FINAL DRAFT

INTERNATIONAL STANDARD

ISO/IEC FDIS 14496-15

ISO/IEC JTC 1/SC 29

Secretariat: JISC

Voting begins on: **2022-07-15**

Voting terminates on:

2022-09-09

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

https://standards.iteh.ai/catalog/standards/sist/00017d28-c50f-40b4-b65d-c57b29045362/iso-iec-fdis-14496-15

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.



iTeh STANDARD PREVIEW (standards.iteh.ai)

ISO/IEC FDIS 14496-15
https://standards.iteh.ai/catalog/standards/sist/00017d28-c50f-40b4-b65d-c57b29045362/iso-iec-fdis-14496-15



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2022

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 Email: copyright@iso.org Website: www.iso.org Published in Switzerland

Contents Page Foreword.......vi Introduction......vii 1 2 3 Terms, definitions, abbreviated terms and conventions......1 3.1 Terms and definitions...... 1 3.2 Conventions.......11 3.3 4 4.1 Overview......12 4.2 Sample and configuration definition......12 Video track structure14 4.3 Template fields used14 4.4 4.5 Visual width and height......14 Decoding time (DTS) and composition time (CTS)......15 4.6 4.7 Sample groups on random access recovery points 'roll' and random access points 'rap '......15 4.8 4.9 On change of sample entry (informative)......16 4.10 4.11 Post-decoder requirements scheme for signalling of SEI......18 4.12 Alternative extraction source track grouping19 NAL unit map entry......19 4.13 4.14 Rectangular region group entry......21 4.15 Layer information sample group......23 5 AVC elementary streams and sample definitions......25 5.1 5.2 Elementary stream structure......25 Sample and configuration definition......28 5.3 5.4 Derivation from ISO base media file format32 SVC elementary stream and sample definitions......44 6 Overview......44 6.1 Elementary stream structure......44 6.2 Use of the plain AVC file format......45 6.3 6.4 Sample and configuration definition......45 Derivation from the ISO base media file format......48 6.5 MVC and MVD elementary stream and sample definitions......54 7 7.1 Overview of MVC or MVD Storage......55 7.2 7.3 MVC and MVD elementary stream structures......57

7.5 Sample and configuration definition	62
	77
7.7 MVC specific information boxes	-
8 HEVC elementary streams and sample definitions	87
8.1 Overview	87
8.2 Elementary stream structure	87
8.3 Sample and configuration definition	88
8.4 Derivation from ISO base media file format	93
9 Layered HEVC elementary stream and sample definitions	102
9.1 Overview	
9.2 Overview of L-HEVC storage	103
9.3 L-HEVC elementary stream structure	104
9.4 Sample and configuration definition	104
9.5 Derivation from the ISO base media file format and the HEVC file format (Clau	se 8)
9.6 L-HEVC specific structures	117
10 Storage of tiled HEVC and L-HEVC video streams	123
10.1 Overview	123
10.2 NAL unit map entry	124
10.3 Tile region group entry	124
10.4 Tile sub track definition	124
10.5 HEVC and L-HEVC tile track	
10.6 HEVC slice segment data track	130
11 VVC elementary streams and sample definitions	
11.1 Overview	131
11.2 Sample and configuration definition	138
11.3 Derivation from ISO base media file format	147
11.4 Sample groups	161
11.5 Entity groups	181
Data sharing and VVC bitstream reconstruction	189
12 EVC elementary streams and sample definitions	200
12.1 Overview	200
12.2 Elementary stream structure	200
12.3 Sample and configuration definition	201
12.4 Derivation from ISO base media file format	204
Annex A (normative) In-stream structures	211
Annex B (normative) SVC, MVC, and MVD sample group and sub-track definitions	229
Annex C (normative) Temporal metadata support	252
Annex D (normative) File format toolsets and brands	261
Annex E (normative) Sub-parameters for the MIME type 'codecs' parameter	265

Annex F (informative)	Unspecified nal_unit_type value management for s	sample entry types of
AVC and HEVC		274
Annex G (informative)	Examples of VVC base and subpicture tracks	276

iTeh STANDARD PREVIEW (standards.iteh.ai)

ISO/IEC FDIS 14496-15
https://standards.iteh.ai/catalog/standards/sist/00017d28-c50f-40b4-b65d-c57b29045362/iso-iec-fdis-14496-15

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 or 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 https://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. In the IEC, see www.iec.ch/understanding-standards.

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 sixth edition cancels and replaces the fifth edition (ISO/IEC 14496-15:2019), which has been technically revised. It also incorporates the Amendment ISO/IEC 14496-15:2019/Amd 1:2020.

