ISO/IEC JTC 1/SC 29/WG 03 N0453

**ISO/IEC JTC 1/SC 29/WG 03  
MPEG Systems   
Convenorship: KATS (Korea, Republic of)**

**Document type:** Output Document

**Title:** Technologies under Consideration for NAL File Format

**Status:** Approved

**Date of document:** 2022-01-2121

**Source:** ISO/IEC JTC 1/SC 29/WG 03

**No. of pages:** 39 (with cover page)

**Email of Convenor:** young.L @ samsung . com

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg3>

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 03 MPEG SYSTEMS**

**ISO/IEC JTC 1/SC 29/WG 03 N0453**

**January 20222, Virtual**

|  |  |
| --- | --- |
| **Title** | **Technologies under Consideration for NAL File Format** |
| **Source** | **WG 03, MPEG Systems** |
| **Status** | **Approved** |
| **Serial Number** | **21169** |

[1 Advanced parameter sets rewriting support (from m56112) 1](#_Toc93672672)

[2 Signalling of non-reference layer (from m56045) 5](#_Toc93672673)

[3 Subpicture treated as picture signalling (from m56044) 6](#_Toc93672674)

[4 VVC independent layers support NALU FF (from m54283) 7](#_Toc93672675)

[5 APS Roll Recovery (from m54403) 30](#_Toc93672676)

[6 14496-15 (NAL video file formats) errata items (from m55192) 31](#_Toc93672677)

[7 Visual width and height (from m56800) 32](#_Toc93672678)

[8 EDRAP and Random access (from m56766) 33](#_Toc93672679)

[9 14496-15 VVC conformance test vectors update (from m57436) 34](#_Toc93672680)

[10 Manifest and Prefix SEIs for AVC (#155) 35](#_Toc93672681)

[11 Generic Codecs Parameter (from m59046) 36](#_Toc93672682)

# Advanced parameter sets rewriting support (from m56112)

To improve the signalling of ‘spor’ sample group to have more information that can further assist rewriting of parameter set, not just subpicture id, in subpicture extraction and merging cases, the following changes are proposed:

1. A flag called subpicture\_param\_rewriting\_flag is signalled in ‘spor’ sample group to specify whether or not parameter set rewriting may be needed.
2. When subpicture\_param\_rewriting\_flag is equal to 0, it is constrained that the value of subpic\_id\_info\_flag is equal to 0.
3. When subpicture\_param\_rewriting\_flag is equal to 1, additional two flags are signalled.
   * sps\_id\_present\_flag to specify whether SPS id information is present or not. When sps\_id\_present\_flag is equal to 1, sps\_id is signalled and the SPS with id equal to sps\_id may need to be rewritten.
   * pps\_id\_present\_flag to specify whether PPS id information is present or not. When pps\_id\_present\_flag is equal to 1, pps\_id is signalled and the PPS with id equal to pps\_id may need to be rewritten.
   * The value of sps\_id\_present\_flag and pps\_id\_present\_flag shall not both be equal to 0 when the value of subpicture\_param\_rewriting\_flag is equal to 1.
4. When both sps\_id\_present\_flag and pps\_id\_present\_flag are equal to 1, the flag pps\_sps\_subpic\_id\_flag is present. Otherwise (either sps\_id\_present\_flag or pps\_id\_present\_flag is equal to 0), the following applies:
   * If pps\_id\_present\_flag is equal to 1, the value of pps\_sps\_subpic\_id\_flag is equal to 1.
   * Otherwise, the value of pps\_sps\_subpic\_id\_flag is equal to 0.

**Spec text for the proposals**

...

aligned(8) class VvcSubpicOrderEntry() extends VisualSampleGroupEntry('spor')  
{  
 unsigned int(1) subpic\_param\_rewriting\_flag;  
 unsigned int(1) subpic\_id\_info\_flag;  
 ~~unsigned int(5) reserved = 0;~~ unsigned int(4) reserved = 0;  
 unsigned int(10) num\_subpic\_ref\_idx;  
 for (i = 0; i < num\_subpic\_ref\_idx; i++)  
 unsigned int(16) subp\_track\_ref\_idx;  
 if (subpic\_param\_rewriting\_flag) {  
 unsigned int(1) sps\_id\_present\_flag;  
 if (sps\_id\_present\_flag)  
 bit(3) reserved = 0;  
 unsigned int(4) sps\_id;  
 else {  
 bit(7) reserved = 0;  
 }  
 unsigned int(1) pps\_id\_present\_flag;  
 if (pps\_id\_present\_flag)  
 bit(1) reserved = 0;  
 unsigned int(6) pps\_id;  
 else {  
 bit(7) reserved = 0;  
 }  
 }  
 if (subpic\_id\_info\_flag) {  
 unsigned int(4) subpic\_id\_len\_minus1;  
 unsigned int(12) subpic\_id\_bit\_pos;  
 unsigned int(1) start\_code\_emul\_flag;  
 if (pps\_id\_present\_flag && sps\_id\_present\_flag) {  
 bit(6) reserved = 0;  
 unsigned int(1) pps\_sps\_subpic\_id\_flag;  
 }  
 else  
 bit(7) reserved = 0;  
 ~~if (pps\_sps\_subpic\_id\_flag)  
 unsigned int(6) pps\_id;  
 else {  
 unsigned int(4) sps\_id;  
 bit(2) reserved = 0;  
 }~~ }  
}

subpic\_param\_rewriting\_flag equal to 0 specifies that no parameter set (i.e., SPS and PPS) rewriting is needed. subpic\_param\_rewriting\_flag equal to 1 specifies that parameter set (i.e., SPS and PPS) rewriting may be needed.

subpic\_id\_info\_flag equal to 0 specifies that the subpicture ID values explicitly signalled in the SPSs and/or PPSs or inferred without explicit signalling are correct for the indicated set of subp\_track\_ref\_idx values and thus no rewriting of SPSs or PPSs is required. subpic\_info\_flag equal to 1 specifies that SPSs and/or PPSs may need rewriting to indicate the subpictures corresponding to the set of subp\_track\_ref\_idx values and the subpicture ID values are not inferred. When the value of subpic\_param\_rewriting\_flag is equal to 0, the value of subpic\_id\_info\_flag shall be equal to 0.

num\_subpic\_ref\_idx greater than 0 specifies the number of reference indices of subpicture tracks or track groups of subpicture tracks referenced by the VVC track. num\_subpic\_ref\_idx equal to 0 specifies that 1) the number of reference indices of subpicture tracks or track groups of subpicture tracks referenced by the VVC track is equal to the number of entries in the 'subp' track reference of the VVC track, and 2) the order of subpicture tracks or track groups of subpicture tracks to be resolved for inclusion into the VVC bitstream reconstructed from the VVC track is the same as the order of the entries in the 'subp' track reference of the VVC track. The value of num\_subpics\_ref\_idx shall be less than or equal to 600.

It is constrained to that the total number of subpictures in the subpicture tracks referred to by the current VVC base track shall be less than or equal to 600.

subp\_track\_ref\_idx, for each value of i, specifies a 'subp' track reference index of the i-th list of one or more subpictures or slices to be included in the VVC bitstream reconstructed from the VVC track.

sps\_id\_present\_flag, when present, specifies whether the syntax element sps\_id is present.

sps\_id, when present, specifies the PPS ID of the PPS applying to the samples mapped to this sample group description entry.

pps\_id\_present\_flag, when present, specifies whether the syntax element pps\_id is present.

pps\_id, when present, specifies the PPS ID of the PPS applying to the samples mapped to this sample group description entry.

subpic\_id\_len\_minus1 plus 1 specifies the number of bits in subpicture identifier syntax elements in PPS or SPS, whichever is referenced by this structure.

subpic\_id\_bit\_pos specifies the bit position starting from 0 of the first bit of the first subpicture ID syntax element in the referenced PPS or SPS RBSP.

start\_code\_emul\_flag equal to 0 specifies that start code emulation prevention bytes are not present before or within subpicture IDs in the referenced PPS or SPS NAL unit. start\_code\_emul\_flag equal to 1 specifies that start code emulation prevention bytes may be present before or within subpicture IDs in the referenced PPS or SPS NAL unit.

pps\_sps\_subpic\_id\_flag, when equal to 1, specifies that the PPS NAL units applying to the samples mapped to this sample group description entry contain subpicture ID syntax elements. pps\_sps\_subpic\_id\_flag, when equal to 0, specifies that the SPS NAL units applying to the samples mapped to this sample group description entry contain subpicture ID syntax elements. When not present the value of pps\_sps\_subpic\_id\_flag is inferred as follows:

* If pps\_id\_present\_flag is equal to 1 and sps\_id\_present\_flag is equal to 0, the value of pps\_sps\_subpic\_id\_flag is inferred to be equal to 1.
* Otherwise, the value of pps\_sps\_subpic\_id\_flag is inferred to be equal to 0.

~~pps\_id, when present, specifies the PPS ID of the PPS applying to the samples mapped to this sample group description entry.~~

~~sps\_id, when present, specifies the SPS ID of the SPS applying to the samples mapped to this sample group description entry.~~

...

In subclause 11.6.3, add the highlighted text showsn below

...

If a reader selects VVC subpicture tracks containing VVC subpictures with a set of subpicture ID values that is the initial selection or differs from the previous selection, the following steps may be taken:

* The 'spor' sample group description entry is studied to conclude whether a PPS or SPS NAL unit needs to be changed.

NOTE 4: An SPS change is only possible at the start of a CLVS.

