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

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

Technologies de l'information — Codage des objets audiovisuels —

*Partie 15: Transport de vidéo structuré en unités NAL au format ISO de base
pour les fichiers médias*

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 14496-15:2014

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 14496-15:2014



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2014

All rights reserved. Unless otherwise specified, 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
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword	v
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	5
4 General Definitions	6
4.1 Introduction	6
4.2 Elementary stream structure	6
4.3 Sample and Configuration definition	6
4.4 Video Track Structure	8
4.5 Template fields used	8
4.6 Visual width and height	9
4.7 Decoding time (DTS) and composition time (CTS)	9
4.8 Sync sample (IDR)	9
4.9 Shadow sync	10
4.10 Sample groups on random access recovery points and random access points	10
4.11 Hinting	10
5 AVC elementary streams and sample definitions	11
5.1 Introduction	11
5.2 Elementary stream structure	11
5.3 Sample and Configuration definition	14
5.4 Derivation from ISO Base Media File Format	18
6 SVC elementary stream and sample definitions	29
6.1 Introduction	29
6.2 Elementary stream structure	30
6.3 Use of the plain AVC file format	31
6.4 Sample and configuration definition	31
6.5 Derivation from the ISO base media file format	33
7 MVC elementary stream and sample definitions	39
7.1 Introduction	39

7.2	Overview of MVC Storage	40
7.3	MVC Track Structure	41
7.4	Use of the plain AVC File Format.....	42
7.5	Sample and configuration definition	42
7.6	Derivation from the ISO base media file format	45
7.7	MVC specific information boxes.....	54
8	HEVC elementary streams and sample definitions.....	63
8.1	Introduction	63
8.2	Elementary Stream Structure	64
8.3	Sample and configuration definition	64
8.4	Derivation from ISO base media file format.....	69
Annex A	(normative) In-stream structures specific to SVC and MVC	76
Annex B	(normative) SVC and MVC sample group and sub-track definitions.....	81
Annex C	(normative) Temporal metadata support.....	102
Annex D	(normative) File format toolsets.....	110
Annex E	(normative) Sub-parameters for the MIME type 'Codecs' parameter	112
Annex F	(Informative) Patent Statements.....	114

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 14496-15:2014

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. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

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.

ISO/IEC 14496-15 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This third edition cancels and replaces the second edition (ISO/IEC 14496-15:2010), which has been technically revised. It also incorporates the Amendment ISO/IEC 14496-15:2010/Amd.1:2011 and the Technical Corrigenda ISO/IEC 14496-15:2010/Cor.1:2011 and ISO/IEC 14496-15:2010/Cor.2:2012.

ISO/IEC 14496 consists of the following parts, under the general title *Information technology — Coding of audio-visual objects*:

- *Part 1: Systems*
- *Part 2: Visual*
- *Part 3: Audio*
- *Part 4: Conformance testing*
- *Part 5: Reference software*
- *Part 6: Delivery Multimedia Integration Framework (DMIF)*
- *Part 7: Optimized reference software for coding of audio-visual objects* [Technical Report]
- *Part 8: Carriage of ISO/IEC 14496 contents over IP networks*
- *Part 9: Reference hardware description* [Technical Report]
- *Part 10: Advanced Video Coding*

- *Part 11: Scene description and application engine*
- *Part 12: ISO base media file format*
- *Part 13: Intellectual Property Management and Protection (IPMP) extensions*
- *Part 14: MP4 file format*
- *Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format*
- *Part 16: Animation Framework eXtension (AFX)*
- *Part 17: Streaming text format*
- *Part 18: Font compression and streaming*
- *Part 19: Synthesized texture stream*
- *Part 20: Lightweight Application Scene Representation (LAsER) and Simple Aggregation Format (SAF)*
- *Part 21: MPEG-J Graphics Framework eXtension (GFX)*
- *Part 22: Open Font Format*
- *Part 23: Symbolic Music Representation*
- *Part 24: Audio and systems interaction*
- *Part 25: 3D Graphics Compression Model*
- *Part 26: Audio conformance*
- *Part 27: 3D Graphics conformance*
- *Part 28: Composite font representation*

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 14496-15:2014

Introduction

This part of ISO/IEC 14496 defines a storage format based on, and compatible with, the ISO Base Media File Format (ISO/IEC 14496-12 and ISO/IEC 15444-12), 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 part of ISO/IEC 14496 enables video streams formatted as Network Adaptation Layer Units (NAL Units) to

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

This part of ISO/IEC 14496 may be used as a standalone specification; it specifies how NAL unit structured video content shall be stored in an ISO Base Media File Format compliant format. However, it is normally used in the context of a specification, such as the MP4 file format, derived from the ISO Base Media File Format, that permits the use of NAL unit structured video such as AVC (ISO/IEC 14496-10) and video and High Efficiency Video Coding (HEVC, ISO/IEC 23008-2) video.

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

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

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

The holder of this patent right has assured the ISO and IEC that he is willing to negotiate licences under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statement of the holder of this patent right is registered with the ISO and IEC. Information may be obtained from the companies listed in Annex F.

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 in Annex F. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

STANDARDSISO.COM : Click to view the full PDF of ISO/IEC 14496-15:2014

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 part of ISO/IEC 14496 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, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

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

ISO/IEC 14496-12, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*¹⁾

ISO/IEC 23008-2, *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 or ISO/IEC 23008-2, and the following apply.

3.1.1

aggregator

in-stream structure using a NAL unit header

NOTE Aggregators are used to group NAL units belonging to the same sample.

¹⁾ ISO/IEC 14496-12 is technically identical to ISO/IEC 15444-12.

3.1.2

AVC base layer

maximum subset of a bitstream that is AVC compatible (i.e. a bitstream not using any of the functionality of ISO/IEC 14496-10 Annex G or Annex H)

NOTE 1 The AVC base layer is represented by AVC VCL NAL units and associated non-VCL NAL units.

NOTE 2 The AVC base layer itself can be a temporal scalable bitstream.

3.1.3

AVC NAL unit

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

3.1.4

AVC VCL NAL unit

NAL unit with type 1 to 5 (inclusive) as specified in ISO/IEC 14496-10

3.1.5

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 An extraction path may 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.

3.1.6

extractor

in-stream structure using a NAL unit header including a NAL unit header extension

NOTE Extractors contain instructions on how to extract data from other tracks. Logically an Extractor can be seen as a 'link'. While accessing a track containing Extractors, the Extractor is replaced by the data it is referencing.

3.1.7

in-stream structure

structure residing within sample data

3.1.8

MVC VCL NAL unit

NAL unit with type 20, and NAL units with type 14, as specified in ISO/IEC 14496-10, when the immediately following NAL units are AVC VCL NAL units.

NOTE MVC VCL NAL units do not affect the decoding process of a legacy AVC decoder.

3.1.9

operating point

subset of a scalable bitstream, representing in SVC a particular spatial resolution, temporal resolution, and quality, or in MVC a set of target output views

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

NOTE 2 In an SVC stream an operating point 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 may be used to define multiple operating points.

NOTE 3 The bitstream subset of an MVC 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.

NOTE 4 An operating point is referred to as an operation point in Annex H of ISO/IEC 14496-10 or in ISO/IEC 23008-2.

3.1.10

parameter set

video parameter set, sequence parameter set, or picture parameter set, as defined in the applicable video standard (e.g. ISO/IEC 14496-10 or ISO/IEC 23008-2)

NOTE This term is used to refer to all types of parameter sets.

3.1.11

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

3.1.12

prefix NAL unit

NAL units with type 14 as specified in ISO/IEC 14496-10

NOTE 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.

3.1.13

scalable layer; layer

set of VCL NAL units with the same values of *dependency_id*, *quality_id*, and *temporal_id*, and the associated non-VCL NAL units as specified in ISO/IEC 14496-10.

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

NOTE 2 SVC uses a “layered” encoder design which 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).

3.1.14

scalable layer representation

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

NOTE A scalable layer representation is also referred to as the representation of the scalable layer.

3.1.15

sub-picture

proper subset of coded slices of a layer representation

3.1.16

sub-picture tier

tier that consists of sub-pictures

NOTE 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.

3.1.17

SVC enhancement layer

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

NOTE 1 An SVC enhancement layer is represented by SVC VCL NAL units and the associated non-VCL NAL units and SEI messages.

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

3.1.18

SVC NAL unit

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

3.1.19

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 The term "SVC stream" is referenced by 'decoding/accessing the entire stream' in this document. There may be NAL units which are not required for decoding this operating point.

3.1.20

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 SVC VCL NAL units do not affect the decoding process of a legacy AVC decoder.

3.1.21

temporal layer representation representation of a temporal layer

temporal layer and all lower temporal layers

3.1.22

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 A tier represents one or more scalable layers of an SVC bitstream.

NOTE 2 The term "tier" is used 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 An MVC tier represents a particular set of temporal subsets of a particular set of views.

3.1.23

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

3.1.24**video elementary stream**

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

3.1.25**virtual base view**

AVC compatible representation of an independently coded non-base view

NOTE The virtual base view of an independently coded non-base view is created according to the process specified in H.8.5.5 of ISO/IEC 14496-10. 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

AVC	Advanced Video Coding. Where contrasted with SVC or MVC in this International Standard, this term refers to the main part of ISO/IEC 14496-10, including neither Annex G (Scalable Video Coding) nor Annex H (Multiview Video 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
MVC	Multiview Video Coding [refers to ISO/IEC 14496-10 when the techniques in Annex H (Multiview Video Coding) are in use]
NAL	Network Abstraction Layer
PPS	Picture Parameter Set
ROI	Region-Of-Interest
SEI	Supplementary Enhancement Information
SPS	Sequence Parameter Set
STSA	Step-wise Temporal Sub-layer Access
SVC	Scalable Video Coding [refers to ISO/IEC 14496-10 when the techniques in Annex G (Scalable Video Coding) are in use]
TSA	Temporal Sub-layer Access
VCL	Video Coding Layer
VPS	Video Parameter Set

4 General Definitions

4.1 Introduction

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

The following table summarizes the correspondences between the sets of terminology used in video specifications and the ISO Base Media File Format.

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 specification 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 Introduction

Sample: A sample is an access unit as defined in the appropriate specification.

Parameter set sample: A parameter set sample is a sample in a parameter set stream which 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.

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

- **Video data NAL units:** All video data NAL units for a single picture shall be contained with the sample whose decoding time and composition time are those of the picture. Each sample shall contain at least one video data NAL unit of the primary 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.
- 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 over-rides 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 may 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 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.

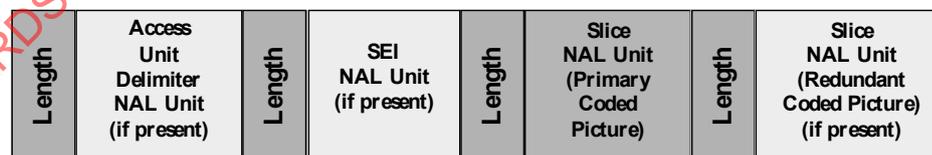


Figure 1 — Example structure of a sample

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 can 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 one byte NAL header and the Rbsp 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.

4.4 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 8.10.3 "Track Selection Box" of ISO/IEC 14496-12 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 which have default values but which 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 5.4.6 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` must correctly document the cropped frame dimensions (visual presentation size) 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. If the width and height of the sequence changes, then a new sample entry is needed.

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 Sync sample (IDR)

A sample is considered as a sync sample if ALL of the following conditions are met:

- The video data NAL units in the sample indicate that the primary 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 an IDR access unit, 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 an IDR access unit), 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 random access point 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.

4.9 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 5.4.6 is recommended.

4.10 Sample groups on random access recovery points and random access points

The video coding system can include the concept of a 'gradual decoding refresh' or random access recovery point. This may be signalled in the bit-stream 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 [4.5]. 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 may not 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 which can be decoded perfectly (either a sync sample, or the end of a pre-roll).

4.11 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.

5 AVC elementary streams and sample definitions

5.1 Introduction

The Advanced Video Coding (AVC) standard, jointly developed by the ITU-T and ISO/IEC JTC 1/SC 29/WG 11 (MPEG), 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, this part of ISO/IEC 14496 defines a storage format for video streams compressed using AVC.

This clause defines the storage for plain AVC streams, where 'plain AVC' refers to the main part of ISO/IEC 14496-10, excluding Annex G (Scalable Video Coding) and Annex H (Multiview Video Coding).

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.

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

- **Video Elementary Streams** shall contain 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. Other NAL units that are not expressly prohibited may be present, and if they are unrecognized should be ignored (e.g. not placed in the output buffer while accessing the file).
- **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.

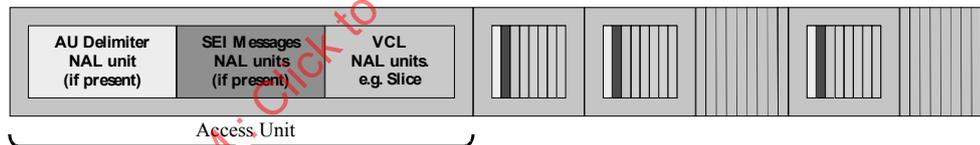
- **Video elementary stream possibly including parameter sets:** In this case, the sample entry indicates whether the stream may contain parameter sets of given types, in addition to other parameters provided in the sample entry. Sequence and picture parameter set NAL units may therefore be part of AVC 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 the following table.

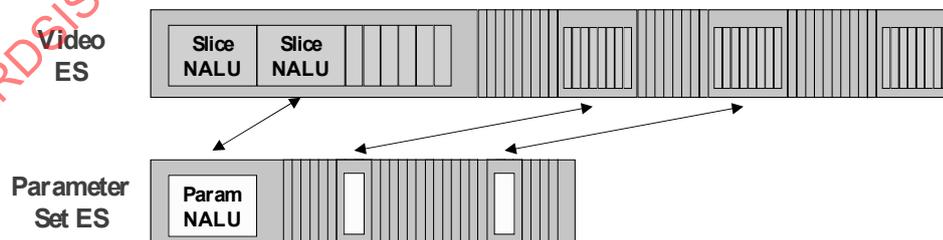
Table 2 – NAL Unit types in elementary Streams

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	Not specified by this part of ISO/IEC 14496	Not specified by this part of ISO/IEC 14496	Not specified by this part of ISO/IEC 14496
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
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 part of ISO/IEC 14496	Not specified by this part of ISO/IEC 14496	Not specified by this part of ISO/IEC 14496
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 part of ISO/IEC 14496	Not specified by this part of ISO/IEC 14496	Not specified by this part of ISO/IEC 14496
24 - 31	Unspecified	Not specified by this part of ISO/IEC 14496	Not specified by this part of ISO/IEC 14496	Not specified by this part of ISO/IEC 14496



(a) Single video elementary stream containing NAL units



(b) Synchronized video and parameter sets with arrows denoting synchronization between streams

Figure 2 — AVC elementary stream structure

5.3 Sample and Configuration definition

5.3.1 Introduction

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

AVC parameter set sample: An AVC parameter set sample is a sample in a parameter set stream which 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 with the sample whose decoding time and composition time are those of the picture.
- **Parameter sets:** If a parameter set elementary stream is used, then the sample in the parameter 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 –** 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 may 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 may 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 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 may 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, 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, 7.4.1.2.3.

- **Sub-sequence and layering SEI messages.** Sub-sequence or layering SEI messages shall not occur in the AVC elementary stream. Specifically, the sub-sequence information, sub-sequence layer characteristics, and sub-sequence characteristics SEI messages shall not occur in the stored AVC video elementary stream. Instead, all such information is stored as external metadata as described in 5.4.5.
- **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 Slice groups are stored in a canonical order to ease hinting, and to make it easier to find a primary 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 The removal or addition of Filler Data NAL units, start codes, SEI messages or Filler Data SEI messages may change the bit-stream characteristics with respect to conformance with the HRD when operating the HRD in CBR mode as specified in ISO/IEC 14496-10, Annex C.

5.3.3 Decoder configuration information

This subclause specifies the decoder configuration information for ISO/IEC 14496-10 video content.

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 must be supplied by the structure which contains it).

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

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

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 which indicate profile compatibility must be valid for all parameter sets of the stream described by this record. The level indication must 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 must 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 must be split into two or more sub-streams with separate configuration records in which these rules can be 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` must 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 specification 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 in profiles other than those indicated for the array specific to such NAL units (`profile_idc` not equal to any of 100, 110, 122, 144), they should be placed in the Sequence Parameter Set Array.

NOTE - the profile identified by `profile_idc` value 144 is deprecated in ISO/IEC 14496-10.

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 ( profile_idc == 100 || profile_idc == 110 ||
        profile_idc == 122 || profile_idc == 144 )
    {
        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 which 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 part of ISO/IEC 14496 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 part of ISO/IEC 14496 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

This subclause defines the sample entry and sample format for AVC video elementary streams.

5.4.2.1 Sample entry name and format

5.4.2.1.1 Definition

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 MPEG4BitRateBox 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 the ISO Base Media File Format specification, 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 usable 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 (see Annex A) may be present but must not be used to access the AVC base data; that is, the AVC data must 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 must be replaced with the data they are referencing, and Aggregators must be examined for contained NAL Units. Tier grouping may be present.

5.4.2.1.2 Syntax

```
// Visual Sequences
class AVCConfigurationBox extends Box('avcC') {
    AVCDecoderConfigurationRecord() AVCConfig;
}

class MPEG4BitRateBox extends Box('btrt') {
    unsigned int(32) bufferSizeDB;
    unsigned int(32) maxBitrate;
    unsigned int(32) avgBitrate;
}

class MPEG4ExtensionDescriptorsBox extends Box('m4ds') {
    Descriptor Descr[0 .. 255];
}

class AVCSampleEntry() extends VisualSampleEntry (type) {
    // type is 'avc1' or 'avc3'
    AVCConfigurationBox config;
    MPEG4BitRateBox (); // optional
    MPEG4ExtensionDescriptorsBox (); // optional
}

class AVC2SampleEntry() extends VisualSampleEntry (type) {
    // type is 'avc2' or 'avc4'
    AVCConfigurationBox avcconfig;
    MPEG4BitRateBox bitrate; // optional
    MPEG4ExtensionDescriptorsBox descr; // optional
    extra_boxes boxes; // 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 5.3.3. If a separate parameter set stream is used, `numOfSequenceParameterSets` and `numOfPictureParameterSets` must both be zero.

`Descr` is a descriptor which 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`.

`bufferSizeDB` gives the size of the decoding buffer for the elementary stream in bytes.

`maxBitrate` gives the maximum rate in bits/second over any window of one second.

`avgBitrate` gives the average rate in bits/second over the entire presentation.

5.4.3 AVC parameter set stream definition

5.4.3.1 Sample entry name and format

5.4.3.1.1 Definition

Box Types: 'avcp'
Container: Sample Description Box ('stsd')
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'){  
    AVCCConfigurationBox 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 5.3.3. `numOfSequenceParameterSets` and `numOfPictureParameterSets` must 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 It is recommended that when several parameter sets are used and parameter set updating is desired, a separate parameter set elementary stream be used.

NOTE 2 Decoders conforming to this specification 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 Layering and sub-sequences

5.4.5.1 Introduction

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 can 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 the Annex D of ISO/IEC 14496-10. 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.

5.4.5.2 Sub-sequence description entry

5.4.5.2.1 Definition

Group Types: '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 it is dependent. A sub-sequence can depend 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.

5.4.5.2.2 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(8)  accurateStatisticsFlag;
        unsigned int(16) avgBitRate;
        unsigned int(16) avgFrameRate;
    }
    unsigned int(8) numReferences;
    DependencyInfo dependency[numReferences];
}
}
```

5.4.5.2.3 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 which 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 which 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.

