0;

opcolor = {0, 0, 0}.

The syntax and values for the following box shall conform to ISO/IEC 14496-12, and the following field of the box should be set to the following specified value.

- Media header box

timescale (when encoded as progressive video)

EA60h (60 000) for 59,94 fps;

C350h (50 000) for 50 fps;

7530h (30 000) for 14,985, or 29,97 fps;

61A8h (25 000) for 25 fps;

5DC0h (24 000) for 23,976[fps];

30D4h (12 500) for 12,5 fps;

timescale (when encoded as interlaced video)

EA60h (60 000) for 29,97 fps (59,94 fields/s);

C350h (50 000) for 25 fps (50 fields/s).

4.6.3.2 AVC video elementary stream

4.6.3.2.1 General

An AVC video elementary stream shall be stored in the track in accordance with ISO/IEC 14496-14 and ISO/IEC 14496-15, basically. The following are limitations for using AVC video elementary streams.

- Only one access unit shall be handled as a sample, and every access unit consisting of an IDR picture shall be handled as a sync sample.
- Sequence and picture parameter set NAL units shall be stored in the sample descriptions of the track.
- One or more sequence and picture parameter set NAL units shall be stored in the AVCDecoderConfigurationRecord.
- The total number of sequence parameter set NAL units or picture parameter set NAL units contained in all sample entries (AVCDecoderConfigurationRecord) in a track shall be 256 maximum, respectively.

4.6.3.2.2 Sample entry

The syntax and values for Sample Entry shall conform to AVCSampleEntry ('avc1') defined in ISO/IEC 14496-15, and the following fields shall be set to the following specified values. There are some "template" fields declared to use; refer to 4.4.16.

The resolution fields shall contain appropriate values from the pixel aspect ratio defined in 4.11. Neither MPEG4BitRateBox ('btrt') nor MPEG4ExtensionDescriptorsBox ('m4ds') shall be used.

compressorname = "AVC Coding";
config = AVC configuration box, see 4.7.2.2;

4.6.4 Thumbnail picture track

4.6.4.1 General

A PCE AV file may contain one or more thumbnail picture tracks. A thumbnail track is a representative still image of presentation (extracted from the stream of a main video track typically). The general nature of ISO/IEC 14496-14 is partly exercised by this format for a thumbnail track structure. It therefore uses the following:

- a) a handler_type of 'vide' in the HandlerBox;
- b) a track reference type 'sync';
- c) a video media header 'vmhd';
- d) a format type of 'mp4v' in the SampleDescriptionBox;
- e) the MP4VisualSampleEntry as defined in MP4 file format for 'mp4v' and
- f) a presentation_type of 00000002h(track for thumbnail picture data) in the track property of metadata box.

The syntax and values for the Track Box and its sub-boxes shall conform to ISO/IEC 14496-12, and the following fields of each box shall be set to the following specified values. There are some "template" fields declared to use; refer to 4.4.16.

- Track header box

flags = 000006h (track_enabled flag = 0);
layer = 0;
alternate_group = 0;
volume = 0100h;

matrix = {00010000h,0,0,0, 00010000h,0,0,0, 40000000h};
width = equal to the width in the visual sample entry;
height = equal to the height in the visual sample entry.

- Track Reference Type Box of 'sync' in Track Reference Box
track_IDs = track ID(s) of the main video track this track represents;

- Handler reference box
name = "Video Media Handler".

- Video media header box
graphicsmode = 0;
opcolor = {0,0,0}.

- Data Reference Box
entry_count = 1;

- Data Entry Url Box
flags = 1 (self-contained);

The syntax and values for the following box shall conform to ISO/IEC 14496-12, and the following field of the box should be set to the following specified value.

- Media HeaderI Box
timescale = 7 530h (30 000).

4.6.4.2 JPEG image sequence for thumbnail pictures

4.6.4.2.1 General

The following are limitations for using JPEG image sequences for thumbnail pictures.

- The parameter values of DecoderConfigDescriptor and Sample Entry shall be consistent.
- Each JPEG image for a thumbnail picture is handled as a sample. All pictures are random access points (sync samples) and therefore, the Sync Sample Box shall not be used.

The total of all sample durations (sample_delta in the Time to Sample Box) shall be equal to or less than the duration of the entire movie converted into the media timescale of this media, which is equivalent to the duration of the longest of all tracks except thumbnail picture tracks themselves.

4.6.4.2.2 Sample entry

The syntax and values for sample entry shall conform to MP4VisualSampleEntry ('mp4v') defined in ISO/IEC 14496-14, and the following fields shall be set to the following specified values. There are some "template" fields declared to use; refer to 4.4.16.

The resolution fields shall contain appropriate values from the pixel aspect ratio defined in 4.11.

ES = ESD Box, see 4.7.3.

4.7 Operational rules for media data

4.7.1 MPEG-4 AAC elementary stream

4.7.1.1 General

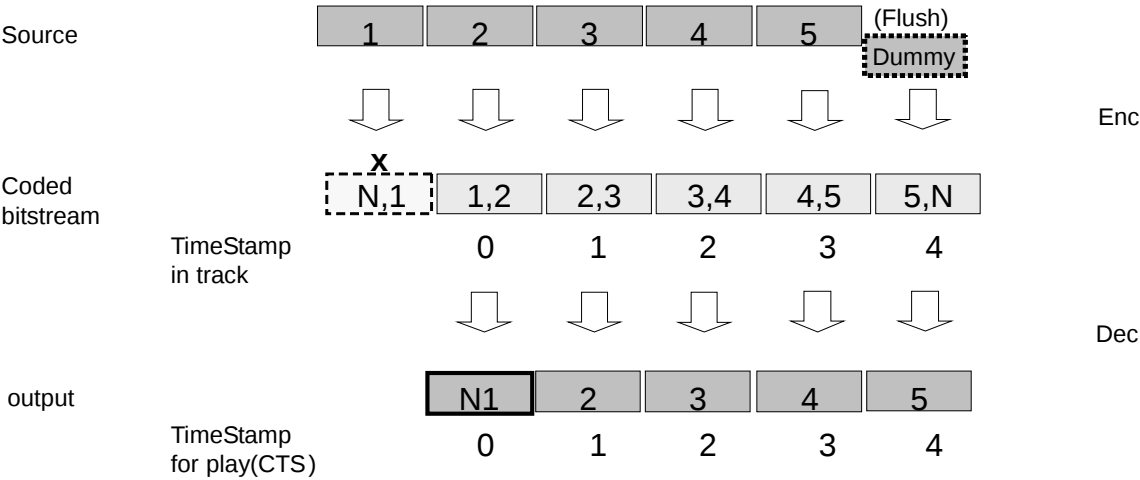
For MPEG-4 AAC elementary streams, a bitstream of AAC LC object type shall be used. It shall be a Raw Data stream, and neither ADTS nor ADIF shall be used.

If the stream contains SBR data, the explicit backward compatible signalling as defined in subpart 1 of ISO/IEC 14496-3 shall be used.

Since the AAC codec is based on overlap transform, and it does not establish a one-to-one relationship between input/output audio frames and audio decoding units (AUs) in bitstreams, it is necessary to be careful in handling timestamps in a track. Figure 4 shows an example of AAC bitstream in the track.

In a single MPEG-4 AAC elementary stream, only one sample entry may exist and the following fields shall not change in the stream. If any of these changes, the function flag of value 40000000h shall be set:

audioObjectType
samplingFrequencyIndex
samplingFrequency (if samplingFrequencyIndex = fh)
channelConfiguration
samplesize



IEC 1902/09

Figure 4 – Example of AAC bitstream

In Figure 4, the first block of the bitstream is AU [1, 2], which is created from input audio frames [1] and [2]. Depending on the encoder implementation, the first block may be AU [N,1] (where N indicates a silent interval inserted by the encoder), but this type of AU might cause failure in synchronization and therefore shall not be included in the file.

To include the last input audio frame (i.e., [5] of source in Figure 4) into the bitstream for encoding, it is necessary to terminate it with a silent interval and include AU [5, N] into the bitstream. This produces the same number of input audio frames, AUs, and output audio frames, eliminating time difference.

When a bitstream is created using the method described above, the decoding result of the first AU does not necessarily correspond to the first input audio frame. This is because of the lack of the first part of the bitstream in overlap transform. Thus, the first audio frame (21[ms] per frame when sampled at 48 kHz, for example) is not guaranteed to play correctly. In this

case, it is up to decoder implementations to decide whether the decoded output audio frame [N1] should be played or muted.

With these things considered, the content should be created by making the first input audio frame a silent interval.

4.7.1.2 ESDescriptor

An ESDescriptor (ESD) is contained in the ESD Box ('esds') in MP4AudioSampleEntry.

The syntax and values for ESDescriptor shall conform to ISO/IEC 14496-1, and the following fields shall be set to the following specified values. They are identical with those defined in ISO/IEC 14496-14. Descriptors other than those below shall not be used.

```
ES_ID = 0;
streamDependenceFlag = 0;
URL_Flag = 0;
OCRstreamFlag = 0 (false);
streamPriority = 0;
decConfigDescr = DecoderConfigDescriptor, see 4.7.1.3;
slConfigDescr = SLConfigDescriptor, predefined type 2.
```

4.7.1.3 DecoderConfigDescriptor

The syntax and values for DecoderConfigDescriptor shall conform to ISO/IEC 14496-1, and the following fields shall be set to the following specified values.

In this descriptor, DecoderSpecificInfo shall always be used, and no ProfileLevelIndicationIndexDescriptor(s) shall be used.

```
objectTypeIndication = 40h (ISO/IEC 14496-3);
streamType = 05h (Audio Stream);
upStream = 0;
decSpecificInfo = DecoderSpecificInfo, see 4.7.1.4.
```

4.7.1.4 DecoderSpecificInfo

A DecoderSpecificInfo consists of AudioSpecificConfig in accordance with ISO/IEC 14496-1.

The syntax and values for the AudioSpecificConfig shall conform to ISO/IEC 14496-3, and the following fields shall be set to the following specified values.

```
audioObjectType = 2 (AAC LC);
channelConfiguration = 0 (for dual mono or 5,1 channels), 1 (for single mono), 2 (for stereo), or 6 (for 5,1 channels);
GASpecificConfig, see 4.7.1.5.
```

If the stream contains SBR data, syncExtensionType and the subsequent syntax elements required shall appear as specified in ISO/IEC 14496-3. Otherwise, they shall not appear.

Only if the stream is encoded as dual monaural or 5,1 channel audio, a channelConfiguration may be set to '0', and a program_config_element that contains program configuration data is used to specify composition of channel elements. See 4.7.1.6 for details on the program_config_element.

Channel assignment shall not be changed over the audio stream that makes up a track.

4.7.1.5 GASpecificConfig

The syntax and values for GASpecificConfig shall conform to ISO/IEC 14496-3, and the following fields shall be set to the following specified values.

```
frameLengthFlag = 0 (1 024 lines IMDCT);
```

```

dependsOnCoreCoder = 0;
extensionFlag = 0;
if (channelConfiguration == 0) {
    program_config_element, see 4.8.1.5;
}

```

4.7.1.6 program_config_element

The syntax and values for `program_config_element` shall conform to ISO/IEC 14496-3, and the following fields shall be set to the following specified values.

```

element_instance_tag = 0;
object_type = 1 (AAC LC);
sampling_frequency_index = 3 (for 48 kHz) or 6 (for 24 kHz);
num_front_channel_elements = 2;
num_side_channel_elements = 0;
num_back_channel_elements = 0 (for dual mono) or 1 (for 5,1 channels);
num_lfe_channel_elements = 0 (for dual mono) or 1 (for 5,1 channels);
num_assoc_data_elements = 0;
num_valid_cc_elements = 0;
mono_mixdown_present = 0;
stereo_mixdown_present = 0;
matrix_mixdown_idx_present = 0 (for dual mono or 5,1 channels) or 1 (for 5,1 channels);
if (matrix_mixdown_idx_present == 1) {
    matrix_mixdown_idx = 0 to 3;
    pseudo_surround_enable = 0 or 1;
}
front_element_is_cpe [0] = 0;
front_element_is_cpe [1] = 0 (for dual mono) or 1 (for 5,1 channels);
back_element_is_cpe [0] = N/A (for dual mono) or 1 (for 5,1 channels).

```

The `program_config_element` is used only if `channelConfiguration` in `DecoderSpecificInfo` is set to '0', which is the case of dual monaural or 5,1 channel audio.

The `program_config_element` should not be contained in a stream. A player shall always use the `program_config_element` in the ESD Box for decoding an audio stream.

If PCE AV file contains one or more 5,1 channel audio tracks but it does not contain a stereo audio track correspondent to those 5,1 channel audio tracks, the mixdown parameters shall be adequately set in the `program_config_element`.

4.7.1.7 Syntactic elements

4.7.1.7.1 General

The syntax and values for syntactic elements shall conform to ISO/IEC 14496-3. The following element is prohibited for use in an MPEG-4 AAC elementary stream:

`coupling_channel_element` (CCE)

In Class HD and Class network HD, `program_config_element` is also prohibited in an MPEG-4 AAC elementary stream. See also 4.7.1.6.

The following elements are allowed in an MPEG-4 AAC elementary stream, but they shall not be interpreted.

`fill_element` (FIL)
`data_stream_element` (DSE)

If the stream is dual mono or 5,1 channel audio, the values of `element_instance_tag` for SCEs or CPEs shall be different.

4.7.1.7.2 Arrangement of syntactic elements

In case of Class HD or Class network HD, syntactic elements shall be arranged in the following order for the channel configurations below. In case of other Classes, syntactic elements should be arranged in the following order.

<CPE><FIL><TERM>... for stereo

<SCE><SCE><FIL><TERM>... for dual mono

<SCE><CPE><CPE><LFE><FIL><TERM>... for 5,1 channels

Note that Angled brackets (<>) are delimiters for syntactic elements.

4.7.1.7.3 individual_channel_stream

The syntax and values for individual_channel_stream shall conform to ISO/IEC 14496-3. The following fields shall be set to the following specified values.

gain_control_data_present = 0.

4.7.1.7.4 ics_info

The syntax and values for ics_info shall conform to ISO/IEC 14496-3. The following fields shall be set to the following specified values.

predictor_data_present = 0.

4.7.1.8 Peak bitrate

The peak bitrate is a bitrate calculated from the maximum number of bits allowed in each raw_data_block().

For example, if the sampling frequency is 48 kHz and the peak bitrate is 384 kbps, the allowed maximum number of bits in a raw_data_block() is calculated as follows:

maximum size of raw_data_block() = $384\ 000 \times 1\ 024 / 48\ 000 = 8\ 192$ [bits]

4.7.2 AVC video elementary stream

4.7.2.1 General

An AVC Video elementary stream shall be encoded in conformity to ISO/IEC 14496-10. The limitations for using AVC Video elementary streams are defined according to the Class. The operational rules for AVC Video elementary streams that conform to PCE AV Classes are defined from 4.7.2.3 through 4.7.2.8.

For each sync(IDR) sample, the decoding delay calculated by the composition time of the first presented picture after decoding from a sync(IDR) sample minus the decoding time of the sync(IDR) sample shall be a period of two frames or less. Note that it may be necessary to set the value of 'media_time' in the Edit List Box ('elst') to the value of composition time of the first sample to be presented if the AVC video elementary stream is encoded the way that requires decoding delay, since the value of 'sample_offset' in the Composition Time to Sample Box ('ctts') is restricted to be non-negative. See 4.9.2.2 for details.

In a single AVC video elementary stream, the following fields shall not change in the stream. If either of these changes, the function flag of value 40000000h shall be set:

- profile_idc
- level_idc
- pic_width_in_mbs_minus1
- pic_height_in_map_units_minus1
- frame_mbs_only_flag
- direct_8x8_inference_flag
- time_scale
- num_units_in_tick
- cpb_cnt_minus1
- bit_rate_scale
- bit_rate_value_minus1
- cpb_size_scale
- cpb_size_value_minus1

entropy_coding_mode_flag

In addition to the above parameters, the values of the following Classes shall not change in the stream.

Class PT and Class MB

all other fields in VUI parameters except aspect_ratio_idc, sar_height and sar_width

all other fields in HRD parameters

Class SD and Class network SD

all other fields in VUI parameters

all other fields in HRD parameters

Class HD and Class network HD

aspect_ratio_idc, sar_height and sar_width

When encoding, a bitstream shall satisfy Type I and Type II HRD conformance point.