NOTE: PPS and / or SPS NAL unit needs to be changed may be indicated by the value of subpic\_param\_rewriting\_flag to be equal to 1.

* If the 'spor' sample group description entry indicates that start code emulation prevention bytes are present before or within the subpicture IDs in the containing NAL unit, an RBSP is derived from the NAL unit (i.e., start code emulation prevention bytes are removed). After the overriding in the next step, start code emulation prevention is re-done.
* The reader uses the bit position and subpicture ID length information in the 'spor' sample group entry to conclude which bits are overwritten to update the subpicture IDs to the selected ones.
* When the subpicture ID values of a PPS or SPS are initially selected, the reader needs to rewrite the PPS or SPS, respectively, with the selected subpicture ID values in the reconstructed access unit.
* When the subpicture ID values of a PPS or SPS are changed compared to the previous PPS or SPS (respectively) with the same PPS ID value or SPS ID value (respectively), the reader needs to include a copy of that previous PPS and SPS (if the PPS or SPS with that same PPS or SPS ID value, respectively, is not present in the access unit otherwise) and rewrite the PPS or SPS (respectively) with the updated subpicture ID values in the reconstructed access unit.

# Signalling of non-reference layer (from m56045)

We propose signaling of a sample group for non-reference samples. Non-reference samples are never used as a reference samples. A layer non reference flag is signaled in the sample group. It is asserted that knowing information about non-reference pictures sample group to indicate samples which are not used as a reference is useful. For example, this information can be used to drop samples and not decode them if the playback is lagging on a resource-constrained system. Additionally, this information can be used by media aware network element to drop samples if there is bandwidth starvation.

The proposed addition is shown below compared to WG03N0035 [1].

* + 1. **Non reference sample group** 
       1. **Definition**

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

This sample group is used to mark non reference samples. A non reference sample is never used as a reference sample for any other sample. An accompanying instance of the SampleGroupDescriptionBox with the same grouping type shall be present. The grouping\_type\_parameter is not defined for the SampleToGroupBox with grouping type 'nref'.

* + - 1. **Syntax**

class NonReferenceEntry() extends VisualSampleGroupEntry ('nref')  
{  
 bit(7) reserved = 0;  
 unsigned int(1) layer\_non\_ref\_only\_flag;  
}

layer\_non\_ref\_only\_flag equal to 1 specifies that for this sample a current picture is never used as a reference picture for any picture within its own layer and may or may not be used a reference picture for inter-layer prediction for pictures of a different layer. layer\_non\_ref\_only\_flag equal to 0 specifies that the current sample is never used as a reference sample.

# Subpicture treated as picture signalling (from m56044)

For subpicture ID sample group signaling, it is proposed to specify for each listed subpicture whether it is treated as a picture. It is asserted that this information is useful to know which subpictures mapped to the sample group description entry could be independently decoded.

The proposed modifications are shown highlighted compared to WG03N0035 [1]

* + 1. **Subpicture ID sample group**
       1. **Definition**

This sample group may be present in VVC subpicture tracks that contain VVC subpictures. This sample group shall not be present in tracks that are not VVC subpicture tracks or do not contain VVC subpictures. Each sample group description entry indicates a list of one or more subpicture ID values in decoding order that are contained in samples mapped to this sample group description entry.

* + - 1. **Syntax**

aligned(8) class VvcSubpicIDEntry() extends VisualSampleGroupEntry('spid')  
{  
 unsigned int(8) num\_subpics\_minus1;  
 for (i = 0; i <= num\_subpics\_minus1; i++) {  
 unsigned int(16) subpic\_id;   
 bit(7) reserved = 0;   
 unsigned int(1) subpic\_treated\_as\_pic\_flag;  
 }  
}

* + - 1. **Semantics**

num\_subpics\_minus1 plus 1 specifies the number of VVC subpictures contained in the sample mapped to this sample group description entry.

subpic\_id for loop counter i specifies the i-th subpicture identifier, in decoding order, contained in the sample mapped to this sample group description entry. The value of subpic\_id shall not be equal to any other value of subpic\_id in the same VvcSubpicIDEntry.

subpic\_treated\_as\_pic\_flag for loop counter i specifies when equal to 1, that the i-th subpicture, in decoding order, contained in the sample mapped to this sample group description entry is treated as a picture in the decoding process excluding in-loop filtering operations. subpic\_treated\_as\_pic\_flag for loop counter i specifies when equal to 0, that the i-th subpicture, in decoding order, contained in the sample mapped to this sample group description entry is not treated as a picture in the decoding process excluding in-loop filtering operations.

# VVC independent layers support NALU FF (from [m54283](http://wg11.sc29.org/doc_end_user/current_document.php?id=75130&id_meeting=183))

## Motivation

### VVC independent layers

At the beginning of the standardisation of VVC, the need of “high-level tiles” was clearly motivated for certain applications e.g.:

* Region of interest
* Multi-party conferencing
* VPCC
* Viewport-dependent

For some applications, the need for different frame rates, resolutions, etc motivated the definition of layers as in the traditional scalability design but layers that mutually intendent from each other and with no enhancement of it. In a way, one could consider a VVC bitstream with multiple VVC independent layers as a bitstream with multiple of traditional base layers with no enhancement layers.

Note that the traditional scalability design is present in VVC which makes the independent layers coincide with the concept of a “base” layer even though the term base layer does not exist in VVC.

Therefore, it is important to observe that those two concepts, scalability and independent layers, intersect in VVC for the definition of an independent layer but they address completely orthogonal use cases. It may thus call for different integrations in the system layer to better suits applications.

### System designs

For the sake of this contribution, we introduce two different types of pipelines that we see coexisting in modern application. The first one we called media-based and the second one is GPU-based.

The following describes each type of pipelines.

### Media-based pipeline

In a media-based processing, the output of the video decoders is fed to the presentation engine. The presentation engine will take care of proper aspect ratio and A/V synchronisation notably, based on a timing information.

ISOBMFF

FF reader

VVC Decoder

Presentation

*Figure 1 - Media-based system architecture*

Note that two or more video sequences may be synchronised by the presentation engines, even up to frame precision.

### GPU-based pipeline

In GPU-based pipeline, the output of a video decoder is processed by a GPU. Then the GPU process is responsible for providing the video buffer for the presentation engine as depicted in Figure 2.

…

ISOBMFF

FF reader

VVC Decoder

Presentation

GPU processing

VVC Decoder

*Figure 2 - GPU-based system architecture*

The decoded video sequence is fundamentally transformed by the GPU process, e.g. omnidirectional video to viewport video, 3D object information (texture, depth, …) to a viewport video. It goes way behind simple upscaling as it may happen in the media-based pipeline.

The video buffer that the GPU feeds to the presentation engine is by nature determined by runtime consideration such as the user’s viewport, the available decoding resources, the available bandwidth resource, the battery level, etc…

As a result, it is by essence impossible for the content creator to anticipate the operating point desired by the application because of the causality issue.

In addition, the output of several video decoders that are to be fed to a GPU process needs to sample-aligned. That is if a source texture was split and encoded as two coded data stream and two decoders decode them, these two pieces need to form the initial source texture with frame accuracy. To be more precise, since this needs to happen before the GPU and after the decoder, the natural criteria for this alignment is the Picture Order Count (POC).

### Summary of proposal

The proposal can be summarised as follows:

* We distinguish scalability and independent layer-based applications which nothing in common besides the use of the concept of layer.
* We define a new sample entry 'vidl' corresponding to a track made of a single VVC independent layer.
* For independent layer-based applications:
  + VVC independent layers are stored in separate tracks (no multiplexing).
  + The operating points is derived by the application at runtime and not present in any track.
  + VVC independent layer tracks are grouped together via their shared VVC base track to allow successful reconstruction into a single VVC bitstream.
  + A new track reference 'indl' is defined for VVC base tracks to reference the VVC independent layer tracks. The VVC base track contain shared parameter sets.
  + The parameter sets present in the VVC independent track overwrite the ones of the same type present in the VVC base track.
  + The VVC base track is also used for independent layers and not only for subpicture tracks.

***Proposed text change (only contains impacted section based on*** [***w19278***](https://wg11.sc29.org/doc_end_user/current_document.php?id=74736&id_meeting=182)***)***

[Editor's note: The following section is out of now out of sync with the latest draft of 14492-15 and should be dealt with at the next meeting.]

* + 1. **Types of VVC tracks**

This specification includes the following tools for carriage of VVC bitstreams:

1. VVC track:  
   A VVC track represents a VVC bitstream by including NAL units in its samples and sample entries and possibly by referencing other VVC tracks containing other sublayers of the VVC bitstream and possibly by referencing VVC subpicture tracks. When a VVC track references VVC subpicture tracks or VVC independent layer tracks, it is referred to as a VVC base track.
2. VVC non-VCL track:  
   APSs, which carry ALF, LMCS, or scaling list parameters, and other non-VCL NAL units can be stored in and transmitted through a track that is separate from the track containing the VCL NAL units; this is the VVC non-VCL track.
3. VVC subpicture track:  
   A VVC subpicture track contains either of the following:

A sequence of one or more VVC subpictures.

A sequence of one or more complete slices forming a rectangular area.

The VVC subpictures or slices included in any sample of a VVC subpicture track are contiguous in decoding order.

NOTE: VVC non-VCL tracks and VVC subpicture tracks enable an optimal delivery of VVC video in streaming applications as follows. These tracks can each be carried in DASH representations of their own, and for decoding and rendering of a subset of the tracks, the DASH representations containing the subset of the VVC subpicture tracks as well as the DASH representation containing the non-VCL tracks can be requested by the client, segment by segment. This way, redundant transmission of APSs and other non-VCL NAL units can be avoided.

