
**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*



Reference number
ISO/IEC 14496-15:2019(E)

Documents
Document Sharing Hub
© ISO/IEC 2019



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

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

Nokia Corporation
PO Box 86
FIN-24101 Salo
Finland

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

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

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.

Note 2 to entry: The AVC base layer is represented by AVC VCL NAL units and associated non-VCL NAL units.

Note 3 to entry: The AVC base layer itself can be a temporal scalable bitstream.

3.1.4

AVC NAL unit

AVC VCL NAL unit (3.1.5) and its associated non-VCL NAL units in a bitstream

3.1.5

AVC VCL NAL unit

NAL unit with type 1 to 5 (inclusive) Note 1 to entry: As specified in ISO/IEC 14496-10.

3.1.6

complete subset

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

3.1.7

cropped frame dimensions

width and height of the decoded frame after applying the output cropping parameters specified by the active SPS

3.1.8

extraction path

set of operations on the original bitstream, each yielding a subset bitstream, ordered such that the complete bitstream is first in the set, and the base layer is last, and all the bitstreams are in decreasing complexity (along one of the scalability axes, such as resolution), and where every bitstream is a valid operating point

Note 1 to entry: An extraction path can be represented by the values of *priority_id* in the NAL unit headers. Alternatively, an extraction path can be represented by the run of tiers or by a set of hierarchically dependent tracks.

3.1.9

extractor

in-stream structure (3.1.11) 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.

3.1.10

implicit reconstruction

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

3.1.11

in-stream structure

structure residing within sample data

3.1.12

layer set

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

Note 1 to entry: As specified in ISO/IEC 23008-2.

3.1.13

MVC NAL unit

MVC VCL NAL unit (3.1.14) and its associated non-VCL NAL units in an MVC stream Note 1 to entry: As specified in ISO/IEC 14496-10:2014, Annex H.

3.1.14**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: As specified in ISO/IEC 14496-10.

Note 2 to entry: MVC VCL NAL units do not affect the decoding process of a legacy AVC decoder.

3.1.15**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 Note 1 to entry: As specified in ISO/IEC 14496-10:2014 Annex I.

3.1.16**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

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

3.1.17**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

Note 1 to entry: As specified in ISO/IEC 14496-10 or ISO/IEC 23008-2.

3.1.18**natively present**

not included in an *aggregator* ([3.1.2](#)) or an *extractor* ([3.1.9](#))

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.

3.1.19**operating point**

<AVC and SVC elementary stream> 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 ISO/IEC 14496-10:2014, Annex H or an output operation point in ISO/IEC 23008-2.

3.1.20

operating point

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

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 ISO/IEC 14496-10:2014, Annex H or an output operation point in ISO/IEC 23008-2.

3.1.21

output layer set

set of layers consisting of the layers of one of the specified *layer sets* ([3.1.12](#)), where one or more layers in the set of layers are indicated to be output layers

Note 1 to entry: As specified in ISO/IEC 23008-2.

3.1.22

parameter set

video parameter set, sequence parameter set or picture parameter set

Note 1 to entry: As defined in the applicable video standard (e.g. ISO/IEC 14496-10 or ISO/IEC 23008-2).

Note 2 to entry: This term is used to refer to all types of parameter sets.

3.1.23

parameter set elementary stream

elementary stream containing samples made up of only sequence and picture parameter set NAL units synchronized with the *video elementary stream* ([3.1.40](#))

3.1.24

picture unit

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

Note 1 to entry: As specified in ISO/IEC 23008-2.

3.1.25

prefix NAL unit

NAL units with type 14

Note 1 to entry: As specified in ISO/IEC 14496-10.

Note 2 to entry: Prefix NAL units provide scalability information about AVC VCL NAL units and filler data NAL units. 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.

3.1.26

reference layer

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

Note 1 to entry: As specified in ISO/IEC 23008-2 and as specified by the 'oinf' sample group defined in subclause [9.6.2](#).

3.1.27 scalable layer layer

<AVC and SVC elementary stream> 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: As specified in ISO/IEC 14496-10.

Note 2 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 3 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).

3.1.28 scalable layer layer

<HEVC, layered HEVC and tiled HEVC elementary stream> set of VCL NAL units with the same value of nuh_layer_id and the associated non-VCL NAL units

Note 1 to entry: As specified in ISO/IEC 23008-2.

3.1.29 scalable layer representation

bitstream subset that is required for decoding the *scalable layer* (3.1.27/3.1.28), consisting of the scalable layer itself and all the scalable layers on which the scalable layer depends

Note 1 to entry: A scalable layer representation is also referred to as the representation of the scalable layer.

3.1.30 sub-picture

proper subset of coded slices of a layer representation

3.1.31 sub-picture tier

tier that consists of *sub-pictures* ([3.1.30](#))

Note 1 to entry: Any coded slice that is not included in the tier representation of a sub-picture tier is not to be referred to in inter prediction or inter-layer prediction for decoding of the sub-picture tier.

3.1.32 sub-layer

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

Note 1 to entry: As specified in ISO/IEC 23008-2.

3.1.33 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).

3.1.34 SVC NAL unit

SVC VCL NAL unit ([3.1.36](#)) and its associated non-VCL NAL units in an *SVC stream* ([3.1.35](#))

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

3.1.35

SVC stream

bitstream represented by the operating point for which `dependency_id` is equal to `mDid`, `temporal_id` is the greatest `temporal_id` value among `mOpSet`, and `quality_id` is the greatest `quality_id` value among `mOpSet`, where the greatest value of `dependency_id` of all the operating points represented by DTQ (`dependency_id`, `temporal_id` and `quality_id`) combinations is equal to `mDid`, and the set of all the operating points with `dependency_id` equal to `mDid` is `mOpSet`

Note 1 to entry: The term “SVC stream” is referenced by ‘decoding/accessing the entire stream’ in this document. There can be NAL units that are not required for decoding this operating point.

3.1.36

SVC 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: As specified in ISO/IEC 14496-10:2014, Annex G.

Note 2 to entry: SVC VCL NAL units do not affect the decoding process of a legacy AVC decoder.

3.1.37

temporal layer representation

representation of a temporal layer

temporal layer and all lower temporal layers

3.1.38

tier

set of operating points within a track, providing information about the operating points and instructions on how to access the corresponding bitstream portions (using maps and groups)

Note 1 to entry: In SVC file format a tier represents one or more scalable layers of an SVC bitstream. In the context of ISO/IEC 23008-2 video, the term tier is used to represent a part of the interoperability point representation consisting of profile, tier, and level. Readers should not be confused about these two different meanings of the word “tier”.

Note 2 to entry: The term “tier” is used in SVC file format to avoid confusion with the frequently used term layer. A tier represents a subset of a track and represents an operating point of an SVC bitstream. Tiers in a track subset the entire track, no matter whether the track references another track by extractors.

Note 3 to entry: An MVC or MVD tier represents a particular set of temporal subsets of a particular set of views.

3.1.39

tier representation

representation of the tier

bitstream subset that is required for decoding the tier, consisting of the tier itself and all the tiers on which the tier depends

3.1.40

video elementary stream

elementary stream containing access units made up of NAL units for coded picture data

3.1.41

video stream

self-contained independently decodable video bitstream

3.1.42

virtual base view

AVC compatible representation of an independently coded non-base view

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

Note 2 to entry: The virtual base view of an independently coded non-base view is created according to the process specified in ISO/IEC 14496-10:2014, H.8.5.5. Samples containing data units of an independently coded non-base view and samples of the virtual base view are aligned by decoding times.

3.2 Abbreviated terms

3D-AVC	three-dimensional advanced video coding (refers to ISO/IEC 14496-10:2014, Annex J)
3D-HEVC	three-dimensional high efficiency video coding (refers to ISO/IEC 23008-2:2017, Annex I)
A3D	three-dimensional advanced video coding (refers to ISO/IEC 14496-10:2014, Annex G)
NOTE 1 The abbreviation A3D is used in terminology related to syntax elements and structures, whereas the abbreviation 3D-AVC is used otherwise.	
AVC	advanced video coding [where contrasted with SVC, MVC, or MVD in this document, this term refers to ISO/IEC 14496-10, excluding its Annex G (Scalable Video Coding), Annex H (Multiview Video Coding), Annex I (Multiview and Depth Video Coding), and Annex J (Multiview and Depth Video with Enhanced Non-Base View Coding)]
BLA	broken link access
CRA	clean random access
CTU	coding tree unit
HEVC	high efficiency video coding
FF	file format
HRD	hypothetical reference decoder
IDR	instantaneous decoding refresh
IRAP	intra random access point
L-HEVC	layered high efficiency video coding
MVC	multiview video coding (refers to ISO/IEC 14496-10:2014, Annex H)
MVCD	multiview video coding plus depth (refers to ISO/IEC 14496-10:2014, Annex I)
MVC+D	multiview video coding plus depth (refers to ISO/IEC 14496-10:2014, Annex I)
NOTE 2 The abbreviation MVCD is used in terminology related to syntax elements and structures, whereas the abbreviation MVC+D is used otherwise.	
MV-HEVC	multiview high efficiency video coding (refers to ISO/IEC 23008-2:2017, Annex G)
MVD	multiview video coding plus depth (refers to ISO/IEC 14496-10:2014, Annexes I and J)

NAL	network abstraction layer
PPS	picture parameter set
RBSP	raw byte sequence payload
ROI	region-of-interest
SAP	stream access point
SEI	supplementary enhancement information
SHVC	scalable high efficiency video coding (refers to ISO/IEC 23008-2:2017, Annex H)
SPS	sequence parameter set
STSA	step-wise temporal sub-layer access
SVC	scalable video coding (refers to ISO/IEC 14496-10:2014, Annex G)
TSA	temporal sub-layer access
VCL	video coding layer
VPS	video parameter set

4 General definitions

4.1 Overview

The specifications in this clause apply to all coding systems identified by this document, unless specifically over-ridden by definitions in the clause for a specific coding system.

[Table 1](#) summarizes the correspondences between the sets of terminology used in video specifications and the ISO base media file format for the AVC file format specified in subclause [4.11](#) and the HEVC file format specified in [Clause 8](#) (it does not apply for all cases for [Clauses 6, 7 and 9](#)).

Table 1 — Correspondence of terms in video and ISO base media file format

Video	ISO base media file format
—	movie
bitstream	track
access unit	sample

4.2 Elementary stream structure

This document concerns video coding systems that specify a set of network abstraction layer (NAL) units, which contain different types of data. This subclause specifies the format of the elementary streams for storing such content.

4.3 Sample and configuration definition

4.3.1 General

Sample: A sample is an access unit or a part of an access unit (e.g., in a track containing a part of a multi-layer video bitstream), where an access unit is as defined in the appropriate specification.

Parameter set sample: A parameter set sample is a sample in a parameter set stream that shall consist 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.

4.3.2 Canonical order and restrictions

The elementary stream is stored in the ISO base media file format in a *canonical* format. The canonical format is as *neutral* as possible so that systems that need to customize the stream for delivery over different transport protocols — MPEG-2 Systems, RTP, and so on — should not have to *remove* information from the stream while being free to *add* to the stream. Furthermore, a canonical format allows such operations to be performed against a known initial state.

When multiple tracks are used to store an elementary stream, as may be the case for [Clauses 6, 7 and 9](#), some tracks may contain canonical streams while others may need to be processed (e.g., when extractors are used or when an implicit reconstruction of access units is needed) to produce a canonical stream.