The main changes are as follows:

- Support for the Versatile Video Coding (ISO/IEC 23090-3) and Essential Video Coding (ISO/IEC 23094-1)
- Addition of sample entry types 'hvc3', 'hev3', 'hvt2', and 'hvt3' targeted at tile-based delivery and merging of High Efficiency Video Coding (ISO/IEC 23008-2) bitstreams

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

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

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 may be used as a standalone document; it specifies how NAL unit structured video content shall be 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 a patent.

ISO and IEC take no position concerning the evidence, validity and scope of this patent right.

The holder of this patent right has 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 statement of the holder of this patent right is registered with ISO and IEC. Information may be obtained from the patent database available at www.iso.org/patents or patents.iec.ch.

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

iTeh STANDARD PREVIEW (standards.iteh.ai)

ISO/IEC FDIS 14496-15

https://standards.iteh.ai/catalog/standards/sist/00017d28-c50f-40b4-b65d-c57b29045362/iso-iec-fdis-14496-15

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. In addition, Annex E specifies parameters and sub-parameters applying when sample entries specified in this document are used as the 'codecs' parameter of a MIME type, as specified in IETF RFC 6381.

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/IEC 14496-12:2020, Information technology — Coding of audio-visual objects — Part 12: ISO base media file format

ISO/IEC 14496-10:2020, Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding

ISO/IEC 23008-2:2020, Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding

ISO/IEC 23090-3:2021, Information technology — Coded representation of immersive media — Part 3: Versatile video coding

ISO/IEC 23094-1:2020, Information technology — General video coding — Part 1: Essential video coding

IETF RFC 4648, The Base16, Base32, and Base64 data encodings

IETF RFC 6381, MIME codecs and profiles

3 Terms, definitions, abbreviated terms and conventions

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, ISO/IEC 23090-3 or ISO/IEC 23094-1, and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

- ISO Online browsing platform: available at https://www.iso.org/obp
- IEC Electropedia: available at https://www.electropedia.org/

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

3.1.2

aggregator

in-stream structure using a NAL unit header for grouping of NAL units belonging to the same sample

3.1.3

alternate region set

set of rectangular regions that are alternatives to be used as a rectangular region when reconstructing a VVC bitstream from a VVC extraction base track

3.1.4

applicable video coding standard

video coding standard for the data carried in the track

Note 1 to entry: The video coding standard can be ISO/IEC 14496-10, ISO/IEC 23008-2, ISO/IEC 23090-3, or ISO/IEC 23094-1.

3.1.5

AU- or picture-level non-VCL NAL unit

non-VCL NAL unit that applies to one or more entire AUs or one or more entire pictures

Note 1 to entry: An AU-level non-VCL NAL unit applies to one or more entire AUs. A picture-level non-VCL NAL unit applies to one or more entire pictures. In VVC, AU-level or picture-level non-VCL NAL units include: 1) all the DCI, OPI, VPS, SPS, PPS, AUD, PH, EOS, and EOB NAL units; 2) APS NAL units that apply to one or more entire AUs or pictures; and 3) SEI NAL units that only contain SEI messages that apply to one or more entire AUs or pictures.

3.1.6

AVC base laver

maximum subset of a bitstream that is AVC compatible

Note 1 to entry: The AVC base layer is represented by AVC VCL NAL units and associated non-VCL NAL units. The AVC base layer is not using any of the functionality of ISO/IEC 14496-10:2020, Annex G, Annex H, Annex I, or Annex J.

Note 2 to entry: The AVC base layer itself can be a temporal scalable bitstream.

3.1.7

AVC parameter set sample

sample in a parameter set elementary stream that consists of those parameter set NAL units that are to be considered as if present in the video elementary stream at the same instant in time

3.1.8

AVC sample

access unit as defined in ISO/IEC 14496-10

3.1.9

AVC NAL unit

AVC VCL NAL unit and its associated non-VCL NAL units in a bitstream

3.1.10

AVC VCL NAL unit

NAL unit with type 1 to 5 (inclusive)

3.1.11

canonical order

order of NAL units that conforms to the applicable video standard

Note 1 to entry: When a single track carries a video bitstream, the NAL units are stored in the canonical order. When multiple tracks are used to a carry a video bitstream, an implicit or explicit video bitstream reconstruction process might be applied to recover the canonical order.