1. VVC independent layer track:  
   A VVC independent layer track contains samples all belonging to the same layer and this layer is independent.
   * 1. **Overview of VVC storage with multiple layers**

The support for VVC bitstream with multiple layers includes a number of tools, and there are various ‘models’ of how they might be used. An VVC stream with multiple layers can be placed in tracks in several ways, among which are the following:

1. all the layers in one track;
2. each layer or sub-layer in its own track;
3. the expected operating points each in a track

The VVC file format allows storage of one or more layers into a track. Storage of multiple layers per track can be used. For example, when a content provider wants to provide a multi-layer bitstream that is not intended for subsetting, or when the bitstream has been created for a few pre-defined sets of output layers where each layer corresponds to a view (for example stereo pair), tracks can be created accordingly.

When a VVC bitstream is represented by multiple tracks and a player uses an operating point for which the layers are stored in multiple tracks, the player must reconstruct VVC access units before passing them to the VVC decoder.

When a VVC bitstream is represented by multiple VVC independent layer track, the player determines the set of layers (tracks) to be selected and the player may reconstruct VVC access units before passing them to the VVC decoder. Alternatively, the player may reconstruct several bitstreams and pass them to several VVC decoders.

A VVC operating point may be explicitly represented by a track, i.e., each sample in the track contains an access unit. The access unit contains NAL units from all the layers and sub-layers that are part of the operating point.

The storage of VVC bitstreams is supported by structures such as the

1. sample entry,
2. Operating Points Information ('vopi') sample group, and
3. Layer Information ('linf') sample group.

The structures within a sample entry provide information for the decoding or use of the samples, in this case coded video and non-VCL data information, that are associated with that sample entry.

The Operating Points Information sample group records information about operating points such as the layers and sub-layers that constitute the operating point, dependencies (if any) between them, the profile, level, and tier parameter of the operating point, and other such operating point relevant information.

The layer information sample group lists all the layers and sub-layers carried in the samples of the track.

The information in these sample groups, combined with using track references to find tracks, is sufficient for a reader to choose an operating point in accordance with its capabilities, identify the tracks that contain the relevant layers and sub-layers needed to decode the chosen operating point, and efficiently extract them.

* 1. ***Elementary stream structure***

Three types of elementary streams are defined for storing VVC 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;

**Non-VCL elementary stream** that contains non-VCL NAL units synchronized with the elementary stream carried in the video track.

* 1. ***Sample and configuration definition***
     1. **Overview**

VVC sample (of a video track): A VVC sample contains an access unit as defined in clause 3.1 of ISO/IEC 23090-3.

VVC layer sample (of a video track): A VVC layer sample contains a picture unit as defined in clause 3.108 of ISO/IEC 23090-3.

VVC subpicture sample: A VVC subpicture sample contains either of the following:

One or more complete subpictures as specified in ISO/IEC 23090-3 that are contiguous in decoding order.

One or more complete slices as specified in ISO/IEC 23090-3 that form a rectangular area and are contiguous in decoding order.

VVC non-VCL sample: A VVC non-VCL sample is a sample in a non-VCL stream that consist of one or more non-VCL NAL units that are to be considered as if present in the video elementary stream at the same instant in time.

* + 1. **Canonical order and restrictions**

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

**Access unit delimiter NAL units**: The constraints obeyed by access unit delimiter NAL units are defined in ISO/IEC 23090-3.

* **DCI NAL units, VPSs, SPSs, and PPSs***:* A VPS, SPS, or PPS 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 DCI NAL units, VPSs, SPSs, and PPSs, shall be stored only in the sample entry when the sample entry name is 'vvc1' or 'vidl', and may be stored in the sample entry and the samples when the sample entry name is 'vvi1'.

NOTE 1: Storing DCI NAL units, VPSs, SPSs, and PPSs in the sample entries of a video stream provides a simple and static way to supply decoding capability information and parameter sets. Storing these NAL units 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 these parameter sets in the sample entry, and then updates using these parameter sets as they occur in the stream, starting from any sample marked as a sync sample. Such updating can replace these parameter sets with a new definition using the same identifier. Each time the sample entry changes, the decoder re-initializes with these parameter sets included in the sample entry.

**APSs***:* An APS to be used in a slice must be sent prior to the VCL NAL unit containing that slice. For a video stream that a particular sample entry applies to, the following applies:

If the track has a track reference of type 'vvcN', the APSs shall be stored only in the VVC non-VCL track that is referred to be the track reference of type 'vvcN'.

Otherwise, if the sample entry name is 'vvc1' or 'vidl', the APSs shall be stored only in the sample entry.

Otherwise (the track does not have a track reference of type 'vvcN' and the sample entry name is 'vvi1'), the APSs shall be stored only in the sample entry and the samples.

**SEI messages:** SEI messages of declarative nature may be stored in the sample entry; there is no prescription about removing such SEI messages from the samples.

**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 in-stream parameter sets.

NOTE 2: The removal or addition of Filler Data NAL units, start codes, SEI messages or Filler Data SEI messages can change the bitstream characteristics with respect to conformance with the HRD when operating the HRD in CBR mode as specified in ISO/IEC 23090-3, Annex C.

* + 1. **Decoder configuration information**
       1. **VVC decoder configuration record**
          1. ***Definition for single-layer bitstream storage***

This subclause specifies the decoder configuration information for ISO/IEC 23090-3 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 is supplied by the structure that 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 shall 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\_idc, general\_tier\_flag, general\_sub\_profile\_idc, general\_constraint\_info, general\_level\_idc, chroma\_format\_idc, bit\_depth\_luma\_minus8, and bit\_depth\_chroma\_minus8 shall 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 profile indication general\_profile\_idc shall indicate a profile to which the stream associated with this configuration record conforms.

NOTE 1: If the SPSs are marked with different profiles, then the stream could 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 shall be split into two or more sub-streams with separate configuration records in which these rules can be met.

The tier indication general\_tier\_flag shall indicate a tier equal to or greater than the highest tier indicated in all the parameter sets.

Each flag in general\_profile\_compatibility\_flags may only be set if all the parameter sets set that bit.

Each bit in general\_constraint\_info may only be set if all the parameter sets set that bit.

The level indication general\_level\_idc shall indicate a level of capability equal to or greater than the highest level indicated for the highest tier in all the parameter sets.

The value of chroma\_format\_idc in all the parameter sets shall be identical.

The value of bit\_depth\_luma\_minus8 in all the parameter sets shall be identical.

The value of bit\_depth\_chroma\_minus8 in all the parameter sets shall be identical.

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

There is a set of arrays to carry initialization NAL units. The NAL unit types are restricted to indicate VPS, SPS, PPS, APS, prefix SEI, and suffix SEI NAL units only. NAL unit types that are reserved in ISO/IEC 23090-3 and in this specification may acquire a definition in future, and readers should ignore arrays with reserved or unpermitted values of NAL unit type. [Ed. (MH): Excluding suffix APS NAL units and suffix SEI NAL units from the allowed NAL units in the decoder configuration record should be considered. If suffix APS NAL units or suffix SEI NAL units are present in the decoder configuration record, how does a reader know what their correct position is in the reconstructed bitstream?]

NOTE 2: 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 DCI, VPS, SPS, PPS, APS, prefix SEI, suffix SEI.

* + - * 1. ***Definition for multi-layer bitstream storage***

When a VVC bitstream with multiple layers is present in a track, the following extensions/restrictions specified in this subclause apply.

The values taken by the following four parameters in the VVC decoder configuration record is defined as below

1. general\_profile\_idc,
2. general\_tier\_flag
3. general\_level\_idc
4. general\_constraint\_info

If the VVC track contains one or more independent layers and there is an output layer set that only contains the layer with the lowest nuh\_layer\_id among these layers, then the values taken by the above parameters shall be the one that applies to the layer with the lowest nuh\_layer\_id. Otherwise, the values taken by the above parameters shall be based on an output layer set (operating point) to which the track is associated.

The values taken by the following three parameters in the VVC decoder configuration record is defined as below

1. chroma\_format\_idc
2. bit\_depth\_luma\_minus8
3. bit\_depth\_chroma\_minus8

If the VVC track contains multiple independent layers, then the values taken by the above parameters shall be the one that applies to the layer with the lowest nuh\_layer\_id.

The semantics of the fields in VVCDecoderConfigurationRecord remain unchanged.

* + - * 1. ***Syntax***

aligned(8) class VvcDecoderConfigurationRecord {  
 unsigned int(8) configurationVersion = 1;  
 unsigned int(7) general\_profile\_idc;  
 unsigned int(1) general\_tier\_flag;  
 unsigned int(24) general\_sub\_profile\_idc;  
 unsigned int(8) num\_bytes\_constraint\_info;  
 unsigned int(8\*num\_bytes\_constraint\_info) general\_constraint\_info;  
 unsigned int(8) general\_level\_idc;  
 bit(6) reserved = '111111'b;  
 unsigned int(2) chroma\_format\_idc;  
 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(16) avgFrameRate;  
 unsigned int(2) constantFrameRate;  
 unsigned int(3) numTemporalLayers;  
 unsigned int(2) lengthSizeMinusOne;   
 unsigned int(8) numOfArrays;  
 for (j=0; j < numOfArrays; j++) {  
 unsigned int(1) array\_completeness;  
 bit(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;  
 }  
 }  
}

* + - * 1. ***Semantics***

general\_profile\_idc, general\_tier\_flag, general\_sub\_profile\_idc, general\_constraint\_info, general\_level\_idc, chroma\_format\_idc, bit\_depth\_luma\_minus8 and bit\_depth\_chroma\_minus8 contain the matching values for the fields general\_profile\_idc, general\_tier\_flag, general\_sub\_profile\_idc, the bits in general\_constraint\_info( ), general\_level\_idc, chroma\_format\_idc, bit\_depth\_luma\_minus8, and bit\_depth\_chroma\_minus8 as defined in ISO/IEC 23090-3, for the stream to which this configuration record applies. [Ed. (MH): Add sub\_layer\_level\_idc[ i ] fields into the syntax?]