The canonical stream format is an elementary stream that satisfies the following conditions:

- **Video data NAL units:** In the context of subclause [4.11](#) or [Clause 8](#), all video data NAL units for a single picture shall be contained within the sample whose decoding time and composition time are those of the picture.
- **SEI NAL units:** All SEI NAL units shall be contained in the parameter set arrays, or in the sample whose decoding time is at the time, or immediately precedes the time (with no intervening samples), when the SEI messages come into effect instantaneously. In general, SEI messages for a picture shall be included in the sample containing that picture and that SEI messages pertaining to a sequence of pictures shall be included in the sample containing the first picture of the sequence to which the SEI message pertains. The order of SEI messages within a sample is as defined in the applicable video coding standard.
- **NAL unit order:** The sequence of NAL units in an elementary stream and within a single sample must be in a valid decoding order for those NAL units as specified in the applicable video coding standard.
- **All timing information is external to stream:** Picture Timing SEI messages that define presentation or composition timestamps may be included in the video elementary stream, as these messages contain other information than timing, and may be required for conformance checking. However, all timing information is provided by the information stored in the various sample metadata tables, and this information overrides any timing provided in the video layer. Timing provided within the video stream in this file format should be ignored as it may contradict the timing provided by the file format and may not be correct or consistent within itself.

NOTE This constraint is imposed due to the fact that post-compression editing, combination, or re-timing of a stream at the file format level can invalidate or make inconsistent any embedded timing information present within the video stream.

- **No start codes:** The elementary streams shall not include start codes. As stored, each NAL unit is preceded by a length field as specified in subclause [4.3.3](#); this enables easy scanning of the sample's NAL units. Systems that wish to deliver, from this file format, a stream using start codes will need to reformat the stream to insert those start codes.

4.3.3 Sample format

4.3.3.1 Definition

This subclause defines the structure of the samples. Samples are externally framed and have a size supplied by that external framing. The syntax of a sample is configured via the decoder specific configuration for the elementary stream. An example of the structure of a video sample is depicted in the following figure.

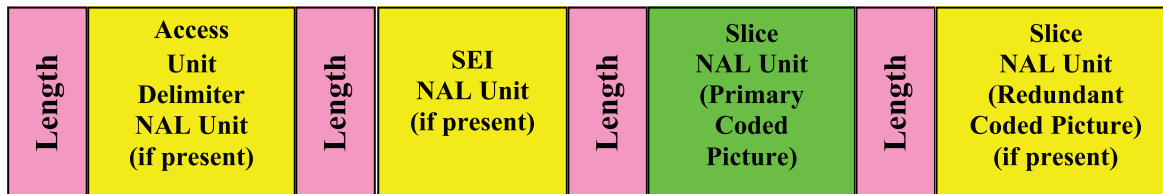


Figure 1 — Example structure of a sample

An access unit is made up of a set of NAL units. Each NAL unit is represented with a:

- *Length*: Indicates the length in bytes of the following NAL unit. The length field may be configured to be of 1, 2, or 4 bytes.
- *NAL Unit*: Contains the NAL unit data as specified in the applicable video coding standard.

4.3.3.2 Syntax

```
aligned(8) class NALUSample
{
    unsigned int PictureLength = sample_size; //Size of Sample from SampleSizeBox
    for (i=0; i<PictureLength; )           // to end of the picture
    {
        unsigned int((DecoderConfigurationRecord.LengthSizeMinusOne+1)*8)
            NALUnitLength;
        bit(NALUnitLength * 8) NALUnit;
        i += (DecoderConfigurationRecord.LengthSizeMinusOne+1) + NALUnitLength;
    }
}
```

4.3.3.3 Semantics

`DecoderConfigurationRecord` indicates the record in the matching sample entry (e.g. `AVCDecoderConfigurationRecord` in the case of AVC).

`NALUnitLength` indicates the size of a NAL unit measured in bytes. The length field includes the size of both the NAL header and the NAL unit payload but does not include the length field itself.

`NALUnit` contains a single NAL unit. The syntax of a NAL unit is defined in the appropriate specification (e.g. ISO/IEC 14496-10) and includes both the one byte NAL header and the variable length encapsulated byte stream payload.

4.3.4 Optional boxes in the sample entry

Unless otherwise explicitly specified, the optional boxes, including the `BitRateBox`, in the sample entry document the information for the output bitstream by the file parser reconstructed from this track and all the tracks this track depends on for decoding.

4.4 Video track structure

In the terminology of ISO/IEC 14496-12, both video and parameter set tracks are video or visual tracks. They therefore use:

- a) a handler_type of 'vide' in the `HandlerBox`;
- b) a video media header 'vmhd';
- c) and, as defined below, a derivative of the `VisualSampleEntry`.

A video stream is represented by one or more video tracks in a file.

If there is more than one track representing scalable aspects of a single stream, then they form alternatives to each other, and the field 'alternate_group' should be used, or the composition system used should select one of them, as appropriate. See ISO/IEC 14496-12:2015, 8.10.3 for informative labelling of why tracks are members of alternate groups.

4.5 Template fields used

The ISO base media file format defines a number of fields that have default values but that may be defined for use by specific sub-systems. Tracks containing video data may use the following template fields:

- a) `alternate_group` in the `TrackHeaderBox` (see subclause 5.4.8 on stream switching).
- b) template field 'depth' in the `VisualSampleEntry` to document the presence of alpha.

`depth` takes one of the following values:

- 0x18 – the video sequence is in colour with no alpha;
- 0x28 – the video sequence is in grayscale with no alpha;
- 0x20 – the video sequence has alpha (gray or colour).

4.6 Visual width and height

The width and height fields in a `VisualSampleEntry` shall correctly document the maximum cropped frame dimensions in any sample of the video stream that is described by that entry.

The width and height fields do *not* reflect any changes in size caused by SEI messages such as pan-scan. The visual handling of SEI messages such as pan-scan is both optional and terminal-dependent.

Note that the visual size in the SPS may be either frame or field size; in the sample entry, it is always the frame size.

The width and height fields in the track header may not be the same as the width and height fields in the one or more `VisualSampleEntry` in the video track. As specified in the ISO base media file format, if normalized visual presentation is needed, all the sequences are normalized to the track width and height for presentation.

4.7 Decoding time (DTS) and composition time (CTS)

Samples are stored in the file format in decoding order. If picture reordering is not used and decoding and composition times are the same, then presentation is the same as decoding order and only the time-to-sample 'stts' table is used. Note that any kind of picture may be reordered, not only B-pictures.

If decoding time and composition time differ, the composition time-to-sample 'ctts' table is also used in conjunction with the 'stts' table.

4.8 Sample groups on random access recovery points 'roll' and random access points 'rap'

This document disallows the presence of `SampleToGroupBox` with `grouping_type` equal to 'roll' or 'rap' and with `version` greater than 0 in tracks with sample entries specified in this document other than [Clause 7](#). Parsers according to this document shall ignore `SampleToGroupBox` with `grouping_type` equal to 'roll' or 'rap' and with `version` greater than 0 in tracks with sample entries specified in this document other than [Clause 7](#).

The video coding system can include the concept of a 'gradual decoding refresh' or random access recovery point. This may be signalled in the bitstream using a mechanism such as the recovery point SEI message. This message is found at the beginning of the random access and indicates how much data must be decoded subsequent to the access unit at the position of the SEI message before the recovery is complete.

When all access units in output order starting from the access unit at the position of the SEI message can be successfully decoded after random access, i.e. when the `recovery_frame_cnt` syntax element of the recovery point SEI message is 0, the random access point ('rap') sample grouping should be used.

This concept of gradual recovery is supported in the file format also by using `RollRecoveryEntry` groups. In order that the group membership marks the sample containing the SEI message the 'roll-distance' is constrained to being only positive (i.e. a post-roll). In other words, `RollRecoveryEntry` groups can be used when the value of the `recovery_frame_cnt` syntax element of the recovery point SEI message is greater than 0.

NOTE The roll-group counts samples in the file format; this does not necessarily match the way that the distances are represented in the SEI message.

Within a stream, it is necessary to mark the beginning of the pre-roll, so that a stream decoder may start decoding there. However, in a file, when performing random access, a deterministic search is desired for the closest preceding frame that can be decoded perfectly (either a sync sample, or the end of a pre-roll).

4.9 Hinting

Note that what the hint tracks call "B frames" are actually 'disposable' pictures or non-reference pictures, for example as defined in ISO/IEC 14496-10.

Care should be taken when the structures in [Annex A](#) (aggregators or extractors) are in use and the track is hinted. These structures are defined only for use in the file format and should not be transmitted. In particular, a hint track that points at an extractor in a video track would cause the extractor itself to be transmitted (which is probably both incorrect and not the desired behaviour), not the data the extractor references. Hint tracks should normally directly reference NAL units specified in the applicable video coding standard.

4.10 On change of sample entry

In this clause, resetting decoding can be regarded as a similar process as starting the decoding from the beginning of a video stream.

A change in the sample entry is caused, for example, by a change of parameter set(s) reusing the same parameter set ID(s) in a `VisualSampleEntry`.

For the single-layer video file formats specified in subclause 4.11 and Clause 8, the video decoder for decoding of the bitstream output from the file parser is expected to be reset at the first sample at which the sample entry changes. When more than one sample entry is used in a track, random accessing is enabled if a video stream is encoded and encapsulated in the track as follows:

- The first sample that a sample entry is associated with (i.e., applies to) is a sync sample.
- Decoding of samples that a sample entry is associated with does not rely on data from any sample not associated with the sample entry or data from any other sample entry.

For the multi-layer video file formats specified in Clauses 6 and 7, the video decoder for decoding of the bitstream output from the file parser is expected to be reset at the first sample in the base track or scalable base track (as defined in Clauses 6 and 7) at which the sample entry changes. When more than one sample entry is used in the base track or scalable base track, random accessing is enabled if a video stream is encoded and encapsulated in the track as follows:

- The first sample that a sample entry in the base track or scalable base track is associated with is a sync sample.
- Decoding of the set of access units from the first access unit containing the first sample in the base track associated with a sample entry to the last access unit containing the last sample in the base track associated with the sample entry, inclusive, does not rely on data from any access unit earlier than the first access unit in decoding order or data from any sample entry associated with a sample contained in an access unit earlier than the first access unit in decoding order.

NOTE When parameter sets are stored other than in the sample entry, i.e. either in a separate parameter set track, or in the samples in the video track, it is possible that there are two parameter sets with the same identifier, with the semantics (defined in the applicable video coding standard) that the second occurrence replaces the first from that time onwards in the stream. Converting such a stream to a form in which all parameter sets are in a single sample entry involves more than moving the applicable NAL units into the sample entry; one of the two occurrences will need a new identifier value, and the video NAL units that refer to it by identifier will also need updating. Alternatively, multiple sample entries can be used for the storage of parameter sets with same ID, in which case sample to sample entry association needs rewriting.

In the following, let the L-HEVC base track be the track referred to by a track reference of type 'oref' (when present) or the present track (when it contains an 'oinf' sample group). For the multi-layer video file formats specified in Clause 9, the video decoder for decoding of the bitstream output from the file parser is expected to be reset at the first sample in the L-HEVC base track at which the sample entry or the 'oinf' sample group description entry changes. When more than one sample entry or 'oinf' sample group description entry is used in the L-HEVC base track, random accessing is enabled if a video stream is encoded and encapsulated in the track as follows:

- The first sample that a sample entry or an 'oinf' sample group description entry in the L-HEVC base track is associated with has an IRAP picture (as specified in ISO/IEC 23008-2) at the lowest layer carried in the track.
- Decoding of the set of access units from the first access unit containing the first sample in the L-HEVC base track associated with a sample entry or an 'oinf' sample group description entry to the last access unit containing the last sample in the L-HEVC base track associated with the sample entry or 'oinf' sample group description entry, inclusive, does not rely on data from any access unit earlier than the first access unit in decoding order or data from any sample entry or 'oinf' sample group description entry associated with a sample contained in an access unit earlier than the first access unit in decoding order.

4.11 SEI information box

4.11.1 Definition

Box Type: 'seii'

Container: Scheme Information box ('schi') or VisualSampleEntry

Mandatory: Yes (in the SchemeInformationBox), No (in a VisualSampleEntry)

Quantity: One (in the SchemeInformationBox), Zero or one (in a VisualSampleEntry)

The SEI Information box documents the SEIs in a stream. When contained in a VisualSampleEntry numRequiredSEIs shall be 0. By inspecting the SEI Information box a player will know which SEI messages it can assume to be present, and which are deemed necessary by the file author for correct playback. There might be other SEIs present in the bitstream that are not documented by this box.