4.7.2.2 AVCDecoderConfigurationRecord

An AVCDecoderConfigurationRecord is contained in the AVC Configuration Box ('avcC') in AVCSampleEntry.

The syntax and values for AVCDecoderConfigurationRecord shall conform to ISO/IEC 14496-15.

An AVCDecoderConfigurationRecord contains Sequence Parameter Sets and Picture Parameter Sets. See 4.7.2.3 through 4.7.2.8 for the setting of Sequence Parameter Set and Picture Parameter Set.

4.7.2.3 Class PT

4.7.2.3.1 General

An AVC video elementary stream compliant with Class PT shall conform to the following rules.

- As entropy coding, only CABAC shall be used.
- NAL unit types other than 1 or 5 to 11 shall not be used.
- One access unit (= 1 picture) consists of one slice.
- One RBSP shall be stored in only one nal_unit().
- One SEI RBSP shall consist of only one SEI message.
- A picture shall be encoded only as a frame.
- The value of primary_pic_type in an access unit delimiter shall be 0, 1, or 2 for I-frames, P-frames and B-frames, respectively.
- For a B-frame, the value of nal_ref_idc shall be 0.
- If two or more B-frames occur successively immediately before an I- or P-frame in display order, they shall be placed immediately after the I- or P-frame in the AVC video elementary stream.
- Decoding order for I- and/or P-frame shall be the same as their display order.
- The decoding order for a B-frame and the subsequent B-frame shall be the same as their display order.
- A B-frame may refer to only I- or P-frame, which immediately occur before and/or after in display order.
- The number of frames to be stored in DPB shall be 3 or less.
- More than three B-frames shall not appear successively in display order.
- The number of reference frames shall be 3 or less.

- The maximum total size of non-VCL NAL units in an access unit shall be 1 024 bytes or smaller.
- The total size of BinCountsInNALunit(s) for all VCL NAL unit(s) composing a frame shall be $1,2 \times 10^6$ bins or less.
- The PPS referred to by a slice and the SPS referred to by that PPS are restricted to those input to the decoder immediately before the slice, that is, the number of Sequence Parameter Set and Picture Parameter Set in a single AVC Configuration Box ('avcC') shall be only one, respectively.
- Encoding shall be done so that no gap occurs in presentation during playback of a stream.
- A display period of the following pictures should be 5 seconds* or less.
 - From an IDR-picture to the picture immediately before the next IDR-picture (in decoding order).
 - From the last IDR-picture to the last picture of the stream (in decoding order).

The decoding time of a sample corresponding to a frame in a track is restricted with the following conditions:

- The decoding times of samples to be decoded equal to or after CT0, where CT0 is the composition time of the first sample to be presented, shall be equal to the composition time of one of the samples in the track.
- No additional restriction is defined for the decoding times of samples before CT0.

End of stream (end_of_stream_rbsp()) should not be present. If it is present, it shall not be handled as a sample by itself and shall be contained in the sample that contains the last access unit in decoding order.

At least one non-reference picture should be present within the interval of 5 seconds* in a stream.

4.7.2.3.2 Structure of an access unit

NAL unit data that compose an access unit (AU) and the order among them shall conform to the following rules. Note that a filler data NAL unit shall not be used.

The structure of AUs shall be as follows:

Access unit delimiter NAL unit;
 SEI NAL unit(s);
 Coded slice of a non-IDR or an IDR picture NAL unit(s);
 End of sequence NAL unit – if exists;
 End of stream NAL unit – if exists.

4.7.2.3.3 Sequence parameter set

The syntax and values for Sequence Parameter Set shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

profile_idc = 77 (Main Profile);
 constraint_set0_flag = 0;
 constraint_set1_flag = 1;
 constraint_set2_flag = 0;
 constraint_set3_flag = 0;
 num_ref_frames = 1 to 3;
 gaps_in_frame_num_value_allowed_flag = 0;
 pic_width_in_mbs_minus1 = 19 (for QVGA), 29 (for 480 × 270) , or 39 (for VGA);
 pic_height_in_map_units_minus1 = 14 (for QVGA), 16 (for 480 × 270), or 29 (for VGA);
 frame_mbs_only_flag = 1;

```

frame_cropping_flag = 0 (for QVGA and VGA) or 1 (for 480 × 270);
if (frame_cropping_flag == 1) {
    frame_crop_left_offset = 0;
    frame_crop_right_offset = 0;
    frame_crop_top_offset = 0;
    frame_crop_bottom_offset = 1.
}
vui_parameters_present_flag = 1;
vui_parameters() = VUI parameters.

```

The value of level_idc should be set to the appropriate code corresponding to the stream.

The syntax and values for VUI parameters shall conform to ISO/IEC 14496-10. The following fields shall be set to the following specified values:

```

aspect_ratio_info_present_flag = 1;
aspect_ratio_idc = 1 or 255;
if (aspect_ratio_idc == 255) {
    sar_width = 4 (for QVGA and VGA) or 3 (for 480 × 270);
    sar_height = 3 (for QVGA and VGA) or 4 (for 480 × 270);
}
if (video_signal_type_present_flag == 1) {
    video_full_range_flag = 0;
    if (colour_description_present_flag == 1) {
        colour_primaries = 5 (if the frame rate is 25 fps) or 6 (if the frame rate is 29,97 fps or 23,976 fps);
        transfer_characteristics = 5 (if the frame rate is 25 fps) or 6 (if the frame rate is 29,97 fps or 23,976 fps);
        matrix_coefficients = 5 (if the frame rate is 25 fps) or 6 (if the frame rate is 29,97 fps or 23,976 fps);
    }
}
chroma_loc_info_present_flag = 0;
timing_info_present_flag = 1;
fixed_frame_rate_flag = 1;
nal_hrd_parameters_present_flag = 1;
vcl_hrd_parameters_present_flag = 1;
    hrd_parameters() = HRD parameters;
low_delay_hrd_flag = 0;
pic_struct_present_flag = 1;
bitstream_restriction_flag = 1;
max_bytes_per_pic_denom = 2 to 16;
num_reorder_frames = 0 to 3;
max_dec_frame_buffering = 0 to 3.

```

The value of aspect_ratio_idc may be set to 255 only in the following cases:

- a) for the visual size of QVGA or VGA, when the AVC video elementary stream is 16:9 squeezed video;
- b) for the visual size of 480 × 270, when the display aspect ratio is 4:3. A widescreen stream whose display aspect ratio is 16:9 should be recorded as squeezed video.

The syntax and values for HRD parameters shall conform to ISO/IEC 14496-10. The following fields shall be set to the following specified values:

```
cpb_cnt_minus1 = 0.
```

4.7.2.3.4 Picture parameter set

The syntax and values for Picture Parameter Set shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```

entropy_coding_mode_flag = 1;
num_ref_idx_l0_default_active_minus1 = 0 to 2;
num_ref_idx_l1_default_active_minus1 = 0 to 2.

```

4.7.2.3.5 Supplemental enhancement information

The syntax and values for Supplemental Enhancement Information (SEI) shall conform to ISO/IEC 14496-10.

A Picture Timing SEI message shall be present in every access unit.

A Buffering Period SEI message shall be present in every IDR access unit.

The following SEI messages shall not be present:

- Pan-scan rectangle SEI message;
- Stereo Video Information SEI message.

The syntax and values for Buffering Period SEI message shall conform to ISO/IEC 14496-10. In Buffering Period SEI messages, the value of initial_cpb_removal_delay shall be 90 000 or less.

The syntax and values for Picture Timing SEI message shall conform to ISO/IEC 14496-10. In Picture Timing SEI messages, the following fields shall be set to the following specified values:

```
pic_struct = 0.
```

It is recommended that the following fields be operated and set to the following specified values:

```

clock_timestamp_flag[0] = 1;
ct_type = 0.

```

4.7.2.3.6 Slice header

The syntax and values for Slice Header shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```

first_mb_in_slice = 0;
slice_type = 5 to 7;
if (num_ref_idx_active_override_flag == 1) {
    num_ref_idx_l0_active_minus1 = 0 to 2;
    num_ref_idx_l1_active_minus1 = 0 to 2;
}

```

4.7.2.3.7 Reference picture list modification

The syntax and values for Reference Picture List Modification shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```

ref_pic_list_modification_flag_l0 = 0;
ref_pic_list_rmodification_flag_l1 = 0.

```

4.7.2.3.8 Prediction weight table

The syntax and values for Prediction Weight Table shall conform to ISO/IEC 14496-10.

When explicit weighted prediction is used, the following fields should be set to the following specified values:

```
luma_log2_weight_denom 0 to 6;
```

chroma_log2_weight_denom $\leq$ 6.

4.7.2.3.9 Decoded reference picture marking

The syntax and values for Decoded Reference Picture Marking shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

long_term_reference_flag = 0;
adaptive_ref_pic_marking_mode_flag = 0.

In the case of an IDR picture, no_output_of_prior_pics_flag shall be set to zero.

4.7.2.4 Class MB

4.7.2.4.1 General

An AVC video elementary streams that conform to the Class MB shall conform to the following rules.

- NAL unit types other than 1 or 5 to 10 shall not be used.
- A display period of the following pictures should be 5 seconds* or less.
 - From an IDR-picture to the picture immediately before the next IDR-picture (in decoding order).
 - From the last IDR-picture to the last picture of the stream (in decoding order).
- The value of primary_pic_type in an access unit delimiter shall be 0 (for I-frames) or 1 (for P-frames).
- Decoding order for I- and/or P-frame shall be the same as their display order.
- The number of frames to be stored in DPB shall be 3 or less.
- The number of reference frames shall be 3 or less.
- Encoding shall be done so that no gap occurs in presentation during playback of a stream.
- The maximum total size of non-VCL NAL units in an access unit shall be 1 024 bytes or smaller.
- One RBSP shall be stored in only one nal_unit().
- The PPS referred to by a slice and the SPS referred to by that PPS are restricted to those input to the decoder immediately before the slice, that is, the number of sequence parameter set and picture parameter set in a single AVC Configuration Box ('avcC') shall be only one, respectively.
- A picture timing SEI message shall be present in every access unit.
- A buffering period SEI message shall be present in every IDR access unit.
- A pan-scan rectangle SEI message shall be present in every IDR access unit only when pan scan (optional) is operated.

It is recommended that the stream be encoded so that input to CPB is decoded within 1,5 s.

4.7.2.4.2 Structure of an access unit

NAL unit data that compose an access unit (AU) and the order among them shall conform to the following rules. Note that a filler data NAL unit shall not be used.

The structure of AUs shall be as follows:

Access unit delimiter NAL unit;
SEI NAL unit(s);
Coded slice of a non-IDR or an IDR picture NAL unit(s);

End of sequence NAL unit – if exists.

4.7.2.4.3 Sequence parameter set

The syntax and values for sequence parameter set shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```

profile_idc = 66 (Baseline Profile);
constraint_set0_flag = 1;
constraint_set1_flag = 1;
constraint_set2_flag = 1;
constraint_set3_flag = 0;
pic_order_cnt_type = 2;
num_ref_frames = 1 to 3;
gaps_in_frame_num_value_allowed_flag = 0;
pic_width_in_mbs_minus1 = 19 (for QVGA and 320 × 180), 29 (for 480 × 270) or 39 (for 640
× 360 and VGA);
pic_height_in_map_units_minus1 = 11 (for 320 × 180), 14 (for QVGA), 16 (for 480 × 270), 22
(for 640 × 360) or 29 (for VGA);
direct_8x8_inference_flag = 1;
frame_cropping_flag = 0 (for QVGA and VGA) or 1 (for 320 × 180, 480 × 270 and 640 × 360);
if (frame_cropping_flag == 1) {
    frame_crop_left_offset = 0;
    frame_crop_right_offset = 0;
    frame_crop_top_offset = 0;
    frame_crop_bottom_offset = 6 (for 320 × 180), 1 (for 480 × 270) or 4 (for 640 × 360);
}
vui_parameters_present_flag = 1;
vui_parameters() = VUI parameters.

```

The value of level_idc should be set to the appropriate code corresponding to the stream.

The syntax and values for VUI parameters shall conform to ISO/IEC 14496-10. The following fields shall be set to the following specified values.

For Level 1:

```

aspect_ratio_info_present_flag = 0;
overscan_info_present_flag = 0;
video_signal_type_present_flag = 0;
chroma_loc_info_present_flag = 0;
timing_info_present_flag = 1;
num_units_in_tick = 1 001 or 1 200;
time_scale = 30 000;
fixed_frame_rate_flag = 1;
nal_hrd_parameters_present_flag = 1;
vcl_hrd_parameters_present_flag = 1;
    hrd_parameters() = HRD parameters;
low_delay_hrd_flag = 0;
pic_struct_present_flag = 0;
if (bitstream_restriction_flag == 1) {
    max_bytes_per_pic_denom = 2 to 16;
    log2_max_mv_length_horizontal = 0 to 9;
    log2_max_mv_length_vertical = 0 to 9;
    num_reorder_frames = 0;
    max_dec_frame_buffering = 0 to 3;
}

```

For Level 2, Level 3 and Level 4:

```

aspect_ratio_info_present_flag = 1;
aspect_ratio_idc = 1, 14 or 255 (for QVGA, 480 × 270 and VGA);
if (aspect_ratio_idc == 255) {

```

```

        sar_width = 4 (for QVGA or VGA) or 3 (for 480 × 270);
        sar_height = 3 (for QVGA or VGA) or 4 (for 480 × 270);
    }
    chroma_loc_info_present_flag = 0;
    timing_info_present_flag = 1;
    num_units_in_tick = 4 004 (for 29,97p), 4 800 (for 25p), 8 008 (for 14,985p) or 9 600 (for
    12,5p);
    time_scale = 240 000;
    fixed_frame_rate_flag = 1;
    nal_hrd_parameters_present_flag = 1;
    vcl_hrd_parameters_present_flag = 1;
    hrd_parameters() = HRD parameters;
    low_delay_hrd_flag = 0;
    pic_struct_present_flag = 0;
    if (bitstream_restriction_flag == 1) {
        max_bytes_per_pic_denom = 2 to 16;
        num_reorder_frames = 0;
        max_dec_frame_buffering = 0 to 3;
    }
}

```

The value of aspect_ratio_idc may be set to 255 only in level 2, 3 or 4:

- a) for the visual size of QVGA or VGA, when the AVC video elementary stream is 16:9 squeezed video;
- b) for the visual size of 480 × 270, when the display aspect ratio is 4:3.

A widescreen stream whose display aspect ratio is 16:9 should be recorded as squeezed video.

The syntax and values for HRD parameters shall conform to ISO/IEC 14496-10. If hrd_parameters() is present, the following fields shall be set to the following specified values:

```

    cpb_cnt_minus1 = 0;
    time_offset_length = 0 (for Level 1).

```

4.7.2.4.4 Picture parameter set

The syntax and values for Picture Parameter Set shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```

    bottom_field_pic_order_in_frame_present_flag = 0;
    num_slice_groups_minus1 = 0;
    num_ref_idx_l0_default_active_minus1 = 0 to 2;
    num_ref_idx_l1_default_active_minus1 = 0;
    pic_init_qs_minus26 = 0;
    constrained_intra_pred_flag = 0;
    redundant_pic_cnt_present_flag = 0.

```

4.7.2.4.5 Supplemental enhancement information

The syntax for Supplemental Enhancement Information (SEI) shall conform to ISO/IEC 14496-10. Only the following SEI messages may be present.