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 sublayer or sublayer in ISO/IEC 23090-3) 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.

lengthSizeMinusOne plus 1 indicates the length in bytes of the NALUnitLength field in a VVC video stream 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 shall be all of that type); it takes a value as defined in ISO/IEC 23090-2; it is restricted to take one of the values indicating a VPS, SPS, PPS, APS, prefix SEI, or suffix 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 shall 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 a DCI, VPS, SPS, PPS, APS or declarative SEI NAL unit, as specified in ISO/IEC 23090-3.

* 1. ***Data sharing and reconstruction VVC bitstream***
     1. **General**

In order to reconstruct an access unit from samples of multiple tracks carrying a multi-layer VVC bitstream, an operating point needs to be determined first.

NOTE: Players can select an operating point from operating point list included in the Operating Points Information sample group. Tracks that carry the relevant layers for an operating point can be obtained by following the 'oref' track references and information in the layer information sample group. Otherwise, players can self determine the adequate set of layers to be selected at run time.

In order to reconstruct a bitstream from multiple VVC tracks carrying a VVC bitstream, the target highest value TemporalId may need to be determined first.

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

When a VVC bitstream is represented by multiple VVC tracks, the decoding times of the samples shall be such that if the tracks were combined into a single stream ordered by increasing decoding time, the access unit order would be correct as specified in ISO/IEC 23090-3.

A sequence of access units is reconstructed from the respective samples in the required tracks according to the implicit reconstruction process as described in 11.6.2.

* + 1. **Implicit reconstruction of a VVC bitstream**

When the Operating Points Information is present, the required tracks are selected based on the layers they carry and their reference layers as indicated by the Operating Points Information and Layer Information sample groups.

When reconstructing a bitstream containing a sublayer for which the VCL NAL units have TemporalId greater than 0, all lower sublayers (i.e., those for which the VCL NAL units have smaller TemporalId) within the same layer are also included in the resulting bitstream and the required tracks are selected accordingly.

When reconstructing an access unit, picture units (as specified in ISO/IEC 23090-3) from samples having the same decoding time are placed into the access unit in increasing order of the nuh\_layer\_id value.

When reconstructing an access unit with dependent layers and max\_tid\_il\_ref\_pics\_plus1 is greater than 0, the sublayers of the reference layers for which the VCL NAL units have TemporalId less than or equal to max\_tid\_il\_ref\_pics\_plus1 - 1 (as indicated in the operating point information sample group) within the same layer are also included in the resulting bitstream and the required tracks are selected accordingly.

When reconstructing an access unit with dependent layers and max\_tid\_il\_ref\_pics\_plus1 is equal to 0, only IRAP picture units of the reference layers are included in the resulting bitstream and the required tracks are selected accordingly.

If a VVC track contains a 'subp' track reference, each picture unit is reconstructed as specified in clause 11.7.3 with the additional constraints on EOS and EOB NAL units as specified below. The process in clause 11.7.3 is repeated for each layer of the target operating point in increasing nuh\_layer\_id order. Otherwise, each picture unit is reconstructed as described below.

The reconstructed access units are placed into the VVC bitstream in increasing order of decoding time, and the duplicates of end of bitstream (EOB) and end of sequence (EOS) NAL units are removed from the VVC bitstream, as described further below.

The access units are placed into the output bitstream in increasing order of the decoding time.

For access units that are within the same coded video sequence of a VVC bitstream and that belong to different sublayers stored in multiple tracks, there may be more than one of the tracks containing an EOS NAL unit with a particular nuh\_layer\_id value in the respective samples. In this case, only one of the EOS NAL units shall be kept in the last of these access units (the one with the greatest decoding time) in the final reconstructed bitstream, placed after all NAL units, except the EOB NAL unit (when present), of the last of these access units, and other EOS NAL units are discarded. Similarly, there may be more than one of such tracks containing an EOB NAL unit in the respective samples. In this case, only one of the EOB NAL units shall be kept in the final reconstructed bitstream, placed at the end of the last of these access units, and other EOB NAL units are discarded.

Since a particular layer or sublayer may be represented by more than one track, when figuring out the required tracks for an operating point, a selection may need to be made among the set of tracks that all carry the particular layer or sublayer.

The final required tracks, after selection among the tracks carrying a same layer or sublayer, may still collectively carry some layers or sublayers that do not belong to the target operating point. The reconstructed bitstream for the target operating point should not contain the layers or sublayers that are carried in the final required tracks but do not belong to the target operating point.

NOTE: Some VVC decoder implementations take as input a bitstream as well as the target output layer set index and the highest TemporalId value of the target operating point, which correspond to the TargetOlsIdx and Htid variables in clause 8 of ISO/IEC 23090-3, respectively. In this case the bitstream reconstructed by the file parser can contain layers or sublayers that do not belong to the target operating point, because these VVC decoders are capable of removing the layers not included in the target output layer set and the sublayers beyond the highest TemporalId value. Some other VVC decoder implementations input a bitstream that is required not to contain any other layers and sublayers than those included in the target operating point. In this case the file parser needs to ensure that the reconstructed bitstream does not contain any other layers and sublayers than those included in the target operating point.

* + 1. **Reconstructing a picture unit from a sample in a VVC track referencing VVC subpicture tracks**

A sample of a VVC track is resolved to an access unit that contains the following NAL units in the order of the bullets:

* The AUD NAL unit, if any, when present (and the first NAL unit) in the sample.
* When the sample is the first sample of a sequence of samples associated with the same sample entry, the parameter set and SEI NAL units contained in the sample entry, if any.
* The NAL units present in the sample up to and including the PH NAL unit.
* The content of the time-aligned (in decoding time) resolved sample from each referenced VVC subpicture track in the order specified in the 'spor' sample group description entry mapped to this sample, excluding all VPS, DCI, SPS, PPS, AUD, PH, EOS, and EOB NAL units, if any. The track references are resolved as specified below.

NOTE 1: If the referenced VVC subpicture track is associated with a VVC non-VCL track, the resolved sample of the VVC subpicture track contains the non-VCL NAL unit(s), if any, of the time-aligned sample in the VVC non-VCL track.

* The remaining NAL units (not included in the access unit already above) in the sample.

The 'subp' track reference indices of a 'spor' sample group description entry are resolved as follows:

* If the track reference points to a track ID of a VVC subpicture track, the track reference is resolved to the VVC subpicture track.
* Otherwise (the track reference points to an 'alte' track group), the track reference is resolved to any of the tracks of the 'alte' track group. If a particular track reference index value was resolved to a particular track in the previous sample, it shall be resolved in the current sample to either of the following:
  + the same particular track, or
  + any other track in the same 'alte' track group that contains a sync sample that is time-aligned with the current sample.

NOTE 2: The VVC subpicture tracks in the same 'alte' track group have to be independent of any other VVC subpicture tracks referenced by the same VVC base track to avoid decoding mismatches and could therefore be constrained as follows:

* + All the VVC subpicture tracks contain VVC subpictures.
  + The subpicture boundaries are like picture boundaries.
  + Loop filtering is turned off across subpicture boundaries.

If a reader selects VVC subpicture tracks containing VVC subpictures with a set of subpicture ID values that differs from the previous selection, the following steps may be taken:

* The 'spor' sample group description entry is studied to conclude whether a PPS or SPS NAL unit needs to be changed.

NOTE: An SPS change is only possible at the start of a CLVS.

* If the 'spor' sample group description entry indicates that start code emulation prevention bytes are present before or within the subpicture IDs in the containing NAL unit, an RBSP is derived from the NAL unit (i.e., start code emulation prevention bytes are removed). After the overriding in the next step, start code emulation prevention is re-done.
* The reader uses the bit position and subpicture ID length information in the 'spor' sample group entry to conclude which bits are overwritten to update the subpicture IDs to the selected ones.
* When the subpicture ID values of a PPS or SPS are changed compared to the previous PPS or SPS (respectively) with the same PPS ID value or SPS ID value (respectively), the reader needs to include a copy of the PPS and SPS with the updated subpicture ID values in the reconstructed access unit.
  1. ***Derivation from ISO base media file format***
     1. **VVC video stream definition**
        1. **Sample entry name and format**
           1. ***Definition***

Sample Entry Types: 'vvc1', 'vvi1', 'vvs1', 'vidl'  
Box Types: 'vvcC', 'vvnC', 'supp'  
Container: Sample Table Box ('stbl')  
Mandatory: An 'vvc1' or 'vvi1' sample entry is mandatory  
Quantity: One or more sample entries may be present

A VVC sample entry has sample entry type equal to 'vvc1','vvi1' or 'vidl'. A VVC visual sample entry shall contain a VVC Configuration Box, as defined below. This includes a VvcDecoderConfigurationRecord, as defined in ‎11.3.3.1.

An optional BitRateBox may be present in the VVC visual sample entry to signal the bit rate information of the VVC 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.