4.11.2 Syntax

```
aligned(8) class SeiInformationBox extends Box('seii') {  
  
    unsigned int(16) numRequiredSEIs;  
  
    for (i = 0; i < numRequiredSEIs; i++) {  
  
        unsigned int(16) requiredSEI_ID;  
  
    }  
  
    unsigned int(16) numNotRequiredSEIs;  
  
    for (i = 0; i < numNotRequiredSEIs; i++) {  
  
        unsigned int(16) notrequiredSEI_ID;  
  
    }  
  
}
```

4.11.3 Semantics

requiredSEI_ID takes on the value “payloadType” of an SEI message present in the bitstream that is deemed necessary by the file author for correct playback.

notrequiredSEI_ID takes on the value “payloadType” of an SEI message present in the bitstream that is not deemed necessary by the file author for correct playback.

4.12 Post-decoder requirements scheme for signalling of SEI

4.12.1 General

In order to handle situations where the file author requires certain actions on the player or renderer, the ISO base media file format specifies the restricted-video mechanism where sample entries are hidden behind the generic sample entry 'resv'. The mechanism applies to all coding systems identified by chapters in this document. For the case of signalling of SEI, a file author can list occurring SEI message IDs (ISO/IEC 14496-10, ISO/IEC 23008-2) and classify them into two categories: those that are deemed required by the file author for correct playback, and others. The occurrence of either type of SEI messages can be signalled in the SEI Information box.

4.12.2 Definition

The scheme for signalling of SEI is defined here.

The SchemeType 'aSEI' is used.

The SEI information box is mandatory in the SchemeInformationBox under the 'aSEI' scheme. In this case, it contains information about the SEI messages present in the bitstream. Although the SEI messages are not required for decoding, the file author may require certain actions for rendering or other purposes. The box distinguishes between an SEI that is required to be understood for correct playback and an SEI that is not required for correct playback (but may enhance playback).

The SEI messages listed here should be stored either in the bitstream or in the Configuration Record. The SEI Information box does not contain the actual SEI messages, it only lists those that occur in the bitstream.

5 AVC elementary streams and sample definitions

5.1 Overview

ISO/IEC 14496-10 offers not only increased coding efficiency and enhanced robustness, but also many features for the systems that use it. To enable the best visibility of, and access to, those features, and to enhance the opportunities for the interchange and interoperability of media, subclause 4.11 defines a storage format for video streams compressed using AVC.

This clause defines the storage for plain AVC streams, where 'plain AVC' refers to ISO/IEC 14496-10, excluding any multi-layer extension of ISO/IEC 14496-10 such as ISO/IEC 14496-10:2014, Annexes G, H, I and J.

This clause specifies the elementary stream and sample structure used to store AVC visual content.

The storage of AVC content uses the existing capabilities of the ISO base media file format but also defines extensions to support the following features of the AVC codec.

d) Switching pictures:

to enable switching between different coded streams and substitution of pictures within the same stream.

e) Sub-sequences and layers:

provides a structuring of the dependencies of a group of pictures to provide for a flexible stream structure (e.g. in terms of temporal scalability and layering).

f) Parameter sets:

the sequence and picture parameter set mechanism decouples the transmission of infrequently changing information from the transmission of coded macroblock data. Each slice containing the coded macroblock data references the picture parameter set containing its decoding parameters. In turn, the picture parameter set references a sequence parameter set that contains sequence level decoding parameter information.

5.2 Elementary stream structure

Two types of elementary streams are defined for storing AVC content (see also [Figure 2](#)):

- **A video elementary stream** contains all video coding related NAL units (i.e. those NAL units containing video data or signalling video structure) and may contain non-video coding related NAL units such as SEI messages and access unit delimiter NAL units. Aggregators and Extractors, when present, shall not be directly output by file parsers.

- **Parameter set elementary streams** shall not contain video coding related NAL units (i.e. those NAL units containing video data or signalling video structure), and would normally contain only sequence parameter sets, picture parameter sets and sequence parameter set extension NAL units.

Using these stream types, AVC content shall be stored in one of these configurations:

- **Video elementary stream with no parameter sets:** In this case, sequence and picture parameter set NAL units shall be stored in the sample entries of this track. Sequence and picture parameter set NAL units shall not be part of AVC samples within the stream itself.
- **Video elementary stream possibly including parameter sets:** In this case, the sample entry indicates whether the stream may contain parameter sets of given types, in addition to other parameters provided in the sample entry. Sequence and picture parameter set NAL units may therefore be part of AVC samples within the stream itself.
- **Video elementary stream and parameter set elementary stream:** In this case, sequence and picture parameter set NAL units shall be transmitted only in the parameter set elementary stream and shall neither be present in the sample entries nor the AVC samples of the video elementary stream.

The types of NAL units that are allowed in each of the video and parameter set elementary streams are specified in [Table 2](#).

Table 2 — NAL unit types in elementary streams

Value of nal_unit_type	Description	Video elementary stream (sample entry 'avc1' or 'avc2')	Video elementary stream (sample entry 'avc3' or 'avc4')	Parameter set elementary stream
0	Unspecified	See Annex F	See Annex F	See Annex F
1	Coded slice of a non-IDR picture slice_layer_without_partitioning_rbsp()	Yes	Yes	No
2	Coded slice data partition A slice_data_partition_a_layer_rbsp()	Yes	Yes	No
3	Coded slice data partition B slice_data_partition_b_layer_rbsp()	Yes	Yes	No
4	Coded slice data partition C slice_data_partition_c_layer_rbsp()	Yes	Yes	No
5	Coded slice of an IDR picture slice_layer_without_partitioning_rbsp()	Yes	Yes	No
6	Supplemental enhancement information(SEI) sei_rbsp()	Yes Except for the sub-sequence, layering or filler SEI messages	Yes Except for the sub-sequence, or layering SEI messages	Only 'declarative' SEIs should be present
7	Sequence parameter set (SPS) seq_parameter_set_rbsp()	No If parameter set elementary stream is not used, SPS shall be stored in the decoder specific information	Yes Parameter set elementary stream shall not be used	Yes

Table 2 (continued)

Value of nal_unit_type	Description	Video elementary stream (sample entry 'avc1' or 'avc2')	Video elementary stream (sample entry 'avc3' or 'avc4')	Parameter set elementary stream
8	Picture parameter set (PPS) pic_parameter_set_rbsp()	No If parameter set elementary stream is not used, PPS shall be stored in the decoder specific information	Yes Parameter set elementary stream shall not be used	Yes
9	Access unit delimiter (AU Delimiter) access_unit_delimiter_rbsp()	Yes	Yes	No
10	End of sequence end_of_seq_rbsp()	Yes	Yes	No
11	End of stream end_of_stream_rbsp()	Yes	Yes	No
12	Filler data (FD) filler_data_rbsp()	No	Yes	No
13	Sequence parameter set extension seq_parameter_set_extension_rbsp()	No If parameter set elementary stream is not used, sequence parameter set extension shall be stored in the decoder specific information	Yes Parameter set elementary stream shall not be used	Yes
14...18	Reserved	Not specified by this document	Not specified by this document	Not specified by this document
19	Coded slice of an auxiliary coded picture without partitioning slice_layer_without_partitioning_rbsp()	Yes	Yes	No
20...23	Reserved	Not specified by this document	Not specified by this document	Not specified by this document
24 – 31	Unspecified	See Annex F	See Annex F	See Annex F

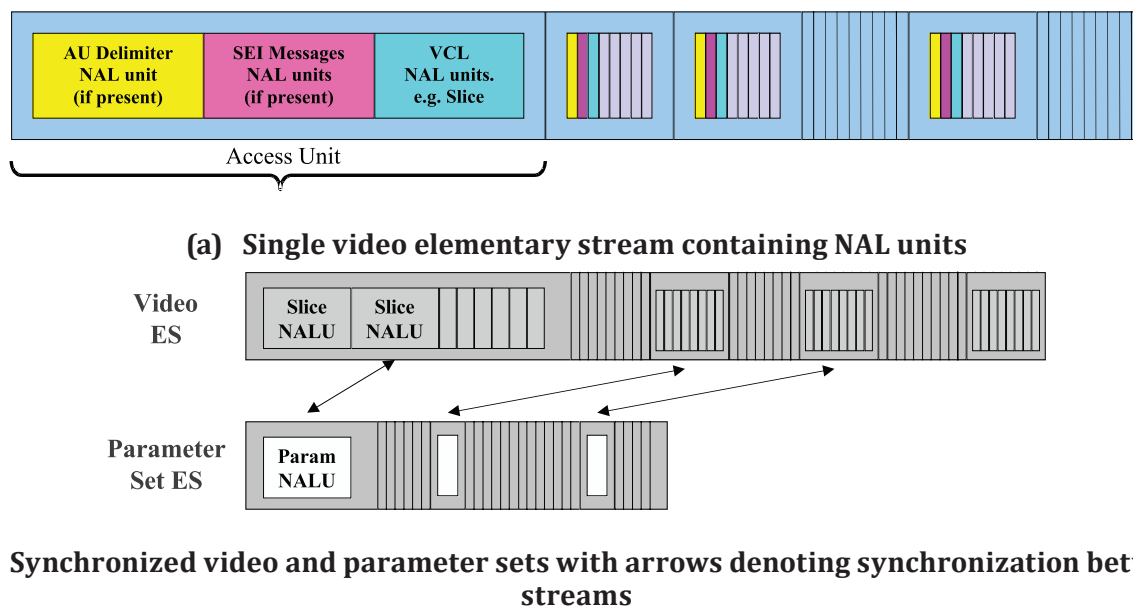


Figure 2 — AVC elementary stream structure

5.3 Sample and configuration definition

5.3.1 Overview

AVC sample: An AVC sample is an access unit as defined in ISO/IEC 14496-10:2014, 7.4.1.2.

AVC parameter set sample: An AVC parameter set sample is a sample in a parameter set stream that shall consist 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.

5.3.2 Canonical order and restrictions

The canonical stream format is an AVC elementary stream that satisfies the following conditions in addition to the general conditions in 4.3.2:

- **Video data NAL units** (Coded Slice, Coded Slice Data Partition A, Coded Slice Data Partition B, Coded Slice Data Partition C, Coded Slice IDR Pictures): All slice and data partition NAL units for a single picture shall be contained within the sample whose decoding time and composition time are those of the picture. Each sample shall contain at least one VCL NAL unit of the primary coded picture.
- **Parameter sets:** If a parameter set elementary stream is used, then the sample in the parameter set stream shall have a decoding time equal or prior to when the parameter set(s) comes into effect instantaneously. This means that for a parameter set to be used in a picture it must be sent prior to the sample containing that picture or in the sample for that picture.

NOTE 1 When the sample entry name is 'avc1' or 'avc2', parameter sets are stored either in the sample entries of the video stream or in the parameter set stream, but never in both. This ensures that it is not necessary to examine every part of the video elementary stream to find relevant parameter sets. It also avoids dependencies of indefinite duration between the sample that contains the parameter set definition and the samples that use it. Storing parameter sets in the sample entries of a video stream provides a simple and static way to supply parameter sets. Parameter set elementary streams on the other hand are more complex but allow for more dynamism in the case of updates. Parameter sets can be inserted into the video elementary stream when the file is streamed over a transport that permits such parameter set updates. When the sample entry name is 'avc3' or 'avc4', then parameter sets can be present both the sample entries and as part of the samples.

- **Parameter set track:** A sync sample in a parameter set track indicates that all parameter sets needed from that (decoding) time forward in the video elementary stream are in that or succeeding parameter stream samples. Also, there shall be a parameter set sample at each point a parameter set is updated. Each parameter set sample shall contain exactly the sequence and picture parameter sets needed to decode the relevant section of the video elementary stream.

NOTE 2 The use of a parameter set track in the file format does not require that a system delivering video content use a separate elementary stream for parameter sets. Instead, implementations can choose to map parameter sets to in-band parameter set NAL units in the video elementary stream or use some out-of-band delivery mechanism defined by the transport layer.

