
**Information technology — Coding of
audio-visual objects —**

**Part 15:
Carriage of network abstraction layer
(NAL) unit structured video in the ISO
base media file format**

Technologies de l'information — Codage des objets audiovisuels —

*Partie 15: Transport de vidéo structurée en unités NAL sur la couche
réseau au format ISO de base pour les fichiers médias*

This document is a preview generated by EMS



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2019

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

	Page
Foreword	vi
Introduction	vii
1 Scope	1
2 Normative references	1
3 Terms, definitions and abbreviated terms	1
3.1 Terms and definitions	1
3.2 Abbreviated terms	7
4 General definitions	8
4.1 Overview	8
4.2 Elementary stream structure	8
4.3 Sample and configuration definition	9
4.3.1 General	9
4.3.2 Canonical order and restrictions	9
4.3.3 Sample format	10
4.3.4 Optional boxes in the sample entry	11
4.4 Video track structure	11
4.5 Template fields used	11
4.6 Visual width and height	11
4.7 Decoding time (DTS) and composition time (CTS)	12
4.8 Sample groups on random access recovery points 'roll' and random access points 'rap'	12
4.9 Hinting	12
4.10 On change of sample entry	12
4.11 SEI information box	14
4.11.1 Definition	14
4.11.2 Syntax	14
4.11.3 Semantics	14
4.12 Post-decoder requirements scheme for signalling of SEI	14
4.12.1 General	14
4.12.2 Definition	15
5 AVC elementary streams and sample definitions	15
5.1 Overview	15
5.2 Elementary stream structure	15
5.3 Sample and configuration definition	18
5.3.1 Overview	18
5.3.2 Canonical order and restrictions	18
5.3.3 Decoder configuration information	19
5.4 Derivation from ISO base media file format	22
5.4.1 AVC file type and identification	22
5.4.2 AVC video stream definition	23
5.4.3 AVC parameter set stream definition	24
5.4.4 Parameter sets	25
5.4.5 Sync sample	26
5.4.6 Shadow sync	26
5.4.7 Layering and sub-sequences	27
5.4.8 Alternate streams and switching pictures	30
5.4.9 Definition of a sub-sample for AVC	33
6 SVC elementary stream and sample definitions	33
6.1 Overview	33
6.2 Elementary stream structure	34
6.3 Use of the plain AVC file format	35
6.4 Sample and configuration definition	35

6.4.1	Overview	35
6.4.2	Canonical order and restrictions	35
6.5	Derivation from the ISO base media file format	37
6.5.1	SVC track structure	37
6.5.2	Data sharing and extraction	37
6.5.3	SVC video stream definition	38
6.5.4	SVC visual width and height	40
6.5.5	Sync sample	40
6.5.6	Shadow sync	41
6.5.7	Independent and disposable samples box	41
6.5.8	Sample groups on random access recovery points 'roll' and random access points 'rap'	41
6.5.9	Definition of a sub-sample for SVC	41
7	MVC and MVD elementary stream and sample definitions	43
7.1	Overview	43
7.2	Overview of MVC or MVD storage	44
7.3	MVC and MVD elementary stream structures	46
7.4	Use of the plain AVC file format	48
7.5	Sample and configuration definition	48
7.5.1	Overview	48
7.5.2	Canonical order and restriction	48
7.5.3	Decoder configuration record	48
7.6	Derivation from the ISO base media file format	51
7.6.1	MVC and MVD track structures	51
7.6.2	Reconstruction of an access unit	51
7.6.3	Sample entry	52
7.6.4	Sync sample	64
7.6.5	Shadow sync	64
7.6.6	Independent and disposable samples box	65
7.6.7	Sample groups on random access recovery points 'roll' and random access points 'rap'	65
7.7	MVC specific information boxes	65
7.7.1	Overview	65
7.7.2	Multiview information box	66
7.7.3	Multiview group box	66
7.7.4	Multiview group relation box	68
7.7.5	Multiview relation attribute box	69
7.7.6	Multiview scene info box	75
7.7.7	MVC view priority assignment box	76
8	HEVC elementary streams and sample definitions	76
8.1	Overview	76
8.2	Elementary stream structure	77
8.3	Sample and configuration definition	77
8.3.1	Overview	77
8.3.2	Canonical order and restrictions	77
8.3.3	Decoder configuration information	78
8.4	Derivation from ISO base media file format	82
8.4.1	HEVC video stream definition	82
8.4.2	Parameter sets in sample entry	83
8.4.3	Sync sample	83
8.4.4	Sync sample sample grouping	84
8.4.5	Temporal scalability sample grouping	85
8.4.6	Temporal sub-layer access sample grouping	87
8.4.7	Step-wise temporal layer access sample grouping	87
8.4.8	Definition of a sub-sample for HEVC	88
8.4.9	Handling non-output samples	90
9	Layered HEVC elementary stream and sample definitions	91

