INTERNATIONAL STANDARD

ISO/IEC 14496-15

Second edition
2010-06-01

Information technology — Coding of audio-visual objects —

Part 15: Advanced Video Coding (AVC) file format

Technologies de l'information — Codage des objets audiovisuels —
Partie 15: Format de fichier de codage vidéo avancé (AVC)

Reference number
ISO/IEC 14496-15:2010(E)

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

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.

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 second edition cancels and replaces the first edition (ISO/IEC 14496-15:2004) which has been technically revised. It also incorporates the Amendments ISO/IEC 14496-15:2004/Amd.1:2006, ISO/IEC 14496-15:2004/Amd.2:2008, and the Technical Corrigenda ISO/IEC 14496-15:2004/Cor.1:2006, ISO/IEC 14496-15;2004/Cor.2:2006 and ISO/IEC 14496-15:2004/Cor.3:2009.

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: Advanced Video Coding (AVC) 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*

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 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 AVC video streams 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 AVC 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 AVC 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.

Extensions to the ISO Base Media File Format are defined here to support the new systems aspects of the AVC codec.

This International Standard defines the storage for plain AVC, SVC, and MVC video 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); SVC refers to ISO/IEC 14496-10 when the techniques in Annex G (Scalable Video Coding) are in use, and MVC refers to ISO/IEC 14496-10 when the techniques in Annex H (Multiview Video Coding) are in use. Specific techniques are introduced for handling of scalable and multiview streams, enabling their use, and assisting the extraction of subsets of scalable and multiview streams.

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 G.

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

Information technology — Coding of audio-visual objects —

Part 15:
Advanced Video Coding (AVC) file format

1 Scope

This part of ISO/IEC 14496 specifies the storage format for AVC (ISO/IEC 14496-10) video streams.

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.

The file format for storage of SVC content, as defined in Annexes A to E, and the file format for storage of MVC content, as defined in Annexes B to F, use the existing capabilities of the ISO base media file format and the plain AVC file format (i.e. the file format specified in Clauses 2 to 5 not including SVC and MVC supports specified in Annexes A to F). In addition, the following new extensions, among others, to support SVC- and/or 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 SVC or 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.

2 Normative references

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

ISO/IEC 14496-1:2001, *Information technology — Coding of audio-visual objects — Part 1: Systems*

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)]

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, this part of ISO/IEC 14496, A.2, F.2 and the following apply.

3.1.1
parameter set
sequence parameter set or a picture parameter set, as defined in ISO/IEC 14496-10

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

3.1.2
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

3.1.3
video elementary stream
elementary stream containing access units made up of NAL units for coded picture data

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)
FF	File Format
HRD	Hypothetical Reference Decoder
IDR	Instantaneous Decoding Refresh
MVC	MultiviewVideo 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

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

ROI	Region-Of-Interest
SEI	Supplementary Enhancement Information
SPS	Sequence Parameter Set
SVC	Scalable Video Coding [refers to ISO/IEC 14496-10 when the techniques in Annex G (Scalable Video Coding) are in use]
VCL	Video Coding Layer

4 Extensions to the ISO Base Media File Format

4.1 Introduction

The technologies originally documented in clause 4 are now defined in ISO/IEC 14496-12:2008 (technically identical to ISO/IEC 15444-12:2008).

4.2 File identification

See subclause 6.3 in ISO/IEC 14496-12.

4.3 Independent and Disposable Samples Box

See subclause 8.6.4 in ISO/IEC 14496-12 for the definition of this box.

4.4 Sample groups

See subclause 8.9 in ISO/IEC 14496-12.

4.5 Random access recovery points

See subclause 10.1 in ISO/IEC 14496-12.

4.6 Representation of new structures in movie fragments

See subclause 8.9.4 in ISO/IEC 14496-12.

5 AVC elementary streams and sample definitions

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

5.1 Elementary stream structure

AVC specifies 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 AVC content. Two types of elementary streams are defined for this purpose (see also Figure 1):

- **Video Elementary Streams** shall contain all video coding related NAL units (i.e. those NAL units containing video data or signaling 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 either one or two elementary streams:

- **Video elementary stream only**: In this case, sequence and picture parameter set NAL units shall be stored in the sample descriptions of this track. Sequence and picture parameter set NAL units shall not be part of AVC samples within the stream itself.

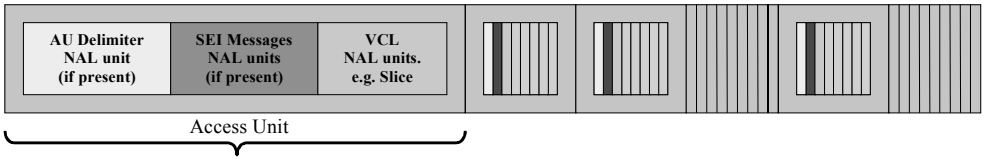
- **Video elementary stream and parameter set elementary stream**: In this case, sequence and picture parameter set NAL units shall be transmitted only in the parameter set elementary stream and shall neither be present in the sample descriptions nor the AVC samples of the video elementary stream.