- **SEI message NAL units:** The order of SEI messages within a sample is as defined in ISO/IEC 14496-10:2014, 7.4.1.2.
- **Access unit delimiter NAL units:** The constraints obeyed by access unit delimiter NAL units are defined in ISO/IEC 14496-10:2014, 7.4.1.2.3.
- **Sub-sequence and layering SEI messages:** Sub-sequence or layering SEI messages should not occur in the AVC elementary stream. Specifically, the sub-sequence information, sub-sequence layer characteristics, and sub-sequence characteristics SEI messages should not occur in the stored AVC video elementary stream. Instead, all such information is stored as external metadata as described in subclause [5.4.7](#).
- **Redundant picture:** NAL units within a single access unit shall be ordered in non-decreasing order of redundant picture count (`redundant_pic_cnt`).
- **Slice groups:** NAL units within a primary coded picture or a redundant coded picture shall be ordered in non-decreasing order of slice group identifier. Within the same slice group, slices shall be ordered by their first Macroblock location (`first_mb_in_slice` in the slice header).

NOTE 3 Slice groups are stored in a canonical order to ease hinting, and to make it easier to find a primary coded picture within a sample.

- **Filler data:** Video data is naturally represented as variable bit rate in the file format and should be filled for transmission if needed. Filler Data NAL units and Filler Data SEI messages shall not be present in the file format stored stream when the sample entry does not also permit parameter sets.

NOTE 4 The removal or addition of Filler Data NAL units, start codes, SEI messages or Filler Data SEI messages can change the bitstream characteristics with respect to conformance with the HRD when operating the HRD in CBR mode as specified in ISO/IEC 14496-10:2014, Annex C.

5.3.3 Decoder configuration information

5.3.3.1 AVC decoder configuration record

5.3.3.1.1 Definition

This record contains the size of the length field used in each sample to indicate the length of its contained NAL units as well as the initial parameter sets. This record is externally framed (its size is supplied by the structure that contains it).

This record contains a version field. This document defines version 1 of this record. Incompatible changes to the record will be indicated by a change of version number. Readers shall not attempt to decode this record or the streams to which it applies if the version number is unrecognized.

Compatible extensions to this record will extend it and will not change the configuration version code. Readers should be prepared to ignore unrecognized data beyond the definition of the data they understand (e.g. after the parameter sets in this document).

When used to provide the configuration of

- a parameter set elementary stream,
- a video elementary stream used in conjunction with a parameter set elementary stream,

the configuration record shall contain no sequence or picture parameter sets (`numOfSequenceParameterSets` and `numOfPictureParameterSets` shall both have the value 0).

When used to provide the configuration of a video elementary stream used without a parameter set elementary stream, the configuration record may or may not contain sequence or picture parameter sets (`numOfSequenceParameterSets` or `numOfPictureParameterSets` may or may not have the value 0).

The values for `AVCProfileIndication`, `AVCLevelIndication`, and the flags that indicate profile compatibility shall be valid for all parameter sets of the stream described by this record. The level indication shall indicate a level of capability equal to or greater than the highest level indicated in the included parameter sets; each profile compatibility flag may only be set if all the included parameter sets set that flag. The profile indication shall indicate a profile to which the entire stream associated with this configuration record conforms. If the sequence parameter sets are marked with different profiles, and the relevant profile compatibility flags are all zero, then the stream may need examination to determine which profile, if any, the entire stream conforms to. If the entire stream is not examined, or the examination reveals that there is no profile to which the entire stream conforms, then the stream shall be split into two or more sub-streams with separate configuration records in which these rules are met.

Explicit indication can be provided in the AVC Decoder Configuration Record about the chroma format and bit depth used by the AVC video elementary stream. The parameter '`chroma_format_idc`' present in the sequence parameter set in AVC specifies the chroma sampling relative to the luma sampling. Similarly, the parameters '`bit_depth_luma_minus8`' and '`bit_depth_chroma_minus8`' in the sequence parameter set specify the bit depth of the samples of the luma and chroma arrays. The values of `chroma_format_idc`, `bit_depth_luma_minus8` and '`bit_depth_chroma_minus8`' shall be identical in all sequence parameter sets in a single AVC configuration record. If two sequences differ in any of these values, two different AVC configuration records will be needed. If the two sequences differ in color space indications in their VUI information, then two different configuration records are also required.

The array of sequence parameter sets, and the array of picture parameter sets, may contain SEI messages of a 'declarative' nature, that is, those that provide information about the stream as a whole. An example of such an SEI is a user-data SEI. Such SEIs may also be placed in a parameter set elementary stream. NAL unit types that are reserved in ISO/IEC 14496-10 and in this document may acquire a definition in future, and readers should ignore NAL units with reserved values of NAL unit type when they are present in these arrays.

NOTE This 'tolerant' behaviour is designed so that errors are not raised, allowing the possibility of backwards-compatible extensions to these arrays in future specifications.

When Sequence Parameter Set Extension NAL units occur in this record with `AVCProfileIndication` equal to 66, 77, or 88, they should be placed in the Sequence Parameter Set Array.

5.3.3.1.2 Syntax

```
aligned(8) class AVCDecoderConfigurationRecord {
    unsigned int(8) configurationVersion = 1;
    unsigned int(8) AVCProfileIndication;
    unsigned int(8) profile_compatibility;
    unsigned int(8) AVCLevelIndication;
    bit(6) reserved = '111111'b;
    unsigned int(2) lengthSizeMinusOne;
    bit(3) reserved = '111'b;
    unsigned int(5) numOfSequenceParameterSets;
    for (i=0; i< numOfSequenceParameterSets; i++) {
        unsigned int(16) sequenceParameterSetLength ;
        bit(8*sequenceParameterSetLength) sequenceParameterSetNALUnit;
    }
    unsigned int(8) numOfPictureParameterSets;
    for (i=0; i< numOfPictureParameterSets; i++) {
        unsigned int(16) pictureParameterSetLength;
        bit(8*pictureParameterSetLength) pictureParameterSetNALUnit;
    }
    if( AVCProfileIndication != 66 && AVCProfileIndication != 77 &&
        AVCProfileIndication != 88 )
    {
        bit(6) reserved = '111111'b;
        unsigned int(2) chroma_format;
        bit(5) reserved = '11111'b;
        unsigned int(3) bit_depth_luma_minus8;
        bit(5) reserved = '11111'b;
        unsigned int(3) bit_depth_chroma_minus8;
        unsigned int(8) numOfSequenceParameterSetExt;
        for (i=0; i< numOfSequenceParameterSetExt; i++) {
            unsigned int(16) sequenceParameterSetExtLength;
            bit(8*sequenceParameterSetExtLength) sequenceParameterSetExtNALUnit;
        }
    }
}
```

5.3.3.1.3 Semantics

`AVCProfileIndication` contains the profile code as defined in ISO/IEC 14496-10.

`profile_compatibility` is a byte defined exactly the same as the byte that occurs between the `profile_IDC` and `level_IDC` in a sequence parameter set (SPS), as defined in ISO/IEC 14496-10.

`AVCLevelIndication` contains the level code as defined in ISO/IEC 14496-10.

`lengthSizeMinusOne` indicates the length in bytes of the `NALUnitLength` field in an AVC video sample or AVC parameter set sample of the associated stream minus one. For example, a size of one byte is

indicated with a value of 0. The value of this field shall be one of 0, 1, or 3 corresponding to a length encoded with 1, 2, or 4 bytes, respectively.

`numOfSequenceParameterSets` indicates the number of SPSs that are used as the initial set of SPSs for decoding the AVC elementary stream.

`sequenceParameterSetLength` indicates the length in bytes of the SPS NAL unit as defined in ISO/IEC 14496-10.

`sequenceParameterSetNALUnit` contains a SPS NAL unit, as specified in ISO/IEC 14496-10. SPSs shall occur in order of ascending parameter set identifier with gaps being allowed.

`numOfPictureParameterSets` indicates the number of picture parameter sets (PPSs) that are used as the initial set of PPSs for decoding the AVC elementary stream.

`pictureParameterSetLength` indicates the length in bytes of the PPS NAL unit as defined in ISO/IEC 14496-10.

`pictureParameterSetNALUnit` contains a PPS NAL unit, as specified in ISO/IEC 14496-10. PPSs shall occur in order of ascending parameter set identifier with gaps being allowed.

`chroma_format` contains the `chroma_format` indicator as defined by the `chroma_format_idc` parameter in ISO/IEC 14496-10.

`bit_depth_luma_minus8` indicates the bit depth of the samples in the Luma arrays. For example, a bit depth of 8 is indicated with a value of zero ($\text{BitDepth} = 8 + \text{bit\_depth\_luma\_minus8}$). The value of this field shall be in the range of 0 to 4, inclusive.

`bit_depth_chroma_minus8` indicates the bit depth of the samples in the Chroma arrays. For example, a bit depth of 8 is indicated with a value of zero ($\text{BitDepth} = 8 + \text{bit\_depth\_chroma\_minus8}$). The value of this field shall be in the range of 0 to 4, inclusive.

`numOfSequenceParameterSetExt` indicates the number of Sequence Parameter Set Extensions that are used for decoding the AVC elementary stream.

`sequenceParameterSetExtLength` indicates the length in bytes of the SPS Extension NAL unit as defined in ISO/IEC 14496-10.

`sequenceParameterSetExtNALUnit` contains a SPS Extension NAL unit, as specified in ISO/IEC 14496-10.

5.4 Derivation from ISO base media file format

5.4.1 AVC file type and identification

Conformance with this document is indicated by the presence of the brand of a specification that permits the inclusion of AVC content, in the compatible brands list of the `FileTypeBox` as defined in ISO/IEC 14496-12. The file extension normally matches the major brand.

AVC content may be used in an MPEG-4 context; in a file with extension “.mp4”, the major brand may be ‘`avc1`’.

Readers conformant to this document should read the file if a suitable brand occurs in the compatible-brands list. Other structures and/or track types, defined in specifications other than that identified by the brand, may be present, and these may be ignored by a reader conformant with the specification identified by the brand.

5.4.2 AVC video stream definition

5.4.2.1 Sample entry name and format

5.4.2.1.1 Definition

Sample Entry and Box Types: 'avc1', 'avc2', 'avc3', 'avc4', 'avcC', 'm4ds', 'btrt'

Container: Sample Description Box ('stsd')

Mandatory: An 'avc1', 'avc2', 'avc3' or 'avc4' sample entry is mandatory

Quantity: One or more sample entries may be present

An AVC visual sample entry shall contain an AVC Configuration Box, as defined below. This includes an `AVCDecoderConfigurationRecord`, as defined in [5.3.3.1](#).

An optional `BitRateBox` may be present in the AVC visual sample entry to signal the bit rate information of the AVC video stream. Extension descriptors that should be inserted into the Elementary Stream Descriptor, when used in MPEG-4, may also be present.

Multiple sample entries may be used, as permitted by ISO/IEC 14496-12, to indicate sections of video that use different configurations or parameter sets.

The sample entry name 'avc1' or 'avc3' may only be used when the stream to which this sample entry applies is a compliant and AVC stream as viewed by an AVC decoder operating under the configuration (including profile and level) given in the `AVCConfigurationBox`. The file format specific structures that resemble NAL units (defined in [Annex A](#)) may be present but shall not be used to access the AVC base data; that is, the AVC data shall not be contained in Aggregators (though they may be included within the bytes referenced by the `additional_bytes` field) nor referenced by Extractors.

The sample entry name 'avc2' or 'avc4' may only be used when Extractors or Aggregators ([Annex A](#)) are required to be supported, and an appropriate Toolset is required (for example, as indicated by the file-type brands). This sample entry type indicates that, in order to form the intended AVC stream, Extractors shall be replaced with the data they are referencing, and Aggregators shall be examined for contained NAL Units. Tier grouping may be present.