```

    buffering period SEI message;
    picture timing SEI message;
    pan-scan rectangle SEI message;

```

The syntax and values for buffering period SEI message shall conform to ISO/IEC 14496-10. In buffering period SEI messages, in Level 1, the value of initial_cpb_removal_delay shall be 135 000 or less, and it is recommended to set the sum of initial_cpb_removal_delay and initial_cpb_removal_delay_offset to 135 000 or less.

The syntax and values for picture timing SEI message shall conform to ISO/IEC 14496-10. In picture timing SEI messages, the following fields shall be set to the following specified value:

dpb_output_delay = 0;

For Level 1, the following field shall be set to the following specified value:

cpb_removal_delay = 0 to 150;

The syntax and values for pan-scan rectangle SEI message shall conform to ISO/IEC 14496-10.

4.7.2.4.6 Slice header

The syntax and values for slice header shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

slice_type = 0, 2, 5 or 7;
 if (num_ref_idx_active_override_flag == 1) {
 num_ref_idx_l0_active_minus1 = 0 to 2;
 }
 disable_deblocking_filter_idc = 0 or 1.

4.7.2.4.7 Reference picture list modification

The syntax and values for reference picture list modification shall conform to ISO/IEC 14496-10.

The value of ref_pic_list_modification_flag_l0 shall be set to 0.

4.7.2.4.8 Decoded reference picture marking

The syntax and values for decoded reference picture marking shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

no_output_of_prior_pics_flag = 0;
 long_term_reference_flag = 0;
 adaptive_ref_pic_marking_mode_flag = 0.

4.7.2.5 Class SD

4.7.2.5.1 General

An AVC video elementary stream compliant with Class SD shall conform to the following rules.

- As entropy coding, only CABAC shall be used.
- NAL unit types other than 1 or 5 to 11 shall not be used.
- One access unit (= 1 picture) consists of one slice.
- One RBSP shall be stored in only one nal_unit().
- One SEI RBSP shall consist of only one SEI message.
- A non-paired field is prohibited.
- The value of primary_pic_type in an access unit delimiter shall be 0, 1, or 2 for I-pictures, P-pictures and B-pictures, respectively.
- For a B-picture, the value of nal_ref_idc shall be 0.
- A complementary field pair that contains a B-picture shall not contain an I- or a P-picture.
- The number of frames or complementary field pairs to be stored in DPB shall be 3 or less.

- More than three B-frames or complementary field pairs of B-pictures shall not appear successively in display order.
- The number of reference frames or complementary field pairs shall be 3 or less.
- The maximum total size of non-VCL NAL units in an access unit shall be 1 024 bytes or smaller.
- The total size of BinCountsInNALunit(s) for all VCL NAL unit(s) composing a frame or complementary field pair shall be $1,2 \times 10^6$ bins or less.
- The PPS referred to by a slice and the SPS referred to by that PPS are restricted to those input to the decoder immediately before the slice, that is, the number of sequence parameter set and picture parameter set in a single AVC Configuration Box ('avcC') shall be only one, respectively.
- A PPS should not be contained in an access unit except the first one in a GOP if that PPS is not referred to by the access unit.
- Encoding shall be done so that no gap occurs in presentation during playback of a stream.
- A display period of the following pictures should be 5 seconds* or less.
 - From an IDR-picture to the picture immediately before the next IDR-picture (in decoding order).
 - From the last IDR-picture to the last picture of the stream (in decoding order).
- The `IDR_intervals_more_than_n_seconds*` flag of the `video_attribute_flags` defined in 4.5.5.3 shall be operated. The value of n is 2. An IDR interval is a display period of the following pictures.
 - From an IDR-picture to the picture immediately before the next IDR-picture (in decoding order).
 - From the last IDR-picture to the last picture of the stream (in decoding order).

The value of `seq_parameter_set_id` shall be always set to 0.

End of sequence (`end_of_sequence_rbsp()`) should not be present. If it is present, it shall not be handled as a sample by itself and shall be contained in the sample that contains the last picture in decoding order.

End of stream (`end_of_stream_rbsp()`) should not be present. If it is present, it shall not be handled as a sample by itself and shall be contained in the sample that contains the last picture in decoding order.

At least one non-reference picture should be present within the interval of 5 seconds* in a stream.

4.7.2.5.2 GOP structure

In this Technical Specification, GOP is defined as a group of pictures that conform to the following constraints. There are two types of GOPs, open GOP (see Figure 6) and closed GOP (see Figure 5). An open GOP starts with a non-IDR I-picture in decoding order. In this type of GOP, pictures that appear before the non-IDR I-picture cannot be correctly decoded when executing random access to this GOP. A closed GOP starts with an IDR picture in decoding order. All pictures in this type of GOP can be correctly decoded when executing random access to this GOP.

The structure of GOP shall conform to the following rules.

- If one or more B-frames or complementary field pairs of B-pictures occur successively immediately before an I- or P-frame or complementary field pair of I- or P-picture of the same GOP in display order, they shall be placed immediately after the I- or P-frame or complementary field pair of I- or P-picture in decoding order.

- Decoding order for I- and/or P-pictures shall be the same as their display order.
- A B-frame or complementary field pairs of B-pictures refer to only frames or complementary reference field pairs of I- or P-pictures, which immediately occur before and/or after in display order.
- The decoding order for a non-reference B-picture and the subsequent non-reference B-picture shall be the same as their display order.
- The first access unit in decoding order shall be an IDR or non-IDR I-picture in a GOP.
- In the case of an open GOP, pictures that appear after the first non-IDR I-picture in display order in the same GOP shall not use past reference to pictures that appear before the first non-IDR I-picture in display order.
- The maximum display period of a GOP shall be 2 second*.
- Decoding delay should be equal to, or less than 2 frames period.

Any frame or complementary field pair in a GOP shall satisfy the following constraint:

- Assume
 uppercase "PICTURE" is a frame or a complementary field pair,
 uppercase "FRAME" is a frame or the first field of a complementary field pair,
 the unit of uppercase "N" is PICTURE,
 $t_{0,dpb}(N_a)$ is the output time of the first FRAME in GOP,
 $tr(N)$ is the smallest of $tr(n)$, where FRAME n satisfies $t_{0,dpb}(N_a) \leq tr(n)$,
 N_a+m is the PICTURE that occurs m PICTUREs forward from the PICTURE N_a in display order,
 $N+m$ is the PICTURE that occurs m PICTUREs forward from the PICTURE N in decoding order,
 N_e is the last PICTURE in the GOP in decoding order;
 and obtain N_f by the following pseudo code:
 if $(N + m < N_e)$
 $N_f = N + m$
 else if $(N + m \geq N_e)$
 $N_f = N_e$.
- For an arbitrary m , PICTUREs ($N_0, \dots, N_f$) in decoding order shall include the PICTURE X , where the PICTURE X is the PICTURE corresponding to N_a+m in display order and N_0 is the first PICTURE in the GOP in decoding order.

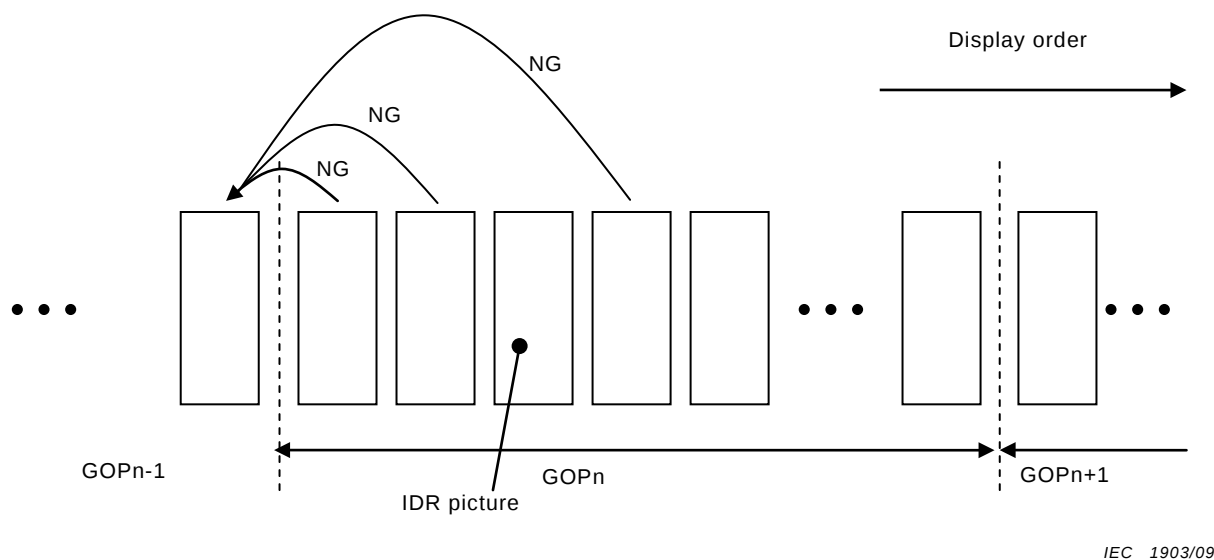


Figure 5 – Example of a closed GOP

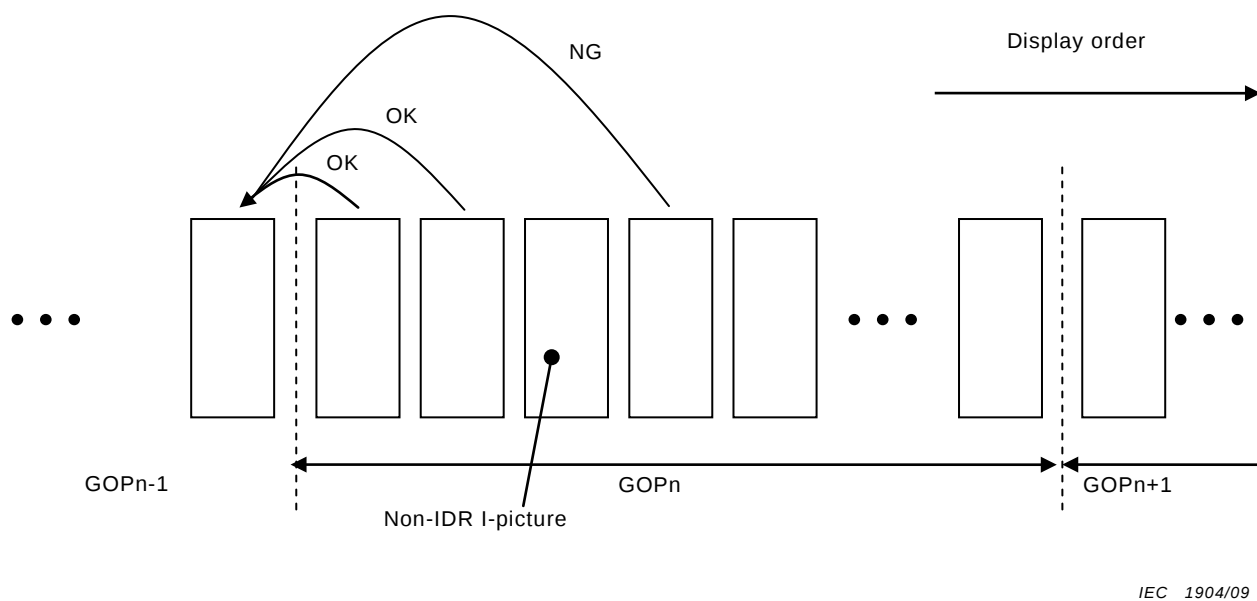


Figure 6 – Example of an open GOP

4.7.2.5.3 Structure of an access unit

NAL unit data that compose an access unit (AU) and the order among them shall conform to the following rules. Note that a filler data NAL unit shall not be used.

The structure of AUs shall be in the following order:

- Access unit delimiter NAL unit;
- SEI NAL unit(s);
- Coded slice of a non-IDR or an IDR picture NAL unit;
- End of sequence NAL unit – if exists;
- End of stream NAL unit – if exists.

As for the location of Sequence Parameter Sets and Picture Parameter Sets, see 4.6.3.2.2.

4.7.2.5.4 Sequence parameter set

The syntax and values for sequence parameter set shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```

profile_idc = 77 (Main Profile);
constraint_set0_flag = 0;
constraint_set1_flag = 1;
constraint_set2_flag = 0;
constraint_set3_flag = 0;
level_idc = 30;
num_ref_frames = 1 to 3;
gaps_in_frame_num_value_allowed_flag = 0;
pic_width_in_mbs_minus1 = 44;
pic_height_in_map_units_minus1 = 14 (for SDTV 59,94 Hz video system) or 17 (for SDTV 50
Hz video system);
frame_mbs_only_flag = 0;
direct_8x8_inference_flag = 1;
frame_cropping_flag = 0;
vui_parameters_present_flag = 1;
vui_parameters() = VUI parameters.

```

The syntax and values for VUI parameters shall conform to ISO/IEC 14496-10. The following fields shall be set to the following specified values.

```

aspect_ratio_info_present_flag = 1;
aspect_ratio_idc = 2 to 5;
if (video_signal_type_present_flag == 1) {
    video_full_range_flag = 0;
    if (colour_description_present_flag == 1) {
        colour_primaries = 5 (for SDTV 50 Hz video system) or 6 (for SDTV 59,94 Hz video
system);
        transfer_characteristics = 5 (for SDTV 50 Hz video system) or 6 (for SDTV 59,94 Hz
video system);
        matrix_coefficients = 5 (for SDTV 50 Hz video system) or 6 (for SDTV 59,94 Hz video
system);
    }
}
chroma_loc_info_present_flag = 0;
timing_info_present_flag = 1;
fixed_frame_rate_flag = 1;
nal_hrd_parameters_present_flag = 1;
vcl_hrd_parameters_present_flag = 1;
hrd_parameters() = HRD parameters;
low_delay_hrd_flag = 0;
pic_struct_present_flag = 1;
bitstream_restriction_flag = 1;
max_bytes_per_pic_denom = 2 to 16;
num_reorder_frames = 0 to 3;
max_dec_frame_buffering = 0 to 3.

```

The syntax and values for HRD parameters shall conform to ISO/IEC 14496-10. The following fields shall be set to the following specified values:

```
cpb_cnt_minus1 = 0.
```

The following fields should be set to the following specified values:

```
cbr_flag = 0;
```

4.7.2.5.5 Picture parameter set

The syntax and values for picture parameter set shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

entropy_coding_mode_flag = 1;
 num_ref_idx_l0_default_active_minus1 = 0 to 2 (if field_pic_flag == 0) or 0 to 5 (if field_pic_flag == 1);
 num_ref_idx_l1_default_active_minus1 = 0 to 2 (if field_pic_flag == 0) or 0 to 5 (if field_pic_flag == 1).

4.7.2.5.6 Supplemental enhancement information

The syntax and values for Supplemental Enhancement Information (SEI) shall conform to ISO/IEC 14496-10. For use of supplemental enhancement information messages shall conform to the following rules.

- A picture timing SEI message shall be present in every access unit.
- A buffering period SEI message shall be present in every IDR access unit.
- When the IDR_intervals_more_than_n_seconds* flag of the video_attribute_flags is set to 1, i.e. the length of an interval between any adjacent IDR-pictures is more than 2 seconds*, a recovery point SEI message shall be present in the first access unit in decoding order in every GOP. In other cases, it is optional. However, it shall not be present in access units other than the first one in decoding order in a GOP.
- The following SEI messages shall not be present:
 pan-scan rectangle SEI message;
 stereo video information SEI message.

The syntax and values for buffering period SEI messages shall conform to ISO/IEC 14496-10. In buffering period SEI messages, the value of initial_cpb_removal_delay shall be 90 000 or less.

The syntax and values for picture timing SEI message shall conform to ISO/IEC 14496-10. In picture timing SEI messages, the following fields shall be operated and set to the following specified values:

clock_timestamp_flag[0] = 1.

The values of pic_struct and ct_type shall be set to the following values according to the frame rate (see Table 14).

Table 14 – The values of pic_struct and ct_type

Frame rate ^a fps	pic_struct	ct_type
29,97	1, 2	1
	3, 4,	0 or 1
	5, 6	0
25	1, 2	1
	3, 4	0 or 1
^a “Frame rate” indicates the value calculated from pic_struct, time_scale and num_units_in_tick.		

4.7.2.5.7 Slice header

The syntax and values for Slice Header shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```

first_mb_in_slice = 0;
slice_type = 5 to 7;
if (num_ref_idx_active_override_flag == 1) {
    num_ref_idx_l0_active_minus1 = 0 to 2(if field_pic_flag == 0) or 0 to 5 (if field_pic_flag
    == 1);
    num_ref_idx_l1_active_minus1 = 0 to 2(if field_pic_flag == 0) or 0 to 5 (if field_pic_flag
    == 1);
}

```

4.7.2.5.8 Reference picture list modification

The syntax and values for reference picture list modification shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```

if (ref_pic_list_modification_flag_l0 == 1) {
    modification_of_pic_nums_idc = 0 or 3;
    if (modification_of_pic_nums_idc == 0) {
        abs_diff_pic_num_minus1 = 0;
    }
}
ref_pic_list_modification_flag_l1 = 0.

```

The value of ref_pic_list_modification_flag_l0 may be set to 1 only for the second field of a non-IDR complementary reference field pair, which appears first in decoding order in an open GOP. Also, such a complementary field pair shall consist of an I-picture and a P-picture as the first and second field in that order. In such a case, the value of ref_pic_list_modification_flag_l0 for the second field shall be always set to 1.