9.1	Overview	91
9.2	Overview of L-HEVC storage	92
9.3	L-HEVC elementary stream structure	92
9.4	Sample and configuration definition	93
9.4.1	Overview	93
9.4.2	Canonical order and restrictions	93
9.4.3	Decoder configuration record	93
9.5	Derivation from the ISO base media file format and the HEVC file format (Clause 8)	94
9.5.1	L-HEVC track structure	94
9.5.2	Data sharing and reconstruction of an L-HEVC bitstream	95
9.5.3	L-HEVC video stream definition	97
9.5.4	L-HEVC visual width and height	100
9.5.5	Sync sample	100
9.5.6	Independent and disposable samples box	101
9.5.7	Stream access point sample group	101
9.5.8	The 'roll', 'rap', 'sync', 'tsas' and 'stsa' sample groups	102
9.5.9	Definition of a sub-sample for L-HEVC	102
9.5.10	Handling non-output samples	102
9.6	L-HEVC specific structures	103
9.6.1	External base layer sample group	103
9.6.2	The operating points information sample group	103
9.6.3	The layer information sample group	107
9.6.4	The layer information sample group	108
10	Storage of tiled HEVC and L-HEVC video streams	109
10.1	Overview	109
10.2	NAL unit map entry	110
10.2.1	Definition	110
10.2.2	Syntax	110
10.2.3	Semantics	111
10.3	Tile region group entry	111
10.3.1	Definition	111
10.3.2	Syntax	112
10.3.3	Semantics	112
10.4	Tile sub track definition	114
10.4.1	Overview	114
10.4.2	TileSubTrackGroupBox	114
10.5	HEVC and L-HEVC tile track	115
10.5.1	Overview	115
10.5.2	Sample entry name and format for HEVC tile tracks	116
10.5.3	Sample entry name and format for L-HEVC tile tracks	117
10.5.4	Bitstream reconstruction from tile base and tile tracks	117
10.5.5	Sample entry names for tile base tracks	118
	Annex A (normative) In-stream structures	119
	Annex B (normative) SVC, MVC, and MVD sample group and sub-track definitions	128
	Annex C (normative) Temporal metadata support	152
	Annex D (normative) File format toolsets and brands	162
	Annex E (normative) Sub-parameters for the MIME type 'codecs' parameter	165
	Annex F (informative) Unspecified nal_unit_type value management	170

Foreword

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

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see <http://patents.iec.ch>).

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

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

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This fifth edition cancels and replaces the fourth edition (ISO/IEC 14496-15:2017), which has been technically revised. It also incorporates Amendments ISO/IEC 14496-15:2017/Amd.1:2018 and ISO/IEC 14496-15:2017/Amd.2:2019.

The main changes compared to the previous edition are as follows:

- additional content incorporated as subclauses [4.11](#), [4.12](#), [9.6.4](#), [D.4.3](#), [D.4.4](#) and [D.4.5](#) and [Annex F](#);
- corrections in [Tables 2](#), [3](#) and [6](#) and subclause [A.1](#);
- deletion of subclause 5.4.10;
- minor editorial changes to align the document with the drafting rules in ISO/IEC Directives Part 2.

A list of all parts in the ISO/IEC 14496 series can be found on the ISO website.

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

Introduction

This document defines a storage format based on, and compatible with, the ISO base media file format (ISO/IEC 14496-12), which is used by the MP4 file format (ISO/IEC 14496-14) and the motion JPEG 2000 file format (ISO/IEC 15444-3) among others. This document enables video streams formatted as network adaptation layer units (NAL units) to

- a) be used in conjunction with other media streams, such as audio,
- b) be used in an MPEG-4 systems environment, if desired,
- c) be formatted for delivery by a streaming server, using hint tracks, and
- d) inherit all the use cases and features of the ISO base media file format on which MP4 and MJ2 are based.

This document can be used as a standalone specification; it specifies how NAL unit structured video content is stored in an ISO base media file format compliant format. However, it is normally used in the context of a specification, such as the MP4 file format, derived from the ISO base media file format, that permits the use of NAL unit structured video such as AVC (ISO/IEC 14496-10) video and high efficiency video coding (HEVC, ISO/IEC 23008-2) video.

The ISO base media file format is becoming increasingly common as a general-purpose media container format for the exchange of digital media, and its use in this context should accelerate both adoption and interoperability.

The International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) draw attention to the fact that it is claimed that compliance with this document may involve the use of patents.

ISO and IEC take no position concerning the evidence, validity and scope of these patent rights. The holders of these patent rights have assured ISO and IEC that they are willing to negotiate licences under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statements of the holders of these patent rights are registered with ISO and IEC. Information may be obtained from:

Fraunhofer-Gesellschaft
Hansastr. 27c
80686 München
Germany

Nokia Corporation
PO Box 86
FIN-24101 Salo
Finland

TDVision Systems, Inc.
8001 Irvine Center Drive, Suite 400
Irvine, CA 92618
USA

Qualcomm Incorporated
5775 Morehouse Drive,
San Diego, CA 92121
USA

Huawei Technologies Co., Ltd.
Bantian Longgang District,
Shenzhen 518129
China

Telefonaktiebolaget LM Ericsson (publ)
Ericsson AB,
SE-164 80 Stockholm
Sweden

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights other than those identified above. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Information technology — Coding of audio-visual objects —

Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format

1 Scope

This document specifies the storage format for streams of video that is structured as NAL units, such as AVC (ISO/IEC 14496-10) and HEVC (ISO/IEC 23008-2) video streams.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 14496-10:2014, *Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding*

ISO/IEC 14496-12:2015, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

ISO/IEC 23008-2:2017, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding*

3 Terms, definitions and abbreviated terms

3.1 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 14496-10, ISO/IEC 23008-2 and the following apply.

3.1.1

3D-AVC NAL unit

3D-AVC VCL NAL unit

NAL unit with type 21 with `avc_3d_extension_flag` equal to 1

Note 1 to entry: As specified in ISO/IEC 14496-10:2014, Annex J.

3.1.2

aggregator

in-stream structure ([3.1.11](#)) using a NAL unit header for grouping of NAL units belonging to the same sample

3.1.3

AVC base layer

maximum subset of a bitstream that is AVC compatible

Note 1 to entry: This can also be expressed as a bitstream not using any of the functionality of ISO/IEC 14496-10:2014 Annex G, Annex H, Annex I, or Annex J.