3.1.12

canonical stream format

elementary stream that contains NAL units in the canonical order and conforms to the constraints specified in this document for carrying an elementary stream of the applicable video standard in one or more tracks

3.1.13

complete subset

minimal set of tracks that contain all the information in the original bitstream

3.1.14

cropped frame dimensions

width and height of the decoded frame after applying the output cropping parameters

3.1.15

default sample group description index 2/iso-iec-fdis-14496-15

 $\label{lem:condition} \mbox{default_group_description_index} \quad \mbox{of SampleGroupDescriptionBox} \quad \mbox{with version} \\ \mbox{greater than or equal to 2}$

3.1.16

elementary stream

sequence of one or more bitstreams of the applicable video standard

Note 1 to entry: The term elementary stream is not directly related to the terms video elementary stream, parameter set elementary stream, and video and parameter set elementary stream.

Note 2 to entry: The applicable video standard can be included as a prefix to the term elementary stream. For example, an AVC elementary stream refers to an elementary stream that is a sequence of one or more bitstreams conforming to ISO/IEC 14496-10.

3.1.17

extractor

in-stream structure using a NAL unit header for extraction of data from other tracks

Note 1 to entry: Extractors contain instructions on how to extract data from other tracks. Logically an Extractor can be seen as a pointer to data. While reading a track containing Extractors, the Extractor is replaced by the data it is pointing to.

3.1.18

HEVC sample

access unit as defined in ISO/IEC 23008-2

3.1.19

implicit reconstruction

reconstruction of a stream of access units from two or more tracks not using extractors

3.1.20

in-stream structure

structure residing within sample data

3.1.21

layer

scalable layer

<SVC, MVC, and MVD> set of VCL NAL units with the same values of dependency_id, quality_id, and temporal_id, and the associated non-VCL NAL units

Note 1 to entry: A scalable layer with any of dependency_id, quality_id, and temporal_id not equal to 0 enhances the video by one or more scalability levels in at least one direction (temporal, quality or spatial resolution)

Note 2 to entry: SVC uses a "layered" encoder design that results in a bitstream representing "coding layers". In some publications the 'base layer' is the first quality layer of a specific coding layer. In some publications the base layer is the scalable layer with the lowest priority. The SVC file format uses "scalable layer" or "layer" in a general way for describing nested bitstreams (using terms like AVC base layer or SVC enhancement layer).

3.1.22

layer

scalable layer

<HEVC and VVC> set of VCL NAL units with the same value of nuh_layer_id and the associated non-VCL

NAL units

https://standards.iteh.ai/catalog/standards/sist/00017d28-c50f-40b4-b65d-

3.1.23

layer set

set of layers represented within a bitstream created from another bitstream by operation of the subbitstream extraction process

3.1.24

L-HEVC sample

picture units that are within an access unit as specified in Annex F of ISO/IEC 23008-2:2020 and are represented by the track

3.1.25

MVC NAL unit

MVC VCL NAL unit and its associated non-VCL NAL units in an MVC stream

Note 1 to entry: The association of non-VCL NAL units with MVC VCL NAL units is specified in ISO/IEC 14496-10:2020, Annex H.

3.1.26

MVC sample

one or more view components as defined in Annex H of ISO/IEC 14496-10:2020 and the associated non-VCL NAL units

3.1.27

MVC VCL NAL unit

NAL unit with type 20, and NAL units with type 14 when the immediately following NAL units are AVC VCL NAL units

Note 1 to entry: MVC VCL NAL units do not affect the decoding process of a legacy AVC decoder.

3.1.28

MVC+D depth NAL unit MVC+D depth VCL NAL unit

NAL unit with type 21 containing a coded slice extension for a depth view component

3.1.29

MVD NAL unit

MVD VCL NAL unit

NAL unit with type 21, containing a coded slice extension for a depth view component coded with MVC+D or 3D-AVC, or a 3D-AVC texture view component

3.1.30

MVD sample

one or more view components as defined in Annex I or Annex J of ISO/IEC 14496-10:2020 and the associated non-VCL NAL units, where each view component contains a texture view component, a depth view component or both

3.1.31

NAL-unit-like structure

data structure that is similar to NAL units in the sense that it also has a NAL unit header and a payload, with a difference that the payload might not follow the start code emulation prevention mechanism required for the NAL unit syntax ii/catalog/standards/sist/00017d28-c50f-40b4-b65d-

3.1.32

natively present

not included in an aggregator or an extractor

Note 1 to entry: Data referred to by (hence not included in) an aggregator is considered as natively present. Data included in an aggregator is not considered as natively present.

3.1.33

operating point

<SVC, MVC, and MVD> independently decodable subset of a layered bitstream

Note 1 to entry: Each operating point consists of all the data needed to decode this particular bitstream subset.