4.7.2.5.9 Prediction weight table

The syntax and values for prediction weight table shall conform to ISO/IEC 14496-10.

When explicit weighted prediction is used, the following fields shall be set to the following specified values:

```

luma_log2_weight_denom  ⌊ to 6;
chroma_log2_weight_denom ⌊ to 6.

```

4.7.2.5.10 Decoded reference picture marking

The syntax and values for decoded reference picture marking shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```

long_term_reference_flag = 0;
adaptive_ref_pic_marking_mode_flag = 0;
no_output_of_prior_pics_flag = 0.

```

4.7.2.6 Class network SD

4.7.2.6.1 General

An AVC video elementary stream compliant with Class network SD shall conform to the following rules.

- As entropy coding, only CABAC shall be used.
- NAL unit types other than 1 or 5 to 11 shall not be used.

- One access unit (= 1 picture) consists of one slice.
- One RBSP shall be stored in only one nal_unit().
- One SEI RBSP shall consist of only one SEI message.
- The value of primary_pic_type in an access unit delimiter shall be 0, 1, or 2 for I-pictures, P-pictures and B-pictures, respectively.
- For a B-picture, the value of nal_ref_idc shall be 0.
- The number of frames to be stored in DPB shall be 3 or less.
- More than three B-frames of B-pictures shall not appear successively in display order.
- The number of reference frames shall be 3 or less.
- The maximum total size of non-VCL NAL units in an access unit shall be 1 024 bytes or smaller.
- The total size of BinCountsInNALunit(s) for all VCL NAL unit(s) composing a frame or complementary field pair shall be $1,2 \times 10^6$ bins or less.
- The PPS referred to by a slice and the SPS referred to by that PPS are restricted to those input to the decoder immediately before the slice, that is, the number of sequence parameter set and picture parameter set in a single AVC Configuration Box ('avcC') shall be only one, respectively.
- A PPS should not be contained in an access unit except the first one in a GOP if that PPS is not referred to by the access unit.
- Encoding shall be done so that no gap occurs in presentation during playback of a stream.
- A display period of the following pictures should be 5 seconds* or less.
 - From an IDR-picture to the picture immediately before the next IDR-picture (in decoding order).
 - From the last IDR-picture to the last picture of the stream (in decoding order).
- The IDR_intervals_more_than_n_seconds* flag of the video_attribute_flags defined in 4.5.5.3 shall be operated. The value of n is 2. An IDR interval is a display period of the following pictures.
 - From an IDR-picture to the picture immediately before the next IDR-picture (in decoding order).
 - From the last IDR-picture to the last picture of the stream (in decoding order).

The value of seq_parameter_set_id shall be always set to 0.

End of sequence (end_of_sequence_rbsp()) should not be present. If it is present, it shall not be handled as a sample by itself and shall be contained in the sample that contains the last picture in decoding order.

End of stream (end_of_stream_rbsp()) should not be present. If it is present, it shall not be handled as a sample by itself and shall be contained in the sample that contains the last picture in decoding order.

At least one non-reference picture should be present within the interval of 5 seconds* in a stream.

4.7.2.6.2 GOP structure

Refer to 4.7.2.5.2.

4.7.2.6.3 Structure of an access unit

Refer to 4.7.2.5.3.

4.7.2.6.4 Sequence parameter set

The syntax and values for sequence parameter set shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```

profile_idc = 77 (Main Profile);
constraint_set0_flag = 0;
constraint_set1_flag = 1;
constraint_set2_flag = 0;
constraint_set3_flag = 0;
level_idc = 30;
num_ref_frames = 1 to 3;
gaps_in_frame_num_value_allowed_flag = 0;
pic_width_in_mbs_minus1 = 44;
pic_height_in_map_units_minus1 = 29 (for 720 × 480) or 35 (for 720 × 576);
frame_mbs_only_flag = 1;
direct_8x8_inference_flag = 1;
frame_cropping_flag = 0;
vui_parameters_present_flag = 1;
vui_parameters() = VUI parameters.

```

The syntax and values for VUI parameters shall conform to ISO/IEC 14496-10. The following fields shall be set to the following specified values.

```

aspect_ratio_info_present_flag = 1;
aspect_ratio_idc = 2 to 5;
if (video_signal_type_present_flag == 1) {
    video_full_range_flag = 0;
    if (colour_description_present_flag == 1) {
        colour_primaries = 5 (for 720 × 576) or 6 (for 720 × 480);
        transfer_characteristics = 5 (for 720 × 576) or 6 (for 720 × 480);
        matrix_coefficients = 5 (for 720 × 576) or 6 (for 720 × 480);
    }
}
chroma_loc_info_present_flag = 0;
timing_info_present_flag = 1;
fixed_frame_rate_flag = 1;
nal_hrd_parameters_present_flag = 1;
vcl_hrd_parameters_present_flag = 1;
    hrd_parameters() = HRD parameters;
low_delay_hrd_flag = 0;
pic_struct_present_flag = 1;
bitstream_restriction_flag = 1;
max_bytes_per_pic_denom = 2 to 16;
num_reorder_frames = 0 to 3;
max_dec_frame_buffering = 0 to 3.

```

The syntax and values for HRD parameters shall conform to ISO/IEC 14496-10. The following fields shall be set to the following specified values:

```
cpb_cnt_minus1 = 0.
```

The following fields should be set to the following specified values:

```
cbr_flag = 0;
```

4.7.2.6.5 Picture parameter set

The syntax and values for picture parameter set shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```

entropy_coding_mode_flag = 1;
num_ref_idx_l0_default_active_minus1 = 0 to 2;

```

num_ref_idx_l1_default_active_minus1 = 0 to 2.

4.7.2.6.6 Supplemental enhancement information

The syntax and values for Supplemental Enhancement Information (SEI) shall conform to ISO/IEC 14496-10. For use of supplemental enhancement information messages shall conform to the following rules.

- A picture timing SEI message shall be present in every access unit.
- A buffering period SEI message shall be present in every IDR access unit.
- When the IDR_intervals_more_than_n_seconds* flag of the video_attribute_flags is set to 1, i.e. the length of an interval between any adjacent IDR-pictures is more than 2 seconds*, a recovery point SEI message shall be present in the first access unit in decoding order in every GOP. In other cases, it is optional. However, it shall not be present in access units other than the first one in decoding order in a GOP.
- The following SEI messages shall not be present:
pan-scan rectangle SEI message;
stereo video information SEI message.

The syntax and values for buffering period SEI messages shall conform to ISO/IEC 14496-10. In buffering period SEI messages, the value of initial_cpb_removal_delay shall be 90 000 or less.

The syntax and values for picture timing SEI message shall conform to ISO/IEC 14496-10. In picture timing SEI messages, the following fields shall be operated and set to the following specified values:

```
clock_timestamp_flag[0] = 1;

pic_struct = 0;

ct_type = 0;
```

4.7.2.6.7 Slice header

The syntax and values for slice header shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```
first_mb_in_slice = 0;
slice_type = 5 to 7;
if (num_ref_idx_active_override_flag == 1) {
    num_ref_idx_l0_active_minus1 = 0 to 2;
    num_ref_idx_l1_active_minus1 = 0 to 2;
}
```

4.7.2.6.8 Reference picture list modification

The syntax and values for reference picture list modification shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```
ref_pic_list_modification_flag_l0 = 0;
ref_pic_list_modification_flag_l1 = 0.
```

4.7.2.6.9 Prediction weight table

Refer to 4.7.2.5.9.

4.7.2.6.10 Decoded reference picture marking

Refer to 4.7.2.5.10.

4.7.2.7 Class HD

4.7.2.7.1 General

An AVC video elementary stream compliant with Class HD shall conform to the following rules.

- NAL unit types other than 1 or 5 to 11 shall not be used.
- A slice shall consist of one or more horizontal lines of macroblocks or macroblock pairs.
- One RBSP shall be stored in only one nal_unit().
- A non-paired field is prohibited.
- The value of primary_pic_type in an access unit delimiter shall be 0, 1, or 2 for I-pictures, P-pictures and B-pictures, respectively.
- A complementary field pair that contains a B-picture shall not contain an I- or a P-picture.
- More than three B-frames or complementary field pairs of B-pictures shall not appear successively in display order.
- The maximum number of frames or complementary field pairs to be stored in DPB shall be 6 (for 1 280 × 720) or 4 (for 1 440 × 1 080 and 1 920 × 1 080).
- A PPS shall not be contained in an access unit except the first one in a GOP if that PPS is not referred to by the access unit.
- Encoding shall be done so that no gap occurs in presentation during playback of a stream.
- A display period of the following pictures should be 5 seconds* or less.
 - From an IDR-picture to the picture immediately before the next IDR-picture (in decoding order).
 - From the last IDR-picture to the last picture of the stream (in decoding order).
- The IDR_intervals_more_than_n_seconds* flag of the video_attribute_flags defined in 4.5.5.3 shall be operated. The value of n is 1. An IDR interval is a display period of the following pictures.
 - From an IDR-picture to the picture immediately before the next IDR-picture (in decoding order).
 - From the last IDR-picture to the last picture of the stream (in decoding order).

The value of seq_parameter_set_id shall be always set to 0.

End of sequence (end_of_sequence_rbsp()) should not be present. If it is present, it shall not be handled as a sample by itself and shall be contained in the sample that contains the last picture in decoding order.

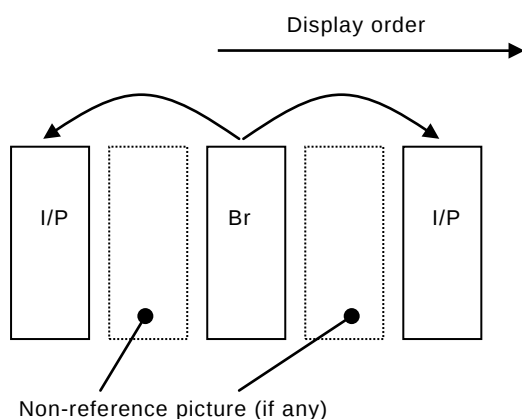
End of stream (end_of_stream_rbsp()) should not be present. If it is present, it shall not be handled as a sample by itself and shall be contained in the sample that contains the last picture in decoding order.

4.7.2.7.2 GOP structure

In this Technical Specification, GOP is defined as a group of pictures that shall conform to the following constraints. There are two types of GOPs: Open GOP (see Figure 10) and Closed GOP (see Figure 9). An open GOP starts with a non-IDR I-picture in decoding order. In this type of GOP, pictures that appear before the non-IDR I-picture cannot be correctly decoded when executing random access to this GOP. A closed GOP starts with an IDR picture in decoding order. All pictures in this type of GOP can be correctly decoded when executing random access to this GOP.

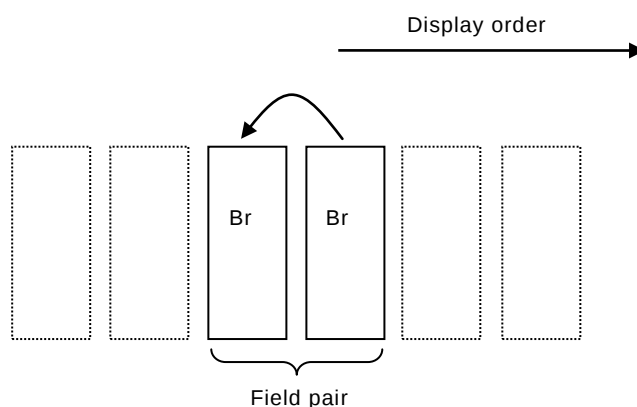
The structure of GOP shall conform to the following rules.

- If one or more B-frames or complementary field pairs of B-pictures occur successively immediately before an I- or P-frame or complementary field pair of I- or P-picture of the same GOP in display order, they shall be placed immediately after the I- or P-frame or complementary field pair of I- or P-picture in decoding order.
- Decoding order for I- and/or P-pictures shall be the same as their display order.
- A P-picture shall not refer to a B-picture.
- The decoding order for a non-reference B-picture and the subsequent non-reference B-picture shall be the same as their display order.
- A reference B-picture may only refer to the following (see Figure 7):
 - a) frames or complementary reference field pairs of I- or P-pictures that appear immediately before and/or after in display order (see Figure 7a);
 - b) if the B-picture is a reference B-field, the reference field that compose a complementary field pair with that B-field (see Figure 7b).



IEC 1905/09

Figure 7a – Reference to I- or P-pictures



Br: Reference B-picture

IEC 1906/09

Figure 7b – Reference between complementary fields

Figure 7 – Reference structure of a reference B-picture

Note that a reference B-picture may appear immediately before or after an I- or P-picture in display order in a sequence such as I B B Br P B Br B P ... or I B Br P Br B P

- A non-reference B-picture shall only refer to the following (see Figure 8):
 - a) frames or complementary reference field pairs of I- or P-pictures that appear immediately before and/or after in display order (pic1 and pic2 in Figure 8);
 - b) a frame or complementary reference field pair of B-pictures that appears immediately after (see Figure 8a) or before (see Figure 8b) that non-reference B-picture in display order and also appears between the pictures that satisfies the condition of a).

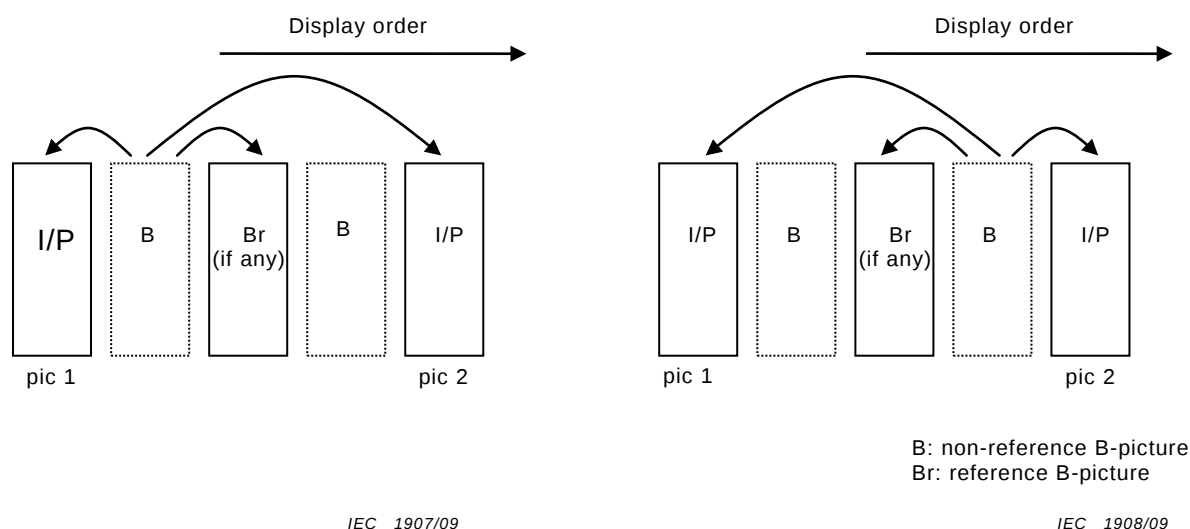


Figure 8a – Non-reference B-picture before reference B-picture

Figure 8b – Non-reference B-picture after reference B-picture

Figure 8 – Reference structure of a non-reference B-picture

- The first access unit in decoding order shall be an IDR or non-IDR I-picture in a GOP.
- In the case of an open GOP, pictures that appear after the first non-IDR I-picture in display order in the same GOP shall not use past reference to pictures that appear before the first non-IDR I-picture in display order.
- The maximum display period of a GOP shall be 1 second*:
- Decoding delay should be equal to or less than 2 frames period.

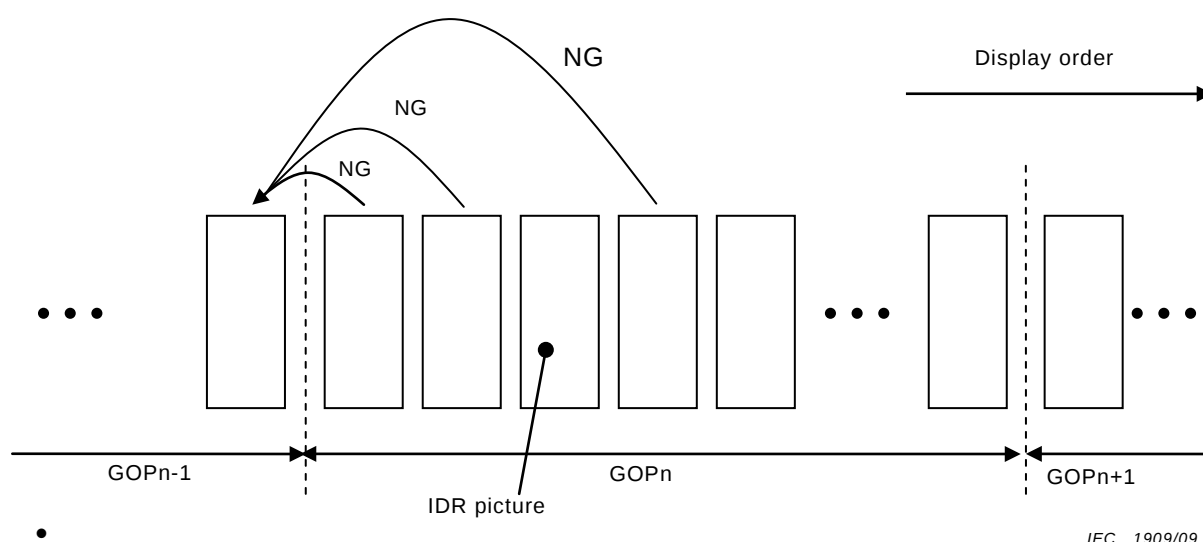
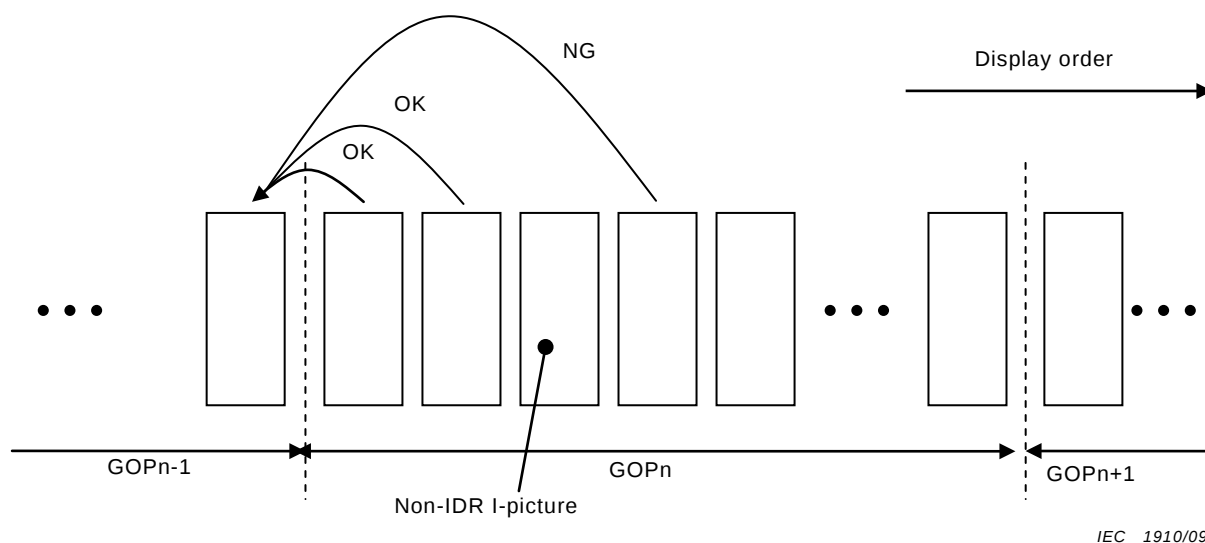


Figure 9 – Example of a closed GOP



IEC 1910/09

Figure 10 – Example of an open GOP

4.7.2.7.3 Structure of an access unit

NAL unit data that compose an access unit (AU) and the order among them shall conform to the following rules. Note that a filler data NAL unit shall not be used.

The structure of AUs shall be in the following order:

- Access unit delimiter NAL unit;
- SEI NAL unit(s);
- Coded slice(s) of a non-IDR or an IDR picture NAL unit;
- End of sequence NAL unit – if exists;
- End of stream NAL unit – if exists.

As for the location of Sequence Parameter Sets and Picture Parameter Sets, see 4.6.3.2.2.

4.7.2.7.4 Sequence parameter set

The syntax and values for the sequence parameter set shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

- profile_idc = 77 (Main Profile) or 100 (High Profile);
- level_idc = 40;
- gaps_in_frame_num_value_allowed_flag = 0;
- frame_cropping_flag = 0 or 1 (for 1 280 × 720), 1 (for 1 440 × 1 080 and 1 920 × 1 080), also see Table 16;
- vui_parameters_present_flag = 1;
- vui_parameters() = VUI parameters.

The values of frame_mbs_only_flag, pic_width_in_mbs_minus1 and pic_height_in_map_units_minus1 shall be set to the values as described in Table 15.

Table 15 – Allowed values of frame_mbs_only_flag, pic_width_in_mbs_minus1 and pic_height_in_map_units_minus1

Visual size	Frame rate fps	frame_mbs_o nly_flag	pic_width_in_mb s_minus1	pic_height_in_map_ units_minus1
1 920 × 1 080	29,97	0	119	33
	25	0	119	33
	23,976	1	119	67
1 440 × 1 080	29,97	0	89	33
	25	0	89	33
	23,976	1	89	67
1 280 × 720	23,976	1	79	44
	50	1	79	44
	59,94	1	79	44

Table 16 – Allowed values of frame cropping related fields

Visual size	frame_mbs_ only_flag	frame_crop_ _left_offset	frame_crop_ _right_offset	frame_crop_ _top_offset	frame_crop_ _bottom_offset
1 920 × 1 080	0	0	0	0	2
	1	0	0	0	4
1 440 × 1 080	0	0	0	0	2
	1	0	0	0	4
1 280 × 720	1	0 ^a	0 ^a	0 ^a	0 ^a
^a Set this value if the value of frame_cropping_flag is 1.					

In addition to above restrictions, the following field shall be set to the following specified value if the bitstream conforms to High Profile:

chroma_format_idc = 1;

The syntax and values for VUI parameters shall conform to ISO/IEC 14496-10. The following fields shall be set to the following specified values:

```

aspect_ratio_info_present_flag = 1;
aspect_ratio_idc = 1 (for 1 280 × 720 and 1 920 × 1 080), 14 or 255 (for 1 440 × 1 080);
if (aspect_ratio_idc == 255) {
    sar_width = 4;
    sar_height = 3;
}
if (video_signal_type_present_flag == 1) {
    video_full_range_flag = 0;
    if (colour_description_present_flag == 1) {
        colour_primaries = 1;
        transfer_characteristics = 1;
        matrix_coefficients = 1;
    }
}
chroma_loc_info_present_flag = 0;
timing_info_present_flag = 1;
fixed_frame_rate_flag = 1;
nal_hrd_parameters_present_flag = 1;
vcl_hrd_parameters_present_flag = 1;
hrd_parameters() = HRD parameters;

```

```
low_delay_hrd_flag = 0;
pic_struct_present_flag = 1
```

The syntax and values for HR.D parameters shall conform to ISO/IEC 14496-10. The following fields shall be set to the following specified values:

```
cbr_flag = 0.
```

4.7.2.7.5 Picture parameter set

The syntax and values for picture parameter set shall conform to ISO/IEC 14496-10.

4.7.2.7.6 Supplemental enhancement information

The syntax and values for Supplemental Enhancement Information (SEI) shall conform to ISO/IEC 14496-10. Supplemental enhancement information messages, shall conform to the following rules.

- A picture timing SEI message shall be present in every access unit.
- A buffering period SEI message shall be present in every IDR access unit.
- When the `IDR_intervals_more_than_n_seconds*` flag of the `video_attribute_flags` is set to 1, i.e. the length of an interval between any adjacent IDR-pictures is more than 1 seconds*, a Recovery Point SEI message shall be present in the first access unit in decoding order in every GOP. In other cases, it is optional. However, it shall not be present in access units other than the first one in decoding order in a GOP.
- The following SEI messages shall not be present:
 - Pan-scan rectangle SEI message;
 - Stereo Video Information SEI message.

The syntax and values for buffering period SEI message shall conform to ISO/IEC 14496-10. In buffering period SEI messages, the value of `initial_cpb_removal_delay` shall be 90 000 or less.

The syntax and values for picture timing SEI message shall conform to ISO/IEC 14496-10. *In* picture timing SEI messages, the following field shall be operated and set to the following specified values:

```
clock_timestamp_flag[0] = 1.
```

4.7.2.7.7 Slice header

The syntax and values for slice header shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```
slice_type = 5 to 7.
```

4.7.2.7.8 Reference picture list modification

The syntax and values for reference picture list modification shall conform to ISO/IEC 14496-10.

4.7.2.7.9 Prediction weight table

The syntax and values for prediction weight table shall conform to ISO/IEC 14496-10.

4.7.2.7.10 Decoded reference picture marking

The syntax and values for decoded reference picture marking shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

no_output_of_prior_pics_flag = 0.

If decoded reference picture marking syntax structure occur in a reference B-picture, it shall be repeated in the I- or P-picture that appear immediately after the B-picture in a decoding order by using decoded reference picture marking repetition SEI message; however, this rule does not apply to a case where the immediately following I- or P-picture belongs to the next GOP.

4.7.2.8 Class network HD

4.7.2.8.1 General

An AVC video elementary stream compliant with Class network HD shall conform to the rules defined in 4.7.2.7.

4.7.2.8.2 GOP structure

Refer to 4.7.2.7.2.

4.7.2.8.3 Structure of an access unit

Refer to 4.7.2.7.3.

4.7.2.8.4 Sequence parameter set

The syntax and values for sequence parameter set shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```

profile_idc = 77 (Main Profile) or 100 (High Profile);
level_idc = 40;
gaps_in_frame_num_value_allowed_flag = 0;
pic_width_in_mbs_minus1 = 79 (for 1 280 × 720), 89 (for 1 440 × 1 080), or 119 (for 1 920 × 1 080);
pic_height_in_map_units_minus1 = 44 (for 1 280 × 720) or 67 (for 1 440 × 1 080 and 1 920 × 1 080);
frame_mbs_only_flag = 1;
frame_cropping_flag = 0 or 1 (for 1 280 × 720) or 1 (for 1 440 × 1 080 and 1 920 × 1 080);
if (frame_cropping_flag = 1) {
    frame_crop_left_offset = 0;
    frame_crop_right_offset = 0;
    frame_crop_top_offset = 0;
    frame_crop_bottom_offset = 0 (for 1 280 × 720) or 4 (for 1 440 × 1 080 and 1 920 × 1 080);
}
vui_parameters_present_flag = 1;
vui_parameters() = VUI parameters.
```

In addition to above restrictions, the following field shall be set to the following specified value if the bitstream conforms to High Profile:

chroma_format_idc = 1.

The syntax and values for VUI parameters shall conform to ISO/IEC 14496-10.

The following fields shall be set to the following specified values:

```

aspect_ratio_info_present_flag = 1;
aspect_ratio_idc = 1 (for 1 280 × 720 and 1 920 × 1 080), 14 or 255 (for 1 440 × 1 080);
if (aspect_ratio_idc = 255) {
    sar_width = 4;
    sar_height = 3;
}
```

```

if (video_signal_type_present_flag == 1) {
    video_full_range_flag = 0;
    if (colour_description_present_flag == 1) {
        colour_primaries = 1;
        transfer_characteristics = 1;
        matrix_coefficients = 1;
    }
}
chroma_loc_info_present_flag = 0;
timing_info_present_flag = 1;
fixed_frame_rate_flag = 1;
nal_hrd_parameters_present_flag = 1;
vcl_hrd_parameters_present_flag = 1;
    hrd_parameters() = HRD parameters;
low_delay_hrd_flag = 0;
pic_struct_present_flag = 1.

```

The syntax and values for HRD parameters shall conform to ISO/IEC 14496-10. For the values, refer to 4.7.2.7.4.

4.7.2.8.5 Picture parameter set

Refer to 4.7.2.7.5.

4.7.2.8.6 Supplemental enhancement information

Refer to 4.7.2.7.6.

4.7.2.8.7 Slice header

Refer to 4.7.2.7.7.

4.7.2.8.8 Reference picture list reordering

Refer to 4.7.2.7.8.

4.7.2.8.9 Prediction weight table

Refer to 4.7.2.7.9.

4.7.2.8.10 Decoded reference picture marking

Refer to 4.7.2.7.10.

4.7.3 JPEG image sequence for thumbnail pictures

4.7.3.1 General

JPEG images as thumbnail pictures shall be encoded in conformity to the data structure of DCF thumbnail file as defined in JEITA CP-3461 based on JEITA CP-3451 and JEITA CP-3451-1. Each JPEG image for a thumbnail picture shall be 16 KB maximum.

Note that the tags defined as recommended for DCF thumbnail files in JEITA CP-3461 are optional in this specification.

4.7.3.2 ESDescriptor

4.7.3.2.1 General

An ESDescriptor (ESD) is contained in the ESD Box ('esds') in MP4VisualSampleEntry.

The syntax and values for ESDescriptor shall conform to ISO/IEC 14496-1, and the following fields shall be set to the following specified values. They are identical with those defined in ISO/IEC 14496-14. Descriptors other than those below shall not be used.

```
ES_ID = 0;
streamDependenceFlag = 0;
URL_Flag = 0;
OCRstreamFlag = 0 (false);
streamPriority = 0;
decConfigDescr = DecoderConfigDescriptor, see 4.7.3.2.2;
slConfigDescr = SLConfigDescriptor, predefined type 2;
```

4.7.3.2.2 DecoderConfigDescriptor

The syntax and values for DecoderConfigDescriptor shall conform to ISO/IEC 14496-1, and the following fields shall be set to the following specified values.

In this descriptor, DecoderSpecificInfo shall always be used, and no ProfileLevelIndicationIndexDescriptor(s) shall be used.

```
objectTypeIndication = 6Ch (Visual ISO/IEC 10918-1);
streamType = 04h (Visual Stream);
upStream = 0;
decSpecificInfo = DecoderSpecificInfo, see 4.7.3.2.3;
```

4.7.3.2.3 DecoderSpecificInfo

A DecoderSpecificInfo consists of JPEG_DecoderConfig in accordance with ISO/IEC 14496-1.

The syntax and values for the JPEG_DecoderConfig shall conform to ISO/IEC 14496-1, and the following fields shall be set to the following specified values.