If the VVC subpicture track is suitable to be decoded with a VVC decoder and to be consumed without other VVC subpicture tracks, a regular VVC sample entry is used ('vvc1' or 'vvi1') for the VVC subpicture track. Otherwise, 'vvs1' sample entry is used for the VVC subpicture track, and the following constraints apply for the track:

* track\_in\_movie flag shall be equal to 0.
* The track shall contain one and only one sample entry.
* VPS, DCI, SPS, PPS, AUD, PH, EOS, and EOB NAL units shall be absent both in the sample entry and in the samples of 'vvs1' tracks.
* Unless indicated otherwise, child boxes of video sample entries (such as CleanApertureBox and PixelAspectRatioBox) should not be present in the sample entry and, if present, shall be ignored.
* A sample shall not be marked as a sync sample unless all the VCL NAL units it contains conform to the sync sample requirements of Table 10.
* Composition time offset information for samples of a 'vvs1' track shall not be present.
* Subsample information for samples of a 'vvs1' track may be present; if present, the subsample information shall follow the definition of sub samples for VVC.

When the sample entry type is equal to 'vvc1', 'vvi1' or 'vidl', VisualSampleEntry.width and VisualSampleEntry.height shall be respectively equal to maximum pic\_width\_max\_in\_luma\_samples and maximum pic\_height\_max\_in\_luma\_samples of the SPSs referenced by the samples mapped to this sample entry.

A VVC track may contain a 'subp' track reference, with entries containing either a track\_ID value of a VVC subpicture track or a track\_group\_id value of an 'alte' track group of VVC subpicture tracks.

A VVC track may contain a 'indl' track reference, with entries containing a track\_ID value of a VVC independent layer track.

When a VVC track contains a 'subp' or a 'indl'track reference, it is referred to as a VVC base track and the following applies:

* The samples of the VVC track shall not contain VCL NAL units.
* A sample group of type 'spor', as specified in clause 11.7.7, shall be present when 'subp' track reference

[Ed. (MH): A dedicated sample entry type for a VVC base track would enable indicating a VVC base track through the codecs MIME parameter of the track type. On the other hand, it might be undesirable to specify a great number of VVC sample entry types.]

The sample entry of type 'vvs1' shall contain VvcNALUConfigBox.

When the VVC subpicture track is referenced by a VVC base track containing a 'spor' sample group description entry having subpic\_id\_info\_flag equal to 1, either the sample entry of a VVC subpicture track shall contain SubpicPropertiesBox or the track shall contain a subpicture ID sample group. When SubpicPropertiesBox is present in a sample entry, all the samples within the scope of the sample entry contain VVC subpictures with the indicated subpicture ID values.

When the sample entry name is 'vvc1', 'vvi1' or 'vidl', the stream to which this sample entry applies shall be a compliant VVC stream as viewed by a VVC decoder operating under the configuration (including profile, tier, and level) given in the VVCConfigurationBox.

When the sample entry name is 'vidl', the track shall be a VVC indepedent layer track and shall not contain a 'vopi' sample group. The num\_layers\_in\_track in the layer information sample group 'linf' shall is equal to 1.When the sample entry name is 'vvc1' or 'vidl', 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 'vvi1', the default value of array\_completeness is 0 for all arrays.

When the samples of a track do not contain the sublayer with TemporalId equal to 0 and contain one or more sublayers with TemporalId greater than 0, the track shall contain an 'oref' track reference to a track that contains the sublayer with TemporalId equal to 0.

* + - * 1. ***Syntax***

class VvcConfigurationBox extends Box('vvcC') {  
 VvcDecoderConfigurationRecord() VvcConfig;  
}

class VvcNALUConfigBox extends Box('vvnC') {  
 unsigned int(6) reserved=0;   
 unsigned int(2) lengthSizeMinusOne;   
}

class SubpicPropertiesBox extends FullBox('supp', 0, 0) {  
 unsigned int(8) num\_subpics\_minus1;  
 for (i = 0; i <= num\_subpics\_minus1; i++)  
 unsigned int(16) subpic\_id;  
}

class VvcSampleEntry() extends VisualSampleEntry ('vvc1' or 'vvi1'){  
 VvcConfigurationBox config;  
 MPEG4ExtensionDescriptorsBox (); // optional  
 SubpicPropertiesBox subpic\_properties; // optional  
}

class VvcSubpicSampleEntry() extends VisualSampleEntry ('vvs1'){  
 VvcNALUConfigBox config;  
 SubpicPropertiesBox subpic\_properties; // optional  
}

* + - * 1. ***Semantics***

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

VvcDecoderConfigurationRecord is defined in ‎11.3.3.

lengthSizeMinusOne plus 1 indicates the length in bytes of the NALUnitLength field in a track that contains the VvcNALUConfigBox. 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.

num\_subpics\_minus1 plus 1 specifies the number of subpicture sequences contained in the VVC subpicture track.

subpic\_id specifies the subpicture identifier of the sequence of subpictures contained in the VVC subpicture track.

* + 1. **VVC non-VCL stream definition**
       1. **Sample entry name and format**
          1. ***Definition***

Sample Entry Type: 'vvcN'  
Container: Sample Description Box ('stsd')  
Mandatory: Yes  
Quantity: One or more sample entries may be present

A VVC non-VCL sample entry shall contain a VvcNALUConfigBox.

When a 'vvcN' sample entry is present in a track, track\_in\_movie flag shall be equal to 0 for the track.

* + - * 1. ***Syntax***

class VvcNonVCLSampleEntry() extends VisualSampleEntry ('vvcN'){  
 VvcNALUConfigBox config;  
}

* + - * 1. ***Semantics***

Compressorname in the base class VisualSampleEntry indicates the name of the compressor used with the value "\013VVC non-VCL" being recommended (\013 is 11, the length of the string as a byte).

* + - 1. **Sample format**
         1. ***Definition***

This subclause defines the sample format for VVC non-VCL streams. A VVC non-VCL sample contains only one or more non-VCL NAL units. [Ed. (MH): It should be clarified which non-VCL NAL units are allowed. Carriage of DCI, VPS, SPS, and PPS NAL units should probably be disallowed.]

* + - * 1. ***Syntax***

aligned(8) class VvcNonVCLSample  
{  
 for (i=0; i<sample\_size; )  
 {  
 unsigned int((VvcDecoderConfigurationRecord.LengthSizeMinusOne+1)\*8)  
 nalUnitLength;  
 bit(nalUnitLength \* 8) nalUnit;  
 i += (VvcDecoderConfigurationRecord.LengthSizeMinusOne+1) +  
 nalUnitLength;  
 }  
}

* + - * 1. ***Semantics***

nalUnitLength indicates the size of a non-VCL NAL unit in bytes. The length field includes the size of both the two-byte NAL unit header and the NAL unit payload but does not include the length field itself.

nalUnit contains a single non-VCL NAL unit. The syntax of a non-VCL NAL unit is defined in ISO/IEC 23090-3 and includes both the two-byte NAL unit header and the NAL unit payload.

* + - 1. **Track reference**

A track reference of type 'vvcN' may be included in a VVC track (containing either a video elementary stream or a video and parameter set elementary stream) or a VVC subpicture track, referencing a VVC non-VCL track. When present, this track reference is used to connect from the VVC track or the VVC subpicture track to the non-VCL track.

When a video track, which may either be a VVC track or a VVC subpicture track, has a track reference of type 'vvcN', the following applies:

* A sample in the video track that is not a sync sample may or may not have a corresponding sample in the associated VVC non-VCL track having the same decoding time. For each sync sample in the video track, there shall be one and only one sample in the associated VVC non-VCL track having the same decoding time, and that sample in the associated VVC non-VCL track shall also be a sync sample.
* For each sample in the associated VVC non-VCL track, there shall be one and only one sample in the video track having the same decoding time. When a sample in the associated VVC non-VCL track is a sync sample, the corresponding sample in the video track having the same decoding time shall also be a sync sample.
  + 1. **Parameter sets in sample entry**

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

Each VVC sample entry, which contains the VVC 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: For VPSs, SPSs, and PPSs, it is recommended that when parameter set updating is desired, the parameter sets are included in the samples of the stream. For APSs, it is recommended that when parameter set updating is desired, which is expected to happen often, the APSs are included in the VVC non-VCL track.

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.

* + 1. **Sync sample**

For each sync sample in a VVC non-VCL track, all APSs needed for decoding of the corresponding video elementary stream from that decoding time forward are in that VVC non-VCL track sample or succeeding VVC non-VCL track samples.

A sync sample in 'vvc1' and 'vvi1' tracks shall contain VCL NAL units indicating that the coded picture in the sample is an Instantaneous Decoding Refresh (IDR) picture or a Clean Random Access (CRA) picture. When the coded picture in a sync sample is a CRA picture, there shall be no RASL pictures associated with that CRA picture.

NOTE: When there is no sync sample in a stream, the sync sample box is present and has entry\_count equal to 0. A VVC stream may contains no IDR or CRA pictures, but starts with a GRA picture. It is also possible that a VVC stream contains some CRA pictures but each of them has at least one associated RASP picture.

Table 10 indicates the mapping between VVC VCL NAL unit types, ISOBMFF sync sample status and SAP types as documented in ISOBMFF.