5.4.5.3 Layer description entry

5.4.5.3.1 Definition

Group Types: 'avll'
 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.

5.4.5.3.2 Syntax

```
class AVCLayerEntry() extends VisualSampleGroupEntry ('av11')
{
    unsigned int(8)    layerNumber;
    unsigned int(8)    accurateStatisticsFlag;
    unsigned int(16)   avgBitRate;
    unsigned int(16)   avgFrameRate;
}
```

5.4.5.3.3 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.

5.4.6 Alternate streams and switching pictures

5.4.6.1 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 can 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.

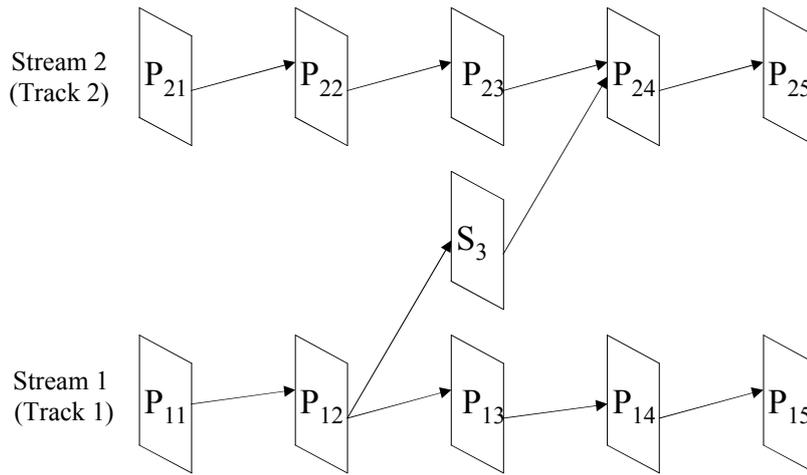


Figure 3 — Stream switching

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 must 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.

5.4.6.2 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 bit-stream can be stored as a separate track and related together as alternate tracks. All the tracks which form a group which may be switched between, but not the tracks containing the switching pictures, must 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.

5.4.6.3 Track references

The switching track must 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 must be linked to the track from which it switches (the source track) by a track reference of type 'swfr' in the switching picture track.

5.4.6.4 Sample dependency

5.4.6.4.1 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').

5.4.6.4.2 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;
    }
  }
}
```

5.4.6.4.3 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 which 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 which immediately follows the decoding time of the switching sample has relative sample number +1, the sample after that +2, and so on.

5.4.7 Definition of a sub-sample for AVC

For the use of the sub-sample information box (8.7.7 of ISO/IEC 14496-12) 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 can still be decoded if this sub-sample is discarded (e.g. the sub-sample consists of an SEI NAL unit, or a redundant coded picture).

The reserved field is defined for AVC as follows:

```
unsigned int(1) RefPicFlag;  
unsigned int(1) RedPicFlag;  
unsigned int(1) VclNalUnitFlag;  
unsigned int(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.

5.4.8 Post-decoder requirements scheme for signalling of SEI for AVC

5.4.8.1 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 AVC and for this case a file author can list occurring SEI message IDs [ISO/IEC 14496-10] 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.

The scheme for signalling of SEI for AVC is defined here.

The SchemeType 'aSEI' (AVC SEI) is used.

5.4.8.2 SEI Information box

5.4.8.2.1 Definition

Box Type: 'seii'
Container: Scheme Information box ('schi')
Mandatory: Yes (when the SchemeType is 'aSEI')
Quantity: One

The SEI Information box is contained in the Scheme Information box when the SchemeType is 'aSEI', and contains information about the SEI messages present in the AVC stream. 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 SEI which is required to be understood for correct playback and SEI which is not required for correct playback (but may enhance playback).

By inspecting the SEI Information box a player will know which SEI messages that occur in the bitstream and whether they are required for rendering.

When the coding system is AVC and the restriction is signaled with SEI messages, then the SEI messages listed here should be stored either in the bitstream or in the AVC Configuration Record. The SEI Information box does not contain the actual SEI messages, it only lists those that occur in the bitstream.

5.4.8.2.2 Syntax

```
aligned(8) class SeiInformationBox extends Box('sei') {
    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;
    }
}
```

5.4.8.2.3 Semantics

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

6 SVC elementary stream and sample definitions

6.1 Introduction

This clause specifies the storage format of SVC data. It extends the definitions of the storage format of AVC in clause 5.

The file format for storage of SVC content, as defined in this clause and Annexes Annex A to Annex 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 clause 5). In addition, the following new extensions, among others, to support SVC-specific features are specified.

- Scalable or multiview grouping:
a structuring and grouping mechanism to indicate the association of NAL units with different types and hierarchy levels of scalability.
- 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.
- Extractor:
a structure to enable efficient extraction of NAL units from other tracks than the one containing the media data.
- Temporal metadata statements:
structures for storing time-aligned information of media samples.
- 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:

- SVC Video Elementary Streams** shall contain 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. Also Aggregators (see A.2) or Extractors (see A.3) may be present. Aggregators and Extractors shall be processed as defined in this International Standard (e.g. shall not directly be placed in the output buffer while accessing the file). Other NAL units that are not expressly prohibited may be present, and if they are unrecognized they should be ignored (e.g. not placed in the output buffer while accessing the file).

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

For SVC streams, Table 2 is updated as follows; 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 – 29	Not specified	Not specified	Not specified	Not specified	Not specified
30	Aggregator	Not specified	Yes	Yes	No
31	Extractor	Not specified	Yes	Yes	No

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 Aggregator NAL units and Extractor NAL units 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 Annex G of ISO/IEC 14496-10 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 this International Standard.

Subclause 5.4.5 is defined for use with plain AVC streams. Its use with SVC streams is deprecated.

6.4 Sample and configuration definition

6.4.1 Introduction

SVC Sample: An SVC sample is also an access unit as defined in 7.4.1.2 of ISO/IEC 14496-10.

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 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 may 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 5.3.3.1 is used for a stream which 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.

Sequence 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;
    bit(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, must have values such that a conforming SVC decoder 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 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 be extractions, subsets, copies or re-orderings of the complete subset.

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

NOTE 2 “A scalable bitstream” may 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 must share the same timescale.

6.5.2 Data sharing and extraction

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

- a) The sample data is copied from one track into another track (and possibly compacted or re-interleaved with other data, such as audio). This creates larger overall files, but the low bit rate data may be compacted and/or interleaved with other material, for ease of extraction.
- b) 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

Types: 'avc2', 'avc4', 'avcC', 'svc1', 'svc2', 'svcC', 'seib'
 Container: Sample Description Box ('std')
 Mandatory: One of the avc1, avc2, avc3, avc4, svc1 and svc2 boxes is mandatory.
 Quantity: One or more sample entries may be present

If an SVC elementary stream contains a usable 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 Parameter Set information pertaining to the entire stream containing the SVC compatible enhancement layers as defined by the `SVCDecoderConfigurationRecord`, stored in the `SVCConfigurationBox`.

For all sample entries except for 'svc1' and 'svc2', i.e. 'avc1', 'avc2', 'avc3' and 'avc4', the width and height fields in the sample entry document the AVC base layer. For an 'svc1' or 'svc2' sample entry, the width and height document the resolution achieved by decoding the entire stream.

If the SVC elementary stream does not contain a usable 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 International Standard.

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. The 'extra_boxes' in an 'avc2' or 'avc4' sample entry may be an `SVCConfigurationBox`, `ScalabilityInformationSEIBox`, `SVCPriorityAssignmentBox` or other extension boxes.

NOTE When AVC compatibility is indicated, it may 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 may 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 `AVCSVCsSampleEntry` definition below applies.

`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).

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).

The following table 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 without SVC NAL units; 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 may be present; Extractors shall not reference AVC NAL units; Aggregators shall not contain but may reference AVC NAL units; Tier grouping may be present.
'avc2' or 'avc4'	AVC Configuration Only	A plain AVC track without SVC NAL units; 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.
'svc1' or 'svc2'	SVC Configuration	An SVC track without AVC NAL units; Extractors may be present and used to reference both AVC and SVC NAL units; Aggregators may be present to contain and reference both AVC and 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 () {
    SVCConfigurationBox svcconfig; // optional
    ScalabilityInformationSEIBox scalability; // optional
    SVCPriorityAssignmentBox method; // optional
}

```

```

class AVC2SVCSampleEntry() extends VisualSampleEntry (type){
    // type is 'avc2' or 'avc4'
    AVCConfigurationBox avcconfig;
    SVCConfigurationBox svcconfig; // optional
    MPEG4BitRateBox bitrate; // optional
    MPEG4ExtensionDescriptorsBox descr; // optional
    ScalabilityInformationSEIBox scalability; // optional
    SVCPriorityAssignmentBox method; // optional
}

// Use this if the track is NOT AVC compatible
class SVCSampleEntry() extends VisualSampleEntry (type){
    // type is 'svc1' or 'svc2'
    SVCConfigurationBox svcconfig;
    MPEG4BitRateBox bitrate; // optional
    MPEG4ExtensionDescriptorsBox descr; // optional
    ScalabilityInformationSEIBox scalability; // optional
    SVCPriorityAssignmentBox method; // optional
}

```

6.5.3.1.3 Semantics

scalinfosei contains an SEI NAL unit containing only a scalability information SEI message as specified in ISO/IEC 14496-10 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 must 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 is the visual width and height of the AVC base layer, if the stream is described by a sample entry of type 'avc1', 'avc2', 'avc3' or 'avc4'; otherwise it is the visual width and height of decoded pictures by decoding the entire stream.

6.5.5 Sync sample (IDR)

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 an IDR access unit, 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 an IDR access unit), 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 random access point 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 must 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 AVC is deprecated.

6.5.7 Independent and disposable samples box

If it is used in a track which is both AVC and SVC compatible, then care should be taken that the statements are 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 and random access points

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 random access recovery sample group identifies random access recovery in the entire SVC bitstream and the random access point sample group identifies random access points in 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 5.4.7.

For the use of the sub-sample information box (8.7.7 of ISO/IEC 14496-12) 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 5.4.7, the subsample_priority field shall be set to a value in accordance with the specification of this field in ISO/IEC 14496-12.

The reserved field is defined for SVC as follows:

```

unsigned int(1) RefPicFlag;
unsigned int(1) RedPicFlag;
unsigned int(1) VclNalUnitFlag;
unsigned int(5) reserved = 0;
unsigned int(1) reserved = 0;
unsigned int(1) IdrFlag;
unsigned int(6) PriorityId;
unsigned int(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;
unsigned int(1) reserved = 0; // corresponding to output_flag
unsigned int(1) StoreBaseRepFlag;
unsigned int(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 that this is not the same definition as the `discardable` field in the sub-sample information box.

StoreBaseRepFlag indicates the `store_base_rep_flag` value of the NAL units in the sub-sample. All the NAL units in the sub-sample shall have the same value of `store_base_rep_flag`.

7 MVC elementary stream and sample definitions

7.1 Introduction

This clause specifies the storage format of MVC data. It extends the definitions of the storage format of AVC in clause 5.

The file format for storage of MVC content, as defined in this clause and Annexes Annex A to Annex 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 clause 5). In addition, the following new extensions, among others, to support MVC-specific features are specified.

- Scalable or multiview grouping:
a structuring and grouping mechanism to indicate the association of NAL units with different types and hierarchy levels of scalability.
- 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.
- Extractor:
a structure to enable efficient extraction of NAL units from other tracks than the one containing the media data.
- Temporal metadata statements:
structures for storing time-aligned information of media samples.
- AVC compatibility:
a provision for storing an MVC bitstream in an AVC compatible manner, such that the AVC compatible base layer can be used by any plain AVC file format compliant reader.

The support for MVC includes a number of tools, and there are various 'models' of how they might be used. In particular, an MVC stream can be placed in tracks in a number of ways, among which are the following:

1. all the views in one track, labelled with sample groups;
2. each view in its own track, labelled in the sample entries;
3. a hybrid, one track containing all views, and one or more single-view tracks each containing a view that can be independently coded;
4. the expected operating points each in a track (e.g. the AVC base, a stereo pair, a multiview scene).

The MVC file format allows storage of one or more views into a track, similarly to the support for SVC in clause 6. Storage of multiple views per track can be used, e.g., when a content provider wants to provide a multiview bitstream that is not intended for subsetting or when the bitstream has been created for a few pre-defined sets of output views (such as 1, 2, 5, or 9 views) where tracks can be created accordingly. If more than one view is stored in a track and there are several tracks (more than one) representing the MVC bitstream, the use of the sample grouping mechanism is recommended. The sample grouping mechanism is used to define tiers identifying the views present in the track and to extract required NAL units for certain operation points conveniently. The sample grouping mechanism is usually used with aggregator NAL units to form regular NAL unit patterns within samples. Thus, SVC-like sample grouping, aggregators, and view definitions for sample groups are specified for MVC.

The Multiview Information box ('mvci') is specified to indicate information that applies to more than one view, such as the target output views in one or more Multiview Group boxes. Characteristics (such as camera parameters) of the respective bitstream subset can also be indicated within the Multiview Group box using the Multiview Relation Attributes box ('mvra'), which is similar to the Track Selection box.

A player should have means to determine which views are preferred for displaying, and select one or more tracks that provide the data for the desired operating point, preferring a track that is specific to that operating point over tracks that also contain other data. The display characteristics of players may differ; for example, the number of simultaneously displayed views and the optimal angle between views can be different. In order to guide a player for selection of output views, alternative groups of output views and the common and differentiating characteristics between them can be indicated with the Multiview Group Relation box ('swtc'), which also includes the Multiview Relation Attributes box ('mvra').

When an MVC bitstream is represented by multiple tracks and a player uses an operating point that contains data in multiple tracks, the player must reconstruct MVC access units before passing them to the MVC decoder. An MVC operating point may be explicitly represented by a track, i.e., an access unit is reconstructed simply by resolving all extractor and aggregator NAL units of a sample. If the number of operating points is large, it may be space-consuming and impractical to create a track for each operating point. In such a case, MVC access units are reconstructed as specified in 7.6.2. The MVC Decoder Configuration record contains a field indicating whether the associated samples use explicit or implicit access unit reconstruction (see the explicit_au_track field).

7.2 Overview of MVC Storage

The storage of MVC streams can be supported by a number of structures, including information in the sample entry, the media information box, and sample groups. The following table provides an overview of the structures provided, their names, and a brief description of their functions.

NOTE – each group of rows starting with an entry in the left column (e.g. 'minf', '?vc?') document a containment structure within that container; the higher level containment is not shown.

Table 5 – Box, sample entry and group structures for MVC Streams

			Box Name	Brief Description
minf			Media Information Box	
	mvci		Multiview Information Box	
		mvcg	Multiview Group Box	Specifies a multiview group for the views of the multiview video stream that are output
		buff	Buffering Information Box	Contains the buffering information of the bitstream subset specified by the multiview group
		mvra	Multiview Relation Attribute Box	Indicates the relation of the tracks or tiers of the respective multiview group with each other (when contained in a Multiview Group box)
		tibr	Tier Bit Rate Box	Provides information about the bit rate values of the bitstream subset specified by the multiview group
		tiri	Tier Information Box	Provides information about the profile, level, frame size, discardability, and frame-rate of the bitstream subset specified by the multiview group
		vwdi	Multiview Scene Information Box	Indicates the maximum disparity in a scene with multiple views
		swtc	Multiview Group Relation Box	Specifies a set of multiview groups from which one multiview group is decoded and played at any time
		mvra	Multiview Relation Attribute Box	Indicates the relation of the multiview groups with each other (when contained in a Multiview Group Relation box)
?vc?			Sample Entry	(Note: various codes are used for sample entries)

	ecam		Extrinsic Camera Parameters Box	Contains camera parameters that define the location and orientation of the camera reference frame with respect to a known world reference frame
	icam		Intrinsic Camera Parameters Box	Contains camera parameters that link the pixel coordinates of an image point with the corresponding coordinates in the camera reference frame
	vwid		View Identifier Box	Indicates the views included in the track (when included in a sample entry)
	mvcP		MVC View Priority Assignment Box	Provides a URI containing a unique name of the method used to assign content_priority_id values for the View Priority sample grouping
	mvcC		MVC Configuration Box	
sgpd			Sample Group Description Box	
	mvif		Multiview Group Entry	Contains the following boxes
	buff		Buffering Information Box	Contains the buffer information of the tier
	ldep		Tier Dependency Box	Identifies the tiers that the current tier is dependent on
	svip		Initial Parameter Sets Box	Contains parameter sets needed for decoding this tier and all the tiers it depends on
	svpr		Priority Range Box	Reports the minimum and maximum priority_id of the NAL units mapped to this tier
	tibr		Tier Bit Rate Box	Provides information about the bit rate values of a tier
	tiri		Tier Information Box	Provides information about the profile, level, frame size, discardability, and frame-rate of a tier
	vipr		View Priority Box	Labels views with priorities based on content
	vwid		View Identifier Box	Indicates the views included in the tier (when included in a Multiview Group entry.)
	dtrt		Decode Re-timing Group Entry	Provides adjusted decoding times when high temporal layers are discarded
	scnm		Sample Map Group Entry	Provides the mapping of NAL units to multiview groups for all samples in the track

The structures within a sample entry provide information for the decoding or use of the samples (video information) that are associated with that sample entry. Sample groups provide time-varying information about the track as a whole, assisting (for example) with the extraction of subsets of the media within a track. Information in the Multiview Information Box (appearing in the media information box) can span several tracks and is descriptive of collections of tracks, even though the Multiview Information Box resides in the track containing the base view of the stream.

7.3 MVC Track Structure

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

- **MVC Video Elementary Streams** shall contain 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. Also Aggregators (see A.2) or Extractors (see A.3) may be present. Aggregators and Extractors shall be processed as defined in this International Standard (e.g. shall not directly be placed in the output buffer while accessing the file). Other NAL units that are not expressly prohibited may be present, and if they are unrecognized they should be ignored (e.g. not placed in the output buffer while accessing the file).