The types of NAL units that are allowed in each of the video and parameter set elementary streams are specified in Table 1.

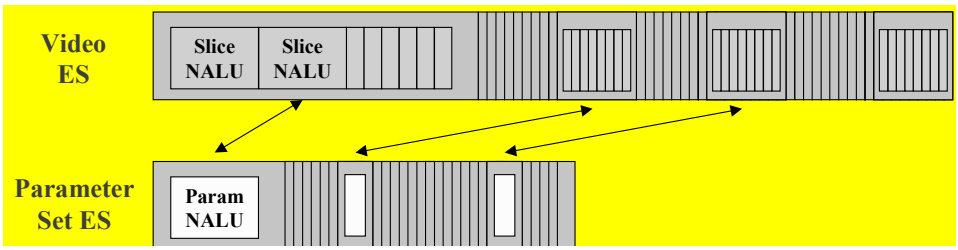
Table 1 — NAL unit types in elementary streams

Value of nal_unit_type	Description	Video elementary stream	Parameter set elementary stream
0	Unspecified	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	No
2	Coded slice data partition A slice_data_partition_a_layer_rbsp( )	Yes	No
3	Coded slice data partition B slice_data_partition_b_layer_rbsp( )	Yes	No
4	Coded slice data partition C slice_data_partition_c_layer_rbsp( )	Yes	No
5	Coded slice of an IDR picture slice_layer_without_partitioning_rbsp( )	Yes	No
6	Supplemental enhancement information(SEI) sei_rbsp( )	Yes. Except for the Sub-sequence, layering or Filler 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
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
9	Access unit delimiter (AU Delimiter) access_unit_delimiter_rbsp( )	Yes	No
10	End of sequence end_of_seq_rbsp()	Yes	No
11	End of stream end_of_stream_rbsp()	Yes	No
12	Filler data (FD) filler_data_rbsp( )	No	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
14…18	Reserved	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	No

20…23	Reserved	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



(a) Single video elementary stream containing NAL units



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

Figure 1 — AVC elementary stream structure

5.2 Sample and Configuration definition

5.2.1 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.

5.2.2 Canonical order and restrictions

The AVC 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 AVC elementary stream that satisfies the following conditions:

- **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. Each AVC sample shall contain at least one video data NAL unit of the primary picture.

- **SEI message NAL units**: All SEI message NAL units shall be contained in the sample whose decoding time is that before which the SEI messages come into effect instantaneously, or in the parameter set arrays. The order of SEI messages within a sample is as defined in ISO/IEC 14496-10, 7.4.1.2. This means that the 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.

- **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.

- **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 Parameter sets are stored either in the sample descriptions 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 descriptions 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.

- 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 ISO/IEC 14496-10.

- **Parameter set track:** A sync sample in a parameter set track indicates that all parameter sets needed from that 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 AVC 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.

- **All timing information is external to stream.** Picture Timing SEI messages that define presentation or composition timestamps may be included in the AVC video elementary stream, as this message contains 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 AVC layer. Timing provided within the AVC 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 AVC stream.

- **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 5.3.12.

- **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.

- **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 5.2.3; 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.

- **No 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.

  NOTE The removal of Filler Data NAL units, start codes, zero_byte syntax elements, 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.

5.2.3 AVC sample structure definition

This subclause defines structure for the samples of AVC streams. Samples are externally framed and have a size supplied by that external framing. An example of the structure of an AVC sample is depicted in Figure 2.

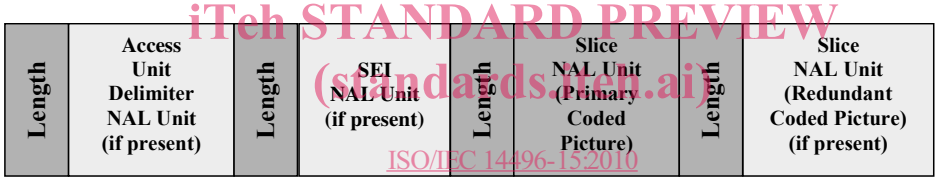


Figure 2 — The structure of an AVC sample

An AVC 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 ISO/IEC 14496-10.

5.2.4 Decoder configuration information

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

5.2.4.1 AVC decoder configuration record

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).

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 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 stream conforms to. If the stream is not examined, or the examination reveals that there is no profile to which the 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.