5.4.2.1.2 Syntax

```
// Visual Sequences
class AVCCConfigurationBox extends Box('avcC') {
    AVCDecoderConfigurationRecord() AVCCConfig;
}
class MPEG4ExtensionDescriptorsBox extends Box('m4ds') {
    Descriptor Descr[0 .. 255];
}
class AVCSampleEntry() extends VisualSampleEntry (type) {
    // type is 'avc1' or 'avc3'
    AVCCConfigurationBox config;
    MPEG4ExtensionDescriptorsBox (); // optional
}
class AVC2SampleEntry() extends VisualSampleEntry (type) {
    // type is 'avc2' or 'avc4'
    AVCCConfigurationBox avcconfig;
    MPEG4ExtensionDescriptorsBox descr; // optional
}
```

5.4.2.1.3 Semantics

Compressorname in the base class VisualSampleEntry indicates the name of the compressor used with the value "\012AVC Coding" being recommended; the first byte is a count of the remaining bytes, here represented by \012, which (being octal 12) is 10 (decimal), the number of bytes in the rest of the string.

config is defined in subclause 5.3.3. If a separate parameter set stream is used, numOfSequenceParameterSets and numOfPictureParameterSets shall both be zero.

Descr is a descriptor that should be placed in the ElementaryStreamDescriptor when this stream is used in an MPEG-4 systems context. This does not include SLConfigDescriptor or DecoderConfigDescriptor, but includes the other descriptors in order to be placed after the SLConfigDescriptor.

5.4.3 AVC parameter set stream definition

5.4.3.1 Sample entry name and format

5.4.3.1.1 Definition

Sample Entry Type: 'avcp'

Container: Sample Description Box ('std')

Mandatory: Yes

Quantity: One or more sample entries may be present

An AVC parameter stream sample entry shall contain an AVC Parameter Stream Configuration Box, as defined below.

5.4.3.1.2 Syntax

```
class AVCPParameterSampleEntry() extends VisualSampleEntry ('avcp'){
```

```

    AVCConfigurationBox config;
}

```

5.4.3.1.3 Semantics

Compressorname in the base class VisualSampleEntry indicates the name of the compressor used with the value "\016AVC Parameters" being recommended (\016 is 14, the length of the string as a byte).

config is defined in subclause 5.3.3. numOfSequenceParameterSets and numOfPictureParameterSets shall both be zero.

5.4.3.2 Sample format

5.4.3.2.1 Definition

This subclause defines the sample format for AVC Parameter set streams. An AVC parameter set sample contains only one or more sequence, picture parameter set, or sequence parameter set extension NAL units.

5.4.3.2.2 Syntax

```

aligned(8) class AVCPParameterSample
{
    unsigned int PictureLength = sample_size;
    //Size of AVCPParameterSample from SampleSizeBox
    for (i=0; i<PictureLength; ) // to end of the picture
    {
        unsigned int ((AVCDecoderConfigurationRecord.LengthSizeMinusOne+1)*8)
            NALUnitLength;
        bit(NALUnitLength * 8) NALUnit;
        i += (AVCDecoderConfigurationRecord.LengthSizeMinusOne+1) + NALUnitLength;
    }
}

```

5.4.3.2.3 Semantics

NALUnitLength indicates the size of a NAL unit measured in bytes. The length field includes the size of both the one byte NAL header and the EBSP payload but does not include the length field itself.

NALUnit contains a single NAL unit. The syntax of a NAL unit is defined in ISO/IEC 14496-10 and includes both the one byte NAL header and the variable length encapsulated byte stream payload.

5.4.3.3 Track reference

A track reference of type 'avcp' in the video elementary stream track reference table, referencing the parameter set stream, is used to connect from the video elementary stream to the parameter set elementary stream.

5.4.4 Parameter sets

This subclause applies when a separate parameter set stream is not used.

Each AVC sample entry, which contains the AVC video stream decoder specific information, includes a group of SPSs and PPSs. This group of parameter sets functions much like a codebook. Each parameter

set has an identifier, and each slice references the parameter set it was coded against using the parameter set's identifier.

When the sample entry name is 'avc1' or 'avc2', the following applies:

- In the file format each configuration of parameter sets is represented separately. A parameter set cannot be updated without causing a different sample entry to be used. For example, suppose that samples 1 to 4 use PPSs identified as 1, 2, 3 and a single SPS identified as 1. At sample 5 a new value of PPS 2 is required but PPSs 1 and 3 remain unaltered and are used until sample 10. In this case, the sample entry for samples 1 through 4 is the same and contains the initial values of PPSs 1, 2, 3 and SPS 1. At sample 5 the sample entry must change to a second sample entry, which contains the updated value for PPS 2 *as well as* the original values of PPSs 1 and 3 and SPS 1. This second sample entry is used for samples 5 through 10.
- Systems wishing to send SPS or PPS updates will need to compare the two configurations to find the differences in order to send the appropriate parameter set updates.

NOTE 1 When several parameter sets are used and parameter set updating is desired, a separate parameter set elementary stream would be used.

NOTE 2 Decoders conforming to this document are required to support both parameter sets stored in separate elementary streams as well as parameter sets stored in the AVC sample entries, unless restricted by another specification using this one.

When the sample entry name is 'avc3' or 'avc4', parameter sets may be present in both sample entry and as part of samples, and an update of a parameter set by a parameter set of the same type that is stored as part of a sample is possible.

5.4.5 Sync sample

A sample is considered as a sync sample if the video data NAL units in the sample indicate that the primary coded picture contained in the sample is an instantaneous decoding refresh (IDR) picture.

When the sample entry name is 'avc1' or 'avc2', all SPSs and PPSs needed to decode the video data NAL units in the sample of the IDR picture and the following samples in decode order are contained in the decoder configuration of the video elementary stream or in a separate parameter set elementary stream sample.

When the sample entry name is 'avc3' or 'avc4', the following applies:

1. If the sample is a sync sample, all parameter sets needed for decoding the sample shall be included either in the sample entry or in the sample itself.
2. Otherwise (the sample is not a sync sample), all parameter sets needed for decoding the same shall be included either in the sample entry or in any of the samples since the previous sync sample to the sample itself, inclusive.

A parameter set elementary stream sample is a sync sample if and only if all parameter sets required by the associated video elementary stream from the time of the parameter set sample forward are supplied, in the parameter set stream, before they are required by the associated video elementary stream.

5.4.6 Shadow sync

The use of the shadow sync table to indicate alternate encodings of a sample for random access are supported as defined in the ISO base media file format. A shadow sync shall indicate a sample that is a random access point as specified in the general requirements and for the specific coding format in the track.

While the use of shadow sync is supported for backward compatibility reasons, this use is deprecated and use of the mechanisms defined in subclause [5.4.8](#) is recommended.

5.4.7 Layering and sub-sequences

5.4.7.1 Overview

Streams may be constructed so that the referential dependencies between samples allow only subsets of the samples to be sent to the decoder. This mechanism is called *thinning* a stream. Thinning discards entire sets of samples using knowledge of what other sets of pictures this set of pictures depends on and what picture sets in turn depend on it.

The referential dependencies between samples in a stream are structured into *layers* and *sub-sequences*. Samples in higher layers only depend on samples in lower layers. Layers are numbered, and the samples are organized such that a sample in layer N has no dependencies on samples in layers greater than N.

Sub-sequences are as defined in ISO/IEC 14496-10:2014, Annex D. Dependency relations between sub-sequences represent the dependency structure of a stream. Each sub-sequence belongs to one and only one layer. A sample shall reside in one layer and in one sub-sequence only.

Layering and sub-sequence information is represented in the file format to allow systems reading the files to understand the ways in which stream thinning may be achieved without having to examine the dependency structure of every sample.

Layer and sub-sequences are represented in the AVC file format as Sample Group. An AVC file shall contain zero or one instance of a `SampleToGroupBox` (per track) with a `grouping_type` equal to 'avll'. This `SampleToGroupBox` instance represents the assignment of samples in a track to layers. An accompanying instance of the `SampleGroupDescriptionBox` with the same grouping type shall, if it exists, contain `AVCLayerEntry` sample group entries describing the layers. Similarly, an AVC file shall contain zero or one instance of a `SampleToGroupBox` (per track) with a `grouping_type` equal to 'avss'. This `SampleToGroupBox` instance represents the assignment of samples in a track to sub-sequences. An accompanying instance of the `SampleGroupDescriptionBox` with the same grouping type shall, if it exists, contain `AVCSubSequenceEntry` sample group entries describing the sub-sequences.

5.4.7.2 Sub-sequence description entry

5.4.7.2.1 Definition

Group Type: 'avss'

Container: Sample Group Description Box ('sgpd')

Mandatory: No

Quantity: Zero or more.

A sub-sequence description entry is a sample group entry that describes a sub-sequence. A sub-sequence is a set of samples in a track belonging to the same layer. A sub-sequence depends on another sub-sequence if and only if there exists a sample in the sub-sequence that is directly referentially dependent on some sample in the other sub-sequence. All samples in a sub-sequence shall directly depend only on (i.e., refer to) other samples within the same sub-sequence or samples in the sub-sequences on which is it dependent. A sub-sequence depends on zero or more sub-sequences in the lower layers. A sub-sequence shall not depend on any other sub-sequence in the same or higher layer.

At most one partition of an AVC stream into layers shall exist in the AVC file format; that is, there is either zero or one instances of the sample group boxes (`SampleToGroupBox`, `SampleGroupDescriptionBox`) per track concerning the grouping of samples into layers and sub-sequences.

The `grouping_type_parameter` is not defined for the `SampleToGroupBox` with grouping type 'avss'.

5.4.7.2.2 Syntax

```
aligned(8) class DependencyInfo
{
    unsigned int(8) subSeqDirectionFlag;
    unsigned int(8) layerNumber;
    unsigned int(16) subSequenceIdentifier;
}

class AVCSubSequenceEntry () extends VisualSampleGroupEntry ('avss')
{
    unsigned int(16) subSequenceIdentifier;
    unsigned int(8) layerNumber;
    unsigned int(1) durationFlag;
    unsigned int(1) avgRateFlag;
    unsigned int(6) reserved = 0;
    if (durationFlag)
        unsigned int(32) duration;
    if (avgRateFlag)
    {
        unsigned int(7) reserved = 0;
        unsigned int(1) accurateStatisticsFlag;
        unsigned int(16) avgBitRate;
        unsigned int(16) avgFrameRate;
    }
    unsigned int(8) numReferences;
    DependencyInfo dependency[numReferences];
}
}
```

5.4.7.2.3 Semantics

subSeqDirectionFlag, layerNumber and subSequenceIdentifier within the DependencyInfo class identify a sub-sequence that is used as a reference for this sub-sequence. Only direct, not indirect, referential dependencies shall be identified. The identified sub-sequence has sub-sequence identifier equal to subSequenceIdentifier and resides in the layer having the layer number equal to layerNumber. If subSeqDirectionFlag is 0, the sub-sequence used as a reference for this sub-sequence is the closest sub-sequence among all the candidate sub-sequences whose first picture precedes the first picture of this sub-sequence in decoding order and that resides in the indicated layer and has the indicated sub-sequence identifier; ‘closest’ means that among all the candidate sub-sequences the first picture of the

referenced sub-sequence is the closest to the first picture of this sub-sequence in decoding order. If `subSeqDirectionFlag` is equal to 1, the sub-sequence used as a reference for this sub-sequence is the closest sub-sequence among all the candidate sub-sequences whose first picture succeeds the first picture of this sub-sequence in decoding order and that resides in the indicated layer and has the indicated sub-sequence identifier; 'closest' has the same meaning as above.

`subSequenceIdentifier` gives the identifier for the sub-sequence.

`layerNumber` gives the layer number to which the sub-sequence belongs.

`durationFlag` equal to 0 indicates that the duration of the target sub-sequence is not specified. Otherwise, a value of 1 indicates that the `duration` field indicates the duration of this sub-sequence.

`avgRateFlag` equal to 0 indicates that the average bit rate and the average frame rate of the target sub-sequence are unspecified. Otherwise, a value of 1 indicates that the average rate characteristics are described by the `accurateStatisticsFlag`, `avgBitRate`, and `avgFrameRate` fields.