MVC streams may also be stored using associated parameter set streams, when needed.

For MVC streams, Table 2 is amended by the following table.

Table 6 – NAL Unit types in MVC and AVC Streams

Value of nal_unit_type	Description	AVC video elementary stream	MVC video elementary stream (sample entry 'avc1', 'avc2', 'mvc1', or 'mvc2')	MVC video elementary stream (sample entry name 'avc3', 'avc4', 'mvc3', or 'mvc4')	Parameter set elementary stream
14	Prefix NAL unit prefix_nal_unit_rbsp()	Not specified	Yes	Yes	No
15	Subset sequence parameter set subset_seq_parameter_set_rbsp()	Not specified	No	Yes Parameter set elementary stream shall not be used	Yes
20	Coded slice extension slice_layer_extension_rbsp()	Not specified	Yes	Yes	No
24 – 29	Not specified	Not specified	Not specified	Not specified	Not specified
30	Aggregator	Not specified	Yes	Yes	No
31	Extractor	Not specified	Yes	Yes	No

There may be AVC VCL NAL units, MVC VCL NAL units and other NAL units, i.e. non-VCL NAL units, present in an MVC video elementary stream. Additionally, there may be Aggregator or Extractor NAL units present in an MVC video elementary stream.

An AVC VCL NAL unit in an MVC video elementary stream conforming to one or more profiles specified in Annex H of ISO/IEC 14496-10 shall be immediately preceded by a prefix 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.

7.4 Use of the plain AVC File Format

The MVC file format is an extension of the plain AVC file format defined in this International Standard.

Clause 5.4.5 is defined for use with plain AVC streams. Its use with MVC streams is deprecated.

7.5 Sample and configuration definition

7.5.1 Introduction

MVC Sample: An MVC sample consists of one or more view components as defined in Annex H of ISO/IEC 14496-10 and the associated non-VCL NAL units.

7.5.2 Canonical Order and Restriction

7.5.2.1 Restrictions

The following restrictions apply to MVC data in addition to the requirements in clause 5.3.2.

- **MVC coded slice NAL units** (Coded slice extension): All MVC 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 MVC sample shall contain at least one AVC or MVC VCL NAL unit.

- **Prefix NAL units:** Each prefix NAL unit is placed immediately before the corresponding AVC VCL NAL unit.
- **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.

7.5.2.2 Decoder Configuration Record

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

A parameter set stream may be used with MVC streams, as with AVC streams. In that case, parameter sets shall not be included in the decoder configuration record.

Sequence parameter sets, including subset sequence 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 `MVCDecoderConfigurationRecord` 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 MVCDecoderConfigurationRecord {
    unsigned int(8) configurationVersion = 1;
    unsigned int(8) AVCProfileIndication;
    unsigned int(8) profile_compatibility;
    unsigned int(8) AVCLevelIndication;
    bit(1) complete_representation;
    bit(1) explicit_au_track;
    bit(4) reserved = '1111'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 for the bitstream represented by this track, i.e., the bitstream that contains all the views of this track and the views required for decoding of this track and wherein all the views in this track are the target output views. If `AVCLevelIndication` is equal to 0, the level that applies to the bitstream defined above operating with all the views of this track being the target output views is unspecified. `AVCProfileIndication`, `profile_compatibility`, and `AVCLevelIndication`, if non-zero, must have values such that a conforming MVC decoder 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, if not present in the following, are as defined for an AVCDecoderConfigurationRecord:

`complete_representation` is set on a minimal set of tracks that contain a portion of the original encoded stream, as defined in 7.6.1. 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.

`explicit_au_track` is set on a track that is 'complete'; it is not necessary to determine the view dependencies, nor calculate whether views not present in this track must be found from other tracks. However, subject to the rules for the sample entry types, extractors may be present and need to be followed to gather all the NAL units needed.

`numOfSequenceParameterSets` indicates the number of SPSs and subset SPSs that are used for decoding the MVC elementary stream.

`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.

7.5.3 Sync Sample (IDR)

A sync sample identifies the presence of an IDR access unit of the MVC bitstream for any sample entry that includes an MVC configuration record.

For video data described by a sample entry of type 'mvc3' or 'mvc4', the following applies:

1. If the sample is an IDR access unit, 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 an IDR access unit), 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 random access point to the sample itself, inclusive.

7.5.4 Shadow sync

A shadow sync box shall not be used for video data described by any MVC sample entry. Its use for AVC is deprecated.

7.5.5 Independent and disposable samples box

If it is used in a track which is both AVC and MVC compatible, then care should be taken that the statements are true no matter what valid subset of the MVC 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.

7.5.6 Sample groups on random access recovery points and random access points

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 MVC decoder (if any) operating on the entire bitstream.

For video data described by an MVC sample entry type, the random access recovery sample group identifies random access recovery in the entire MVC bitstream and the random access point sample group identifies random access points in the entire MVC bitstream.

7.6 Derivation from the ISO base media file format

7.6.1 MVC track structure

A multi-view video stream is represented by one or more video tracks in a file. Each track represents one or more views of the stream.

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".

The track that has the flag "complete_representation" set and contains NAL units of the base view with temporal_id equal to 0 shall be nominated as the 'base view track'. All the other tracks that are part of the same stream shall be linked to this base track by means of a track reference of type 'sbas' (view base). The complete encoded information can be retained when the tracks included in the "complete subset" are retained; all other tracks shall be subsets, copies or re-orderings of the complete subset.

All the tracks sharing the same base view track must share the same timescale.

If a view represented by a track uses another view represented by another track as an inter-view prediction reference, a track reference of type 'scal' shall be included in the track referring to the source track for inter-view prediction.

If edits are applied to tracks that contain view components of an MVC bitstream, edit lists shall be consistent over all tracks affected by the edits.

NOTE If a track containing a part of an MVC bitstream is removed from a file, care should be taken to remove also those tracks that contain 'scal' and 'sbas' track references to the removed track and references to the multiview groups that include the removed track.

7.6.2 Reconstruction of an access unit

In order to reconstruct an access unit from samples of one or more MVC tracks, the target output views may need to be determined first, by examining the Multiview Group box (7.7.3) and the Multiview Group Relation box (7.7.4). The explicit_au_track flag states that this track is a complete operating point; nonetheless, the track should be examined to determine which views delivered by this track are the output views.

If the target output views are not exactly represented by any track marked with explicit_au_track equal to 1 in the MVC decoder configuration record, access units are reconstructed as follows.

The views that are required for decoding the determined target output views can be concluded from reference view identifiers included in the View Identifier box, the 'scal' track references, or Tier Dependency boxes.

If several tracks contain data for the access unit, the alignment of respective samples in tracks is performed on decoding time, i.e. using the time-to-sample table only without considering edit lists.

An access unit is reconstructed from the respective samples in the required tracks and tiers by arranging their NAL units in an order conforming to ISO/IEC 14496-10. The following order provides an outline of the procedure to form a conforming access unit:

- All parameter set NAL units (from the associated parameter set tracks and from the associated elementary stream tracks).
- All SEI NAL units (from the associated parameter set tracks and from the associated elementary stream tracks).
- View components in ascending order of view order index value. NAL units within a view component are in their appearance order within the sample.

7.6.3 Sample Entry

7.6.3.1 Boxes for Sample Entry

7.6.3.1.1 Intrinsic Camera Parameters Box

7.6.3.1.1.1 Definition

Box Type: 'icam'
Container: Sample Entry ('avc1', 'avc2', 'avc3', 'avc4',
'mvc1', 'mvc2', 'mvc3', 'mvc4')
Mandatory: No
Quantity: Zero or more

This subclause specifies intrinsic camera parameters that link the pixel coordinates of an image point with the corresponding coordinates in the camera reference frame. A specification of focal length and parameters related to the geometric distortion due to camera optics is given in Annex H of ISO/IEC 14496-10.

7.6.3.1.1.2 Syntax

```
class IntrinsicCameraParametersBox extends FullBox ('icam', version=0, flags) {
    unsigned int(6)    reserved=0;
    unsigned int(10)   ref_view_id;
    unsigned int(32)   prec_focal_length;
    unsigned int(32)   prec_principal_point;
    unsigned int(32)   prec_skew_factor;
    unsigned int(8)    exponent_focal_length_x;
    signed   int(64)   mantissa_focal_length_x;
    unsigned int(8)    exponent_focal_length_y;
    signed   int(64)   mantissa_focal_length_y;
    unsigned int(8)    exponent_principal_point_x;
    signed   int(64)   mantissa_principal_point_x;
    unsigned int(8)    exponent_principal_point_y;
    signed   int(64)   mantissa_principal_point_y;
    unsigned int(8)    exponent_skew_factor;
    signed   int(64)   mantissa_skew_factor;
}
```

7.6.3.1.1.3 Semantics

reserved this field shall be equal to zero

ref_view_id indicates the view_id identifying a view for which intrinsic camera parameters are indicated in this Intrinsic Camera Parameters Box

prec_focal_length specifies the exponent of the maximum allowable truncation error for focal_length_x and focal_length_y as given by $2^{-\text{prec_focal_length}}$. The value of prec_focal_length shall be in the range of 0 to 31, inclusive.

prec_principal_point specifies the exponent of the maximum allowable truncation error for principal_point_x and principal_point_y as given by $2^{-\text{prec_principal_point}}$. The value of prec_principal_point shall be in the range of 0 to 31, inclusive.

prec_skew_factor specifies the exponent of the maximum allowable truncation error for skew factor as given by $2^{-\text{prec_skew_factor}}$. The value of prec_skew_factor shall be in the range of 0 to 31, inclusive.

exponent_focal_length_x specifies the exponent part of the focal length in the horizontal direction. The value of exponent_focal_length_x shall be in the range of 0 to 62, inclusive. The value 63 is reserved for future use by ITU-T | ISO/IEC. Decoders shall treat the value 63 as indicating an unspecified focal length.

mantissa_focal_length_x specifies the mantissa part of the focal length of the i-th camera in the horizontal direction.

exponent_focal_length_y specifies the exponent part of the focal length in the vertical direction. The value of exponent_focal_length_y shall be in the range of 0 to 62, inclusive. The value 63 is reserved for future use by ITU-T | ISO/IEC. Decoders shall treat the value 63 as indicating an unspecified focal length.

mantissa_focal_length_y specifies the mantissa part of the focal length in the vertical direction.

mantissa_principal_point_x specifies the mantissa part of the principal point in the horizontal direction.

exponent_principal_point_y specifies the exponent part of the principal point in the vertical direction. The value of exponent_principal_point_y shall be in the range of 0 to 62, inclusive. The value 63 is reserved for future use by ITU-T | ISO/IEC. Decoders shall treat the value 63 as indicating an unspecified principal point.

`mantissa_principal_point_y` specifies the mantissa part of the principal point in the vertical direction.

`exponent_skew_factor` specifies the exponent part of the skew factor. The value of `exponent_skew_factor` shall be in the range of 0 to 62, inclusive. The value 63 is reserved for future use by ITU-T | ISO/IEC. Decoders shall treat the value 63 as indicating an unspecified skew factor.

`mantissa_skew_factor` specifies the mantissa part of the skew factor.

The intrinsic matrix A for the camera associated to the view indicated by `ref_view_id` is represented as follows:

$$\begin{bmatrix} \text{focalLengthX} & \text{skewFactor} & \text{principalPointX} \\ 0 & \text{focalLengthY} & \text{principalPointY} \\ 0 & 0 & 1 \end{bmatrix}$$

Each component of the intrinsic matrix is obtained from the variables specified in Table 7 as the variable x computed as follows.

- If $0 < e < 63$, $x = 2^{e-31} * (1 + n \div 2v)$, with $v = \max(0, e + p - 31)$ [Eq. F-1]
- If e is equal to 0, $x = 2^{-(30+v)} * n$, with $v = \max(0, p - 30)$ [Eq. F-2]

Table 7 – Association between camera parameter variables and syntax elements

x	e	n	p
<code>focalLengthX</code>	<code>exponent_focal_length_x</code>	<code>mantissa_focal_length_x</code>	<code>prec_focal_length</code>
<code>focalLengthY</code>	<code>exponent_focal_length_y</code>	<code>mantissa_focal_length_y</code>	<code>prec_focal_length</code>
<code>principalPointX</code>	<code>exponent_principal_point_x</code>	<code>mantissa_principal_point_x</code>	<code>prec_principal_point</code>
<code>principalPointY</code>	<code>exponent_principal_point_y</code>	<code>mantissa_principal_point_y</code>	<code>prec_principal_point</code>
<code>skewFactor</code>	<code>exponent_skew_factor</code>	<code>mantissa_skew_factor</code>	<code>prec_skew_factor</code>

7.6.3.1.2 Extrinsic Camera Parameters Box

7.6.3.1.2.1 Definition

Box Type: `'ecam'`
 Container: Sample Entry (`'avc1'`, `'avc2'`, `'avc3'`, `'avc4'`, `'mvc1'`, `'mvc2'`, `'mvc3'`, `'mvc4'`,)
 Mandatory: No
 Quantity: Zero or more

This subclause specifies extrinsic camera parameters that define the location and orientation of the camera reference frame with respect to a known world reference frame. A specification of extrinsic camera parameters including translation vector and rotation matrix is given in Annex H of ISO/IEC 14496-10.

The extrinsic camera parameters are specified according to a right-handed coordinate system, where the upper left corner of the image is the origin, i.e., the (0, 0) coordinate, with the other corners of the image having non-negative coordinates. With these specifications, a 3-dimensional world point, $wP = [x \ y \ z]$ is mapped to a 2-dimensional camera point, $cP = [u \ v \ 1]$, according to:

$$s * cP = A * R^{-1} * (wP - T)$$

where A denotes the intrinsic camera parameter matrix which can be indicated by an intrinsic camera parameters box (see 7.6.3.1.1), R^{-1} denotes the inverse of the rotation matrix R, T denotes the translation vector, and s (a scalar value) is an arbitrary scale factor chosen to make the third coordinate of cP equal to 1. The elements of A, R, T are determined according the syntax elements signalled in this box and as specified below.

7.6.3.1.2.2 Syntax

```
class ExtrinsicCameraParametersBox extends FullBox ('ecam', version=0, flags) {
    unsigned int(6)    reserved=0;
    unsigned int(10)   ref_view_id;
    unsigned int(8)    prec_rotation_param;
    unsigned int(8)    prec_translation_param;
    for (j=1; j<=3; j++) { /* row */
        for (k=1; k<=3; k++) { /* column */
            unsigned int(8)    exponent_r[j][k];
            signed    int(64)   mantissa_r [j][k];
        }
        unsigned int(8)    exponent_t[j];
        signed    int(64)   mantissa_t[j];
    }
}
```

7.6.3.1.2.3 Semantics

reserved this field shall be equal to zero

ref_view_id indicates the view_id identifying a view for which intrinsic camera parameters are indicated in this Intrinsic Camera Parameters Box

prec_rotation_param specifies the exponent of the maximum allowable truncation error for $r[j][k]$ as given by $2^{-\text{prec_rotation_param}}$. The value of prec_rotation_param shall be in the range of 0 to 31, inclusive.

prec_translation_param specifies the exponent of the maximum allowable truncation error for $t[j]$ as given by $2^{-\text{prec_translation_param}}$. The value of prec_translation_param shall be in the range of 0 to 31, inclusive.

exponent_r[j][k] specifies the exponent part of (j, k) component of the rotation matrix. The value of exponent_r[j][k] shall be in the range of 0 to 62, inclusive. The value 63 is reserved for future use by ITU-T | ISO/IEC. Decoders shall treat the value 63 as indicating an unspecified rotation matrix.

mantissa_r[j][k] specifies the mantissa part of (j, k) component of the rotation matrix.

exponent_t[j] specifies the exponent part of the j-th component of the translation vector. The value of exponent_t[j] shall be in the range of 0 to 62, inclusive. The value 63 is reserved for future use by ITU T | ISO/IEC. Decoders shall treat the value 63 as indicating an unspecified translation vector.

mantissa_t[j] specifies the mantissa part of the j-th component of the translation vector.

The rotation matrix R is represented as follows:

$$\begin{bmatrix} rE[0][0] & rE[0][1] & rE[0][2] \\ rE[1][0] & rE[1][1] & rE[1][2] \\ rE[2][0] & rE[2][1] & rE[2][2] \end{bmatrix}$$

The translation vector T is represented as follows:

$$\begin{bmatrix} tE[0] \\ tE[1] \\ tE[2] \end{bmatrix}$$

Each component of the rotation matrix and the translation vector is obtained from the variables specified in Table 8 as the variable x computed as follows.

- If $0 < e < 63$, $x = 2^{e-31} * (1 + n \div 2^v)$, with $v = \max(0, e + p - 31)$ [Eq. F-3]

- If e is equal to 0, $x = 2^{-(30+v)} * n$, with $v = \max(0, p - 30)$ [Eq. F-4]

Table 8 - Association between camera parameter variables and syntax elements

x	e	n	p
rE[j][k]	exponent_r[j][k]	mantissa_r[j][k]	prec_rotation_param
tE[j]	exponent_t[j]	mantissa_t[j]	prec_translation_param

7.6.3.1.3 View Identifier Box

7.6.3.1.3.1 Definition

Box Type: 'vwid'

Container: Sample Entry ('avc1', 'avc2', 'avc3', 'avc4',
'mvc1', 'mvc2', 'mvc3', 'mvc4',)
or MultiviewGroupEntry

Mandatory: Yes (for sample entries and the primary group definition in Multiview Group entries)

Quantity: Exactly one (for sample entries and the primary group definition in Multiview Group entries)

Zero for non-primary group definitions in Multiview Group entries

When included in a sample entry, this box indicates the views included in the track. When included in a Multiview Group entry, this box indicates the views included in the respective tier. This box also indicates the view order index for each listed view. Additionally, the box includes the minimum and maximum values of temporal_id included in the track or tier when the View Identifier box is included in a sample entry or Multiview Group entry, respectively. Moreover, the box indicates the referenced views required for decoding the views included in the track or tier.