```
numComponents = 3 (Y, Cr, Cb);
```

4.8 Logical structure of media data

4.8.1 General

This subclause describes layout of media data in a movie and a fragmented movie. Figure 2 and Figure 3 show the examples of both movies.

If decoding order and display order of sample data are not identical, decoding order is used in this subclause.

4.8.2 Interleaving of main audio and video

When a file has the `additional_track_exists` flag and `main_AV_track_is_edited` flag in the Profile Box set to '0', the sample data of main audio and video tracks, which are supposed to be handled approximately at the same time and for the same period (see NOTE), shall be interleaved in decoding order at an interval of 1 second* maximum if one or more main video tracks contain AVC video elementary streams.

An interleave unit contains one or more consecutive data chunks of such tracks. In the case of an AVC video elementary stream, an IDR access unit should be placed at the top of the video part in an interleave unit within the interval of 5 seconds* or less. The temporal offset between the first audio and video sample data of an interleave unit with an IDR access unit at the top of the video part shall be less than the duration of the first sample of interleave unit even if the stream is edited at the top of this interleave unit.

NOTE The boundaries of audio and video sample data are not necessarily coincident.

It is highly recommended that the intervals of interleaving be approximately regular intervals within a file.

4.8.3 Arrangement of data chunks

For main video tracks, data duration of a single chunk in each track shall be 1 second* maximum for main video tracks that contain AVC video elementary streams.

For main audio tracks, data duration of a single chunk in each track shall be 1 second* maximum if the file contains one or more main video tracks composed of AVC video elementary streams.

For all tracks, the chunks shall be located in decoding order when a file has the `additional_track_exists` flag and `main_AV_track_is_edited` flag in the Profile Box set to '0'.

4.8.4 Placement of thumbnail picture

The thumbnail picture image sample data shall be contained in an 'exclusive' Media Data Box ('mdat'), and any other sample data, such as main audio or video, shall not be contained in this exclusive box.

The exclusive Media Data Box shall be placed immediately after the Movie Box or the initial Movie Box. Note that a Media Data Box that contains any other data except for the thumbnail picture sample data may be placed anywhere after the File Type Box and the Profile Box in a (regular) movie.

Figure 2 and Figure 3 show a layout of thumbnail picture image sample data.

4.8.5 Media data structure in movie fragment

A movie fragment may consist of the main audio data, main video data, which are supposed to be handled approximately at the same time and for the same period in decoding order.

When one or more main audio and/or video tracks are self-contained in a fragmented movie, each movie fragment shall be created as an integral multiple of the interleave unit. For information about the rules for the data size of each fragment, see A.3.2. The interleave units shall satisfy the conditions described in 4.8.2 and 4.8.3. It is recommended that the intervals of the movie fragments be approximately the same within a file.

A movie fragment should have at least one sync sample of each track that composes the movie fragment. A sync sample does not necessarily have to be the first sample in each track fragment.

4.8.6 Exception for logical structure of media data

When a main video track contains a sample whose duration is over 1 second* if one or more main video tracks contain AVC video elementary streams, the limitations of interleaving, arrangement of data chunks and media data structure in movie fragment are not applied to any tracks, which are described in 4.8.2, 4.8.3 and 4.8.5, respectively.

4.9 Presentation arrangement for tracks

4.9.1 Alternate group support

An alternate group specifies a group or collection of tracks that contain alternate data for one another.

Making an alternate group is allowed for the main video track and main audio track. When multiple tracks as an alternate group for such track types are contained in a file, they shall be identified as a group of alternate tracks by assigning the same ID (nonzero) in each track's alternate_group field in their track header box.

The track_enabled flag of exactly one track belonging to an alternate group shall be set to ON (= 1) in the flags field of its track header box. For the other track(s), the track_enabled flag shall be set to OFF (= 0). If no track whose track_enabled flag is set to ON exists, it indicates that no track is set as default in the alternate group.

The other flags in the track header box shall be set to a default value defined in ISO/IEC 14496-12.

4.9.2 Edit list support

4.9.2.1 General

A PCE AV file structured as a (regular) movie can be re-scheduled to present the media data by using the edit list entries in the Edit List Box contained in the Edit Box, and therefore it causes a capability of non-linear editing such as random presentation or cutting part.

All tracks shall contain an Edit Box and an Edit List Box even though such editing is not necessary (the presentation time-line is one-to-one mapped to the media time-line, and presentation of the track starts from the beginning).

Refer to ISO/IEC 14496-12 for details on the Edit Box and the Edit List Box.

4.9.2.2 Invariable edit entry

Only one 'invariable' edit entry shall exist in the Edit List Box of a track when editing using the edit list entries is not needed. When only 'initial-offset' edit entry defined in 4.9.2.3 is needed for a main audio or video track, only one 'invariable' edit entry that follows the 'initial-offset' edit entry shall exist in the Edit List Box of such a track.

The 'invariable' edit entry has the following specified values:

segment_duration = the sum of all sample durations in the track, converted into the timescale in the Movie Header Box;

media_time = 0 (if no decoding delay is necessary) or ct_offset (if decoding delay is necessary, see 4.6.3.2);

media_rate_integer = 1;

media_rate_fraction = 0.

For the definition of 'ct_offset', refer to 4.4.15. The value of 'ct_offset' shall be calculated from the following formula:

$$ct_offset = T \times N$$

where

T = 1 000 (if the frame rate is 25 fps or 50 fps) or 1 001 (if the frame rate is 23,976 fps, 29,97 fps or 59,94 fps)

N = Period of decoding delay in units of frames.

Note that the value of 'N' for interlaced video is twice as much as that for progressive video.

4.9.2.3 Initial-offset edit entry for synchronizing main audio with main video

The boundary of audio frame and video frame is not necessarily coincident. Starting offsets for the main audio or video track (stream), therefore, may be necessary to start the presentation synchronously.

In the case where such starting offsets are necessary, the recorders that have a capability of editing shall use an 'initial-offset' edit entry for the main audio or video track to maintain a sync point as the first edit entry in the Edit List Box. The entry shall be attached to only main audio or video track and shall be made so that the value of segment_duration is less than the duration of the first sample of the track that the track with this entry synchronizes with.

The 'initial-offset' edit entry is the specific 'empty' edit entry that has the following specified values:

```
segment_duration = duration of starting offset;
media_time = -1 (FFFFFFFFh);
media_rate_integer = 1;
media_rate_fraction = 0.
```

4.9.2.4 Allowed edit entries

In case of regular movies, there is no restriction in use of edit lists. However, if edit lists except for one consisting of one 'invariable' edit entry or one 'initial-offset' edit entry followed by an 'invariable' edit entry exist, the function flag of value 40000000h shall be set.

In case of fragmented movies, the only allowed edit entry is an 'invariable' edit entry or an 'initial-offset' edit entry followed by an 'invariable' edit entry that contains the segment duration equal to the sum of all sample durations managed by an initial 'moov' for each track.

4.10 Metadata

4.10.1 Metadata types

4.10.1.1 General

Subclause 4.10.1 defines the structure of data_type_ID and currently defined metadata types.

4.10.1.2 Data type ID

This subclause discusses the structure of the data_type_ID that identifies the type of metadata.

The data_type_ID has the following structure:

```
unsigned int(16) reserved = 0000h;
unsigned int(16) data_type.
```

The semantics of these fields are defined as follows:

data_type A 16-bit integer that identifies the metadata type. The value shall be one or higher.

4.10.1.3 Metadata units

The metadata in a metadata unit can be human-readable information to present or binary data. The metadata are formatted with more specific data_type_ID in a specific media file format to be added to the files conforming to the format.

In each metadata type of text data (human-readable information), each element is delimited by TAB character in a metadata field unless the specific media file format states otherwise, when a metadata field has several elements.

Table 17 shows currently defined data_type_IDs of metadata to be added to the files that shall conform to this Technical Specification.

Table 17 – data_type_ID

Values	Description	Form of meta_data
00000001h	Title: Title of content	Characters; free form
00000003h	ProductionDate: Date and time the “content” is created on (not a “file”)	19 alphanumeric characters (“0”-“9”, “/”, “ ”, and “:”), see 4.10.2.1.
00000004h	Software: The name of software that generated this movie file. This metadata should be used as the supplement information to distinguish the software.	Characters; free form
00000005h	Product: The name of a product, software or hardware, that generated this movie file.	Characters; free form
00000006h	Description: Description about content, such as outline, production note, or story.	Characters; free form
00000008h	Keywords: Keywords of content (for search)	Characters; free form
00000017h	Collection: The name(s) of the collection(s) to which the content is a member.	Characters; free form
00000018h	Star Rating: This metadata contains the rating information for user	A 8-bit binary data contained rating information;
0000000Ah	TrackProperty: Purpose of the track to present.	8-byte binary data contained presentation type and priority of track; see 4.10.2.2.
0000000Bh	TimeZoneOffset: The time offset from UTC.	2-byte binary data contained time offset from UTC; see 4.10.2.3.
0000000Ch	ModificationDate: Date and time the “content” is modified on (not a “file”)	19 alphanumeric characters (“0”-“9”, “/”, “ ”, and “:”), see 4.10.2.4.
00000A01h	CopyrightNotice: Descriptions about copyright restrictions	Characters; free form
00000E01h	CameraPerson: The name of person(s), shooting the content.	Characters; free form
00000E02h	Subject/Object: The name(s) of recorded subject, object or person.	Characters; free form
00000E03h	ShootingLocation: The place of shooting.	Characters; free form

4.10.2 Metadata types definition

4.10.2.1 Production date

This metadata contains the date and time in UTC, on which the content is created (note that not a “file”). It should be the date and time when the recording of the content is started.

This metadata has 19 alphanumeric characters conforming to the following form that consists of “0”-“9”, “/”, “ ” (space character), and “.” (note that when such alphanumeric characters are encoded in UTF-16BE (ISO/IEC 10646), only the code from 0020h and/or 002Fh to 003Ah in Unicode shall be used):

“yyyy/mm/dd_hh:mm:ss”

where “yyyy/mm/dd” indicates year/month/day,

“ ” represents a space character,

“hh:mm:ss” indicates hour:minute:second.

Space characters shall be used for numerical values that are unavailable. For example:

"2003/10/20_ _ _:_:_ _"

where an underscore (_) represents a space character.

One-digit month, day, hour, minute, and second shall be preceded by zero. For example:

"2003/01/02_01:02:03"

where an underscore (_) represents a space character.

4.10.2.2 Track property

This metadata contains 8-byte attribute information for every track. In a presentation, it may be able to judge for what purpose each track is used and what intention each track is created, etc. This metadata shall be attached to each track ('trak') and shall not be attached to entire presentation ('moov').

This metadata has the following structure:

```
unsigned int(32)  presentation_type;
unsigned int(8)  track_attribute_flags;
unsigned int(8)  reserved = 0;
unsigned int(16) priority.
```

The semantics of these fields are defined as follows:

presentation_type A 32-bit integer that identifies the presentation type. The following values are currently defined in this Technical Specification (see Table 18).

Table 18 – Presentation types

Values	Description
00000001h	an original or main track
00000002h	a track for thumbnail picture data.
others	reserved for future use in this Technical Specification

track_attribute_flags An 8-bit flag space that indicates an attribute of a track. The following flags are currently defined in this Technical Specification (see Table 19).

Table 19 – Track attribute flags

Values	Description
80h	track_is_read_only: indicates the creator of this file requires that the track with this flag set to ON (= 1) be not modified or deleted intentionally by a device or software with a function to modify or delete tracks. When editing the track for which this flag is set to ON, warning should be displayed to notify users that the track is intended to be read only.
Others	reserved for future use in this Technical Specification.

priority A 16-bit integer that specifies the track priority. The smaller number indicates a higher priority, and 1 indicates the highest priority. The value 0 specifies un-setting up and is the normal value.

4.10.2.3 Time zone offset

This metadata contains the time zone offset from Coordinated Universal Time (UTC). The local time is calculated by this metadata and time stamp such as creation_time in UTC.

Recorders that have a capability to treat time zone shall use this metadata and record the time zone offset in it.

This metadata has the following structure:

int(16) time_zone_offset;

The semantics of the field is defined as follows:

time_zone_offset A 16-bit integer that specifies time zone offset from UTC in minutes. For example, the value is set to 021Ch (540 minutes = 9 hours) for UTC+09:00.

4.10.2.4 Modification date

This metadata contains the date and time in UTC, on which the content is modified (note that not a “file”).

This metadata has 19 alphanumeric characters conforming to the following form that consists of “0”-“9”, “/”, “ ” (space character), and “:” (note that when such alphanumeric characters are encoded in UTF-16BE (ISO/IEC 10646), only the code from 0020h and/or 002Fh to 003Ah in Unicode shall be used):

“yyyy/mm/dd_hh:mm:ss”
where “yyyy/mm/dd” indicates year/month/day,
“ ” represents a space character,

“hh:mm:ss” indicates hour:minute:second.

Space characters shall be used for numerical values that are unavailable. For example:

“2003/10/20_ _ _ : _ _ : _ _ ”
where an underscore (_) represents a space character.

One-digit month, day, hour, minute, and second shall be preceded by zero. For example:

“2003/01/02_01:02:03”
where an underscore (_) represents a space character.

4.10.2.5 Star rating

This metadata contains the rating information for users.

This metadata has the following structure:

```
unsigned int(8) star_rating;
```

The semantics of the field is defined as follows:

star_rating A 8-bit unsigned integer. Either the value 00h (indicating no rating assigned) or a value in the range 0Ah through 32h inclusive, indicating a rating between 1,0 (1 star) and 5,0 (5 stars) inclusive.

4.10.2.6 Usage of metadata

Shows usage of the metadata types. In Table 20, the data size of a metadata field indicates the size including termination (null).

For other operational rules on the use of metadata, see Annex A.

Table 20 – Usage of the metadata

data_type_ID	Description	Data type	Data size of a meta_data	Attached to content	Attached to track	Multi-language	Multi-element
00000001h	Title	text	Max 256 bytes	O	P	Yes	No
00000003h	ProductionDate	text	Max 40 bytes	O	O	'und' only	No
00000004h	Software	text	Max 256 bytes	O	P	'eng' only	No
00000005h	Product	text	Max 256 bytes	O	P	'eng' only	No
00000006h	Description	text	Max 2 000 bytes	O	O	Yes	No
00000008h	Keyeords	text	Max 256 bytes	O	P	Yes	Yes
00000017h	Collection	text	Max 1 024 bytes	O	P	Yes	Yes
00000018h	Star Rating	binary	1 byte	O	P	'und' only	–
0000000Ah	TrackProperty	binary	8 bytes	P	M	'und' only	–
0000000Bh	TimeZoneOffset	binary	2 bytes	O	P	'und' only	–
0000000Ch	ModificationDate	text	Max 40 bytes	O	O	'und' only	No
00000E01h	CameraPerson	text	Max 1 024 bytes	O	P	Yes	Yes
00000E02h	Subject/Object	text	Max 1 024 bytes	O	P	Yes	Yes
00000E03h	ShootingLocation	text	Max 256 bytes	O	P	Yes	No
M Mandatory P Prohibited for use O Optional							

4.11 Pixel aspect ratios

The horizontal and vertical resolution fields in a Visual Sample Entry of a track specify the pixel aspect ratio of image data contained in the track, and the pixel_aspect_ratio field of a Video Profile Entry in the Profile Box (Private Extension) indicates informative pixel aspect ratio of visual/image data.

The horizontal and vertical spacing portion in the pixel_aspect_ratio field shall be set in accordance with the aspect ratio specified by the horizontal and vertical resolution fields in the Visual Sample Entry, which means that the ratio of the two spacing values shall be identical to any of the ones specified in the resolution fields. See 4.5.1.2 and 4.4.16 for information on these fields.

Table 21 and Table 22 show the aspect ratios currently defined in this Technical Specification and the settable values in the resolution fields or spacing portions. Note that other values may be defined in a specific media file format.

Table 21 – Resolution values in visual sample entry

Video types			Values	
Video system	Visual size	Display aspect ratio	horizresolution	vertresolution
HD	1 920 × 1 080	16:9	00480000h (72,0)	00480000h (72,0)
	1 440 × 1 080	16:9 squeezed	00480000h (72,0)	00600000h (72,0×4/3)
	1 280 × 720	16:9	00480000h (72,0)	00480000h (72,0)
SDTV 59,94 Hz	720 × 480	4:3	004F3333h (72,0 × 11/10)	00480000h (72,0)
		16:9 squeezed	00480000h (72,0)	005745D1h (72,0×40/33)
SDTV 50 Hz	720 × 576	4:3	00480000h (72,0)	004E8BA2h (72,0×12/11)
		16:9 squeezed	00480000h (72,0)	0068BA2Eh (72,0×16/11)
VGA	640 × 480 320 × 240	4:3	00480000h (72,0)	00480000h (72,0)
		16:9 squeezed	00480000h (72,0)	00600000h (72,0×4/3)
	640 × 360 320 × 180	16:9	00480000h (72,0)	00480000h (72,0)
-	480 × 270	4:3	00600000h (72,0 × 4/3)	00480000h (72,0)
		16:9	00480000h (72,0)	00480000h (72,0)
Note that the minimum value of the resolution fields is 72,0.				

Table 22 – spacing values in pixel_aspect_ratio

Video types			Values	
Video system	Visual size	Display aspect ratio	Horizontal spacing	Vertical spacing
HD	1 920 × 1 080	16:9	0001h (1)	0001h (1)
	1 440 × 1 080	16:9 squeezed	0004h (4)	0003h (3)
	1 280 × 720	16:9	0001h (1)	0001h (1)
SDTV 59,94 Hz	720 × 480	4:3	000Ah (10)	000Bh (11)
		16:9 squeezed	0028h (40)	0021h (33)
SDTV 50 Hz	720 × 576	4:3	000Ch (12)	000Bh (11)
		16:9 squeezed	0010h (16)	000Bh (11)
VGA	640 × 480 320 × 240	4:3	0001h (1)	0001h (1)
		16:9 squeezed	0004h (4)	0003h (3)
	640 × 360 320 × 180	16:9	0001h (1)	0001h (1)
-	480 × 270	4:3	0003h (3)	0004h (4)
		16:9	0001h (1)	0001h (1)

5 PCE AV classes

5.1 Class

In this Technical Specification, Class indicates the video and audio data formats and constraints on the values of the syntax defined for those formats, with which recorders shall comply. Classes are defined for specific purposes based on use cases.

Additionally, level indicates the restrictions on the parameters for media data such as bitrate and the visual size. Note that it is also intended that Class and level indicate the required capability for players. See also Annex A and Annex B for the required capability for players.

5.2 Class PT

5.2.1 General

Class PT is provided for playback of higher quality video than Class MB on portable CE products that have a higher playback capability.

For this Class, AVC video Main Profile is employed as the video format and MPEG-4 AAC LC as the audio format. This Class is currently classified into three levels.

Recorders compliant with this Class shall make a recording within the bounds of this Class complying with Annex B. The recommended recording modes for this Class are listed in Annex C.

Video

AVC video ISO/IEC 14496-10:2009

Main Profile at Level 3

With the following restrictions:

- Visual Size
QVGA, 480 × 270, or VGA
- Frame Rate
Progressive video:
29,97 fps, 25 fps or 23,976 fps
- Pixel Aspect Ratio
1:1, 3:4 or 4:3
Refer to 4.11 for details.

Audio

MPEG-4 ISO/IEC 14496-3:2005

AAC Profile at Level 2 / High Quality Audio Profile at Level 2 (AAC LC only)

With the following restrictions:

- Bitrate
average bitrate
256 kbps max (stereo / dual mono)
peak bitrate
The stream should be encoded at equal to or less than 384 kbps max (stereo / dual mono).
- Sampling Frequency
48 kHz only

This Class is classified into three levels according to bitrate and other parameters.

5.2.2 Level 1

Video

Equivalent to Main Profile at Level 2.1 (ISO/IEC 14496-10)

With the following restrictions:

- Visual Size
QVGA only
- Frame Rate
Progressive video:
29,97 fps or 25 fps
- Pixel Aspect Ratio
1:1 or 4:3

Audio

- Bitrate
average bitrate
128 kbps max (stereo / dual mono)
peak bitrate
The stream should be encoded at equal to or less than 384 kbps max (stereo / dual mono).

5.2.3 Level 2

Video

Equivalent to Main Profile at Level 2.1 (ISO/IEC 14496-10)

With the following restrictions:

- Visual Size
QVGA or 480 × 270

Audio

- Bitrate
average bitrate
128 kbps max (stereo / dual mono)
peak bitrate
The stream should be encoded at equal to or less than 384 kbps max (stereo / dual mono).

5.2.4 Level 3

Video

No further restriction as this Class PT defined in 5.2.1.

Audio

No further restriction as this Class PT defined in 5.2.1.

5.3 Class MB

5.3.1 General

Class MB is provided for playback of video that is smaller than SDTV size on general portable CE products. It is intended to be the basic Class to secure interoperability among various products.

For this Class, AVC video baseline profile is employed as the video format. MPEG-4 AAC LC as the audio format. SBR tool may be optionally used.

Recorders compliant with this Class shall make a recording within the bounds of this Class complying with Annex B. The recommended recording modes for this Class are listed in Annex C.

Video

AVC video ISO/IEC 14496-10:2009

Baseline Profile at Level 3

With the following restrictions:

- Visual Size
320 × 180, QVGA, 480 × 270, 640 × 360 or VGA
- Frame Rate
Progressive video:
29,97 fps, 25 fps, 14,985 fps, or 12,5 fps
- Pixel Aspect Ratio
1:1 or 4:3
Refer to 4.11 for detail.

Audio

MPEG-4 ISO/IEC 14496-3:2005

AAC Profile at Level 2 or High Quality Audio Profile at Level 2 (AAC LC only)

With the following restrictions:

- CODEC
Support of SBR tool is optional.
- Bitrate
average bitrate
128 kbps max (stereo/dual mono) or 64 kbps max (single mono) (if SBR tool is not used)
96 kbps max (stereo/dual mono) or 48 kbps max (single mono) (if SBR tool is used)
peak bitrate
In case of stereo / dual mono
The stream should be encoded at equal to or less than 384 kbps max (if SBR tool is not used).
The stream should be encoded at equal to or less than 144 kbps max (if SBR tool is used).
In case of single mono
The stream should be encoded at equal to or less than 96 kbps max (if SBR tool is not used).
The stream should be encoded at equal to or less than 72 kbps max (if SBR tool is used).
- Sampling Frequency
48 kHz (if SBR tool is not used) or 24 kHz (if SBR tool is used).
Note that above values are for the core coder and the output sampling frequency will be 48 kHz when playing back SBR data.

This Class is classified into four levels according to bitrate and other parameters.

5.3.2 Level 1

Video

Equivalent to Baseline Profile at Level 1.2. (ISO/IEC 14496-10)

With the following restrictions:

- Visual Size
320 × 180 or QVGA
- Frame Rate
14,985 fps or 12,5 fps only
- Pixel Aspect Ratio
1:1 only

Audio

No further restriction as this Class MB defined in 5.3.1.

5.3.3 Level 2

Video

Equivalent to Baseline Profile at Level 1.3. (ISO/IEC 14496-10)

With the following restrictions:

- Visual Size
320 × 180 or QVGA
- Pixel Aspect Ratio
1:1 or 4:3

Audio

No further restriction as this Class MB defined in 5.3.1.

5.3.4 Level 3

Video

Equivalent to Baseline Profile at Level 2.1. (ISO/IEC 14496-10)

With the following restrictions:

- Bitrate
Maximum value of MaxBR is 1 000
Note that for both of VCL HRD parameters and NAL HRD parameters,
Bitrate[SchedSelIdx] ≤ 1 000 × MaxBR,
i.e. MaxBR is specified in units of 1 000 bits/s.
- Visual Size
320 × 180, QVGA or 480 × 270

Audio

No further restriction as this Class MB defined in 5.3.1.

5.3.5 Level 4

Video

No further restriction as this Class MB defined in 5.3.1.

Audio

No further restriction as this Class MB defined in 5.3.1.

5.4 Class SD

Class SD is provided for SDTV quality video. It is intended to be commonly used on portable and stationary products.

For this Class, AVC video Main Profile is employed as the video format and MPEG-4 AAC LC as the audio format.

Recorders compliant with this Class shall make a recording within the bounds of this Class complying with Annex B. The recommended recording modes for this Class are listed in Annex C.

Video

AVC video ISO/IEC 14496-10:2009
Main Profile at Level 3

With the following restrictions:

- Visual Size
720 × 480 for SDTV 59,94 Hz video system
720 × 576 for SDTV 50 Hz video system
- Frame Rate
Interlaced video:
29,97 fps (59,94 fields/sec) for SDTV 59,94 Hz video system
25 fps (50 fields/sec) for SDTV 50 Hz video system

- Pixel Aspect Ratio
The values defined for SDTV 59,94 Hz and SDTV 50 Hz video systems in 4.11.

Audio

MPEG-4 ISO/IEC 14496-3:2005

AAC Porfile at Level 2 / High Quality Audio Profile at Level 2 (AAC LC only)

With the following restrictions:

- Bitrate
average bitrate
256 kbps max (stereo / dual mono)
peak bitrate
The stream should be encoded at equal to or less than 384 kbps max (stereo / dual mono).
- Sampling Frequency
48 kHz.

5.5 Class network SD

Class network SD is provided for SDTV quality video to be used for interchange of content on the internet.

For this Class, AVC video Main Profile is employed as the video format and MPEG-4 AAC LC as the audio format.

Recorders compliant with this Class shall make a recording within the bounds of this Class complying with Annex B. The recommended recording modes for this Class are listed in Annex C.

Video

AVC video ISO/IEC 14496-10:2009

Main Profile at Level 3

With the following restrictions:

- Visual Size
720 × 480
720 × 576
- Frame Rate
Progressive video:
29,97 fps or 23,976 fps for 720 × 480
25 fps for 720 × 576
- Pixel Aspect Ratio
10:11 or 40:33 for 720 × 480
12:11 or 16:11 for 720 × 576

Audio

MPEG-4 ISO/IEC 14496-3:2005

AAC Porfile at Level 2 / High Quality Audio Profile at Level 2 (AAC LC only)

With the following restrictions:

- Bitrate
average bitrate
256 kbps max (stereo / dual mono)
peak bitrate
The stream should be encoded at equal to or less than 384 kbps max (stereo / dual mono).
- Sampling Frequency
48 kHz

5.6 Class HD

5.6.1 General

Class HD is provided for HDTV quality video to be displayed on HDTV products.

For this Class, AVC video Main Profile or AVC video High Profile is employed as the video format and MPEG-4 AAC LC as the audio format.

Recorders compliant with this Class shall make a recording within the bounds of this Class complying with Annex B. The recommended recording modes for this Class are listed in Annex C.

Video

AVC video ISO/IEC 14496-10:2009

Main Profile at Level 4 / High Profile at Level 4

With the following restrictions:

- Bitrate
 - Main Profile:
 - Maximum value of Max BR is 20 000
 - Note that for both of VCL HRD parameters and NAL HRD parameters,
 $\text{Bitrate}[\text{SchedSelIdx}] \leq 1\,000 \times \text{MaxBR}$
 i.e. MaxBR is specified in units of 1 000 bits/s.
 - High Profile:
 - Maximum value of Max BR is 24 000
 - Note that both of VCL HRD parameters and NAL HRD parameters,
 $\text{Bitrate}[\text{SchedSelIdx}] \leq 1\,000 \times \text{MaxBR}$
 i.e. MaxBR is specified in units of 1 000 bits/s.
- CPB Size
 - Main Profile:
 - No restriction
 - High Profile:
 - Maximum size of MaxCPB is 30 000
 - Note that for both of VCL HRD parameters and NAL HRD parameters,
 $\text{CpbSize}[\text{SchedSelIdx}] \leq 1\,000 \times \text{MaxCPB}$,
 i.e. MaxCPB is specified in units of 1 000 bits/s.
- Visual Size
 - Interlaced video:
 - 1 440 × 1 080 or 1 920 × 1 080
 - Progressive video:
 - 1 280 × 720, 1 440 × 1 080 or 1 920 × 1 080
- Frame Rate
 - Interlaced video:
 - 29,97 fps (59,94 fields/s) or 25 fps (50 fields/s) for 1 440 × 1 080 or 1 920 × 1 080
 - Progressive video:
 - 23,976 fps for 1 280 × 720, 1 440 × 1 080 or 1 920 × 1 080
 - 50 fps or 59,94 fps for 1 280 × 720
- Pixel Aspect Ratio
 - 1:1 for 1 280 × 720 and 1 920 × 1 080
 - 4:3 for 1 440 × 1 080

Audio

MPEG-4 ISO/IEC 14496-3:2005

AAC Profile at Level 4 / High Quality Audio Profile at Level 4 (AAC LC only)

With the following restrictions:

- Bitrate
 - average bitrate
 - 256 kbps max (stereo / dual mono) or 640 kbps max (5,1 channels)

peak bitrate

In case of stereo / dual mono, the stream shall be encoded at equal to or less than 384 kbps max.

In case of 5,1 channels, the stream shall be encoded at equal to or less than 960 kbps max.

- Sampling Frequency
48 kHz

This Class is classified into two levels according to bitrate and other parameters.

5.6.2 Level 1

Video

Equivalent to Main Profile at Level 4. (ISO/IEC 14496-10)

With the following restrictions:

- Visual Size
Interlaced video:
1 440 × 1 080
Progressive video:
1 280 × 720

Audio

No further restriction as this Class HD defined in 5.6.1.

5.6.3 Level 2

Video

No further restriction as this Class HD defined in 5.6.1.

Audio

No further restriction as this Class HD defined in 5.6.1.

5.7 Class network HD

5.7.1 General

Class network HD is provided for HDTV quality video to be used for interchange of content on the internet.

For this Class, AVC video Main Profile or AVC video High Profile is employed as the video format and MPEG-4 AAC LC as the audio format.

Recorders compliant with this Class shall make a recording within the bounds of this Class complying with Annex B. The recommended recording modes for this Class are listed in Annex C.

Video

AVC video ISO/IEC 14496-10:2009

Main Profile at Level 4 /High Profile at Level 4

With the following restrictions:

- Bitrate
Main Profile:
Maximum value of Max BR is 20 000
Note that for both of VCL HRD parameters and NAL HRD parameters,
Bitrate[SchedSelIdx] ≤ 1 000 × MaxBR
i.e. MaxBR is specified in units of 1 000 bits/s.
High Profile:
Maximum value of Max BR is 24 000
Note that for both of VCL HRD parameters and NAL HRD parameters,
Bitrate[SchedSelIdx] ≤ 1 000 × MaxBR
i.e. MaxBR is specified in units of 1 000 bits/s.

- CPB Size

Main Profile:

No restriction

High Profile:

Maximum size of MaxCPB is 30 000

Note that for both the VCL HRD and NAL HRD parameters,

$CpbSize[SchedSelIdx] \leq 1\,000 \times \text{MaxCPB}$,

i.e. MaxCPB is specified in units of 1 000 bits/s.

- Visual Size

Progressive video:

1 280 × 720, 1 440 × 1 080 or 1 920 × 1 080

- Frame Rate

Progressive video:

29,97 fps or 25 fps

- Pixel Aspect Ratio

1:1 for 1 280 × 720 and 1 920 × 1 080

4:3 for 1 440 × 1 080

Audio

MPEG-4 ISO/IEC 14496-3:2005

AAC Profile at Level 4 / High Quality Audio Profile at Level 4 (AAC LC only)

With the following restrictions:

- Bitrate

average bitrate

256 kbps max (stereo / dual mono) or 640 kbps max (5,1 channels)

peak bitrate

In case of stereo or dual mono, the stream shall be encoded at equal to or less than 384 kbps max (stereo / dual mono).

In case of 5,1 channels, the stream shall be encoded at equal to or less than 960 kbps max.

- Sampling frequency

48 kHz only

This Class is classified into three levels according to bitrate and other parameters.

5.7.2 Level 1

Video

Equivalent to Main Profile at Level 4. (ISO/IEC 14496-10)

With the following restrictions:

- Visual Size

Progressive video:

1 280 × 720

Audio

Equivalent to AAC Profile at Level 2 / High Quality Audio Profile at Level 2

With the following restrictions:

- Bitrate

average bitrate

256 kbps max (stereo / dual mono)

peak bitrate

The stream shall be encoded at equal to or less than 384 kbps max (stereo / dual mono).

5.7.3 Level 2

Video

Equivalent to Main Profile at Level 4. (ISO/IEC 14496-10)

With the following restrictions:

- Visual Size

Progressive video:

1 280 × 720

Audio

No further restriction as this Class network HD defined in 5.7.1.

5.7.4 Level 3

Video

No further restriction as this Class network HD defined in 5.7.1.

Audio

No further restriction as this Class network HD defined in 5.7.1.

6 Definitions for other files

A thumbnail picture file stores JPEG image for thumbnail use, in conformance with the DCF thumbnail file as defined in JEITA CP-3461 based on JEITA CP-3451 and JEITA CP-3451-1. Note that this file is not based on ISO/IEC 14496-14 and is optional.

This file corresponds to the PCE AV file that has the same file name (not including the file extension). The file shall be 16 KB maximum.

Annex A (normative)

Operation rules for compliant products

A.1 General

This annex discusses the operation rules for products compliant with this Technical Specification.

A.2 Rules for players

A.2.1 General policy