`duration` indicates the duration of the target sub-sequence in clock ticks of a 90-kHz clock.

`accurateStatisticsFlag` indicates how reliable the values of `avgBitRate` and `avgFrameRate` are. `accurateStatisticsFlag` equal to 1 indicates that `avgBitRate` and `avgFrameRate` are rounded from statistically correct values. `accurateStatisticsFlag` equal to 0 indicates that `avgBitRate` and `avgFrameRate` are estimates and may deviate somewhat from the correct values.

`avgBitRate` gives the average bit rate in (1000 bits)/second of this sub-sequence. All NAL units of this sub-sequence are taken into account in the calculation. In the following, B is the number of bits in all NAL units in the sub-sequence. t_1 is the decoding timestamp of the first picture of the sub-sequence (in decoding order), and t_2 is the decoding timestamp of the last picture of the sub-sequence (in decoding order). Then, the `avgBitRate` is calculated as follows provided that $t_1 \neq t_2$: $\text{avgBitRate} = \text{round}(B \div ((t_2 - t_1) * 1000))$. If $t_1 = t_2$, `avgBitRate` shall be 0.

`avgFrameRate` gives the average frame rate in units of frames/(256 seconds) of this sub-sequence. All NAL units of this sub-sequence are taken into account in the calculation. The average frame rate is calculated according to the presentation timestamp of the frame. In the following, C is the number of frames in the sub-sequence. t_1 is the presentation timestamp of the first picture of the sub-sequence (in decoding order), and t_2 is the presentation timestamp (in seconds) of the last picture of the sub-sequence (in decoding order). Then, the `avgFrameRate` is calculated as follows provided that $t_1 \neq t_2$: $\text{avgFrameRate} = \text{round}(C * 256 \div (t_2 - t_1))$. If $t_1 = t_2$, `avgFrameRate` shall be 0. Value zero indicates an unspecified frame rate.

`numReferences` gives the number of sub-sequences directly referenced in this sub-sequence. `dependency` is an array of `DependencyInfo` structures giving the identifying referenced sub-sequences.

5.4.7.3 Layer description entry

5.4.7.3.1 Definition

Group Type: 'av11'

Container: Sample Group Description Box ('sgpd')

Mandatory: No

Quantity: Zero or more.

A layer sample group entry defines the layer information for all samples in a layer. Layers are numbered with non-negative integers. Layers are ordered hierarchically based on their dependency on each other: A layer having a larger layer number is a higher layer than a layer having a smaller layer number. The layers are ordered hierarchically based on their dependency on each other so that a layer does not depend on any higher layer and may depend on lower layers. The lowest layer is numbered as zero and

other layers are given consecutive numbers. In other words, layer 0 is independently decodable, pictures in layer 1 may be predicted from layer 0, pictures in layer 2 may be predicted from layers 0 and 1, etc.

The `grouping_type_parameter` is not defined for the `SampleToGroupBox` with grouping type 'av11'.

5.4.7.3.2 Syntax

```
class AVCLayerEntry() extends VisualSampleGroupEntry ('av11')
{
    unsigned int(8)    layerNumber;

    unsigned int(7)    reserved = 0;

    unsigned int(1)    accurateStatisticsFlag;

    unsigned int(16)   avgBitRate;

    unsigned int(16)   avgFrameRate;
}
```

5.4.7.3.3 Semantics

`layerNumber` gives the number of this layer with the base layer being numbered as zero and all enhancement layers being numbered as one or higher with consecutive numbers.

`accurateStatisticsFlag` indicates how reliable the values of `avgBitRate` and `avgFrameRate` are. `accurateStatisticsFlag` equal to 1 indicates that `avgBitRate` and `avgFrameRate` are rounded from statistically correct values. `accurateStatisticsFlag` equal to 0 indicates that `avgBitRate` and `avgFrameRate` are estimates and may deviate somewhat from the correct values.

`avgBitRate` gives the average bit rate in units of 1000 bits per second. All NAL units in this and lower sub-sequence layers are taken into account in the calculation. The average bit rate is calculated according to the decoding timestamp. In the following, B is the number of bits in all NAL units in this and lower sub-sequence layers. t_1 is the decoding timestamp of the first picture in this and lower sub-sequence layers in the presentation order, and t_2 is the decoding timestamp of the latest picture in this and lower sub-sequence layers in the presentation order. Then, `avgBitRate` is calculated as follows provided that $t_1 \neq t_2$: $\text{avgBitRate} = \text{round}(B \div ((t_2 - t_1) * 1000))$. If $t_1 = t_2$, `avgBitRate` shall be 0. Value zero indicates an unspecified bit rate.

`avgFrameRate` gives the average frame rate in units of frames/(256 seconds). All NAL units in this and lower sub-sequence layers are taken into account in the calculation. In the following, C is the number of frames in this and lower sub-sequence layers. t_1 is the presentation timestamp of the first picture in this and lower sub-sequence layers in presentation order, and t_2 is the presentation timestamp of the latest picture in this and lower sub-sequence layers in the presentation order. Then, the `avgFrameRate` is calculated as follows provided that $t_1 \neq t_2$: $\text{avgFrameRate} = \text{round}(C * 256 \div (t_2 - t_1))$. If $t_1 = t_2$, `avgFrameRate` shall be 0. Value zero indicates an unspecified frame rate.

5.4.8 Alternate streams and switching pictures

5.4.8.1 Switching pictures

In typical streaming scenarios, one of the key requirements is to scale the bit rate of the compressed data in response to changing network conditions. The simplest way to achieve this is to encode multiple streams with different bandwidths and quality settings for representative network conditions. The server can then switch amongst these pre-coded streams in response to network conditions. In earlier standards, switching between streams is only possible at I-pictures, because the pictures may only be switched when there are no dependencies on prior pictures for reconstruction.

AVC includes supports for SP-pictures and SI-pictures ("switching pictures") that allow switching from one stream to another while still supporting inter coding of switching pictures. The following figure shows how SP pictures are used to switch between two different bit streams.

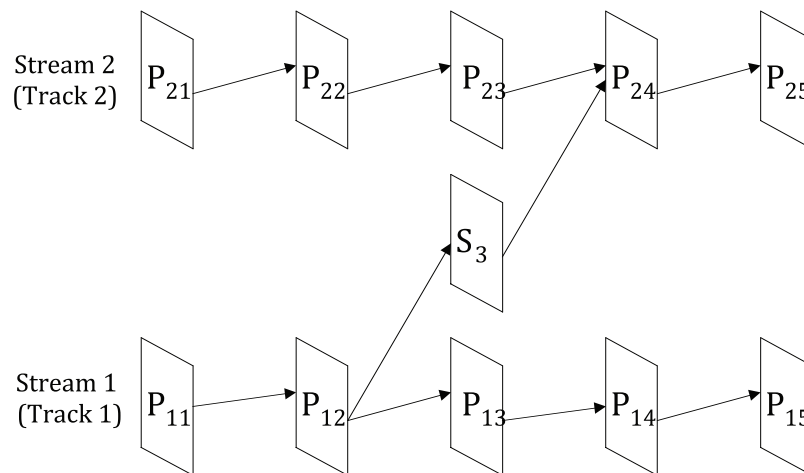


Figure 3 — Stream switching

In the file format, switching pictures are stored in *switching picture* tracks, which are tracks separate from the track that is being switched from and the track being switched to. Switching picture tracks can be identified by the existence of a specific required track reference in that track. A switching picture is an alternative to the sample in the destination track that has exactly the same decoding time. If all switching pictures are SI pictures, then no further information is needed.

If any of the pictures in the switching track are SP pictures, then two extra pieces of information may be needed. First, the source track that is being switched from shall be identified by using a track reference (the source track may be the same track as the destination track). Second, the dependency of the switching picture on the samples in the source track may be needed, so that a switching picture is only used when the pictures on which it depends have been supplied to the decoder.

This dependency is represented by means of an optional extra sample table. There is one entry per sample in the switching track. Each entry records the relative sample numbers in the source track on which the switching picture depends. If this array is empty for a given sample, then that switching sample contains an SI picture. If the dependency box is not present, then only SI-frames shall be present in the track.

A switching sample may have multiple coded representations with different dependencies. For AVC video, the multiple representations of a switching sample are stored in different switching tracks (i.e. access unit). For example, one switch track might contain a SP-picture representation dependent on some earlier samples, used for stream switching, while another switch track may contain another representation as an SI-picture, used for random access.

5.4.8.2 Alternate group

The ISO base media file format (but not the version one specification of the MPEG-4 file format, which is branded as 'mp41') supports the use of what is called *alternate tracks*. Each track can optionally specify an *alternate group* (in the track header box) that groups together alternate encodings of the same content. Thus, each alternate bitstream can be stored as a separate track and related together as alternate tracks. All the tracks that form a group that may be switched between, but not the tracks containing the switching pictures, shall be a member of an `alternate_group` with a non-zero group identifier.

An alternate group is not needed if there is only one primary track, with a switching track. This switching track may contain SI pictures, or SP pictures for trick modes or error resilience, which predict both from and to the same track.

5.4.8.3 Track references

The switching track shall be linked to the track into which it switches (the destination track) by a track reference of type 'swto' in the switching picture track.

If the switching track contains SP pictures, the switching track shall be linked to the track from which it switches (the source track) by a track reference of type 'swfr' in the switching picture track.

5.4.8.4 Sample dependency

5.4.8.4.1 Definition

Box Type: 'sdep'

Container: Sample Table 'stbl'

Mandatory: No

Quantity: Zero or exactly one.

This subclause defines the dependencies of each switching sample on sample(s) in the source track. This table is only needed in a switching track that has a source ('swfr') track dependency.

This box contains the sample dependencies for each switching sample. The dependencies are stored in the table, one record for each sample. The size of the table, `sample_count` is taken from the `sample_count` in the Sample Size Box ('stsz') or Compact Sample Size Box ('stz2').

5.4.8.4.2 Syntax

```
aligned(8) class SampleDependencyBox
{
    extends FullBox('sdep', version = 0, 0) {
        for (i=0; i < sample_count; i++){
            unsigned int(16) dependency_count;

            for (k=0; k < dependency_count; k++) {
                signed int(16) relative_sample_number;
            }
        }
    }
}
```

5.4.8.4.3 Semantics

`dependency_count` is an integer that counts the number of samples in the source track on which this switching sample directly depends.

`relative_sample_number` is an integer that identifies a sample in the source track. The relative sample numbers are encoded as follows. If there is a sample in the source track with the same decoding time, it has a relative sample number of 0. Whether or not this sample exists, the sample in the source track that immediately precedes the decoding time of the switching sample has relative sample number -1, the sample before that -2, and so on. Similarly, the sample in the source track that immediately follows the decoding time of the switching sample has relative sample number +1, the sample after that +2, and so on.

5.4.9 Definition of a sub-sample for AVC

For the use of the sub-sample information box (ISO/IEC 14496-12:2015, 8.7.7) in an AVC stream, a sub-sample is defined as one or more contiguous NAL units within a sample and having the same value of the following fields; `RefPicFlag`, `RedPicFlag` and `VclNalUnitFlag`. Each sub-sample includes both NAL unit(s) and their preceding NAL unit length field(s). The presence of this box is optional; however, if present in a track containing AVC data, it shall have the semantics defined here.

The `subsample_priority` field shall be set to a value in accordance with the specification of this field in ISO/IEC 14496-12.

The `discardable` field shall be set to 1 only if this sample is still decodable if this sub-sample is discarded (e.g. the sub-sample consists of an SEI NAL unit, or a redundant coded picture).

The `codec_specific_parameters` field of the Subsample Information box is defined for AVC as follows:

```
unsigned int(1) RefPicFlag;

unsigned int(1) RedPicFlag;

unsigned int(1) VclNalUnitFlag;

bit(29) reserved = 0;
```

`RefPicFlag` equal to 0 indicates that all the NAL units in the sub-sample have `nal_ref_idc` equal to 0. `RefPicFlag` equal to 1 indicates that all the NAL units in the sub-sample have `nal_ref_idc` greater than 0.