7.6.3.1.3.2 Syntax

```
class ViewIdentifierBox extends FullBox ('vwid', version=0, flags)
{
    unsigned int(2) reserved6 = 0;
    unsigned int(3) min_temporal_id;
    unsigned int(3) max_temporal_id;
    unsigned int(16) num_views;
    for (i=0; i<num_views; i++) {
        unsigned int(6) reserved1 = 0;
        unsigned int(10) view_id[i];
        unsigned int(6) reserved2 = 0;
        unsigned int(10) view_order_index;
        unsigned int(4) reserved3 = 0;
        unsigned int(2) base_view_type;
        unsigned int(10) num_ref_views;
        for (j = 0; j < num_ref_views; j++) {
            unsigned int(6) reserved5 = 0;
            unsigned int(10) ref_view_id[i][j];
        }
    }
}
```

7.6.3.1.3.3 Semantics

`min_temporal_id`, `max_temporal_id` take the minimum and maximum value, respectively, of the `temporal_id` syntax element that is present in the NAL unit header extension of the NAL units mapped to the track or tier when the View Identifier box is included in a sample entry or Multiview Group entry, respectively. For AVC streams this takes the value that is, or would be, in the prefix NAL unit.

`num_views`, when the View Identifier box is present in a sample entry, indicates the number of views included in the track. When the View Identifier box is present in a Multiview Group entry, `num_views` indicates the number of views included in the respective tier.

`view_id[i]` indicates the value of the `view_id` syntax element in the NAL unit header extension of a view included in the track or tier when the View Identifier box is included in a sample entry or Multiview Group entry, respectively.

`view_order_index` indicates the value of the `VOIdx` variable, as specified in Annex H of ISO/IEC 14496-10, for a view included in the track or tier when the View Identifier box is included in a sample entry or Multiview Group entry, respectively.

`base_view_type` indicates whether the view is a base view (virtual or not). It takes the following values:

0 indicates that the view is neither a base view nor virtual base view.

1 shall be used to label the non-virtual base view of the MVC bitstream.

2 is a reserved value and shall not be used.

3 indicates that the view with `view_id[i]` is a virtual base view. The respective independently coded non-base view with `view_id[i]` resides in another track. When `base_view_type` is equal to 3, the subsequent `num_ref_views` shall be equal to 0.

`num_ref_views` indicates the number of views which are referenced by the view with `view_id[i]`.

`ref_view_id[i][j]` indicates the view identifier of the `j`-th view that may be directly or indirectly referenced by the view with `view_id[i]`, i.e., that may be required for decoding of the view with `view_id[i]`. If a view is required for decoding the view with `view_id[i]`, it shall be listed as one of `ref_view_id[i][j]`. When the View Identifier box is included in a sample entry, it is recommended

to indicate the referenced views for both anchor and non-anchor access units in the same sample entry.

7.6.3.2 Sample Entry Definition

Sample Entry Types: 'avc1', 'avc2', 'avc3', 'avc4', 'mvc1', 'mvc2', 'mvc3', 'mvc4',
 Container: Sample Description Box ('stsd')
 Mandatory: One of the avc1, avc2, avc3, avc4, mvc1, mvc2, mvc3, and mvc4 boxes is mandatory
 Quantity: One or more sample entries may be present

If an MVC elementary stream contains a usable AVC compatible base view, 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 MVC Configuration Box as defined below. The AVC Configuration Box documents the Profile, Level and Parameter Set information pertaining to the AVC compatible base view as defined by the `AVCDecoderConfigurationRecord`. The MVC Configuration Box documents the Profile, Level and Parameter Set information pertaining to the entire stream containing the non-base views as defined by the `MVCDecoderConfigurationRecord`, stored in the `MVCConfigurationBox`.

For all sample entries 'avc1', 'avc2', 'avc3' and 'avc4', the width and height fields in the sample entry document the AVC base layer. For an 'mvc' sample entry ('mvc1', 'mvc2', 'mvc3', or 'mvc4'), the width and height document the resolution achieved by decoding any single view of the entire stream.

If the MVC elementary stream does not contain a usable AVC base view, then an MVC visual sample entry ('mvc1', 'mvc2', 'mvc3', or 'mvc4') shall be used. The MVC visual sample entry shall contain an MVC Configuration Box, as defined below. This includes an `MVCDecoderConfigurationRecord`, as defined in this International Standard.

The `lengthSizeMinusOne` field in the MVC 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 MVC 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.

The requirements for the sample entry types 'avc1' and 'avc2' as documented in 6.5.3.1.1 also apply here.

Any of the boxes `MVCConfigurationBox`, `ViewScalabilityInfoSEIBox`, `IntrinsicCameraParametersBox`, and `ExtrinsicCameraParametersBox` may be present in an 'avc1', 'avc2', 'avc3', or 'avc4' sample entry. In these cases, the `AVCMVCSampleEntry` or `AVC2MVCSampleEntry` definition below applies, respectively. The `ScalabilityInformationSEIBox` is defined in 6.5.3.1.2.

`Compressorname` in the base class `VisualSampleEntry` indicates the name of the compressor used, with the value "\012MVC Coding" being recommended (\012 is 10, the length of the string "MVC coding" in bytes).

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).

The following table shows for a video track all the possible uses of sample entries when an MVC elementary stream is stored in one or more tracks, configurations, and the MVC tools (excluding timed metadata, which is always used in another track).

Table 9 – Use of sample entries for AVC and MVC tracks

sample entry name	with configuration records	Meaning
'avc1' or 'avc3'	AVC and MVC Configurations	An MVC track with both AVC and MVC NAL units; Aggregators and Extractors may be present; Aggregators shall not contain but may reference AVC NAL units; Tier grouping may be present.
'avc2' or 'avc4'	AVC and MVC Configurations	An MVC track with both AVC NAL units and MVC NAL units; Extractors may be present and used to reference both AVC and MVC NAL units; Aggregators may be present to contain and reference both AVC and MVC NAL units; Tier grouping may be present.
'mvc1' or 'mvc3'	MVC Configuration	An MVC track without a usable AVC base layer; Aggregators may be present to contain and reference both AVC and MVC NAL units; Tier grouping may be present.
'mvc2' or 'mvc4'	MVC Configuration	An MVC track without a usable AVC base layer; Extractors may be present and used to reference MVC NAL units; Aggregators may be present to contain and reference MVC NAL units; Tier grouping may be present.

The sample entry `mvc-type` in the following is one of {`mvc1`, `mvc2`, `mvc3`, `mvc4`}.

7.6.3.3 Syntax

```
class MVCConfigurationBox extends Box('mvcC') {
    MVCCodecConfigurationRecord() MVCConfig;
}

class ViewScalabilityInformationSEIBox extends Box('vsib', size)
{
    unsigned int(8*size-64) viewscalinfosei;
}

class AVCMVCSEISampleEntry() extends AVCSampleEntry (type){
    // type is 'avc1' or 'avc3'
    ViewScalabilityInformationSEIBox scalability; // optional
    ViewIdentifierBox view_identifiers; // optional
    MVCConfigurationBox mvcconfig; // optional
    MVCViewPriorityAssignmentBox view_priority_method; // optional
    IntrinsicCameraParametersBox intrinsic_camera_params; // optional
    ExtrinsicCameraParametersBox extrinsic_camera_params; // optional
}
```

```

class AVC2MVCSampleEntry() extends AVC2SampleEntry (type){
    // type is 'avc2' or 'avc4'
    ViewScalabilityInformationSEIBox scalability; // optional
    ViewIdentifierBox view_identifiers; // optional
    MVCConfigurationBox mvccconfig; // optional MVCViewPriorityAssignmentBox
    view_priority_method; // optional
    IntrinsicCameraParametersBox intrinsic_camera_params; // optional
    ExtrinsicCameraParametersBox extrinsic_camera_params // optional
}

// Use this if the track is NOT AVC compatible
class MVCSampleEntry() extends VisualSampleEntry (mvc-type){
    MVCConfigurationBox mvccconfig; // mandatory
    ViewScalabilityInformationSEIBox scalability; // optional
    ViewIdentifierBox view_identifiers; // mandatory
    MPEG4BitRateBox bitrate; // optional
    MPEG4ExtensionDescriptorsBox descr; // optional
    MVCViewPriorityAssignmentBox view_priority_method; // optional
    IntrinsicCameraParametersBox intrinsic_camera_params; // optional
    ExtrinsicCameraParametersBox extrinsic_camera_params // optional
}

```

7.6.3.4 Semantics

viewscalinfosei contains an SEI NAL unit containing only a view scalability information SEI message as specified in ISO/IEC 14496-10 Annex H. The 'size' field of the container box ViewScalabilityInformationSEIBox shall not be equal to 0 or 1.

7.7 MVC specific information boxes

7.7.1 Introduction

The following boxes specify information that relate to more than one output view of an MVC elementary stream. As any subset of views of an MVC elementary stream can be chosen for output, the information carried in these boxes is not necessarily specific to any track and thus contained separately. The information can be specified for different groups of output views.

7.7.2 Multiview Information Box

7.7.2.1 Definition

Box Type: 'mvci'
 Container: Media Information Box ('minf')
 Mandatory: No
 Quantity: Zero or one

Located in the Media Information Box of the base view track indicated by the 'sbas' track reference, this box contains Multiview Group boxes, and Multiview Group Relation boxes.

7.7.2.2 Syntax

```

aligned(8) class MultiviewInformationBox
    extends FullBox('mvci', version = 0, flags) {
}

```

7.7.3 Multiview Group Box

7.7.3.1 Definition

Box Type: 'mvcg'
 Container: Multiview Information box ('mvci')
 Mandatory: No
 Quantity: Zero or more

This box specifies a multiview group for the views of the multiview video stream that are output. Target output views can be indicated on the basis of track_id, tier_id, or view_id. When the views included in a track match an operating point, it is recommended to use track_id (i.e., entry_type equal to 0) within the Multiview Group box. When multiview sample grouping is in use, and tiers cover more than one view or some tiers contain a temporal subset of the bitstream, it is recommended to use tier_id (i.e., entry_type equal to 1) within the Multiview Group box. Otherwise, it is recommended to use one of the view_id based indications (i.e., entry_type equal to 2 or 3).

Each view in a track or tier that is included in this box is a target output view. If a track or tier included in this box contains multiple views, all the contained views are target output views.

Decoding of the output views may require decoding of other views that are not target output views. The views that are required for decoding but are not target output views can be concluded from reference view identifiers included in the View Identifier box, the 'scal' track references, or from the Tier Dependency box.

If the box contains a track_id or tier_id that is not present or refers to a view_id of a view that is not present, the respective view should be considered removed and the multiview group should be ignored.

7.7.3.2 Syntax

```
aligned(8) class MultiviewGroupBox extends FullBox('mvcg', version = 0, flags) {
    unsigned int(32) multiview_group_id;
    unsigned int(16) num_entries;
    unsigned int(8) reserved = 0;
    for(i=0; i<num_entries; i++) {
        unsigned int(8) entry_type;
        if (entry_type == 0)
            unsigned int(32) track_id;
        else if (entry_type == 1) {
            unsigned int(32) track_id;
            unsigned int(16) tier_id;
        }
        else if (entry_type == 2) {
            unsigned int(6) reserved1 = 0;
            unsigned int(10) output_view_id;
        }
        else if (entry_type == 3) {
            unsigned int(6) reserved2 = 0;
            unsigned int(10) start_view_id;
            unsigned int(16) view_count;
        }
    }
    TierInfoBox subset_stream_info; // optional
    MultiviewRelationAttributeBox relation_attributes; // optional
    TierBitRateBox subset_stream_bit_rate; // optional
    BufferingBox subset_stream_buffering; // optional
    MultiviewSceneInfoBox multiview_scene_info; // optional
}
```

7.7.3.3 Semantics

`multiview_group_id` provides a unique identifier for the multiview group within the file.

`num_entries` is the number of tracks and tiers included in this multiview group.

`entry_type` specifies how the target output views are indicated. The following values of `entry_type` are specified:

- 0 – all the views included in an indicated track are target output views
- 1 – the view(s) of an indicated tier within an indicated track are target output views
- 2 – the views with `view_id` equal to `output_view_id` are target output views
- 3 – the views having `view_id` within the range of `start_view_id` to $(\text{start_view_id} + \text{view_count} - 1)$, inclusive, are target output views

`track_id` indicates a track containing target output views.

`tier_id` indicates a tier within a track where all views within the tier are target output views.

`output_view_id` indicates a `view_id` of a target output view.

`start_view_id` indicates the first `view_id` in a range of contiguous values of `view_id` all being target output views.

`view_count` indicates the number of contiguous values of `view_id` all being target output views.

`track_id` indicates a track.

`tier_id` indicates a tier within a track.

`subset_stream_info` indicates the characteristics of the bitstream subset containing the indicated output views and the views they depend on.

`relation_attributes` indicate the relations between output views. If 'ecam' is used as a common attribute, all the output views are associated with extrinsic camera parameters indicating that the cameras have identical rotation and constant spacing. If 'ecam' is used as a differentiating attribute, at least one output view is associated with extrinsic camera parameters with different rotation from the others or the output views are associated with extrinsic camera parameters not having a constant spacing.

`subset_stream_bit_rate` indicates the bit rate statistics of the bitstream subset containing the indicated output views and the views they depend on. The values of `tierBaseBitRate`, `tierMaxBitRate`, and `tierAvgBitRate` within the `TierBitRateBox` are unspecified.

`subset_stream_buffering` indicates the HRD parameters that apply to the bitstream subset containing the indicated output views and the views they depend on and operating with the indicated target output views.

`multiview_scene_info` contains the maximum disparity in units of integer pixel resolution between any spatially adjacent output views in any access unit.

7.7.4 Multiview Group Relation Box

7.7.4.1 Definition

Box Type: `swtc`
 Container: Multiview Information box (`mvei`)
 Mandatory: No
 Quantity: Zero or more

This box specifies a set of multiview groups from which one multiview group is decoded and played at any time. The given relation attributes specify which features are common in all associated multiview groups and which factors make the multiview groups differ from each other. The relation attributes can be used to select a suitable set of multiview groups for playback, e.g., based on the number of output views. The differentiating attributes can be used to select which multiview group within the set is suitable for the player, e.g., based on the required level for decoding.

7.7.4.2 Syntax

```
aligned(8) class MultiviewGroupRelationBox() extends FullBox('swtc', version = 0,
flags) {
    unsigned int(32) num_entries;
    for (i=0; i<num_entries; i++)
        unsigned int(32) multiview_group_id;
    MultiviewRelationAttributeBox relation_attributes;
}
```

7.7.4.3 Semantics

`num_entries` indicates the number of associated multiview groups.

`multiview_group_id` is the identifier of an associated multiview group.

`relation_attributes` indicate the relations between the associated multiview groups.

7.7.5 Multiview Relation Attribute Box

7.7.5.1 Definition

Box Type: `mvra`
 Container: MultiviewGroupBox or MultiviewGroupRelationBox
 Mandatory: No in MultiviewGroupBox, Yes in MultiviewGroupRelationBox
 Quantity: Zero or One in MultiviewGroupBox
 One in MultiviewGroupRelationBox

When the Multiview Relation Attribute box is contained in a Multiview Group box, it indicates the relation of the output views of the respective multiview group with each other. When the Multiview Relation Attribute box is contained in a Multiview Group Relation box, it indicates the relation of the multiview groups with each other.

The Multiview Relation Attribute box contains common and differentiating attributes. When the Multiview Relation Attribute box is included in a Multiview Group box, a common attribute indicates a characteristic that is common for each one of the target output views of the multiview group and a differentiating attribute indicates a characteristic which differs in at least one of one of the target output views of the multiview group. When Multiview Relation Attribute box is included in a Multiview Group Relation box, a common attribute indicates a characteristic that is common for the indicated multiview groups or for the respective target output views in each one of the indicated multiview groups, whereas

a differentiating attribute indicates a characteristic that differs in at least one of the indicated multiview groups or at least one of the respective target output views in the indicated multiview groups.

A common attribute is associated with an additional parameter, which carries the value of the common attribute. The syntax and semantics of the additional parameter depend on the attribute in question.

For example, a file writer can create a Multiview Group for each stereo pair suitable for display from a multiview bitstream. Furthermore, a file writer can create a Multiview Group Relation box listing all the multiview groups for stereo pair output and including a Multiview Relation Attribute box with common attributes number of views (equal to 2) and in-line camera arrangement. A file reader can study the Multiview Group Relation box to find the options for stereo pair output and choose one multiview group for processing. Note that the presence of views in a group does not necessarily imply they are all suggested as output views at any given time – the terminal may choose which views to output, and it is not limited by the group information.

7.7.5.2 Syntax

```
aligned(8) class MultiviewRelationAttributeBox
  extends FullBox('mvra', version = 0, flags) {
  unsigned int(16) reserved1 = 0;
  unsigned int(16) num_common_attributes;
  for (i=0; i<num_common_attributes; i++) {
    unsigned int(32) common_attribute;
    bit(32) common_value;
  }
  unsigned int(16) reserved2 = 0;
  unsigned int(16) num_differentiating_attributes;
  for (i=0; i<num_differentiating_attributes; i++)
    unsigned int(32) differentiating_attribute;
}
```

7.7.5.3 Semantics

`common_attribute` and `differentiating_attribute` are selected from the list below. Attributes that can be used as a differentiating attribute are associated with a distinguishing pointer to the field or information.

`common_value` specifies the value for the common attribute. Its syntax and semantics depend on the common attribute and are specified in the table below.

Name	Attribute	Pointer and semantics	<i>common_value</i> syntax and semantics
Profile	'prfl'	This attribute shall not be included in the Multiview Group box. When included in the Multiview Group Relation box, the attribute refers to the profile required for decoding the bitstream subset corresponding to the multiview group. The attribute points to the <code>profileIndication</code> field of the <code>subset_stream_info</code> element of the Multiview Group box.	unsigned int(24) reserved = 0; unsigned int(8) <code>profileIndication</code> ; <code>profileIndication</code> is the profile sufficient for decoding the bitstream subset corresponding to all indicated multiview groups.
Level	'levl'	This attribute shall not be included in the Multiview Group box. When included in the Multiview Group Relation box, the attribute refers to the	unsigned int(24) reserved = 0; unsigned int(8) <code>levelIndication</code> ; <code>profileIndication</code> is the level sufficient for decoding the

		level required for decoding the bitstream subset corresponding to the multiview group. The attribute points to the levelIndication field of the subset_stream_info element of the Multiview Group box.	bitstream subset corresponding to all indicated multiview groups, or 0 if the level is unspecified.
Bitrate	'bitr'	This attribute shall not be included in the Multiview Group box. When included in the Multiview Group Relation box, the attribute refers to the total size of bitstream subset required for decoding of the multiview group divided by the duration in the track header box. The attribute points to the avgBitRate field of the subset_stream_bit_rate element of the Multiview Group box, if present, or a value that would be contained in the avgBitRate field of the subset_stream_bit_rate element of the Multiview Group box, if it were present.	unsigned int(32) bitrate; bitrate indicates the average bit rate in bits per second of the bitstream subset required for decoding the multiview group. The bitrate may be rounded up.
Frame rate	'frar'	This attribute shall not be included in the Multiview Group box. When included in the Multiview Group Relation box, the attribute refers the number of samples in the track divided by duration in the track header box.	unsigned int(16) integer_part; unsigned int(16) reserved = 0; integer_part shall be equal to the output rate of decoded access units in second rounded to the closest integer using the Round function specified in ISO/IEC 14496-10.
Number of output views	'nvws'	Number of target output views indicated in the Multiview Group Box ('mvcg') If this attribute is included in the Multiview Group box, it shall be a common attribute and merely documents the number of output views in the respective multiview group.	unsigned int(32) num_views; num_views indicates the number of views in the multiview group.
Intrinsic camera parameters	'icam'	The intrinsic camera parameters are stored in 'avc1', 'avc2', 'avc3', 'avc4', 'mvc1', 'mvc2', 'mvc3' or 'mvc4' Sample Entry (in Sample entry box of media track). If this attribute is included in the Multiview Group box and used as a common attribute, the intrinsic camera parameters of the target output views are identical. If this attribute is included in the Multiview Group box and used as a differentiating attribute, the intrinsic camera parameters of the target output views differ at least partly. If this attribute is included in the	Unspecified.