Note 2 to entry: In an SVC stream an operating point represents a particular spatial resolution, temporal resolution, and quality, and can be represented either by (i) specific values of DTQ (dependency_id, temporal_id and quality_id) or (ii) specific values of P (priority_id) or (iii) combinations of them (e.g. PDTQ). Note that the usage of priority_id is defined by the application. In an SVC file a track represents one or more operating points. Within a track tiers can be used to define multiple operating points.

Note 3 to entry: The bitstream subset of an MVC or MVD operating point represents a particular set of target output views at a particular temporal resolution, and consists of all the data needed to decode this particular bitstream subset. In MVD each target output view in the bitstream subset of an MVD operating point can contain a texture view, a depth view or both.

Note 4 to entry: An operating point is referred to as an operation point in Annex H of ISO/IEC 14496-10.

3.1.34

operating point

<HEVC> independently decodable subset of a layered bitstream, where one or more layers in the set of layers are indicated to be output layers

Each operating point consists of all the data needed to decode this particular bitstream subset. Note 1 to entry:

Note 2 to entry: An operating point is referred to as an output operation point in ISO/IEC 23008-2.

3.1.35

operating point

<VVC> temporal subset of an output layer set (OLS), identified by an output layer set (OLS) index and a highest value of TemporalId

Note 1 to entry: Each operating point consists of all the data needed to decode this particular bitstream subset.

Note 2 to entry: An operating point is referred to as an operation point in ISO/IEC 23090-3.

3.1.36

output layer set

set of layers consisting of the layers of one of the specified layer sets, where one or more layers in the set of layers are indicated to be output layers, as specified in ISO/IEC 23008-2

3.1.37

parameter set

video parameter set, sequence parameter set, picture parameter set, or adaptation parameter set as defined in the applicable video standard and a live a live in a live in a live in the applicable video standard

This term is used to refer to all types of parameter sets. Note 1 to entry:

3.1.38 parameter set elementary stream c57b29045362/iso-iec-fdis-14496-15

elementary stream containing samples made up of only sequence and picture parameter set NAL units synchronized with the video elementary stream

3.1.39

picture unit

set of VCL NAL units and their associated non-VCL NAL units

The association of VCL NAL units and non-VCL NAL units with picture units is specified in the Note 1 to entry: applicable video standard.

3.1.40

prefix NAL unit

NAL units with type 14

Prefix NAL units provide scalability information about AVC VCL NAL units and filler data NAL units. Note 1 to entry: Prefix NAL units do not affect the decoding process of a legacy AVC decoder. The behaviour of a legacy AVC file reader as a response to prefix NAL units is undefined.

3.1.41

rectangular region

rectangle that does not contain holes and does not overlap with any other rectangular region of the same picture

3.1.42

reference layer

layer that is indicated as possibly needed for decoding of another layer

Note 1 to entry: For layered HEVC, reference layers can be indicated by the 'oinf' sample group defined in clause 9.6.2.

3.1.43

scalable layer representation

bitstream subset that is required for decoding the scalable layer, consisting of the scalable layer itself and all the scalable layers on which the scalable layer depends

3.1.44

sub-picture

<SVC> proper subset of coded slices of a scalable layer representation

3.1.45

sub-picture tier

<SVC> tier that consists of sub-pictures that are constrained so that any coded slice that is not included in the tier representation of this sub-picture tier is not referred to in inter prediction or inter-layer prediction for decoding of this sub-picture tier

3.1.46

sub-layer

temporal sub-layer

<HEVC> set of VCL NAL units with a particular value of TemporalId and the associated non-VCL NAL units

3.1.47

sublayer

ISO/IEC FDIS 14496-15

temporal sublayer indards.iteh.ai/catalog/standards/sist/00017d28-c50f-40b4-b65d-

< VVC > set of VCL NAL units with a particular value of Temporalld and the associated non-VCL NAL units

3.1.48

SVC enhancement layer

layer that specifies a part of a scalable bitstream that enhances the video

Note 1 to entry: An SVC enhancement layer is represented by SVC VCL NAL units and the associated non-VCL NAL units and SEI messages.

Note 2 to entry: Usually an SVC enhancement layer represents a spatial or coarse-grain scalability (CGS) coding layer (identified by a specific value of dependency_id).

3.1.49

SVC NAL unit

SVC VCL NAL unit and its associated non-VCL NAL units in an SVC stream

Note 1 to entry: The association of non-VCL NAL units with svc VCL NAL units is specified in ISO/IEC 14496-10:2020, Annex G.

3.1.50

SVC sample

NAL units that belong to an access unit as defined in ISO/IEC 14496-10:2020, subclause 7.4.1.2, and are represented by the track