`RedPicFlag` equal to 0 indicates that all the NAL units in the sub-sample have `redundant_pic_cnt` equal to 0. `RedPicFlag` equal to 1 indicates that all the NAL units in the sub-sample have `redundant_pic_cnt` greater than 0.

`VclNalUnitFlag` equal to 0 indicates that all NAL units in the sub-sample are non-VCL NAL units. Value 1 indicates that all NAL units in the sub-sample are VCL NAL units.

6 SVC elementary stream and sample definitions

6.1 Overview

This clause specifies the storage format of SVC data. It extends the definitions of the storage format of AVC in subclause 4.11.

The file format for storage of SVC content, as defined in this clause and Annexes A to D, uses the existing capabilities of the ISO base media file format and the plain AVC file format (i.e. the file format specified in subclause 4.11). In addition, the following new extensions, among others, to support SVC-specific features are specified.

a) Scalable or multiview grouping:

a structuring and grouping mechanism to indicate the association of NAL units with different types and hierarchy levels of scalability.

b) Aggregator:

a structure to enable efficient scalable grouping of NAL units by changing irregular patterns of NAL units into regular patterns of aggregated data units.

c) Extractor:

a structure to enable efficient extraction of NAL units from other tracks than the one containing the media data.

d) Temporal metadata statements:

structures for storing time-aligned information of media samples.

e) AVC compatibility:

a provision for storing an SVC bitstream in an AVC compatible manner, such that the AVC compatible base layer can be used by any plain AVC file format compliant reader.

6.2 Elementary stream structure

SVC streams are stored in accordance with 5.2, with the following definition of an SVC video elementary stream:

- **An SVC Video Elementary Streams** contains all video coding related NAL units (i.e. those NAL units containing video data or signalling video structure, possibly after resolution of extractors and aggregators) and may contain non-video coding related NAL units such as SEI messages and access unit delimiter NAL units. Aggregators and Extractors, when present, shall not be directly output by file parsers.

SVC streams may also be stored using associated parameter set streams, if needed.

For SVC streams, Table 2 is updated as follows in Table 3; only entries where the definition for SVC differs from AVC, are shown.

Table 3 — NAL unit types in SVC and AVC streams

Value of nal_unit_ type	Description	AVC video elementary stream	SVC video elementary stream (sample entry 'avc1', 'avc2', or 'svc1')	SVC video elementary stream (sample entry 'avc3', 'avc4', or 'svc2')	Parameter set elementary stream
14	Prefix NAL unit in scalable extension prefix_nal_unit_rbsp()	Not specified	Yes	Yes	No
15	Subset sequence parameter set subset_seq_parameter_set_rbsp()	Not specified	No. If parameter set elementary stream is not used, Subset SPS shall be stored in the Decoder Specific Information.	Yes. Parameter set elementary stream shall not be used.	Yes
20	Coded slice in scalable extension slice_layer_extension_rbsp()	Not specified	Yes	Yes	No
24 – 31	Not specified	See Annex F	See Annex F	See Annex F	See Annex F

NOTE slice_layer_extension_rbsp was previously called slice_layer_in_scalable_extension_rbsp.

There may be AVC VCL NAL units, SVC VCL NAL units and other NAL units, i.e. non-VCL NAL units, present in an SVC video elementary stream. Additionally, there may be Aggregators and Extractors present in an SVC video elementary stream.

An AVC VCL NAL unit in an SVC video elementary stream conforming to one or more profiles specified in ISO/IEC 14496-10:2014, Annex G shall be immediately preceded by a prefix NAL unit containing the scalability information for the AVC VCL NAL unit. In this file format an AVC VCL NAL unit and the

immediately preceding prefix NAL unit are logically seen as one NAL unit: the prefix NAL unit provides the scalability information and the AVC VCL NAL unit provides the NAL unit type and payload.

6.3 Use of the plain AVC file format

The SVC file format is an extension of the plain AVC file format defined in subclause 4.11.

Subclause 5.4.7 is defined for use with plain AVC streams. Its use with SVC streams is deprecated.

6.4 Sample and configuration definition

6.4.1 Overview

SVC Sample: An SVC sample consists of the NAL units that belong to an access unit as defined in ISO/IEC 14496-10:2014, 7.4.1.2 and that are represented by the track.

6.4.2 Canonical order and restrictions

6.4.2.1 Restrictions

The following restrictions apply to SVC data in addition to the requirements in subclause 5.3.2.

- **SVC coded slice NAL units** (Coded slices in scalable extension): All SVC coded slice NAL units for a single instant in time shall be contained in the sample whose composition time is that of the picture represented by the access unit. An SVC sample shall contain at least one AVC or SVC VCL NAL unit.
- **Prefix NAL units** (Prefix NAL unit in scalable extension): Each prefix NAL unit is placed immediately before the corresponding AVC VCL NAL unit, providing scalability information about the AVC VCL NAL unit.

NOTE Prefix NAL units can also be associated with filler data NAL units.

- **Aggregators/Extractors:** The order of all NAL units included in an Aggregator or referenced by an Extractor is exactly the decoding order as if these NAL units were present in a sample not containing aggregators or extractors. After processing the Aggregator or the Extractor, all NAL units must be in valid decoding order as specified in ISO/IEC 14496-10.

6.4.2.2 Decoder configuration record

When the decoder configuration record defined in subclause 5.3.3.1 is used for a stream that can be interpreted as either an SVC or AVC stream, the AVC decoder configuration record shall reflect the properties of the AVC compatible base layer, e.g. it shall contain only parameter sets needed for decoding the AVC base layer.

If the sample entry name is 'svc1', a parameter set stream may be used with SVC streams, as with AVC streams, in which case, parameter sets shall not be included in the decoder configuration record. Otherwise (the sample entry name is 'svc2'), parameter sets may be stored in both the decoder configuration record or as part of samples while a parameter set elementary stream shall not be used.

Sequence or picture parameter sets are numbered in order of storage from 1 to `numOfSequenceParameterSets` or `numOfPictureParameterSets` respectively. Sequence and Picture parameter sets stored in this record in a file may be referenced using this 1-based index by the `InitialParameterSetBox`.

The `SVCDecoderConfigurationRecord` is structurally identical to an `AVCDecoderConfigurationRecord`. However, the reserved bits preceding and succeeding the `lengthSizeMinusOne` field are re-defined. The syntax is as follows:

```

aligned(8) class SVCDecoderConfigurationRecord {

    unsigned int(8) configurationVersion = 1;

    unsigned int(8) AVCProfileIndication;

    unsigned int(8) profile_compatibility;

    unsigned int(8) AVCLevelIndication;

    unsigned int(1) complete_representation;

    bit(5) reserved = '11111'b;

    unsigned int(2) lengthSizeMinusOne;

    bit(1) reserved = '0'b;

    unsigned int(7) numOfSequenceParameterSets;

    for (i=0; i< numOfSequenceParameterSets; i++) {

        unsigned int(16) sequenceParameterSetLength ;

        bit(8*sequenceParameterSetLength) sequenceParameterSetNALUnit;

    }

    unsigned int(8) numOfPictureParameterSets;

    for (i=0; i< numOfPictureParameterSets; i++) {

        unsigned int(16) pictureParameterSetLength;

        bit(8*pictureParameterSetLength) pictureParameterSetNALUnit;

    }

}

```

The semantics of the fields `AVCProfileIndication`, `profile_compatibility`, and `AVCLevelIndication` differ from the `AVCDecoderConfigurationRecord` as follows:

The fields `AVCProfileIndication`, `AVCLevelIndication` carry the profile and level indications, respectively, indicating the profile and level of the entire scalable stream in this track. They, and the `profile_compatibility` field, shall have values such that an SVC decoder supporting these values is able to decode bitstreams conforming to the profile, level and profile compatibility flags indicated in any of the sequence parameter sets or subset sequence parameter sets contained in this record.

The semantics of other fields are as follows, or are as defined for an `AVCDecoderConfigurationRecord`:

- `complete_representation` is set on a minimal set of tracks that contain a portion of the original encoded scalable stream, as defined in subclause 6.5.1. Other tracks may be removed from the file without loss of any portion of the original encoded bitstream, and, once the set of tracks has been reduced to only those in the complete subset, any further removal of a track removes a portion of the encoded information.
- `numOfSequenceParameterSets` indicates the number of SPSs and subset SPSs that are used for decoding the SVC elementary stream. The value of `numOfSequenceParameterSets` shall be in the range of 0 to 64, inclusive.
- `SequenceParameterSetLength` indicates the length in bytes of the SPS or subset SPS NAL unit.

- `SequenceParameterSetNALUnit` contains a SPS or subset SPS NAL unit. SPSs shall occur in order of ascending parameter set identifier with gaps being allowed. Subset SPSs shall occur in order of ascending parameter set identifier with gaps being allowed. Any SPS shall occur before all the subset SPSs, if any.

6.5 Derivation from the ISO base media file format

6.5.1 SVC track structure

A scalable video stream is represented by one or more video tracks in a file. Each track represents one or more operating points of the scalable stream. A scalable stream may, of course, be further thinned, if desired.

There is a minimal set of one or more tracks that, when taken together, contain the complete set of encoded information. All these tracks shall have the flag “`complete_representation`” set in all their sample entries. This group of tracks that form the complete encoded information are called the “complete subset”.

Let the lowest operating point be the one of all the operating points represented by DTQ (`dependency_id`, `temporal_id` and `quality_id`) combinations that has the least values of `dependency_id`, `temporal_id` and `quality_id`, respectively. The track that has the flag “`complete_representation`” set and contains the lowest operating point shall be nominated as the ‘scalable base track’. All the other tracks that are part of the same scalable encoded information shall be linked to this base track by means of a track reference of type ‘`sbas`’ (scalable base). The complete encoded information can be retained when the tracks included in the “complete subset” are retained; all other tracks shall represent subsets, copies (i.e. the same media data) or re-orderings of the complete subset.

NOTE 1 An alternate group can also include completely independent bitstreams, as well as alternative operating points of the same bitstream. The SVC tracks in the alternate group shall be examined to see how many scalable base tracks are identified.

NOTE 2 “A scalable bitstream” can require more than one track to represent it (consider a stream with a low-resolution, low-frame-rate base layer, and a high resolution enhancement layer, and a high frame-rate enhancement layer, but missing the data for high resolution high frame-rate). However, such a scalable bitstream is typically a non-conforming bitstream.

All the tracks sharing the same scalable base track shall share the same timescale as the scalable base track.

6.5.2 Data sharing and extraction

Different tracks may logically share data. This sharing can take one of the following two forms:

- The sample data is copied (i.e., the same media data is present) in different tracks.
- There may be instructions on how to perform this copy at the time that the file is read.

For the second case, Extractors (defined in [A.3](#)) are used.

6.5.3 SVC video stream definition

6.5.3.1 Sample entry name and format

6.5.3.1.1 Definition

Sample Entry and Box Types: 'svc1', 'svc2', 'svcC', 'seib'

Container: Sample Description Box ('stsd')

Mandatory: One of the 'avc1', 'avc2', 'avc3', 'avc4',
'svc1', and 'svc2' sample entries is mandatory.

Quantity: One or more sample entries may be present

If an SVC elementary stream contains an AVC compatible base layer, then an AVC visual sample entry ('avc1', 'avc2', 'avc3', or 'avc4') shall be used. Here, the entry shall contain initially an AVC Configuration Box, possibly followed by an SVC Configuration Box as defined below. The AVC Configuration Box documents the Profile, Level and Parameter Set information pertaining to the AVC compatible base layer as defined by the `AVCDecoderConfigurationRecord`. The SVC Configuration Box documents the Profile, Level, and possibly also parameter sets pertaining to the entire stream containing the SVC compatible enhancement layers as defined by the `SVCDecoderConfigurationRecord`, stored in the `SVCConfigurationBox`.

If the SVC elementary stream does not contain an AVC base layer, then an SVC visual sample entry ('svc1' or 'svc2') shall be used. The SVC visual sample entry shall contain an SVC Configuration Box, as defined below. This includes an `SVCDecoderConfigurationRecord`, as defined in this document.