Multiview Group Relation box and used as a common attribute, the number of target output views in all indicated multiview groups shall be the same and the intrinsic camera parameters of the respective target output views in all indicated multiview groups are identical. If this attribute is included in the Multiview Group Relation box and used as a differentiating attribute, the intrinsic camera parameters of the respective target output views differ at least partly.

Extrinsic camera parameters 'ecam'

The extrinsic camera parameters are stored in 'avc1', 'avc2', 'avc3', 'avc4', 'mvc1', 'mvc2', 'mvc3' or 'mvc4' Sample Entry (in Sample entry box of media track).

Unspecified.

If this attribute is included in the Multiview Group box and used as a common attribute, the rotation of the cameras for all the target output views is the same and, if the cameras are arranged in linear, elliptical, or rectangular arrangement, the distance of adjacent cameras is the same. If this attribute is included in the Multiview Group box and used as a differentiating attribute, the rotation or the distance of adjacent cameras in linear, elliptical, or rectangular arrangement differs.

If the attribute is included in the Multiview Group Relation box and used as a common attribute, the relative extrinsic camera parameters target output views in all indicated multiview groups are identical. That is, the distance of cameras relative to each other and their rotation matches in the indicated multiview groups. If the attribute is included in the Multiview Group Relation box and used as a differentiating attribute, the relative extrinsic camera parameters of respective target output views differ at least partly.

Inline view array 'ilvi'

If used as a common attribute, the associated cameras are located on a straight line.

When included in a Multiview Group box, the attribute shall be a common attribute.

```
unsigned int(28) reserved = 0;
unsigned int(2)
horizontal_order;
unsigned int(2) vertical_order;
```

horizontal_order
indicates the horizontal
order of the views:

0: the views are in the

			<p>same horizontal location</p> <p>1: the views are ordered left-to-right</p> <p>2: the views are ordered right-to-left</p> <p>3: the order of the views is undefined, or left and right are not well-defined.</p> <p>vertical_order indicates the vertical order of the views:</p> <p>0: the views are in the same vertical location</p> <p>1: the views are ordered bottom-to-top</p> <p>2: the views are ordered top-to-bottom</p> <p>3: the order of the views is undefined, or top and bottom are not well-defined.</p>
Rectangular view array	'rtvi'	<p>If used as a common attribute, the associated cameras form a rectangular shape and are regularly spaced along the orthogonal coordinate axes.</p> <p>When included in a Multiview Group box, the attribute shall be a common attribute.</p>	<p>unsigned int (16)</p> <p>row_view_count;</p> <p>unsigned int (16)</p> <p>col_view_count;</p> <p>row_view_count specifies the number of rows in the rectangular array.</p> <p>col_view_count specifies the number of columns in the rectangular array.</p> <p>The views are indicated in raster scan order in the Multiview Group box.</p>
Planar view array	'plvi'	<p>If used as a common attribute, the associated cameras are located on a plane, but may be irregularly spaced.</p> <p>When included in a Multiview Group box, the attribute shall be a common attribute.</p>	<p>Unspecified.</p>

Elliptical view array	`elvi`	<p>If used as a common attribute, the associated cameras are located on the arc of an ellipse.</p> <p>When included in a Multiview Group box, the attribute shall be a common attribute.</p>	<pre>unsigned int(28) reserved = 0; unsigned int(2) horizontal_order; unsigned int(2) vertical_order;</pre> <p>The semantics are identical to those for the Inline view array.</p>
Spherical view array	`spvi`	<p>If used as a common attribute, the associated cameras are located on the surface of a sphere.</p> <p>When included in a Multiview Group box, the attribute shall be a common attribute.</p>	<p>Unspecified.</p>
Stereo view array	`stvi`	<p>If used as a common attribute, the associated cameras are a pair of views suitable for stereo viewing.</p> <p>When included in a Multiview Group box, the attribute shall be a common attribute.</p>	<pre>unsigned int(6) reserved1 = 0; unsigned int(10) left_view_id; unsigned int(6) reserved2 = 0; unsigned int(10) right_view_id;</pre>
Geometry	`geom`	<p>If used as a differentiating attribute, indicates that the views or groups of views belong to different view arrangements (e.g. inline, planar, etc.)</p>	<p>Unspecified.</p>

7.7.6 Multiview Scene Info Box

7.7.6.1 Definition

Box Type: `vwdi`
 Container: Multiview Group box ('mvcg')
 Mandatory: No
 Quantity: Zero or one

An optional Multiview Scene Info Box includes the maximum disparity between the adjacent views of the respective multiview group. This information can be used for processing the multiview video prior to rendering on a 3D display.

NOTE A Multiview Scene Information SEI message, as specified in MVC H.12.1.5, can indicate the maximum disparity between any adjacent views in the bitstream. Thus, the Multiview Scene Info Box represents similar information as carried in the Multiview Scene Information SEI message but is limited to a certain set of views rather than concerns all the views in the bitstream.

The Multiview Scene Info Box shall not be present for multiview groups associated with cameras that do not form a one-dimensional arrangement, such as a line or an arc of an ellipse.

7.7.6.2 Syntax

```
class MultiviewSceneInfoBox extends Box (`vwdi`)
{
    unsigned int(8) max_disparity;
}
```

7.7.6.3 Semantics

`max_disparity` specifies the maximum disparity in units of integer luma samples between the spatially adjacent view components (within an access unit) in this multiview group. This information can be used for processing the multiview video prior to rendering on a 3D display.

7.7.7 MVC View priority Assignment Box

7.7.7.1 Definition

Box Type: `'mvcP'`
 Container: Sample Entry (`'avc1'`, `'avc2'`, `'avc3'`, `'avc4'`,
`'mvc1'`, `'mvc2'`, `'mvc3'`, `'mvc4'`)
 Mandatory: No
 Quantity: Zero or one

A priority assignment URI provides the name (in the URI space) of a method used to assign content priority values in the View Priority sample grouping. 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 or selection of output views based on particular content priority values would do.

7.7.7.2 Syntax

```
class MVCViewPriorityAssignmentBox extends Box('mvcP')
{
    unsigned int(8)    method_count;
    string PriorityAssignmentURI[method_count];
}
```

7.7.7.3 Semantics

`method_count` provides a count of the number of following URIs.

`PriorityAssignmentURI` provides a unique name of the method used to assign `content_priority_id` values in View Priority sample groupings. In the case of absence of this box, the priority assignment method is unknown.

8 HEVC elementary streams and sample definitions

8.1 Introduction

The High Efficiency Video Coding (HEVC) standard, jointly developed by the ITU-T and ISO/IEC JTC 1/SC 29/WG 11 (MPEG), 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, this part of ISO/IEC 14496-15 defines a storage format for video streams compressed using HEVC.

This clause of ISO/IEC 14496-15 specifies the storage format for HEVC (ISO/IEC 23008-2) video streams.

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

— Parameter sets:

The video, sequence and picture parameter set mechanism decouples the transmission of infrequently changing information from the transmission of coded block data. Each slice containing the coded block 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, and the sequence parameter set references a video parameter set that contains global decoding parameter information (across layers or view in potential scalable and 3DV extensions).

This specification includes the following tools for supporting of HEVC contents:

— Temporal scalability sample grouping:

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

— Temporal sub-layer access sample grouping:

a structuring and grouping mechanism to indicate the identification of access units as temporal sub-layer access (TSA) samples.

— Step-wise temporal sub-layer access sample grouping:

a structuring and grouping mechanism to indicate the identification of access units as step-wise temporal sub-layer access (STSA) samples.

8.2 Elementary Stream Structure

A video stream is represented by one video track in a file.

Two types of elementary streams are defined for storing HEVC content:

- **Video Elementary Stream** that does not contain any parameter sets; all parameter sets are stored in a sample entry or sample entries;
- **Video and Parameter set elementary stream** that may contain parameter sets, and may also have parameter sets stored in their sample entry or sample entries.

8.3 Sample and configuration definition

8.3.1 Introduction

HEVC sample: An HEVC sample is an access unit as defined in ISO/IEC 23008-2.

8.3.2 Canonical order and restrictions

The canonical stream format is an HEVC elementary stream that satisfies the following conditions in addition to the general conditions in <<ed: x-ref>> 4.3.2:

- **Access unit delimiter NAL units:** The constraints obeyed by access unit delimiter NAL units are defined in ISO/IEC 23008-2.
- **Parameter sets:** A parameter set to be used in a picture must be sent prior to the sample containing that picture or in the sample for that picture. For a video stream that a particular sample entry applies to, the video parameter set, sequence parameter sets, and picture parameter sets, shall be stored only in the sample entry when the sample entry name is 'hvc1', and may be stored in the sample entry and the samples when the sample entry name is 'hev1'.

NOTE Storing parameter sets in the sample entries of a video stream provides a simple and static way to supply parameter sets. Storing parameters in samples on the other hand is more complex but allows for more dynamism in the case of parameter set updates (a particular parameter set's content is changed but using the same ID) and in the case of adding additional parameter sets. A decoder initializes with the parameter sets in the sample entry, and then updates using the parameter sets as they occur in the stream, starting from any sample marked as a sync sample. Such updating may replace parameter sets with a new definition using the same identifier. Each time the sample entry changes, the decoder re-initializes with the parameter sets included in the sample entry.

- **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 The removal or addition of Filler Data NAL units, start codes, SEI messages or Filler Data SEI messages may change the bitstream characteristics with respect to conformance with the HRD when operating the HRD in CBR mode as specified in ISO/IEC 23008-2, Annex C.

8.3.3 Decoder configuration information

8.3.3.1 HEVC decoder configuration record

8.3.3.1.1 Definition

This subclause specifies the decoder configuration information for ISO/IEC 23008-2 video content.

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 parameter sets, if stored in the sample entry. This record is externally framed (its size must be supplied by the structure which contains it).

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

Compatible extensions to this record will extend it and will not change the configuration version code. Readers should be prepared to ignore unrecognised data beyond the definition of the data they understand.

The values for `general_profile_space`, `general_tier_flag`, `general_profile_idc`, `general_profile_compatibility_flags`, `general_constraint_indicator_flags`, `general_level_idc`, and `min_spatial_segmentation_idc` must be valid for all parameter sets that are activated when the stream described by this record is decoded (referred to as "all the parameter sets" in the following sentences in this paragraph). Specifically, the following restrictions apply:

- The value of `general_profile_space` in all the parameter sets must be identical.
- The tier indication `general_tier_flag` must indicate a tier equal to or greater than the highest tier indicated in all the parameter sets.
- The profile indication `general_profile_idc` must indicate a profile to which the stream associated with this configuration record conforms.

NOTE If the sequence parameter sets are marked with different profiles, 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 entire stream must be split into two or more sub-streams with separate configuration records in which these rules can be met.

- Each bit in `general_profile_compatibility_flags` may only be set if all the parameter sets set that bit.
- Each bit in `general_constraint_indicator_flags` may only be set if all the parameter sets set that bit.
- The level indication `general_level_idc` must indicate a level of capability equal to or greater than the highest level indicated for the highest tier in all the parameter sets.
- The `min_spatial_segmentation_idc` indication must indicate a level of spatial segmentation equal to or less than the lowest level of spatial segmentation indicated in all the parameter sets.

Explicit indication can be provided in the HEVC Decoder Configuration Record about the chroma format and bit depth as well as other important format information used by the HEVC video elementary stream. Each type of such information must be identical in all parameter sets, if present, in a single HEVC configuration record. If two sequences differ in any type of such information, two different HEVC configuration records must be used. If the two sequences differ in color space indications in their VUI information, then two different configuration records are also required.

There is a set of arrays to carry initialization NAL units. The NAL unit types are restricted to indicate SPS, PPS, VPS, and SEI NAL units only. NAL unit types that are reserved in ISO/IEC 23008-2 and in this specification may acquire a definition in future, and readers should ignore arrays with reserved or unpermitted values of NAL unit type.

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.

It is recommended that the arrays be in the order VPS, SPS, PPS, SEI.

When `general_non_packed_constraint_flag` (bit 3 of the 6-byte `general_constraint_indicator_flags`) is equal to 0 and some of the samples referring to this sample entry represent frame-packed content and any of the default display windows specified by the active SPSs for the samples referring to this sample entry covers more than one constituent frame of the frame-packed content, the techniques described in 8.15 of ISO/IEC 14496-12 (‘Post-decoder requirements on media’) using the scheme type “stvl” shall be used. In this case, the `stereo_scheme` in the Stereo Video Box should be set to 1, to indicate that the frame packing scheme used in HEVC is the same as in AVC.

8.3.3.1.2 Syntax

```
aligned(8) class HEVCDecoderConfigurationRecord {
    unsigned int(8) configurationVersion = 1;
    unsigned int(2) general_profile_space;
    unsigned int(1) general_tier_flag;
    unsigned int(5) general_profile_idc;
    unsigned int(32) general_profile_compatibility_flags;
    unsigned int(48) general_constraint_indicator_flags;
    unsigned int(8) general_level_idc;
    bit(4) reserved = '1111'b;
    unsigned int(12) min_spatial_segmentation_idc;
    bit(6) reserved = '111111'b;
    unsigned int(2) parallelismType;
    bit(6) reserved = '111111'b;
    unsigned int(2) chromaFormat;
    bit(5) reserved = '11111'b;
    unsigned int(3) bitDepthLumaMinus8;
    bit(5) reserved = '11111'b;
    unsigned int(3) bitDepthChromaMinus8;
    bit(16) avgFrameRate;
    bit(2) constantFrameRate;
    bit(3) numTemporalLayers;
    bit(1) temporalIdNested;
    unsigned int(2) lengthSizeMinusOne;
    unsigned int(8) numOfArrays;
    for (j=0; j < numOfArrays; j++) {
        bit(1) array_completeness;
        unsigned int(1) reserved = 0;
        unsigned int(6) NAL_unit_type;
        unsigned int(16) numNalus;
        for (i=0; i < numNalus; i++) {
            unsigned int(16) nalUnitLength;
            bit(8*nalUnitLength) nalUnit;
        }
    }
}
```

8.3.3.1.3 Semantics

`general_profile_space`, `general_tier_flag`, `general_profile_idc`, `general_profile_compatibility_flags`, `general_constraint_indicator_flags`, `general_level_idc`, and `min_spatial_segmentation_idc` contain the matching values for the fields `general_profile_space`, `general_tier_flag`, `general_profile_idc`, `general_profile_compatibility_flag[i]` for i from 0 to 31, inclusive, the 6 bytes starting with the byte containing the `general_progressive_source_flag`, `general_level_idc`, and `min_spatial_segmentation_idc` as defined in ISO/IEC 23008-2, for the stream to which this configuration record applies.

`parallelismType` indicates the type of parallelism that is used to meet the restrictions imposed by `min_spatial_segmentation_idc` when the value of `min_spatial_segmentation_idc` is greater than 0. Value 1 indicates that the stream to which this configuration record applies supports slice based parallel decoding. Value 2 indicates that the stream to which this configuration record applies supports tile based parallel decoding. Value 3 indicates that the stream to which this configuration record applies supports entropy coding synchronization based parallel decoding. Value 0 indicates that the stream supports mixed types of parallel decoding or that the parallelism type is unknown.

`chromaFormat` contains the `chroma_format` indicator as defined by the `chroma_format_idc` parameter in ISO/IEC 23008-2, for the stream to which this configuration record applies.

`bitDepthLumaMinus8` contains the luma bit depth indicator as defined by the `bit_depth_luma_minus8` parameter in ISO/IEC 23008-2, for the stream to which this configuration record applies.

`bitDepthChromaMinus8` contains the chroma bit depth indicator as defined by the `bit_depth_chroma_minus8` in ISO/IEC 23008-2, for the stream to which this configuration record applies.

`avgFrameRate` gives the average frame rate in units of frames/(256 seconds), for the stream to which this configuration record applies. Value 0 indicates an unspecified average frame rate.

`constantFrameRate` equal to 1 indicates that the stream to which this configuration record applies is of constant frame rate. Value 2 indicates that the representation of each temporal layer in the stream is of constant frame rate. Value 0 indicates that the stream may or may not be of constant frame rate.

`numTemporalLayers` greater than 1 indicates that the stream to which this configuration record applies is temporally scalable and the contained number of temporal layers (also referred to as temporal sub-layer or sub-layer in ISO/IEC 23008-2) is equal to `numTemporalLayers`. Value 1 indicates that the stream is not temporally scalable. Value 0 indicates that it is unknown whether the stream is temporally scalable.

`temporalIdNested` equal to 1 indicates that all SPSs that are activated when the stream to which this configuration record applies is decoded have `sps_temporal_id_nesting_flag` as defined in ISO/IEC 23008-2 equal to 1 and temporal sub-layer up-switching to any higher temporal layer can be performed at any sample. Value 0 indicates that at least one of the SPSs that are activated when the stream to which this configuration record applies is decoded has `sps_temporal_id_nesting_flag` as defined in ISO/IEC 23008-2 equal to 0.

`lengthSizeMinusOne` plus 1 indicates the length in bytes of the `NALUnitLength` field in an HEVC video sample in the stream to which this configuration record applies. 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.

`numArrays` indicates the number of arrays of NAL units of the indicated type(s)

`array_completeness` when equal to 1 indicates that all NAL units of the given type are in the following array and none are in the stream; when equal to 0 indicates that additional NAL units of the indicated type may be in the stream; the default and permitted values are constrained by the sample entry name;

`NAL_unit_type` indicates the type of the NAL units in the following array (which must be all of that type); it takes a value as defined in ISO/IEC 23008-2; it is restricted to take one of the values indicating a VPS, SPS, PPS, or SEI NAL unit;

`numNalus` indicates the number of NAL units of the indicated type included in the configuration record for the stream to which this configuration record applies. The SEI array must only 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 could be a user-data SEI.