*Table 10 – Mapping of sync sample status and SAP types to NAL unit type*

|  |  |  |
| --- | --- | --- |
| NAL Unit Type | ISOBMFF sync sample status | DASH SAP type |
| IDR\_N\_LP | true | 1 |
| IDR\_W\_RADL | true | 2 (if the IRAP has associated RADL pictures) 1 (if the IRAP has no associated RADL pictures) |
| CRA | false  true  true | 3 (if the IRAP has associated RASL pictures)  2 (if the IRAP has no associated RASL pictures but has associated RADL pictures)  1 (if the IRAP has no associated leading pictures) |
| GDR | false | 4 |

When the sample entry name is 'vvc1' and the track does not have a track reference of type 'vvcN', the following applies:

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

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) shall not be used in a VVC 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 GDR based 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 pictures.

In the context of this clause, the leading samples, defined as part of the definition of the 'rap ' sample group in ISO/IEC 14496-12, contain Random Access Skipped Leading (RASL) access units as defined in ISO/IEC 23090-3.

* + 1. **Definition of a sub-sample for VVC**

For the use of the SubSampleInformationBox (8.7.7 of ISO/IEC 14496-12) in a VVC 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 VVC data, the ‘codec\_specific\_parameters’ field in the box 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: Tile-based sub-samples. A sub-sample contains one tile and the associated non-VCL NAL units, if any, of the VCL NAL unit(s) containing the tile.

2: CTU-row-based sub-samples. A sub-sample either contains one CTU row within a tile and the associated non-VCL NAL units, if any, of the VCL NAL unit(s) containing the CTU row. This type of sub-sample information shall not be used when entropy\_coding\_sync\_enabled\_flag is equal to 0.

3: Slice-based sub-samples. A sub-sample either contains one slice (i.e., one VCL NAL unit) and the associated non-VCL NAL units, if any.

4: Subpicture based sub-samples. A sub-sample contains one coded subpicture and the associated non-VCL NAL Units, if any, of the VCL NAL unit(s) containing the subpicture.

5: Picture-based sub-samples. A sub-sample contains one coded picture and the associated 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 is still decodable 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 shall also be included in the same sub-sample.

The codec\_specific\_parameters field of the SubSampleInformationBox is defined for VVC as follows:

if (flags == 0) {  
 unsigned int(1) RapNalUnitFlag;  
 unsigned int(1) GraNalUnitFlag;  
 unsigned int(1) VclNalUnitFlag;  
 bit(29) reserved = 0;  
 else if (flags == 1) {  
 unsigned int(2) vcl\_idc;  
 bit(2) reserved = 0;  
 unsigned int(4) log2\_min\_luma\_ctb;  
 unsigned int(12) ctb\_x;  
 unsigned int(12) ctb\_y;  
 } else if (flags == 2 || flags == 3) {  
 unsigned int(2) vcl\_idc;  
 bit(30) reserved = 0;  
 } else if (flags == 4) {  
 unsigned int(2) vcl\_idc;  
 unsigned int(5) vcl\_nalu\_type;  
 bit(5) reserved = 0;  
 unsigned int(2) log2\_min\_luma\_ctb;  
 unsigned int(8) ctb\_x;  
 unsigned int(8) ctb\_y;  
 unsigned int(1) DiscardableFlag;   
 unsigned int(1) NoInterLayerPredFlag;   
 } else if (flags == 5){  
 unsigned int(1) DiscardableFlag;   
 unsigned int(1) NoInterLayerPredFlag;  
 bit(30) reserved = 0;  
 }

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, or CRA\_NUT as specified in ISO/IEC 23090-3. Value 1 indicates that all NAL units in the sub-sample have nal\_unit\_type equal to IDR\_W\_RADL, IDR\_N\_LP, or CRA\_NUT as specified in ISO/IEC 23090-3.

GraNalUnitFlag equal to 0 indicates that none of the NAL units in the sub-sample has nal\_unit\_type equal to GRA\_NUT as specified in ISO/IEC 23090-3. Value 1 indicates that all NAL units in the sub-sample have nal\_unit\_type equal to GRA\_NUT as specified in ISO/IEC 23090-3.

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.

DiscardableFlag indicates the non\_reference\_picture\_flag value of the VCL NAL units in the sub-sample. All the VCL NAL units in the sub-sample shall have the same non\_reference\_picture\_flag value.

NOTE: This is not the same definition as the discardable field in the SubSampleInformationBox.

NoInterLayerPredFlag equal to 1 indicates that the current subpicture or picture has no inter-layer reference pictures as active entries in reference picture lists 0 and 1. NoInterLayerPredFlag shall be equal to 1 for an independent layer. NoInterLayerPredFlag equal to 0 indicates that the current subpicture or picture may have inter-layer reference pictures as active entires in reference picture list 0 or 1.

* + 1. **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 that 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'),
   1. version 1 of the CompositionOffsetBox shall be used,
   2. the value of sample\_offset shall be set equal to -231 for each non-output sample,
   3. the CompositionToDecodeBox ('cslg') should be contained in the SampleTableBox ('stbl') of the track, and
   4. 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 -231.

* 1. ***Sample groups***
     1. **Stream access point sample group**

The stream access point (SAP) sample group 'sap ' specified in ISO/IEC 14496-12 is used to provide information of all SAPs.

The semantics of layer\_id\_method\_idc equal to 0 are specified in ISO/IEC 14496-12.

When layer\_id\_method\_idc is equal to 0, a SAP is interpreted as follows:

* If the sample entry type is 'vvc1' or 'vvi1', and the track does not contain any sublayer with TemporalId equal to 0, a SAP specifies access to all the sublayers present in the track.
* Otherwise, a SAP specifies access to all layers present in the track.

NOTE: If the sample entry type is 'vvc1' or 'vvi1', and the track does not contain any sub-layer with TemporalId equal to 0, an STSA picture with TemporalId equal to the lowest TemporalId present in the track serves as a SAP.

The semantics of layer\_id\_method\_idc equal to 1 are specified in clause 9.5.7.

When SAP type 4 is indicated in the 'sap ' sample group, the respective picture in the VVC bitstream shall be a GDR picture.

When SAP sample groups are used, they shall be used on all tracks carrying the same VVC bitstream. The SAP documentation shall be consistent for any given layer.

* + 1. **Random access recovery point sample group**

The random access recovery point sample group 'roll' specified in ISO/IEC 14496-12 is used to provide information on recovery points for gradual decoding refresh.

When a 'roll' sample group is used with VVC tracks, the syntax and semantics of grouping\_type\_parameter are specified identically to those for the 'sap ' sample group in ISO/IEC 14496-12.

When layer\_id\_method\_idc is equal to 0, the 'roll' sample group specifies the behaviour for all layers present in the track.

The semantics of layer\_id\_method\_idc equal to 1 are specified in clause 9.5.7.

When a 'roll' sample group concerns a dependent layer but not its reference layer(s), the sample group indicates characteristics that apply when all the reference layers of the dependent layer are available and decoded. The sample group can be used to initiate decoding of the predicted layer.

* + 1. **Random access point sample group**

The random access point sample group 'rap ' specified in ISO/IEC 14496-12 should not be used for VVC tracks.

* + 1. **Temporal level sample group**

The temporal level sample group 'tele' specified in ISO/IEC 14496-12 should be used to indicate the TemporalId value. When the 'tele' sample group is present in a VVC track, the pictures with TemporalId equal to tidValue shall be mapped to the sample group description index equal to tidValue + 1.

* + 1. **Step-wise sublayer access sample group**

The step-wise sublayer access sample group 'stsa' specified in clause 8.4.7 is used to mark step-wise sublayer access (STSA) pictures.

When an 'stsa' sample group is used with VVC tracks, the syntax and semantics of grouping\_type\_parameter are specified identically to those for the 'stsa' sample group in ISO/IEC 14496-12.

When layer\_id\_method\_idc is equal to 0, the 'stsa' sample group specifies the behaviour for all layers present in the track.

The semantics of layer\_id\_method\_idc equal to 1 are specified in clause 9.5.7.

* + 1. **Operating points information sample group**
       1. **Definition**

Applications are informed about the different operating points provided by a given VVC bitstream and their constitution by using the Operating Points Information sample group ('vopi'). Each operating point is related to an output layer set, a max TemporalId value, and a profile, level and tier signalling. All this information is captured by the 'vopi' sample group. Apart from this information, this sample group also provides the dependency information between layers.

Among all tracks of the same VVC bitstream, there shall be at most one track in this set (in case of subpictures it will be the VVC base track of the subpicture) that carries a 'vopi' sample group. All other tracks of the VVC bitstream in this set (in case of subpictures it will be the other VVC base tracks of the subpicture for dependent and independent layers) shall have a track reference of type 'oref' to the track that carries the 'vopi' sample group when present

For any specific sample in a given track, the temporally collocated sample in another track is defined as the one with the same decoding time as that of this specific sample. For each sample SN in a track TN that has an 'oref' track reference to the track Tk that carries the 'vopi' sample group, the follow applies:

* If there is a temporally collocated sample Sk in the track Tk, then the sample SN is associated with the same 'vopi' sample group entry as the sample Sk.
* Otherwise, the sample SN is associated with the same 'vopi' sample group entry as the last of the samples in the track Tk that precede the sample SN in decoding time.

When several VPSs are referenced by a VVC bitstream, it may be needed to include several entries in the sample group description box with grouping\_type 'vopi'. For more common cases where a single VPS is present, it is recommended to use the default sample group mechanism defined in ISO/IEC 14496-12 and include the operating points information sample group in the sample table box, rather than including it in each track fragment.

The grouping\_type\_parameter is not defined for the SampleToGroupBox with grouping type 'vopi'.