Players shall be capable of parsing any files that are recorded in accordance with this Technical Specification. Playback of such files may be done depending on their capabilities. When reading the files, players shall ignore and skip boxes and descriptors that are unrecognized or prohibited for use in this Technical Specification. Malfunctions or hang-ups shall not occur, and parsing should be continued by ignoring such unknown boxes or descriptors while aborting playback is allowed.

A.2.2 Judgement of compatibility

To judge if the player can play back a file compliant with this Technical Specification, the following information should be used.

File type box:

- major-brand
- compatible-brand

Profile box:

Profile entries (FileGlobalProfileEntry/AudioProfileEntry/VideoProfileEntry)

Additionally, the following should be considered.

File type box:

- It is recommended to use the value of compatible-brand rather than major-brand for better interoperability.
- Note that the values of minor-version may be different. They shall not be used for judgement if the file is compatible.

Profile box:

- To judge compatibility of the audio stream, the value of audioObjectType should be used. Since some audio codec types such as AAC-LC are contained in multiple profiles defined by MPEG or other standards, the value of audioProfileLevelIndication should not have effect on judgement of compatibility if the player supports the audio object type.

A.2.3 Handling of profile box

A.2.3.1 Profile entries

A.2.3.1.1 General

At playback, the profile entries for main audio and video shall be used only as a clue by which a player judges whether a file is within the bounds of the level that the product supports. The player shall not judge if it can play back the file normally solely from this value when the file has parameter settings that are not playable on the player even though those settings are within the bounds of the level it supports. In such a case, detailed parameters shall be confirmed from other information sources such as the DecoderConfigDescriptor. If it is judged from this value that the file does not belong to the level / Class that the player supports, playback need not be done. However, a message or other means should be used to notify users.

A.2.3.1.2 Frame rate

Players that use the `average_frame_rate` field should not interpret the value of this field as the actual average frame rate.

Players that use the `max_frame_rate` field should allow the values even if they are not set to the closest one.

A.2.3.2 Function flags

Players should be able to play back a file with the `function_flags` set to '0' in their Profile Box. See 4.5.3 for details on the `function_flags`. See Annex B for details on codec support.

A.2.4 Handling of data streams

A.2.4.1 General

The operational rules for handling of data streams are described in the following sub-clauses for each Class.

A.2.4.2 MPEG-4 AAC elementary stream

It is optional to have capability to decode SBR data.

If the stream contains SBR data, `fill_element` (FIL) may be interpreted for AAC SBR decoding.

The `program_config_element` in the ESD Box shall be used for decoding an audio stream.

A.2.4.3 AVC video elementary stream

A.2.4.3.1 General

Streams should be displayed according to their pixel aspect ratio, for example un-squeeze the squeezed image, manually or automatically.

A.2.4.3.2 Class MB

In Level 1, the values of `colour_primaries`, `transfer_characteristics` and `matrix_coefficients` shall be assumed as set to 1 (ITU-R BT.709).

Pan-scan rectangle SEI messages need not be used when decoding.

A.2.4.3.3 Class PT

If the value of `colour_description_present_flag` is set to 0, the values of `colour_primaries`, `transfer_characteristics` and `matrix_coefficients` shall be assumed as set to the following values:

- 5 (if the frame rate is 25 fps);
- 6 (if the frame rate is 29,97 fps or 23,976 fps)

A.2.4.3.4 Class SD

If the value of `colour_description_present_flag` is set to 0, the values of `colour_primaries`, `transfer_characteristics` and `matrix_coefficients` shall be assumed as set to the following values:

- 5 (if the frame rate is 25 fps)
- 6 (if the frame rate is 29,97 fps).

It is not mandatory to use the values of `pic_struct` and `ct_type` for display. Note that progressive video can be identified by the parameter setting of Table C.4.

A.2.4.3.5 Class network SD

If the value of `colour_description_present_flag` is set to 0, the values of `colour_primaries`, `transfer_characteristics` and `matrix_coefficients` shall be assumed as set to the following values:

- 5 (if the frame rate is 25 fps)
- 6 (if the frame rate is 29,97 fps or 23,976 fps).

It is not mandatory to use the values of `pic_struct` and `ct_type` for display.

A.2.4.3.6 Class HD

If the value of `colour_description_present_flag` is set to 0, the values of `colour_primaries`, `transfer_characteristics` and `matrix_coefficients` shall be assumed as set to value '1'.

A.2.4.3.7 Class network HD

If the value of `colour_description_present_flag` is set to 0, the values of `colour_primaries`, `transfer_characteristics` and `matrix_coefficients` shall be assumed as set to value '1'.

A.3 Rules for recorders**A.3.1 Metadata**

To identify the product that created a file easier, recorders shall set the following data types of metadata:

- product;
- software.

The content of each data type should be set as follows.

- Product

The manufacturer and the model name / number of the product, software or hardware.

- Software

The version of the software product, or the name of the correspondent component of the software product.

A.3.2 Fragmented movie

A.3.2.1 General

For a fragmented movie, it is recommended that either of the following rules be satisfied.

A.3.2.2 data duration of movie fragment

If a PCE AV file belongs to a Class other than Class HD, the data duration of each movie fragment should be 1 second* maximum if one or more main video tracks contain AVC Video elementary streams.

A.3.2.3 Size of fragment header

The maximum size of a fragment header should be 32 KB (that is, the total size of all Track Boxes in the Movie box in the case of the first fragment, and the size of the Movie Fragment Box in the case of other fragments should be 32 KB or less, respectively).

The data duration of each fragment, therefore, should be limited to its corresponding maximum value that satisfies the condition above.

Annex B (normative)

Capabilities required for compliant products

B.1 General

This annex discusses the capabilities required for products compliant with this Technical Specification.

B.2 Interoperability

B.2.1 Rules for recorders

Recorders compliant with this Technical Specification shall have at least one recording mode that conforms to one of the levels of Classes defined in this Technical Specification. It is not mandatory to support the recording modes defined in Annex C, though it is recommended.

B.2.2 Rules for players

B.2.2.1 General

Playback ability of a player is expressed as a combination of the Class and the level.

It is strongly desired that players compliant with this Technical Specification shall be able to play back files that meet the requirements of the level of the Class that they support.

It is recommended to support playback of files created in the following recording modes.

B.2.2.2 Products supporting Class SD

Players that support Class SD shall support either video system (NTSC or PAL), but it is not necessary to support both.

B.2.2.3 Products supporting Class network SD

Players that support Class network SD shall support either video system (NTSC or PAL), but it is not necessary to support both.

B.2.2.4 Products supporting Class HD

Players that support Class HD shall be able to play back PCE AV files in the form of a fragmented movie as well as those in the form of a regular movie.

Players that support Class HD shall be able to play back PCE AV files of which frame rate is 29,97 fps. Playback of PCE AV files of which frame rate is 25 fps is optional.

B.2.2.5 Products supporting Class network HD

Players that support Class network HD shall be able to play back PCE AV files of which frame rate is 29,97 fps. Playback of PCE AV files of which frame rate is 25 fps is optional.

B.3 Codec

Players and recorders that support one or more of PCE AV Classes (see Clause 5) shall support AVC video codec. If they have audio capability, MPEG-4 AAC-LC codec shall be supported. Players shall support MPEG-4 AAC-LC regardless of the profile for AAC defined by ISO/IEC 14496-3, that is high quality audio profile or AAC Profile; recorders should use AAC profile when encoding. An SBR tool is optional. JPEG codec is optional.

Annex C (informative)

Recommended recording modes

C.1 General

A recording mode is a combination of visual size, frame rate, bitrate, and number of channels.

C.2 Class PT

In this Class, the following recording modes shown below are recommended for use when recording PCE AV files for interoperability with other products (see Table C.1).

Table C.1 – Recommended recording modes for Class PT

Level	Video				Audio		
	Visual size	Pixel aspect ratio	Bitrate kbps	Frame rate fps	Sampling frequency kHz	Bitrate kbps	Channel
3	640 × 480	1:1 / 4:3	3 000	29,97/ 25	48	256	stereo
			3 000			128	
			1 500			128	
2	480 × 270	1:1 / 3:4	1 000	29,97	48	128	stereo
			768			128	
1	320 × 240	1:1 / 4:3	768	29,97 / 25	48	128	stereo
			384			128	
NOTE 1 The “Level” means the level in Class PT to which the recording mode conforms.							
NOTE 2 The “Bitrate” means the target value of the average bitrate of NAL CPB.							

C.3 Class MB

In this Class, the recording modes shown below are recommended for use when recording PCE AV files for interoperability with other products (see Table C.2).

Table C.2 – Recommended recording modes for Class MB

Level	Video				Audio			
	Visual size	Pixel aspect ratio	Bitrate kbps	Frame rate fps	Sampling frequency kHz	Bitrate kbps	Channel	SBR
4	640 × 480	1:1 / 4:3	3 000	29,97 / 25	48	256	stereo	No
			3 000			128		No
			2 000			128		No
	640 × 360	1:1	3 000	29,97 / 25	48	256	stereo	No
			2 000			128		No
			2 000			128		No
3	480 × 270	1:1 / 3:4	768	29,97 / 25	48	128	stereo	No
2	320 × 240	1:1 / 4:3	768	29,97 / 25	48	128	stereo	No
			384			128		No
1	320 × 240	1:1	384	14,985 / 12,5	48	128	stereo	No
	320 × 180		384			128		No
NOTE 1 The “Level” means the level in Class MB to which the recording mode conforms.								
NOTE 2 The “Bitrate” means the target value of the average bitrate of NAL CPB.								

C.4 Class SD

In this Class, the recording modes shown below are recommended for use when recording PCE AV files for interoperability with other products (see Table C.3).

Table C.3 – Recommended recording modes for Class SD

Video					Audio		
Video system	Visual size	Pixel aspect ratio	Bitrate kbps	Frame rate fps	Sampling frequency kHz	Bitrate kbps	Channel
SDTV 50 Hz	720 × 576	12:11 / 16:11	9 000	25	48	256	stereo or dual mono
			9 000			128	
			6 000			256	
			6 000			128	
			3 000			128	
			2 000			128	
SDTV 59,94 Hz	720 × 480	10:11 / 40:33	9 000	29,97	48	256	stereo or dual mono
			9 000			128	
			6 000			256	
			6 000			128	
			3 000			128	
			2 000			128	

NOTE 1 The “Level” means the level in Class SD to which the recording mode conforms.

NOTE 2 The “Bitrate” means the target value of the average bitrate of NAL CPB.

With this Class, progressive video 25[fps] and 29,97[fps] can be achieved by using the parameters shown in Table C.4

Table C.4 – Parameters for progressive video 25[fps] and 29.97[fps]

Frame rate fps	interlaced_video flag in video profile entry	frame_mbs_only_f lag in Sequence Parameter Set	pic_struct in Picture Timing SEI Message	ct_type in Picture Timing SEI Message
25	1	0	3 (top field first and then bottom) or 4	0
29,97	1	0	3 (top field first and then bottom) or 4	0

C.5 Class network SD

In this Class, the recording modes shown below are recommended for use when recording PCE AV files for interoperability with other products (see Table C.5).

Table C.5 – Recommended recording modes for Class network SD

Video					Audio		
Video system	Visual size	Pixel aspect ratio	Bitrate kbps	Frame rate fps	Sampling frequency kHz	Bitrate kbps	Channel
SDTV 50 Hz	720 × 576	12:11 / 16:11	9 000	25	48	256	stereo or dual mono
			9 000			128	
			6 000			256	
			6 000			128	
			3 000			128	
			2 000			128	
SDTV 59,94 Hz	720 × 480	10:11 / 40:33	9 000	29,97/23,976	48	256	stereo or dual mono
			9 000			128	
			6 000			256	
			6 000			128	
			3 000			128	
			2 000			128	

NOTE 1 The “Level” means the level in Class network SD to which the recording mode conforms.

NOTE 2 The “Bitrate” means the target value of the average bitrate of NAL CPB.

C.6 Class HD

In this Class, the recording modes shown below are recommended for use when recording PCE AV files for interoperability with other products (see Table C.6).

Table C.6 – Recommended recording modes for Class HD

Video					Audio		
Level	Visual size	Pixel aspect ratio	Bitrate kbps	Frame rate fps	Sampling frequency kHz	Bitrate kbps	Channel
2	1 920 × 1 080	1:1	15 000	29,97 / 25 / 23,976	48	384	5,1 channels
			15 000			256	stereo or
			15 000			128	dual mono
			12 000			384	5,1 channels
			12 000			256	stereo or
			12 000			128	dual mono
			9 000			384	5,1 channels
			9 000			256	stereo or
			9 000			128	dual mono
1	1 280 × 720	1:1	12 000	59,94 / 50 / 23,976	48	384	5,1 channels
			12 000			256	stereo or
			12 000			128	dual mono
			9 000			384	5,1 channels
			9 000			256	stereo or
			9 000			128	dual mono
			6 000			384	5,1 channels
			6 000			256	stereo or
			6 000			128	dual mono
	1 440 × 1 080	4:3	12 000	29,97 / 25 / 23,976	48	384	5,1 channels
			12 000			256	stereo or
			12 000			128	dual mono
			9 000			384	5,1 channels
			9 000			256	stereo or
			9 000			128	dual mono
			6 000			384	5,1 channels
			6 000			256	stereo or
			6 000			128	dual mono

NOTE 1 The "Level" means the level in Class HD to which the recording mode conforms.

NOTE 2 The "Bitrate" means the target value of the average bitrate of NAL CPB.

C.7 Class network HD

In this Class, the recording modes shown below are recommended for use when recording PCE AV files for interoperability with other products (see Table C.7).

Table C.7 – Recommended recording modes for Class network HD

Video					Audio		
Level	Visual size	Pixel aspect ratio	Bitrate kbps	Frame rate fps	Sampling frequency kHz	Bitrate kbps	Channel
3	1 920 × 1 080	1:1	15 000	29,97 / 25	48	384	5,1 channels
			15 000			256	stereo or dual mono
			15 000			128	
			12 000			384	5,1 channels
			12 000			256	stereo or dual mono
			12 000			128	
			9 000			384	5,1 channels
			9 000			256	stereo or dual mono
			9 000			128	
	1 440 × 1 080	4:3	12 000			384	5,1 channels
			12 000			256	stereo or dual mono
			12 000			128	
			9 000			384	5,1 channels
			9 000			256	stereo or dual mono
			9 000			128	
			6 000			384	5,1 channels
			6 000			256	stereo or dual mono
			6 000			128	
2	1 280 × 720	1:1	9 000	384	5,1 channels		
			6 000	384			
			4 000	384			
1	1 280 × 720	1:1	9 000	256	stereo or dual mono		
			9 000	128			
			6 000	256			
			6 000	128			
			4 000	256			
			4 000	128			
NOTE 1 The “Level” means the level in Class network HD to which the recording mode conforms.							
NOTE 2 The “Bitrate” means the target value of the average bitrate of NAL CPB.							

Annex D (normative)

Operation rules for rotated video file

D.1 General

This annex defines the rules to contain video that is recorded at a rotated angle. To contain rotated video, an existing Class shall be used for a video elementary stream, while manipulating the matrix parameter in the Track Header Box.

Note that the rotated angle cannot be changed within a stream.

D.2 Class

The Classes that allow the containing of rotated video are Class PT, Class MB, Class HD, and Class network HD.

D.3 File structure – Movie structure

The structure of a rotated video file shall conform to the following rule:

- A subtitle track is not contained.

D.4 Operational rules for tracks – Main video track

To contain rotated video, the following values shall be used.

Track Header Box

matrix = 90 degree clockwise rotation {0, 00010000h,0, FFFF0000h,0,0, height,0, 40000000h}

180 degree clockwise rotation {FFFF0000h,0,0,0, FFFF0000h,0, width, height, 40000000h}

270 degree clockwise rotation {0, FFFF0000h,0, 00010000h,0,0,0, width, 40000000h}

where width and height shall be the values specified in the Track Header Box. Other settings are prohibited.

width = the value converted for square pixels.

If not square pixels, an appropriate value shown in the table below shall be used.

height = the value converted for square pixels.

If not square pixels, an appropriate value shown in the Table D.1 shall be used.

Table D.1 – Width and height in the track header box

Visual sample entry		Pixel aspect ratio	Track header box	
Width	Height		Width	Height
1 440	1 080	4:3	1 920	1 080
640	480	4:3	854	480
480	270	3:4	360	270
320	240	4:3	426	240

D.5 Operational rules for products

D.5.1 Rules for recorders – Thumbnail

When there is a video track with a rotated matrix, a thumbnail image for it can be any of the following:

- a rotated image (with black bars or transparent color inserted);
- a image extracted as is from the video elementary stream (i.e. not rotated image);
- in the case of JPEG, a picture using the orientation tag.

D.5.2 Rules for players

The devices supporting rotation are recommended to be able to handle rotated video for all Classes/levels they support.

Also, the devices that have a feature to display JPEG thumbnail or JPEG artwork images, it is recommended that they support the JPEG orientation tag.

Bibliography

ISO/IEC 14496 (all parts), *Information technology – Coding of audio-visual objects*

JEITA CP-3451:2002, *Exchangeable image file format for digital still cameras: Exif Version 2.2*

JEITA CP-3451-1:2003, *Exchangeable image file format for digital still cameras: Exif Version 2.21 (Amendment Ver2.2)*

JEITA CP-3461:2003, *Design rule for Camera File system, version 2.0*

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

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

INTERNATIONAL
ELECTROTECHNICAL
COMMISSION

3, rue de Varembé
PO Box 131
CH-1211 Geneva 20
Switzerland

Tel: + 41 22 919 02 11
Fax: + 41 22 919 03 00
info@iec.ch
www.iec.ch