`nalUnitLength` indicates the length in bytes of the NAL unit.

`nalUnit` contains an SPS, PPS, VPS or declarative SEI NAL unit, as specified in ISO/IEC 23008-2.

8.4 Derivation from ISO base media file format

8.4.1 HEVC video stream definition

8.4.1.1 Sample entry name and format

8.4.1.1.1 Definition

Box Types: 'hvc1', 'hev1', 'hvcC'
 Container: Sample Table Box ('stbl')
 Mandatory: An 'hvc1' or 'hev1' sample entry is mandatory
 Quantity: One or more sample entries may be present

An HEVC visual sample entry shall contain an HEVC Configuration Box, as defined below. This includes an HEVCDecoderConfigurationRecord, as defined in 5.3.3.1.

An optional MPEG4BitRateBox may be present in the HEVC visual sample entry to signal the bit rate information of the HEVC 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 the ISO Base Media File Format specification, to indicate sections of video that use different configurations or parameter sets.

When the sample entry name is 'hvc1' or 'hev1', the stream to which this sample entry applies shall be a compliant and usable HEVC stream as viewed by an HEVC decoder operating under the configuration (including profile, tier, and level) given in the HEVCConfigurationBox.

When the sample entry name is 'hvc1', the default and mandatory value of `array_completeness` is 1 for arrays of all types of parameter sets, and 0 for all other arrays. When the sample entry name is 'hev1', the default value of `array_completeness` is 0 for all arrays.

8.4.1.1.2 Syntax

```
class HEVCConfigurationBox extends Box('hvcC') {
    HEVCDecoderConfigurationRecord() HEVCConfig;
}

class HEVCSESampleEntry() extends VisualSampleEntry ('hvc1' or 'hev1'){
    HEVCConfigurationBox config;
    MPEG4BitRateBox (); // optional
    MPEG4ExtensionDescriptorsBox (); // optional
    extra_boxes boxes; // optional
}
```

8.4.1.1.3 Semantics

Compressorname in the base class VisualSampleEntry indicates the name of the compressor used with the value "\013HEVC Coding" being recommended (\013 is 11, the length of the string in bytes).

HEVCDecoderConfigurationRecord is defined in 5.3.3..

8.4.2 Parameter sets in sample entry

This subclause applies to a particular type of parameter sets (VPSs, SPSs, or PPSs) when the particular type of parameter sets is included in the sample entry.

Each HEVC sample entry, which contains the HEVC video stream decoder specific information, includes a group of the particular type of parameter sets. 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.

In the file format each configuration of parameter sets is represented separately. When the value of the applicable `array_completeness` is 1, a parameter set cannot be updated without causing a different sample entry to be used.

Systems wishing to send parameter set updates will need to compare the two configurations to find the differences in order to send the appropriate parameter set updates.

NOTE 1 It is recommended that when several parameter sets are used and parameter set updating is desired, the parameter sets are included in the samples of the stream.

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

8.4.3 Sync sample

An HEVC sample is considered as a sync sample if the VCL NAL units in the sample indicate that the coded picture contained in the sample is an Instantaneous Decoding Refresh (IDR) picture, a Clean Random Access (CRA) picture, or a Broken Link Access (BLA) picture.

When the sample entry name is 'hev1', the following applies:

- If the sample is a random access point, all parameter sets needed for decoding that sample shall be included either in the sample entry or in the sample itself.
- Otherwise (the sample is not a random access point), 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 random access point to the sample itself, inclusive.

For signalling of various types of random access points, the following guidelines are recommended:

- The sync sample table (and the equivalent flag in movie fragments) must be used in an HEVC track unless all samples are sync samples. Note that track fragment random access box refers to the presence of signalled sync samples in a movie fragment.
- The 'roll' sample group is recommended to be used only for gradual decoding refresh (GDR) based random access points, i.e. those that contain non-intra coded slices.
- The use of the 'rap ' or 'sync' sample group is optional, depending on the need of either the information on leading samples associated with the random access points or the picture types (e.g. IDR, CRA, or BLA) of the random access points.
- The use of the Alternative Startup Sequences (ISO/IEC 14496-12 section 10.3) sample group is recommended to be used only with random access points consisting of CRA and BLA pictures.

In the context of this clause, the leading samples, defined as part of the definition of the 'rap ' sample group in ISO/IEC 14496-2, contain Random Access Skipped Leading (RASL) access units as defined in ISO/IEC 23008-2.

8.4.4 Sync sample sample grouping

8.4.4.1 Introduction

Sync samples in HEVC may be of various types. These sample groups may be used to identify the sync samples of a specific type. If a sample group is given for a specific type of sync sample, then all samples (if any) containing that type of sync sample are marked by the group. If the group is absent (there is no sample to group mapping for that type), it is unknown which samples contain a sync sample of that type.

8.4.4.2 Sync sample sample group entry

8.4.4.2.1 Definition

Group Types: 'sync'
 Container: Sample Group Description Box ('sgpd')
 Mandatory: No
 Quantity: Zero or more

A sync sample sample group entry identifies samples containing a sync sample of a specific type.

8.4.4.2.2 Syntax

```
class SyncSampleEntry() extends VisualSampleGroupEntry ('sync')
{
    unsigned int(2) reserved = 0;
    unsigned int(6) NAL_unit_type;
}
```

8.4.4.2.3 Semantics

`NAL_unit_type` must be a type that identifies a valid sync sample (e.g. IDR).

8.4.5 Temporal scalability sample grouping

8.4.5.1 Introduction

An HEVC video track may contain zero or one instance of a `SampleToGroupBox` with a `grouping_type` 'tscl'. This `SampleToGroupBox` instance represents the assignment of samples in the track to temporal layers (referred to as temporal sub-layers or sub-layers in ISO/IEC 23008-2). An accompanying instance of the `SampleGroupDescriptionBox` with the same grouping type shall be present, and contain `TemporalLayerEntry` sample group entries describing the temporal layers.

There may also be a set of Temporal Level sample group definitions (ISO/IEC 14496-12 section 10.5). A sample mapped to a sample group description entry with index A of a Temporal Level sample grouping shall also be considered mapped to the sample group description entry in the temporal scalability sample group (defined here) having `temporalLayerId` equal to A (i.e. a sample to group mapping for the temporal scalability sample group may also be present but is not required).

8.4.5.2 Temporal layer sample group entry

8.4.5.2.1 Definition

Group Types: 'tscl'
 Container: Sample Group Description Box ('sgpd')

Mandatory: No
 Quantity: Zero or more

A temporal layer sample group entry defines the temporal layer information for all samples in a temporal layer. Temporal layers are numbered with non-negative integers, and each temporal layer is associated with a particular value of `TemporalId` as defined in ISO/IEC 23008-2. A temporal layer associated with a `TemporalId` value greater than 0 depends on all temporal layers associated with lower `TemporalId` values. A temporal layer representation, also referred to as the representation of a temporal layer, associated with a particular `TemporalId` value consists of all temporal layers associated with `TemporalId` values less than or equal to the given `TemporalId` value.

8.4.5.2.2 Syntax

```
class TemporalLayerEntry() extends VisualSampleGroupEntry ('tscl')
{
    unsigned int(8)    temporalLayerId;
    unsigned int(2)    tlprofile_space;
    unsigned int(1)    tltier_flag;
    unsigned int(5)    tlprofile_idc;
    unsigned int(32)   tlprofile_compatibility_flags;
    unsigned int(48)   tlconstraint_indicator_flags;
    unsigned int(8)    tllevel_idc;
    unsigned int(16)   tlMaxBitRate;
    unsigned int(16)   tlAvgBitRate;
    unsigned int(8)    tlConstantFrameRate;
    unsigned int(16)   tlAvgFrameRate;
}
```

8.4.5.2.3 Semantics

`temporalLayerId` gives the ID of this temporal layer. For all samples that are members of this sample group, the VCL NAL units shall have `temporal_id`, as defined in ISO/IEC 23008-2, equal to `temporalLayerId`.

`tlprofile_space`, `tltier_flag`, `tlprofile_idc`, `tlprofile_compatibility_flags`, `tlconstraint_indicator_flags`, and `tllevel_idc` contain the values of `general_profile_space`, `general_tier_flag`, `general_profile_idc`, `general_profile_compatibility_flag[i]` for `i` from 0 to 31, inclusive, the 6 bytes starting with the byte containing the `general_progressive_source_flag`, and `general_level_idc`, respectively, for the representation of the temporal layer identified by `temporalLayerId`.

`tlMaxBitrate` gives the maximum rate in 1000 bits per second over any window of one second, for the representation of the temporal layer identified by `temporalLayerId`.

`tlAvgBitRate` gives the average bit rate in units of 1000 bits per second, for the representation of the temporal layer identified by `temporalLayerId`.

`tlConstantFrameRate` equal to 1 indicates that the representation of the temporal layer identified by `temporalLayerId` is of constant frame rate. Value zero indicates that the representation of the temporal layer identified by `temporalLayerId` may or may not be of constant frame rate.

`tlAvgFrameRate` gives the average frame rate in units of frames/(256 seconds), for the representation of the temporal layer identified by `temporalLayerId`.

8.4.6 Temporal sub-layer access sample grouping

8.4.6.1 Introduction

An HEVC video track may contain zero or one instance of a `SampleToGroupBox` with a `grouping_type` 'tsas'. This `SampleToGroupBox` instance represents the marking of samples as temporal sub-layer

access points. An accompanying instance of the SampleGroupDescriptionBox with the same grouping type shall be present.

8.4.6.2 Temporal sub-layer sample group entry

8.4.6.2.1 Introduction

Group Types: 'tsas'
 Container: Sample Group Description Box ('sgpd')
 Mandatory: No
 Quantity: Zero or one

This sample group is used to mark temporal layer access (TSA) samples.

8.4.6.2.2 Syntax

```
class TemporalSubLayerEntry() extends VisualSampleGroupEntry ('tsas')
{
}
```

8.4.7 Step-wise temporal layer access sample grouping

8.4.7.1 Introduction

An HEVC video track may contain zero or one instance of a SampleToGroupBox with a grouping_type 'stsa'. This SampleToGroupBox instance represents the marking of samples as step-wise temporal layer access points. An accompanying instance of the SampleGroupDescriptionBox with the same grouping type shall be present.

When temporalIdNested in the applicable sample entry is equal to 1, the quantity of step-wise temporal sub-layer access sample group entry shall be zero.

8.4.7.2 Step-wise temporal layer sample group entry

8.4.7.2.1 Definition

Group Types: 'stsa'
 Container: Sample Group Description Box ('sgpd')
 Mandatory: No
 Quantity: Zero or one

This sample group is used to mark step-wise temporal layer access (STSA) samples.

8.4.7.2.2 Syntax

```
class StepwiseTemporalLayerEntry() extends VisualSampleGroupEntry ('stsa')
{
}
```

8.4.8 Definition of a sub-sample for HEVC

For the use of the sub-sample information box (8.7.7 of ISO/IEC 14496-12) in an HEVC stream, a sub-sample is defined on the basis of the value of the flags field of the sub-sample information box as specified below. The presence of this box is optional; however, if present in a track containing HEVC data, it shall have the semantics defined here.

flags specifies the type of sub-sample information given in this box as follows:

- 0: NAL-unit-based sub-samples. A sub-sample contains one or more contiguous NAL units.
- 1: Decoding-unit-based sub-samples. A sub-sample contains exactly one decoding unit.
- 2: Tile-based sub-samples. A sub-sample either contains one tile and the associated non-VCL NAL units, if any, of the VCL NAL unit(s) containing the tile, or contains one or more non-VCL NAL units.
- 3: CTU-row-based sub-samples. A sub-sample either contains one CTU row within a slice and the associated non-VCL NAL units, if any, of the VCL NAL unit(s) containing the CTU row or contains one or more non-VCL NAL units. This type of sub-sample information shall not be used when entropy_coding_sync_enabled_flag is equal to 0.
- 4: Slice-based sub-samples. A sub-sample either contains one slice (where each slice may contain one or more slice segments, each of which is a NAL unit) and the associated non-VCL NAL units, if any, or contains one or more non-VCL NAL units.

Other values of flags are reserved.

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 can still be decoded if this sub-sample is discarded (e.g. the sub-sample consists of an SEI NAL unit).

When the first byte of a NAL unit is included in a sub-sample, the preceding length field must also be included in the same sub-sample.

```

if (flags == 0) {
    unsigned int(1) SubLayerRefNalUnitFlag;
    unsigned int(1) RapNalUnitFlag;
    unsigned int(1) VclNalUnitFlag;
    unsigned int(29) reserved = 0;
} else if (flags == 1) {
    unsigned int(32) reserved = 0;
} else if (flags == 2) {
    unsigned int(2) vcl_idc;
    unsigned int(2) reserved = 0;
    unsigned int(4) log2_min_luma_ctb;
    unsigned int(12) ctb_x;
    unsigned int(12) ctb_y;
} else if (flags == 3 || flags == 4) {
    unsigned int(2) vcl_idc;
    unsigned int(30) reserved = 0;
}
    
```

SubLayerRefNalUnitFlag equal to 0 indicates that all NAL units in the sub-sample are VCL NAL units of a sub-layer non-reference picture as specified in ISO/IEC 23008-2. Value 1 indicates that all NAL units in the sub-sample are VCL NAL units of a sub-layer reference picture as specified in ISO/IEC 23008-2.

RapNalUnitFlag equal to 0 indicates that none of the NAL units in the sub-sample has nal_unit_type equal to IDR_W_RADL, IDR_N_LP, CRA_NUT, BLA_W_LP, BLA_W_RADL, BLA_N_LP, RSV_IRAP_VCL22, or RSV_IRAP_VCL23 as specified in ISO/IEC 23008-2. Value 1 indicates that all NAL units in the sub-sample have nal_unit_type equal to IDR_W_RADL,

I_{DR_N_LP}, C_{RA_NUT}, B_{LA_W_LP}, B_{LA_W_RADL}, B_{LA_N_LP}, R_{SV_IRAP_VCL22}, or R_{SV_IRAP_VCL23} as specified in ISO/IEC 23008-2.

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.

vcl_idc indicates whether the sub-sample contains Video Coding Layer (VCL) data, non-VCL data, or both, as follows:

- 0: the sub-sample contains VCL data and does not contain non-VCL data
- 1: the sub-sample contains no VCL data and contains non-VCL data
- 2: the sub-sample may contain both VCL and non-VCL data, which shall be associated with each other. For example, a sub-sample may contain a decoding unit information SEI message followed by the set of NAL units associated with the SEI message.
- 3: reserved

log2_min_luma_ctb indicates the unit of ctb_x and ctb_y, specified as follows:

- 0: 8 luma samples
- 1: 16 luma samples
- 2: 32 luma samples
- 3: 64 luma samples

ctb_x specifies the 0-based coordinate of the right-most luma samples of the tile associated with the sub-sample when flags is equal to 2 and vcl_idc is equal to 1 or 2, in units derived from log2_min_luma_ctb as specified above.

ctb_y specifies the 0-based coordinate the bottom-most luma samples of the tile associated with the sub-sample when flags is equal to 2 and vcl_idc is equal to 1 or 2, in units derived from log2_min_luma_ctb as specified above.

8.4.9 Handling non-output samples

HEVC allows for file format samples that are used only for reference and not output (e.g. a non-displayed reference picture in video). When any such non-output sample is present in a track, the file shall be constrained as follows:

- 1) A non-output sample shall be given a composition time which is outside the time-range of the samples that are output;
- 2) An edit list shall be used to exclude the composition times of the non-output samples.
- 3) When the track includes a CompositionOffsetBox ('ctts'),
 - a. version 1 of the CompositionOffsetBox shall be used,
 - b. the value of sample_offset shall be set equal to -2^{31} for each non-output sample,
 - c. the CompositionToDecodeBox ('cslg') should be contained in the SampleTableBox ('stbl') of the track, and
 - d. when the CompositionToDecodeBox is present for the track, the value of leastDecodeToDisplayDelta field in the box shall be equal to the smallest composition offset in the CompositionOffsetBox excluding the sample_offset values for non-output samples.

NOTE: Thus, leastDecodeToDisplayDelta is greater than -2^{31} .

Annex A (normative)

In-stream structures specific to SVC and MVC

A.1 Introduction

Aggregators and Extractors are file format internal structures enabling efficient grouping of NAL units or extraction of NAL units from other tracks.

Aggregators and Extractors use the NAL unit syntax. These structures are seen as NAL units in the context of the sample structure. While accessing a sample, Aggregators must be removed (leaving their contained or referenced NAL Units) and Extractors must be replaced by the data they reference. Aggregators and Extractors must not be present in a stream outside the file format.

These structures use NAL unit types reserved for the application/transport layer by ISO/IEC 14496-10.

NOTE The following is from ISO/IEC 14496-10:

“NOTE – NAL unit types 0 and 24..31 may be used as determined by the application. No decoding process for these values of nal_unit_type is specified in this Recommendation | International Standard.”

A.2 Aggregators

A.2.1 Definition

This subclause describes Aggregators, which enable NALU-map-group entries to be consistent and repetitive. (See Annex B).

Aggregators are used to group NAL units belonging to the same sample. Aggregators use the same NAL unit header as SVC VCL NAL units or MVC VCL NAL units, but with a different value of NAL unit type. When the svc_extension_flag of the NAL unit syntax (specified in 7.3.1 of ISO/IEC 14496-10) of an aggregator is equal to 1, the NAL unit header of SVC VCL NAL units is used for the aggregator. Otherwise, the NAL unit header of MVC VCL NAL units is used for the aggregator.

Aggregators can both aggregate, by *inclusion*, NAL units within them (within the size indicated by their length) and also aggregate, by *reference*, NAL units that follow them (within the area indicated by the additional_bytes field within them). When the stream is scanned by an AVC file reader, only the included NAL units are seen as “within” the aggregator. This permits an AVC file reader to skip a whole set of un-needed SVC VCL NAL units or MVC VCL NAL units when they are aggregated by inclusion. This also permits an AVC reader not to skip AVC NAL units but let them remain in-stream when they are aggregated by reference.

Aggregators can be used to group AVC base view NAL units. If these Aggregators are used in an ‘avc1’ track then an aggregator shall not use inclusion but reference of AVC base view NAL units (the length of the Aggregator includes only its header and the NAL units referenced by the Aggregator are specified by additional_bytes).

When the aggregator is referenced by either an extractor with data_length equal to zero, or by a Map sample group, the aggregator is treated as aggregating both the included and referenced bytes.

An Aggregator may include or reference Extractors. An Extractor may extract from Aggregators. An aggregator must not include or reference another aggregator directly; however, an aggregator may include or reference an extractor which references an aggregator.