* + - 1. **Syntax**

class OperatingPointsRecord {  
 bit(2) reserved = 0;  
 unsigned int(6) num\_profile\_tier\_level;  
 for (i=1; i<=num\_profile\_tier\_level; i++) {  
 unsigned int(7) general\_profile\_idc;  
 unsigned int(1) general\_tier\_flag;  
 unsigned int(8) num\_bytes\_constraint\_info;  
 unsigned int(8\*num\_bytes\_constraint\_info) general\_constraint\_info;  
 unsigned int(8) general\_level\_idc;  
 unsigned int(8) num\_sub\_profiles;  
 for (i=1; i<= num\_sub\_profiles; i++)  
 unsigned int(32) general\_sub\_profile\_idc;  
 }  
 unsigned int(1) all\_independent\_layers\_flag;  
 bit(7) reserved = 0;  
 if (all\_independent\_layers\_flag){  
 unsigned int(1) each\_layer\_is\_an\_ols\_flag;  
 bit(7) reserved = 0;  
 } else  
 unsigned int(8) ols\_mode\_idc;   
 unsigned int(16) num\_operating\_points;  
 for (i=0; i<num\_operating\_points) {  
 unsigned int(16) output\_layer\_set\_idx;  
 unsigned int(8) ptl\_idx;  
 unsigned int(8) max\_temporal\_id;  
 unsigned int(8) layer\_count;  
 for (j=0; j<layer\_count; j++) {  
 unsigned int(6) layer\_id;  
 unsigned int(1) is\_outputlayer;  
 bit(1) reserved = 0;  
 }  
 bit(6) reserved = 0;  
 unsigned int(1) frame\_rate\_info\_flag  
 unsigned int(1) bit\_rate\_info\_flag  
 if (frame\_rate\_info\_flag) {  
 unsigned int(16) avgFrameRate;  
 bit(6) reserved = 0;  
 unsigned int(2) constantFrameRate;  
 }  
 if (bit\_rate\_info\_flag) {  
 unsigned int(32) maxBitRate;  
 unsigned int(32) avgBitRate;  
 }  
 }  
 unsigned int(8) max\_layer\_count;  
 for (i=0; i<max\_layer\_count; i++) {  
 unsigned int(8) layerID;  
 unsigned int(8) num\_direct\_ref\_layers;  
 for (j=0; j<num\_direct\_ref\_layers; j++)  
 unsigned int(8) direct\_ref\_layerID;  
 unsigned int(8) max\_tid\_il\_ref\_pics\_plus1;  
 }  
}

class OperatingPointsInformation extends VisualSampleGroupEntry ('vopi') {  
 OperatingPointsRecord oinf;  
}

* + - 1. **Semantics**

num\_profile\_tier\_level: Gives the number of following profiles, tier, and level combinations as well as the associated fields.

general\_profile\_idc, general\_tier\_flag, general\_constraint\_info, general\_level\_idc, num\_sub\_profiles, general\_sub\_profile\_idc, all\_independent\_layers\_flag, each\_layer\_is\_an\_ols\_flag, ols\_mode\_idc and max\_tid\_il\_ref\_pics\_plus1 are defined in ISO/IEC 23090-3.

num\_bytes\_constraint\_info indicates the number of bytes used for specifying general\_constraint\_info

num\_operating\_points: Gives the number of operating points for which the information follows.

output\_layer\_set\_idx is the index of the output layer set that defines the operating point. The mapping between output\_layer\_set\_idx and the layer\_id values shall be the same as specified by the VPS for an output layer set with index output\_layer\_set\_idx.

ptl\_idx: Signals the one-based index of the listed profile, level, and tier flags for a layer with identifier equal to layer\_id. When the value of ptl\_idx equals zero for a layer, that layer shall be assumed to have no profile, level, and tier signalled and that layer shall not be an output layer or a layer that is a direct or indirect reference layer of any output layer of the operating point.

max\_temporal\_id: Gives the maximum TemporalId of NAL units of this operating point.

Note: The maximum TemporalId value indicated in the layer information sample group has different semantics from the maximum TemporalId indicated here. However, they may carry the same literal numerical values.

layer\_count: This field indicates the number of necessary layers, as defined ISO/IEC 23090-3, of this operating point.

layer\_id: provides the nuh\_layer\_id values for the layers of the operating point.

is\_outputlayer: A flag that indicates if the layer is an output layer or not. A one indicates an output layer.

frame\_rate\_info\_flag equal to 0 indicates that no frame rate information is present for the operating point. The value 1 indicates that frame rate information is present for the operating point.

bit\_rate\_info\_flag equal to 0 indicates that no bitrate information is present for the operating point. The value 1 indicates that bitrate information is present for the operating point.

avgFrameRate gives the average frame rate in units of frames/(256 seconds) for the operating point. Value 0 indicates an unspecified average frame rate.

constantFrameRate equal to 1 indicates that the stream of the operating point is of constant frame rate. Value 2 indicates that the representation of each temporal layer in the stream of the operating point is of constant frame rate. Value 0 indicates that the stream of the operating point may or may not be of constant frame rate.

maxBitRate gives the maximum bit rate in bits/second of the stream of the operating point, over any window of one second.

avgBitRate gives the average bit rate in bits/second of the stream of the operating point.

max\_layer\_count: The count of all unique layers in all of the operating points that relate to this associated base track.

layerID: nuh\_layer\_id of a layer for which the all the direct reference layers are given in the following loop of direct\_ref\_layerID.

num\_direct\_ref\_layers: The number of direct reference layers for the layer with nuh\_layer\_id equal to layerID.

direct\_ref\_layerID: nuh\_layer\_id of the direct reference layer.

# APS Roll Recovery (from [m54403](http://wg11.sc29.org/doc_end_user/current_document.php?id=75250&id_meeting=183))

11.8.X APS Roll Recovery

11.8.X.1 Definition

The 'apsr' sample group indicates that a VVC sync sample from a VVC track requires additional gathering of prefix and suffix APS NAL units from preceding samples and rewriting these as APS prefix NAL units to be a self-contained sync sample. This avoids having to duplicate APS information at each sync sample in the track.

A sync sample not belonging to an 'apsr' sample group does not require any additional processing to gather the dependent APS (i.e., all APS required are in the sample entry or in the sample). The 'apsr' sample group shall only be present in a VVC track or a VVC base track with no dependencies to a VVC non-VCL track; it shall not be present in VVC subpicture tracks or VVC non-VCL tracks.

The aps\_roll\_count shall be such that all samples described by the roll operation are available in the track, track fragment or ISOBMFF segment (as indicated by the roll\_type) being processed.

A sample associated to an 'apsr' sample group description entry shall be a sync sample or a sample with SAP type 3 or 4 (potentially associated to a 'roll' sample group description entry).

A sample associated to an 'apsr' sample group description entry with aps\_roll\_count not equal to 0 shall be considered as a SAP type 4.

11.8.X.2 Syntax

aligned(8) class APSRollRecoveryEntry () extends VisualSampleGroupEntry('apsr')  
{  
 unsigned int(2) roll\_type;  
 unsigned int(2) roll\_mode;  
 unsigned int(4) reserved=0;  
 if (roll\_type == 0) {  
 unsigned int(16) aps\_roll\_count;  
 }

X.3 Semantics

roll\_type indicates the pre-roll distance for APS NAL units when producing a sync sample with sample number N belonging to this group. The following values are defined:

0: APS NAL units are gathered starting from the sample located aps\_roll\_count samples before the sample belonging to the group

1: APS NAL units are gathered starting from the first sample of the track or track fragment

2: APS NAL units are gathered starting from the first sample of the associated ISOBMFF segment

3: reserved

roll\_mode indicates which samples in the identified roll sample window should be analyzed for APS NAL unit gathering. The following values are defined:

0: the required APS NAL units may be present in any sample

1: the required APS NAL units are only present in the first sample

2: the required APS NAL units are only present in samples that are either sync samples or samples marked as 'rap '

3: the required APS NAL units are only present in samples of the current track fragment

aps\_roll\_count indicates the number of samples to rewind for APS gathering; a value of 0 indicates that the associated sample contains all APS NAL units for its processing.

# 14496-15 (NAL video file formats) errata items (from [m55192](https://wg11.sc29.org/doc_end_user/current_document.php?id=76249&id_meeting=184))

[Editor’s note] The following changes have been captured in the current TuC document in lack of a Defect Report on 14496-15 issued at MPEG #132. This paragraph is thus meant to be moved in a more appropriate document at the next meeting.

1. Search and replace "RBSP payload" (1 instance) with "NAL unit payload", because in the context where the phase is used, the start code emulation bytes that are part of "NAL unit payload" but not part of "RBSP payload" should be considered.
2. Search and replace "byte stream payload" (2 instances) with "NAL unit payload", for similar reason as above.
3. Change, in clause 8.4.1.1.1, the following:

"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."

to

"When the sample entry name is 'hvc1', the ~~default and mandatory~~ value of array\_completeness ~~is~~ shall be equal to 1 for the 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 shall be equal to 1 for the arrays of all types of parameter sets ~~is 0 for all arrays~~."

1. Change, in clause 9.5.3.1.1, the following:

"When the sample entry name is 'lhv1', 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 'lhe1', the default value of array\_completeness is 0 for all arrays."

to

"When the sample entry name is 'lhv1', the ~~default and mandatory~~ value of array\_completeness ~~is~~ shall be equal to 1 for the arrays of all types of parameter sets~~, and 0 for all other arrays~~. When the sample entry name is 'lhe1', the ~~default~~ value of array\_completeness shall be equal to 1 for the arrays of all types of parameter sets ~~is 0 for all arrays~~."

1. Change, in clause 8.3.3.1.3, the following:

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.

to the following:

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.