The `lengthSizeMinusOne` field in the SVC and AVC configurations in any given sample entry shall have the same value.

A priority assignment URI provides the name (in the URI space) of a method used to assign `priority_id` values. When it occurs in an AVC or SVC sample entry, exactly one URI shall be present, that documents the `priority_id` assignments in the stream. The URI is treated here as a name only; it should be de-referenceable, though this is not required. File readers may be able to recognize some methods and thereby know what stream extraction operations based on `priority_id` would do.

Extractors or aggregators may be used for SVC VCL NAL units in 'avc1', 'avc2', 'avc3', 'avc4', 'svc1' or 'svc2' tracks.

NOTE When AVC compatibility is indicated, it can be necessary to indicate an unrealistic level for the AVC base layer, to accommodate the bit rate of the entire stream, because all the NAL units are considered as included in the AVC base layer and hence can be fed to the decoder, which is expected to discard those NAL unit it does not recognize. This case happens when the 'avc1' or 'avc3' sample entry is used and both AVC and SVC configurations are present.

Either or both of a `ScalabilityInformationSEIBox` or `SVCConfigurationBox` may be present in an 'avc1' or 'avc3' sample entry. In this case the `AVCSVCSampleEntry` definition below applies.

The parameter sets required to decode a NAL unit that is present in the sample data of a video stream, either directly or by reference from an Extractor, shall be present in the decoder configuration of that video stream or in the associated parameter set stream (if used).

[Table 4](#) shows for a video track all the possible uses of sample entries, configurations and the SVC tools (excluding timed metadata, which is always used in another track):

Table 4 — Use of sample entries for AVC and SVC tracks

sample entry name	with configuration records	meaning
'avc1' or 'avc3'	AVC Configuration Only	A plain AVC track with AVC NAL units only; Extractors, aggregators, and tier grouping shall not be present.
'avc1' or 'avc3'	AVC and SVC Configurations	An SVC track with both AVC and SVC NAL units; Extractors and aggregators shall not be present.
'avc2' or 'avc4'	AVC Configuration Only	A plain AVC track with AVC NAL units only; Extractors may be present and used to reference AVC NAL units; Aggregators may be present to contain and reference AVC NAL units; Tier grouping may be present.
'avc2' or 'avc4'	AVC and SVC Configurations	An SVC track with both AVC and SVC NAL units; Extractors and aggregators may be present; Extractors may reference both AVC and SVC NAL units; Aggregators shall not contain but may reference AVC NAL units, and may both contain and reference SVC NAL units; Tier grouping may be present.
'svc1' or 'svc2'	SVC Configuration	An SVC track without AVC NAL units; Extractors shall be present to reference AVC NAL units (from a different track) and may also be used to reference SVC NAL units; Aggregators may be present to contain and reference SVC NAL units; Tier grouping may be present.

6.5.3.1.2 Syntax

```

class SVCConfigurationBox extends Box('svcC') {
    SVCDecoderConfigurationRecord() SVCConfig;
}

class ScalabilityInformationSEIBox extends Box('seib', size)
{
    unsigned int(8*size-64) scalinfosei;
}

class SVCPriorityAssignmentBox extends Box('svcP')
{
    unsigned int(8)    method_count;
    string PriorityAssignmentURI[method_count];
}

class AVCSVCSampleEntry() extends AVCSampleEntry ('avc1' or 'avc3') {
    SVCConfigurationBox  svcconfig;          // optional
    ScalabilityInformationSEIBox  scalability; // optional
    SVCPriorityAssignmentBox  method;         // optional
}

class AVC2SVCSampleEntry() extends AVC2SampleEntry('avc2' or 'avc4') {

```

```

    SVCConfigurationBox  svcconfig;          // optional

    ScalabilityInformationSEIBox  scalability; // optional

    SVCPriorityAssignmentBox  method;          // optional
}

// Use this if the track is NOT AVC compatible

class SVCSampleEntry() extends VisualSampleEntry ('svc1' or 'svc2') {

    SVCConfigurationBox      svcconfig;

    MPEG4ExtensionDescriptorsBox  descr; // optional

    ScalabilityInformationSEIBox  scalability; // optional

    SVCPriorityAssignmentBox  method;          // optional
}

```

6.5.3.1.3 Semantics

When the sample entry is 'svc1' or 'svc2', Compressorname in the base class VisualSampleEntry indicates the name of the compressor used, with the value "\012SVC Coding" being recommended (\012 is 10, the length of the string "SVC coding" in bytes).

scalinfosei contains an SEI NAL unit containing only a scalability information SEI message as specified in ISO/IEC 14496-10:2014, Annex G. The 'size' field of the container box ScalabilityInformationSEIBox shall not be equal to 0 or 1.

method_count provides a count of the number of following URIs. This field shall take the value 1 in an 'avc1', 'avc2', 'avc3', 'avc4', 'svc1' or 'svc2' sample entry.

PriorityAssignmentURI provides a unique name of the method used to assign priority_id values. In the case of absence of this box, the priority assignment method is unknown.

6.5.4 SVC visual width and height

The visual width and height documented in a VisualSampleEntry of a stream containing SVC VCL NAL unit are the cropped frame dimensions of the AVC base layer, if the stream is described by a sample entry of type 'avc1', 'avc2', 'avc3' or 'avc4'; otherwise they are the cropped frame dimensions of decoded pictures by decoding the entire stream.

6.5.5 Sync sample

For video data described by a sample entry of type 'avc1', 'avc2', 'avc3', or 'avc4', the sync sample table identifies IDR access units for both an AVC decoder, and an SVC decoder (if any) operating on the entire bitstream.

For video data described by a sample entry of type 'svc1', the sync sample table identifies IDR access units in the entire SVC bitstream.

For video data described by a sample entry of type 'svc2', the sync sample table identifies IDR access units in the entire SVC bitstream, and additionally the following applies:

1. If the sample is a sync sample, all parameter sets needed for decoding that sample shall be included either in the sample entry or in the sample itself.

2. Otherwise (the sample is not a sync sample), all parameter sets needed for decoding the sample shall be included either in the sample entry or in any of the samples since the previous sync sample to the sample itself, inclusive.

NOTE The sync sample table, if present, documents only access units that are IDR access units for both the AVC compatible base layer and the layer corresponding to decoding the entire bitstream contained in the track. In case documenting of layer-specific IDR access units is desired, the stream should be stored in separate tracks, e.g. two tracks, one containing the AVC base layer with a sample entry of type 'avc1' or 'avc3', and the other containing the SVC enhancement layers with a sample entry of type 'svc1' or 'svc2'. However, extractors shall then be used for tracks that are not the scalable base track.

6.5.6 Shadow sync

A shadow sync box shall not be used for video data described by an 'svc1' or 'svc2' sample entry. Its use for SVC is deprecated.

6.5.7 Independent and disposable samples box

If the `SampleDependencyTypeBox` is used in a track that is both AVC and SVC compatible, then care should be taken that the information provided by this box is true no matter what valid subset of the SVC data (possibly only the AVC data) is used. The 'unknown' values (value 0 of the fields `sample_depends_on`, `sample_is_depended_on`, and `sample_has_redundancy`) may be needed if the information varies.

6.5.8 Sample groups on random access recovery points 'roll' and random access points 'rap'

For video data described by a sample entry of type 'avc1', 'avc2', 'avc3' or 'avc4', the random access recovery sample group and the random access point sample group identify random access recovery points and random access points, respectively, for both an AVC decoder, and an SVC decoder (if any) operating on the entire bitstream.

NOTE If the random access recovery points or the random access points for the AVC decoder and the SVC decoder operating on the entire bitstream are not all aligned, the random access recovery points table or the random access point table, respectively, will not document all of them. In this case, the stream can be stored in multiple tracks, e.g. two tracks, one containing the AVC base layer with a sample entry of type 'avc1' or 'avc3', and the other containing the SVC enhancement layers with a sample entry of type 'svc1' or 'svc2'.

For video data described by a sample entry of type 'svc1' or 'svc2', the information provided by the random access recovery sample group and the random access point sample group is true for any valid subset of the entire SVC bitstream.

6.5.9 Definition of a sub-sample for SVC

This subclause extends the definition of a sub-sample for AVC in [5.4.9](#).

For the use of the sub-sample information box (ISO/IEC 14496-12:2015, 8.7.7) in an SVC stream, a sub-sample is defined as one or more contiguous whole NAL units having the same values of the following fields: `RefPicFlag`, `RedPicFlag`, `VclNalUnitFlag`, `IdrFlag`, `PriorityId`, `DependencyId`, `QualityId`, `TemporalId`, `UseRefBasePicFlag`, `DiscardableFlag` and `StoreBaseRepFlag`, specified subsequently. Each sub-sample includes both NAL unit(s) and their preceding NAL unit length field(s). The presence of this box is optional; however, if present in a track containing SVC data, it shall have the semantics defined here.

As required in [5.4.9](#), the `subsample_priority` field shall be set to a value in accordance with the specification of this field in ISO/IEC 14496-12.

The `codec_specific_parameters` field of the Subsample Information box is defined for SVC as follows:

```
unsigned int(1) RefPicFlag;

unsigned int(1) RedPicFlag;

unsigned int(1) VclNalUnitFlag;
```

```

bit(6) reserved = 0;

unsigned int(1) IdrFlag;

unsigned int(6) PriorityId;

bit(1) reserved = 0; // corresponding to no_inter_layer_pred_flag

unsigned int(3) DependencyId;

unsigned int(4) QualityId;

unsigned int(3) TemporalId;

unsigned int(1) UseRefBasePicFlag;

unsigned int(1) DiscardableFlag;

bit(1) reserved = 0; // corresponding to output_flag

unsigned int(1) StoreBaseRepFlag;

bit(1) reserved = 0;

```

For an AVC VCL NAL unit in an SVC context, the prefix NAL unit shall be grouped with the AVC VCL NAL unit in the same sub-sample, and its fields values apply to the AVC VCL NAL unit.

RefPicFlag equal to 0 indicates that all the NAL units in the sub-sample have **nal_ref_idc** equal to 0. **RefPicFlag** equal to 1 indicates that all the NAL units in the sub-sample have **nal_ref_idc** greater than 0.

RedPicFlag equal to 0 indicates that all the NAL units in the sub-sample have **redundant_pic_cnt** equal to 0. **RedPicFlag** equal to 1 indicates that all the NAL units in the sub-sample have **redundant_pic_cnt** greater than 0.

VclNalUnitFlag equal to 0 indicates that all NAL units in the sub-sample are non-VCL NAL units. Value 1 indicates that all NAL units in the sub-sample are VCL NAL units.

IdrFlag indicates the **idr_flag** value of the NAL units in the sub-sample. All the NAL units in the sub-sample shall have the same value of **idr_flag**.

PriorityId indicates the **priority_id** value of the NAL units in the sub-sample. All the NAL units in the sub-sample shall have the same value of **priority_id**.

NoInterLayerPredFlag indicates the **no_inter_layer_pred_flag** of the NAL units in the sub-sample. All the NAL units in the sub-sample shall have the same value of **no_inter_layer_pred_flag**.

DependencyId indicates the **dependency_id** value of the NAL units in the sub-sample. All the NAL units in the sub-sample shall have the same **dependency_id** value.

QualityId indicates the **quality_id** value of the NAL units in the sub-sample. All the NAL units in the sub-sample shall have the same **quality_id** value.

TemporalId indicates the **temporal_id** value of the NAL units in the sub-sample. All the NAL units in the sub-sample shall have the same **temporal_id** value.

UseRefBasePicFlag indicates the **use_ref_base_pic_flag** value of the NAL units in the sub-sample. All the NAL units in the sub-sample shall have the same value of **use_ref_base_pic_flag**.

DiscardableFlag indicates the **discardable_flag** value of the NAL units in the sub-sample. All the NAL units in the sub-sample shall have the same **discardable_flag** value.

NOTE This is not the same definition as the **discardable** field in the sub-sample information box.