When scanning the stream:

- a) if the aggregator is unrecognized (e.g. by an AVC reader or decoder) it is easily discarded with its included content;
- b) if the aggregator is not needed (i.e. it belongs to an undesired layer) it and its contents both by inclusion and reference are easily discarded (using its length and additional_bytes fields);
- c) if the aggregator is needed, its header is easily discarded and its contents retained.

An aggregator is stored within a sample like any other NAL unit.

All NAL units remain in decoding order within an aggregator.

A.2.2 Syntax

```
class aligned(8) Aggregator (AggregatorSize) {
    NALUnitHeader();
    unsigned int i = sizeof(NALUnitHeader());
    unsigned int((lengthSizeMinusOne+1)*8)
        additional_bytes;
    i += lengthSizeMinusOne+1;
    while (i<AggregatorSize) {
        unsigned int((lengthSizeMinusOne+1)*8)
            NALUnitLength;
        unsigned int(NALUnitLength*8) NALUnit;
        i += NALUnitLength+lengthSizeMinusOne+1;
    }
};
```

A.2.3 Semantics

The value of the variable `AggregatorSize` is equal to the size of the aggregator NAL unit, and the function `sizeof(X)` returns the size of the field `X` in bytes.

`NALUnitHeader()`: the first four bytes of SVC and MVC VCL NAL units.

`nal_unit_type` shall be set to the aggregator NAL unit type (type 30).

For an aggregator including or referencing SVC NAL units, the following shall apply.

`forbidden_zero_bit` and `reserved_three_2bits` shall be set as specified in ISO/IEC 14496-10.

Other fields (`nal_ref_idc`, `idr_flag`, `priority_id`, `no_inter_layer_pred_flag`, `dependency_id`, `quality_id`, `temporal_id`, `use_ref_base_pic_flag`, `discardable_flag`, and `output_flag`) shall be set as specified in A.4.

For an aggregator including or referencing MVC NAL units, the following shall apply.

`forbidden_zero_bit` and `reserved_one_bit` shall be set as specified in ISO/IEC 14496-10.

Other fields (`nal_ref_idc`, `non_idr_flag`, `priority_id`, `view_id`, `temporal_id`, `anchor_pic_flag`, and `inter_view_flag`) shall be set as specified in A.5.

`additional_bytes`: The number of bytes following this aggregator NAL unit that should be considered as aggregated when this aggregator is referenced by an extractor with `data_length` equal to zero or Map sample group.

`NALUnitLength`: Specifies the size, in bytes, of the NAL unit following. The size of this field is specified with the `lengthSizeMinusOne` field.

`NALUnit`: a NAL unit as specified in ISO/IEC 14496-10, including the NAL unit header. The size of the NAL unit is specified by `NALUnitLength`.

A.3 Extractors

A.3.1 Definition

This subclause describes Extractors, which enable compact formation of tracks that extract, by reference, NAL unit data from other tracks.

An Aggregator may include or reference Extractors. An Extractor may reference Aggregators. When an extractor is processed by a file reader that requires it, the extractor is logically replaced by the bytes it references. Those bytes must not contain extractors; an extractor must not reference, directly or indirectly, another extractor.

NOTE The track that is referenced may contain extractors even though the data that is referenced by the extractor must not.

An extractor contains an instruction to extract data from another track, which is linked to the track in which the extractor resides, by means of a track reference of type `scal`.

The bytes copied shall be one of the following:

- a) One entire NAL unit; note that when an Aggregator is referenced, both the included and referenced bytes are copied
- b) More than one entire NAL unit

In both cases the bytes extracted start with a valid length field and a NAL unit header.

The bytes are copied only from the single identified sample in the track referenced through the indicated `scal` track reference. The alignment is on decoding time, i.e. using the time-to-sample table only, followed by a counted offset in sample number. Extractors are a media-level concept and hence apply to the destination track before any edit list is considered. (However, one would normally expect that the edit lists in the two tracks would be identical).

A.3.2 Syntax

```
class aligned(8) Extractor () {
    NALUnitHeader();
    unsigned int(8) track_ref_index;
    signed int(8) sample_offset;
    unsigned int((lengthSizeMinusOne+1)*8)
        data_offset;
    unsigned int((lengthSizeMinusOne+1)*8)
        data_length;
}
```

A.3.3 Semantics

`NALUnitHeader()`: the first four bytes of SVC and MVC VCL NAL units.

`nal_unit_type` shall be set to the extractor NAL unit type (type 31).

For an extractor referencing SVC NAL units, the following shall apply.

`forbidden_zero_bit` and `reserved_three_2bits` shall be set as specified in ISO/IEC 14496-10.

Other fields (`nal_ref_idc`, `idr_flag`, `priority_id`, `no_inter_layer_pred_flag`, `dependency_id`, `quality_id`, `temporal_id`, `use_ref_base_pic_flag`, `discardable_flag`, and `output_flag`) shall be set as specified in A.4.

For an extractor referencing MVC NAL units, the following shall apply.

`forbidden_zero_bit` and `reserved_one_bit` shall be set as specified in ISO/IEC 14496-10.

Other fields (`nal_ref_idc`, `non_idr_flag`, `priority_id`, `view_id`, `temporal_id`, `anchor_pic_flag`, and `inter_view_flag`) shall be set as specified in A.5.

`track_ref_index` specifies the index of the track reference of type 'scal' to use to find the track from which to extract data. The sample in that track from which data is extracted is temporally aligned or nearest preceding in the media decoding timeline, i.e. using the time-to-sample table only, adjusted by an offset specified by `sample_offset` with the sample containing the Extractor. The first track reference has the index value 1; the value 0 is reserved.

`sample_offset` gives the relative index of the sample in the linked track that shall be used as the source of information. Sample 0 (zero) is the sample with the same, or the closest preceding, decoding time compared to the decoding time of the sample containing the extractor; sample 1 (one) is the next sample, sample -1 (minus 1) is the previous sample, and so on.

`data_offset`: The offset of the first byte within the reference sample to copy. If the extraction starts with the first byte of data in that sample, the offset takes the value 0. The offset shall reference the beginning of a NAL unit length field.

`data_length`: The number of bytes to copy. If this field takes the value 0, then the entire single referenced NAL unit is copied (i.e. the length to copy is taken from the length field referenced by the data offset, augmented by the `additional_bytes` field in the case of Aggregators).

NOTE If the two tracks use different `lengthSizeMinusOne` values, then the extracted data will need re-formatting to conform to the destination track's length field size.

A.4 NAL unit header values for SVC

Both extractors and aggregators use NAL unit headers with the NAL unit header SVC extension. The NAL units extracted by an extractor or aggregated by an aggregator are all those NAL units that are referenced or included by recursively inspecting the contents of aggregator or extractor NAL units. The fields `nal_ref_idc`, `idr_flag`, `priority_id`, `no_inter_layer_pred_flag`, `dependency_id`, `quality_id`, `temporal_id`, `use_ref_base_pic_flag`, `discardable_flag`, and `output_flag` shall take the following values:

The fields below shall take the following values:

`nal_ref_idc` shall be set to the highest values of the fields, respectively, in all the extracted or aggregated NAL units.

`idr_flag` shall be set to the highest values of the fields, respectively, in all the extracted or aggregated NAL units.

`priority_id`, `temporal_id`, `dependency_id`, and `quality_id` shall be set to the lowest values of the fields, respectively, in all the extracted or aggregated NAL units.

`discardable_flag` shall be set to 1 if and only if all the extracted or aggregated NAL units have the `discardable_flag` set to 1, and set to 0 otherwise.

`output_flag` should be set to 1 if at least one of the aggregated or extracted NAL units has this flag set to 1, and otherwise set to 0.

`use_ref_base_pic_flag` shall be set to 1 if and only if at least one of the extracted or aggregated VCL NAL units have the `use_ref_base_pic_flag` set to 1, and set to 0 otherwise.

`no_inter_layer_pred_flag` shall be set to 1 if and only if all the extracted or aggregated VCL NAL units have the `no_inter_layer_pred_flag` set to 1, and set to 0 otherwise.

If the set of extracted or aggregated NAL units is empty, then each of these fields takes a value conformant with the mapped tier description.

NOTE Aggregators could group NAL units with different scalability information.

NOTE Aggregators could be used to group NAL units belonging to a level of scalability which may not be signalled by the NAL unit header SVC extension (e.g. NAL units belonging to a region of interest). The description of such Aggregators may be done with the tier description and the NAL unit map groups. In this case more than one Aggregator with the same scalability information may occur in one sample.

NOTE If multiple scalable tracks reference the same media data, then an aggregator should group NAL units with identical scalability information only. This ensures that the resulting pattern can be accessed by each of the tracks.

NOTE If no NAL unit of a particular layer exist in an access unit then an empty Aggregator (in which the length of the Aggregator includes only the header, and `additional_bytes` is zero) may exist.

A.5 NAL unit header values for MVC

Both Aggregators and Extractors use NAL unit headers with the NAL unit header extension. The NAL units extracted by an extractor or aggregated by an aggregator are all those NAL units that are referenced or included by recursively inspecting the contents of aggregator or extractor NAL units.

The fields `nal_ref_idc`, `non_idr_flag`, `priority_id`, `view_id`, `temporal_id`, `anchor_pic_flag`, and `inter_view_flag` shall take the following values:

`nal_ref_idc` shall be set to the highest values of the field in all the aggregated or extracted NAL units.

`non_idr_flag` shall be set to the lowest values of the field in all the aggregated or extracted NAL units.

`priority_id` and `temporal_id` shall be set to the lowest values of the fields, respectively, in all the aggregated or extracted NAL units.

`view_id` shall be set to the `view_id` value of the VCL NAL unit with the lowest view order index among all the aggregated or extracted VCL NAL units.

`anchor_pic_flag` and `inter_view_flag` shall be set to the highest value of the fields, respectively, in all the aggregated or extracted VCL NAL units.

Annex B (normative)

SVC and MVC sample group and sub-track definitions

B.1 Introduction

The following sample groups may be used in an SVC or MVC track to document the structure of the SVC or MVC stream and to ease obtaining information of subsets of the stream and extraction of any of the subsets.

If views from the same MVC bitstream are stored in multiple MVC tracks and one or more of these tracks contain multiple views, sample group entries and map groups can be used for these tracks containing multiple views.

There are a number of boxes, defined below, which may occur in the sample group description, namely the Scalable Group Entry for an SVC stream or the Multiview Group Entry for an MVC stream.

Each Scalable Group Entry or Multiview Group Entry documents a subset of the SVC stream or the MVC stream, respectively. Each of the subsets is associated with a tier and may contain one or more operating points. A grouping type of 'scif' or 'mvif' is used to define Scalable Group Entries or Multiview Group Entries, respectively.

For each tier, there may be more than one Scalable Group Entry or Multiview Group Entry in the SampleGroupDescriptionBox of grouping type 'scif' or 'mvif', respectively. Only one of those entries is the primary definition of the tier.

Though the Scalable and Multiview Group Entries are contained in the SampleGroupDescription box, the grouping is not a true *sample* grouping as each sample may be associated with more than one tier, as these groups are used to describe *sections* of the samples – the NAL units. As a result, it is possible that there may not be a SampleToGroup box of the grouping type 'scif' or 'mvif', unless it happens that a group does, in fact, describe a whole sample. Even if a SampleToGroup box of the grouping type 'scif' or 'mvif' is present, the information is not needed for extraction of NAL units of tiers; instead, the map groups must always document the 'pattern' of NAL units within the samples and provide the NAL-unit-to-tier mapping information that may be needed for extraction of NAL units.

A multiview group specifies an MVC operating point and is therefore associated with the target output views of the MVC operating point. The Multiview Group box (7.7.3), is used to specify a multiview group. Many of the boxes used to characterize SVC and MVC tiers are also used to characterize MVC operating points and can therefore be contained in the Multiview Group box too.

B.2 Definition

B.2.1 Tier information box

B.2.1.1 Definition

Box Type: 'tiri'
 Container: ScalableGroupEntry or MultiviewGroupEntry or MultiviewGroupBox

Mandatory: Yes
 Quantity: Zero or One // depends on primary_definition

The tier information box provides information about the profile, level, frame size, discardability, and frame-rate of a covered bitstream subset. If the Tier Information box is included in a Scalable Group entry or a Multiview Group entry, the covered bitstream subset consists of the tier and tiers it depends upon. If the Tier Information box is included in a Multiview Group box, the covered bitstream subset consists of the target output views of the multiview group and all the views required for decoding the target output views.

B.2.1.2 Syntax

```
class TierInfoBox extends Box('tiri') { //Mandatory Box
  unsigned int(16) tierID;
  unsigned int(8) profileIndication;
  unsigned int(8) profile_compatibility;
  unsigned int(8) levelIndication;
  unsigned int(8) reserved = 0;

  unsigned int(16) visualWidth;
  unsigned int(16) visualHeight;

  unsigned int(2) discardable;
  unsigned int(2) constantFrameRate;
  unsigned int(4) reserved = 0;
  unsigned int(16) frameRate;
}
```

B.2.1.3 Semantics

`tierID` gives the identifier of the tier, when the Tier Information box is included a Scalable Group entry or a Multiview Group entry. Otherwise, the semantics of `tierID` are unspecified, and in this case, `tierID` must be set to the reserved value 0.

`profileIndication` contains the `profile_idc` as defined in ISO/IEC 14496-10, when the parameter applies to the covered bitstream subset.

`profile_compatibility` is a byte defined exactly the same as the byte which occurs between the `profile_idc` and `level_idc` in a sequence parameter set or a subset sequence parameter set, as defined in ISO/IEC 14496-10 Annex G or Annex H, when the parameters apply to the covered bitstream subset.

`levelIndication` contains the `level_idc` as defined in ISO/IEC 14496-10, when the parameter applies to the covered bitstream subset. If the Tier Information Box is included in a Multiview Group Entry, `levelIndication` shall be valid when all the views of the covered bitstream subset are target output views. If the Tier Information Box is included in a Multiview Group Box, `levelIndication` shall be valid when the views specified by the respective multiview group are the target output views. If `levelIndication` is equal to 0 for an MVC stream, the level that applies to the covered bitstream subset and operating with all the views being target output views is unspecified.

The profile, profile compatibility flags and level indicated by the fields `profileIndication`, `profile_compatibility`, and `levelIndication` specifies an interoperability point with which the covered bitstream subset, and, for MVC, operating with the target output views as specified in the semantics of `levelIndication`, is compatible.

`visualWidth` gives the value of the width of the coded picture (of an SVC stream), coded sub-picture (of an SVC stream), or coded view component (of an MVC stream) in luma pixels of the representation of this tier in the stream or any view component of the covered bitstream subset. A coded sub-picture consists of a proper subset of coded slices of a coded picture. A tier may

consist of only sub-pictures. In this case, the tier is referred to as a sub-picture tier. A sub-picture tier may represent a region-of-interest part of the region represented by the entire stream.

NOTE The tier representation of a sub-picture tier might not be a valid stream. One example is as follows. An AVC bitstream is encoded using two slice groups. The first slice group includes the macroblocks representing a region-of-interest and is coded without referring to slices in the other slice group for inter prediction over all the access units. The slices of the first slice group in each access unit then form a sub-picture and a sub-picture tier can be specified to include all the sub-pictures over all the access units.

`visualHeight` gives the value of the height of the coded picture (of an SVC stream), coded sub-picture (of an SVC stream), or coded view component (of an MVC stream) in luma pixels of the representation of this tier in the stream or any view component of the covered bitstream subset.

`discardable` takes one of the following values; the value 02 is reserved.

- 00 this tier does not contain NAL units with `discardable_flag` (for SVC) equal to 1 or `inter_view_flag` (for MVC) equal to 0.
- 01 this tier contains both NAL units with `discardable_flag` (for SVC) equal to 1 or `inter_view_flag` (for MVC) equal to 0 and `discardable_flag` (for SVC) equal to 0 or `inter_view_flag` (for MVC) equal to 1.
- 03 all NAL units in this tier are with `discardable_flag` (for SVC) equal to 1 or `inter_view_flag` (for MVC) equal to 0.

`constantFrameRate` specifies if the frame rate of this tier is constant. A value of 0 denotes a non-constant frame rate, a value of 1 denotes a constant frame rate and a value of 2 denotes that it is not clear whether the frame rate is constant. A value of 3 is reserved.

`frameRate` gives the frame rate when the bitstream corresponding to this tier and all the lower tiers that this tier depends on is decoded in frames per second rounded to the closest integer using the Round function specified in ISO/IEC 14496-10. If `constantFrameRate` has a value of 0 or 2 then `frameRate` gives the average frame rate. If `constantFrameRate` has a value of 1 then `frameRate` gives the constant frame rate. `frameRate` equal to 0 indicates an unspecified frame rate. For SVC streams, decoded frames, complementary field pairs and non-paired fields are regarded as frames when deriving the value of `frameRate`. For MVC streams, decoded view components of any single view only are regarded as frames when deriving the value of `frameRate`, regardless of the total number of the views, since all output views are required to have simultaneous view components.

B.2.2 Tier bit rate box

B.2.2.1 Definition

Box Type: 'tibr'
 Container: ScalableGroupEntry or MultiviewGroupEntry or MultiviewGroupBox
 Mandatory: No
 Quantity: Zero or One

When included in a Scalable Group entry or a Multiview Group entry, the tier bit rate box provides information about the bit rate values of a tier. Two sets of information are provided: for the tier representation, including all the tiers on which the current tier depends, and for the tier alone. Similarly, for each set of information, the following values are supplied:

for SVC streams, the lowest long-term average bit rate that this tier could deliver. Let `maxDid` be the greatest `dependency_id` for all NAL units of the tier, and `minQid` be the least `quality_id` for all the NAL units of the tier and having `dependency_id` equal to `maxDid`. The following NAL units of this tier are not

considered in calculating this bit rate value: those having `dependency_id` equal to `maxDid` and `quality_id` greater than `minQid`. For MVC streams, the lowest long-term average bit rate that this tier could deliver is equal to the long-term average bit rate of the tier, when all NAL units of the tier are considered.

the long-term average bit rate of the tier; all NAL units of the tier are considered.

the maximum, or peak, bit rate of the tier; all NAL units of the tier are considered.

When included in a Multiview Group box, the tier bit rate box provides information about the bit rate values of the covered bitstream subset consisting of the target output views indicated by the multiview group and all the views required for decoding of the target output views. The maximum and long-term average bit rate for the covered bitstream subset are provided.

B.2.2.2 Syntax

```
class TierBitRateBox extends Box('tibr') {
    unsigned int(32) baseBitRate;
    unsigned int(32) maxBitRate;
    unsigned int(32) avgBitRate;

    unsigned int(32) tierBaseBitRate;
    unsigned int(32) tierMaxBitRate;
    unsigned int(32) tierAvgBitRate;
}
```

B.2.2.3 Semantics

`baseBitRate` gives the lowest long-term average bit rate in bits/second of the stream made from this tier and the lower tiers this tier depends on over the entire stream.