1. Change, in clause 8.3.3.1.1, the following:

The level indication general\_level\_idc shall indicate a level of capability equal to or greater than the highest level indicated for the highest tier in all the parameter sets.

to the following:

The level indication general\_level\_idc shall indicate a level of capability equal to or greater than the highest level ~~indicated for the highest tier~~ in all the parameter sets.

Because the highest level of the highest tier could be lower than the highest level of the lowest tier, while level determines spatial resolution etc., which is of vital importance for determining the required decoding capability.

# Visual width and height (from m56800)

Background

The following paragraph has been removed from ISO/IEC 14496-15 (MDS19981/WG03N00149).

Text

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

[Ed. (MH): FI\_14-015 & FR02-018 resolution: Remove the following paragraph and add it to the 14496-15 TuC.:

When there are samples whose exact cropped frame dimensions are less than the maximum cropped framed dimensions indicated in the associated VisualSampleEntry, the picture dimensions sample group specified in ISO/IEC 14496-12 should be present.

]

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

Note that the visual size in the SPS or PPS 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.

# EDRAP and Random access (from m56766)

The ‘edrp’ sample group from ISO/IEC 14496-12 TuC [1] proposes to signal additional random access point in bitstream that could outperform DRAP in coding efficiency thank to the possibility to refer to previous EDRAP for inter prediction as represented in the figure below:

Logo

Description automatically generated

*Figure 1: Inter prediction reference relationship among the RAP pictures in the EDRAP case.*

In current version of the Part-12 TuC, no information is provided in case an EDRAP depends on APS NAL units present in prior samples.

For example, the figure below is an example of such VVC bitstream.

A picture containing logo

Description automatically generated

When Random Accessing from the EDRAP6 sample, the ‘edrp’ sample group indicates that only IDR0 and EDRAP2 are needed for reference to decode the bitstream (yellow or grey arrows). The ‘edrp’ sample group is silent about possible references to APS NAL units in previous samples (red arrows). For example, EDRAP6 may reference APSs that are not part of the sample used as references in the EDRAP as per section 11.3.4 of ISO/IEC 14496-15 (recalled hereafter for convenience):

*When the sample entry name is 'vvc1'and the track does not have a track reference of type 'vvcN', the following applies:*

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

A screenshot of a video game

Description automatically generated

From the discussion at MPEG #134 (<http://mpegx.int-evry.fr/software/MPEG/Systems/FileFormat/NALuFF/-/issues/133>), the following observation has been made:

All non-VCL NAL units needed to decode the EDRAP should be referenced from the EDRAP sample group.

# 14496-15 VVC conformance test vectors update (from m57436)

The following table gives the list of contributed conformance test vectors. The VVC encoded bitstreams are conforming to v12.0 and/or v13.0 of the VTM software. The conformance files have been updated to fix some bugs and errors.

Conformance bitstreams can be accessed via the following link:

[**https://mpegfs.int-evry.fr/mpegcontent/**](https://mpegfs.int-evry.fr/mpegcontent/)under

**/MPEG-04/Part15-VVC\_File\_Format/ConformanceTestVectors/Nokia/**

Graphical user interface, text

Description automatically generated

Discussion is ongoing to cross-check the contributed test vectors: <http://mpegx.int-evry.fr/software/MPEG/Systems/FileFormat/NALuFF/-/issues/154>

# Manifest and Prefix SEIs for AVC (#155)

Tracked at: <http://mpegx.int-evry.fr/software/MPEG/Systems/FileFormat/NALuFF/-/issues/155>

## Initial question

The latest amendment to 14496-15 provides a recommendation for the carriage of SEI manifest and SEI prefix indication SEI NAL units for HEVC. However, AVC also allows those SEIs to be present in the elementary stream. Should we add the same recommendation for AVC carriage to 14496-15?

We could add the following paragraph to clause 5.3.3.1:

When one or more SEI NAL units containing an SEI manifest SEI message and/or an SEI prefix indication SEI message are available, they should be stored in the array of picture parameter sets. Such SEIs should also be placed in a parameter set elementary stream.

Or maybe it is even better to put such a recommendation in a general section of part 15 as this seems to apply for AVC, HEVC and maybe VVC?

## Discussion conclusion

Since both, Manifest and Prefix SEI apply to multiple codecs it was agreed to put this into a generic section.

# Generic Codecs Parameter (from m59046)

## EVC Codecs Parameter

DASH and other applications require defined values for the Codecs parameter specified in IETF RFC 6381 for ISO BMFF Media tracks. The 'codecs' parameter string for the EVC codec is defined as follows:

<sample entry 4CC>.<key1><value1>.<key2><value2>.….<keyN><valueN>

Keys are defined as 4CCs. An set of keys and the associated value pairs are defined in Table 1. Additional keys may be specified as 4CCs. Preferably, keys are aligned with ISO/IEC 23091-2. [Ed. (MH/KB): If the keys derived from ISO/IEC 23091-2 are kept in this document, a normative reference to ISO/IEC 23091-2 has to be added.]

If a specific key is not provided, then the value takes the default value specified in the table, or the value is unknown if no default is specified.

NOTE: The parameters from 'vbit' onwards in Table 1 are not EVC specific and are applicable to any video codec. It is foreseen that these values are moved to ISO/IEC 14496-12 in future versions and may then be applicable to other video codecs as well. [Ed. (MH): FI\_106-207 resolution: Keys from vbit onwards moved to ISO/IEC 14496-12.]

**Table 1 - Definition of Set of Keys and Values defined for EVC**

|  |  |  |  |
| --- | --- | --- | --- |
| **Key** | **Key Definition** | **Value** | **Default Value** |
| 'vprf' | Defines the video profile | profile\_idc | 1 |
| 'vlev' | Defines the video level | level\_idc | 51 |
| 'vtoh' | Defines the toolset | toolset\_idc\_h in hex decimal | x1FFFFF |
| 'vtol' | Defines the toolset | toolset\_idc\_l in hex decimal | x000000 |
| 'vbit' | Defines the video bit depth for luma and chroma | Value is a 2 digit decimal . The first digit is the luma bit depth minus 8. The second digit is the chroma bit depth minus 8 | unknown |
| 'vcss' | Defines the chroma subsampling | The subsampling scheme is expressed as a three part ratio J:a:b (e.g. 4:2:2) that describes the number of luminance and chrominance samples in a conceptual region that is J pixels wide, and 2 pixels high. The parts are (in their respective order):   * J: horizontal sampling reference (width of the conceptual region). Usually, 4. * a: number of chrominance samples (Cr, Cb) in the first row of J pixels. * b: number of changes of chrominance samples (Cr, Cb) between first and second row of J pixels.   Then the value is as follows. The first digit is J, the second digit is a and the third digit is b. | 420 |
| 'vcpr' | Defines colour primaries (ColourPrimaries) as defined in ISO/IEC 23091-2. | Defines colour primaries (ColourPrimaries) values as defined in ISO/IEC23091-2. Value is a 2 digit decimal with a possible preceding 0. | 01 |
| 'vtrc' | Defines transfer characteristics (TransferCharacteristics) as defined in ISO/IEC23091-2. | Defines transfer characteristics (TransferCharacteristics) values as defined in ISO/IEC23091-2. Value is a 2 digit decimal with a possible preceding 0. | 01 |
| 'vmac' | Defines matrix coefficients (MatrixCoefficients) as defined in ISO/IEC23091-2. | Defines matrix coefficients (MatrixCoefficients) values as defined in ISO/IEC23091-2. Value is a 2 digit decimal with a possible preceding 0. | 01 |
| 'vfrf' | Defines VideoFullRangeFlag as defined in ISO/IEC23091-2. | Defines VideoFullRangeFlag as defined in ISO/IEC23091-2. Value is a 1 digit decimal restricted to values 1 or 0. | 0 |
| 'vfpq' | Defines video frame packing type (VideoFramePackingType) as defined in ISO/IEC23091-2 together with the QuincunxSamplingFlag. | Defines video frame packing type (VideoFramePackingType) values as defined in ISO/IEC23091-2. Value is a 2 digit decimal with the first one being the QuincunxSamplingFlag and the second digit the VideoFramePackingType. | If not present, then no frame packing is used. |
| 'vpci' | Defines Packed content interpretation type (PackedContentInterpretationType) as defined in ISO/IEC23091-2. | Packed content interpretation type (PackedContentInterpretationType) as defined in ISO/IEC23091-2. Value is a 1 digit. | If not present, then no packed content is used. |
| 'vsar' | Defines Sample aspect ratio indicator (SampleAspectRatio) as defined in ISO/IEC23091-2. | Defines Sample aspect ratio indicator (SampleAspectRatio) values as defined in ISO/IEC23091-2. Value is a 2 digit decimal with a possible preceding 0. | 01 |

For example, codecs="evc1.vprf3.vlev51.vto1FFFFF.vtoh000000.vbit20.vcss420.‌vcpr09.vtrc16.vmac09.vsar01" represents EVC Main Profile, level 5.1, with 4:2:0 chroma subsampling co-located with (0, 0) luma sample, a restricted tool set, ITU-R BT.2100 color primaries, ITU-R BT.2100 PQ transfer characteristics, ITU-R BT.2100 YCbCr color matrix and sample aspect ratio 1:1. [Ed. (KB): If the keys derived from ISO/IEC 23091-2 are kept in this document, these ITU-Rs should be added to the Bibiography.]

All keys in Table 1 must be recognized, if the 'evc1' sample entry is recognized.

If a key is not recognized, the key value pair is ignored.