For SVC streams, `baseBitRate` is derived as follows. Let `maxDid` be the greatest `dependency_id` for all NAL units of the tier, and `minQid` be the least `quality_id` for all NAL units of the tier and having `dependency_id` equal to `maxDid`. The NAL units that are taken into account when calculating this bit rate value are as follows: 1) all NAL units of the tier except for those having `dependency_id` equal to `maxDid` and `quality_id` greater than `minQid`; 2) all NAL units of the lower tiers the current tier depends on.

For MVC streams, `baseBitRate` shall be equal to `avgBitRate`.

`maxBitRate` gives the maximum bit rate in bits/second of the stream containing all NAL unit mapped to this tier and the lower tiers this tier depends on, over any window of one second. All NAL units in this tier and the lower tiers this tier depends on are taken into account.

`avgBitRate` gives the long-term average bit rate in bits/second of the stream containing all NAL unit mapped to this tier and the lower tiers this tier depends on, averaged over the entire stream. All NAL units in this tier and the lower tiers this tier depends on are taken into account.

`tierBaseBitRate` gives the lowest long-term average bit rate in bits/second of the stream made from only this tier over the entire stream. For SVC streams, the set of NAL units that are taken into account when calculating this bit rate value is the same as for `baseBitRate` but excluding all NAL units of the lower tiers this tier depends on. For MVC streams, `tierBaseBitRate` shall be equal to `tierAvgBitRate`.

`tierMaxBitRate` gives the maximum bit rate in bits/second that is provided by only this tier over any window of one second. All NAL units mapped to this tier are taken into account. All NAL units of the lower tiers this tier depends on are not considered.

`tierAvgBitRate` - gives the long-term average bit rate in bits/second that is provided by only this tier, averaged over the entire stream. All NAL units mapped to this tier are taken into account. All NAL units of the lower tiers this tier depends on are not considered.

B.2.3 Priority range

B.2.3.1 Definition

Box Type: 'svpr'
 Container: ScalableGroupEntry or MultiviewGroupEntry
 Mandatory: Yes
 Quantity: Exactly One

NOTE - this box was previously called SVCPriorityRangeBox.

This box reports the minimum and maximum `priority_id` of the NAL units mapped to this tier.

B.2.3.2 Syntax

```
class PriorityRangeBox extends Box('svpr') {
    unsigned int(2) reserved1 = 0;
    unsigned int(6) min_priorityId;
    unsigned int(2) reserved2 = 0;
    unsigned int(6) max_priorityId;
}
```

B.2.3.3 Semantics

`min_priority_id`, `max_priority_id` take the minimum or maximum value of the `priority_id` syntax element that is present in the NAL unit header extension of the SVC or MVC NAL units mapped to the tier. For AVC streams this takes the value that is, or would be, in the prefix NAL unit.

B.2.4 SVC dependency range

B.2.4.1 Definition

Box Types: 'svdr'
 Container: ScalableGroupEntry
 Mandatory: Yes
 Quantity: Exactly One

This box reports the minimum and maximum `dependency_id` of the NAL units mapped to this tier.

The field `min_temporal_id` reports the minimum value of `temporal_id` of the NAL units in the tier having `dependency_id` equal to `min_dependency_id`. Similarly the field `min_quality_id` reports the minimum `quality_id` of those NAL units. The fields `max_temporal_id` and `max_quality_id` similarly report on the maximum values of the respective fields in those NAL units having `dependency_id` equal to `max_dependency_id`.

B.2.4.2 Syntax

```
class SVCDependencyRangeBox extends Box('svop') {
    unsigned int(3) min_dependency_id;
    unsigned int(3) min_temporal_id;
    unsigned int(6) reserved = 0;
    unsigned int(4) min_quality_id;
    unsigned int(3) max_dependency_id;
    unsigned int(3) max_temporal_id;
    unsigned int(6) reserved = 0;
    unsigned int(4) max_quality_id;
}
```

B.2.4.3 Semantics

`min_dependency_id`, `max_dependency_id` take the minimum or maximum value of the `dependency_id` syntax element that is present in the scalable extension NAL unit header defined in the SVC video specification of the NAL units mapped to the tier. For AVC streams this takes the value that is, or would be, in the prefix NAL unit (note: this value is zero).

`min_temporal_id`, `max_temporal_id` take the minimum or value of the `temporal_id` syntax element that is present in the scalable extension NAL unit header defined in the SVC video specification of the NAL units mapped to the tier having `dependency_id` equal to `min_dependency_id` and `max_dependency_id` respectively. For AVC streams this takes the value that is, or would be, in the prefix NAL unit.

`min_quality_id`, `max_quality_id` take the minimum or value of the `quality_id` syntax element that is present in the scalable extension NAL unit header defined in the SVC video specification of the NAL units mapped to the tier having `dependency_id` equal to `min_dependency_id` and `max_dependency_id` respectively. For AVC streams this takes the value that is, or would be, in the prefix NAL unit.

B.2.5 Initial parameter sets box

B.2.5.1 Definition

Box Type: 'svip'
 Container: ScalableGroupEntry or MultiviewGroupEntry
 Mandatory: No
 Quantity: Zero or One

The SVC initial parameter sets box documents which parameter sets are needed for decoding this tier and all the lower tiers it depends on.

B.2.5.2 Syntax

```
class InitialParameterSetBox extends Box('svip') {
    unsigned int(8) sps_id_count;
    for (i=0; i< sps_id_count; i++)
        unsigned int(8) SPS_index;
    unsigned int(8) pps_id_count;
    for (i=0; i< pps_id_count; i++)
        unsigned int(8) PPS_index;
}
```

B.2.5.3 Semantics

`sps_id_count`, `pps_id_count` gives the number of entries in the following tables.

`SPS_index` specifies that the SPS or subset SPS with this index is needed for decoding this tier and all the lower tiers it depends on. These are 1-based indices into the arrays in the `SVCDecoderConfigurationRecord` or the `MVCDecoderConfigurationRecord`.

`PPS_index` specifies that the PPS with this index is needed for decoding this tier and all the lower tiers it depends on. These are 1-based indices into the arrays in the `SVCDecoderConfigurationRecord` or the `MVCDecoderConfigurationRecord`.

B.2.6 SVC rect region box

B.2.6.1 Definition

Box Type: 'rrgn'
 Container: ScalableGroupEntry
 Mandatory: No
 Quantity: Zero or One

The SVC rect region box documents the geometry information of the region represented by the current tier relative to the region represented by another tier. When extended spatial scalability was used to encode in the current tier a cropped region of another tier, then the geometry information of the cropped region can be signaled by this box. This box can also be used to signal the geometry information of a region-of-interest (ROI) when the current tier is a sub-picture tier. This area can either be static for all samples or vary at sample-by-sample basis. Note that it is possible that independent sub-pictures do not depend on all the tiers with lower tierID. In this case dependencies can be given with the tier dependency box.

B.2.6.2 Syntax

```
class RectRegionBox extends Box('rrgn') {
    unsigned int(16) base_region_tierID;
    unsigned int(1) dynamic_rect;
    unsigned int(7) reserved = 0;
    if(dynamic_rect == 0) {
        unsigned int(16) horizontal_offset;
        unsigned int(16) vertical_offset;
        unsigned int(16) region_width;
        unsigned int(16) region_height;
    }
}
```

B.2.6.3 Semantics

`base_region_tierID` gives the tierID value of the tier wherein the represented region is used as the base region for derivation of the region represented by the current tier.

`dynamic_rect` equal to 1 indicates that the region represented by the current tier is a dynamically changing rectangular part of the base region. Otherwise the region represented by the current tier is a fixed rectangular part of the base region.

`horizontal_offset` and `vertical_offset` give respectively the horizontal and vertical offsets of the top-left pixel of the rectangular region represented by the tier, in relative to the top-left pixel of the base region, in luma samples of the base region.

`region_width` and `region_height` give respectively the width and height of the rectangular region represented by the tier, in luma samples of the base region.

B.2.7 Buffering information box

B.2.7.1 Definition

Box Type: 'buff'
 Container: ScalableGroupEntry or MultiviewGroupEntry or MultiviewGroupBox
 Mandatory: No
 Quantity: Zero or One

The BufferingBox contains the buffer information of the covered bitstream subset. If the Buffering box is included in a Scalable Group entry or a Multiview Group entry, the covered bitstream subset consists of the tier and all tiers on which it depends. If the Buffering box is included in a Multiview Group box, the covered bitstream subset consists of the target output views of the multiview group and all the views required for decoding the target output views.

B.2.7.2 Syntax

```
class BufferingBox extends Box('buff') {
  unsigned int(16)    operating_point_count
  for (i = 0; i < operating_point_count; i++){
    unsigned int (32)  byte_rate
    unsigned int (32)  cpb_size
    unsigned int (32)  dpb_size
    unsigned int (32)  init_cpb_delay
    unsigned int (32)  init_dpb_delay
  }
}
```

B.2.7.3 Semantics

`operating_point_count` specifies the number of HRD operating points for the covered bitstream subset. Values of the HRD parameters are specified separately for each operating point. The value of `operating_point_count` shall be greater than 0.

`byte_rate` specifies the input byte rate (in bytes per second) to the coded picture buffer (CPB) of the HRD. The covered bitstream subset is constrained by the value of `BitRate` equal to `byte_rate * 8` for NAL HRD parameters as specified in ISO/IEC 14496-10. For VCL HRD parameters, the value of `BitRate` is equal to `byte_rate * 40 / 6`. The value of `byte_rate` shall be greater than 0.

`cpb_size` specifies the required size of the coded picture buffer in bytes. The covered bitstream subset is constrained by the value of `CpbSize` equal to `cpb_size * 8` for NAL HRD parameters as specified in ISO/IEC 14496-10. For VCL HRD parameters, the value of `CpbSize` is equal to `cpb_size * 40 / 6`.

At least one pair of values of `byte_rate` and `cpb_size` of the same operating point shall conform to the maximum bit rate and CPB size allowed by profile and level of the covered bitstream subset.

`dpb_size` specifies the required size of the decoded picture buffer (DPB), in unit of bytes. The covered bitstream subset is constrained by the value of `max_dec_frame_buffering` equal to $\text{Min}(16, \text{Floor}(\text{dpb_size}) / (\text{PicWidthMbs} * \text{FrameHeightInMbs} * 256 * \text{ChromaFormatFactor}))$ as specified in ISO/IEC 14496-10.

`init_cpb_delay` specifies the required delay between the time of arrival in the CPB of the first bit of the first access unit and the time of removal from the CPB of the first access unit. It is in units of a 90 kHz clock. The covered bitstream subset is constrained by the value of the nominal removal time of the first access unit from the CPB, $t_{r,n}(0)$, equal to `init_cpb_delay` as specified in ISO/IEC 14496-10.

`init_dpb_delay` specifies the required delay between the time of arrival in the DPB of the first decoded picture and the time of output from the DPB of the first decoded picture. It is in units of a 90 kHz clock. The covered bitstream subset is constrained by the value of `dpb_output_delay` for the first decoded picture in output order equal to `init_dpb_delay` as specified in ISO/IEC 14496-10 assuming that the clock tick variable, t_c , is equal to 1 / 90 000.

B.2.8 Tier dependency box

B.2.8.1 Definition

Box Type: 'ldep'
 Container: ScalableGroupEntry or MultiviewGroupEntry
 Mandatory: No for ScalableGroupEntry, Yes for MultiviewGroupEntry
 Quantity: Zero or One

The TierDependencyBox identifies the tiers that the current tier is dependent on.

B.2.8.2 Syntax

```
class TierDependencyBox extends Box('ldep'){
    unsigned int(16) entry_count;
    for (i=0; i < entry_count; i++)
        unsigned int(16) dependencyTierId;
}
```

B.2.8.3 Semantics

`dependencyTierId` gives the `tierId` of a tier on which the current tier is directly or indirectly dependent. Tier A is directly dependent on tier B if there is at least one NAL unit in tier A using inter prediction, inter-layer prediction, or inter-view prediction from tier B. Tier A is indirectly dependent on tier B if tier A is not directly dependent on tier B while decoding of tier A requires the presence of tier B. The value of `dependencyTierId` shall be smaller than the `tierId` of the current tier. The decoding of the current tier requires the presence of the tier indicated by `dependencyTierId`. All dependencies up to the tier with the lowest `tierId` shall be given with the TierDependencyBox.

B.2.9 SVC region of interest box

B.2.9.1 Definition

Box Type: 'iroi'
 Container: ScalableGroupEntry
 Mandatory: No
 Quantity: Zero or One

This box provides the geometry information of region-of-interest (ROI) divisions of the current tier, when the current tier is encoded to multiple (typically a large number of) independent rectangular ROIs.

NOTE This box is typically used for interactive ROI use cases, where the server can interactively transmit only the NAL units belonging to the ROIs requested by a client.

The assignment of NAL units to a ROI is done in a time parallel metadata track as specified in Annex C.

A ROI ID, denoted as `roi_id`, is specified for each ROI in a tier. If `iroi_type` is equal to 0, `roi_id` is equal to the index of a ROI in a ROI raster scan (see ISO/IEC 14496-10 for the definition of "raster scan" and the use of "macroblock raster scan") of the region represented by the tier starting with zero for the top-left ROI in the region. If `iroi_type` is equal to 1, `roi_id` is equal to the entry index `i` in the syntax of `IroiInfoBox()`. If `iroi_type` is equal to 2, `roi_id` is set to a number identifying the region of interest. In this case, the temporal metadata must contain statements mapping NAL units to `roi_ids`.

B.2.9.2 Syntax

```
class IroiInfoBox extends Box('iroi'){
    unsigned int(2) iroi_type;
    unsigned int(6) reserved = 0;
    if(iroi_type == 0) {
        unsigned int(8) grid_roi_mb_width;
        unsigned int(8) grid_roi_mb_height;
    }
    else if(iroi_type == 1){
        unsigned int(24) num_roi;
        for(int i=1; i<= num_roi; i++) {
            unsigned int(32) top_left_mb;
            unsigned int(8) roi_mb_width;
            unsigned int(8) roi_mb_height;
        }
    }
}
```

B.2.9.3 Semantics

`iroi_type` indicates the types of region division for all the ROIs. The value 0 indicates that all the ROIs (except possibly the right-most ones and the bottom-most ones) are of identical width and height. The value 1 indicates that the geometry information for each ROI is separately signalled. The value 2 indicates that the geometry can not be given. Values greater than 2 are reserved.

`grid_roi_mb_width` and `grid_roi_mb_height` indicate the width and height, respectively, in units of macroblocks, of the ROIs. All the ROIs have identical width and height, with the following exceptions.

When $(\text{PicWidthInMbs} \% \text{grid_roi_mb_width})$ is not equal to 0, the right-most ROIs have a width equal to $(\text{PicWidthInMbs} \% \text{grid_roi_mb_width})$ macroblocks.

When $(\text{PicHeightInMbs} \% \text{grid_roi_mb_height})$ is not equal to 0, the bottom-most ROIs have a height equal to $(\text{PicHeightInMbs} \% \text{grid_roi_mb_height})$ macroblocks.

Where `PicWidthInMbs` and `PicHeightInMbs` are the visual width and height of a decoded picture of the tier representation in units of macroblocks, respectively, as specified in ISO/IEC 14496-10, $(x \% y)$ returns the remainder of x divided by y .

`num_roi` indicates the number of ROIs in a coded picture of the tier representation.

`top_left_mb` specifies the macroblock address of the first macroblock in raster scan order in the ROI of the current entry. The value of `top_left_mb` shall be equal to the syntax element `first_mb_in_slice` of the coded slices that belong to the current tier and that cover the top-left macroblock of the ROI of the current entry.

`roi_mb_width` and `roi_mb_height` indicate the width and height, respectively, in unit of macroblocks, of the ROI of the current entry.

B.2.10 SVC lightweight transcoding Box

B.2.10.1 Definition

Box Type: 'tran'
 Container: ScalableGroupEntry
 Mandatory: No
 Quantity: Zero or One

The presence of the box indicates that the bitstream represented by this tier (and tiers it depends upon) can be transcoded from an SVC stream to an AVC stream as indicated, and that the transcoded bitstream can be given the indicated profile and level indicators, with the indicated bit rates. The information on the resulting profile, level, and bit rate may be given for either of the entropy coding systems, or both.

B.2.10.2 Syntax

```
class TranscodingInfoBox extends Box('tran'){
  unsigned int(4) reserved = 0;
  unsigned int(2) conversion_idc;
  unsigned int(1) cavlc_info_present_flag;
  unsigned int(1) cabac_info_present_flag;
  if(cavlc_info_present_flag){
    unsigned int(24) cavlc_profile_level_idc;
    unsigned int(32) cavlc_max_bitrate;
    unsigned int(32) cavlc_avg_bitrate;
  }
  if(cabac_info_present_flag){
    unsigned int(24) cabac_profile_level_idc;
    unsigned int(32) cabac_max_bitrate;
    unsigned int(32) cabac_avg_bitrate;
  }
}
```

B.2.10.3 Semantics

`conversion_idc` equal to 0, 1, or 2 indicates that the representation of the current tier can be translated to an AVC bit-stream as specified in the semantics of the scalability information SEI message in ISO/IEC 14496-10 Annex G. `conversion_idc` equal to 3 is reserved.

`cavlc_info_present_flag` specifies whether the transcoding information of the translated bitstream using the Context-based Adaptive Variable Length Coding (CAVLC) entropy coder (i.e. when the syntax element `entropy_coding_mode_flag` in the translated bitstream is equal to 0) as specified in ISO/IEC 14496-10 Annex G is present.

`cabac_info_present_flag` specifies whether the transcoding information of the translated bitstream using the Context-based Adaptive Binary Arithmetic Coding (CABAC) entropy coder (i.e. when `entropy_coding_mode_flag` in the translated bitstream is equal to 1) as specified in ISO/IEC 14496-10 Annex G is present.

`cavlc_profile_level_idc` is the exact copy of the three bytes comprised of `profile_idc`, `constraint_set0_flag`, `constraint_set1_flag`, `constraint_set2_flag`, `constraint_set3_flag` and `level_idc`, if these syntax elements were used to specify the profile and level compliancy of the transcoded bitstream using the CAVLC entropy coder.

`cavlc_max_bitrate` specifies the maximum bit rate in bits/second (units of 1000 bits/sec), over any window of one second, that is provided by the transcoded bitstream using the CAVLC entropy coder.

`cavlc_avg_bitrate` specifies the average bit rate in bits/second (units of 1000 bits/sec) that is provided by the transcoded bitstream using the CAVLC entropy coder.