3GPP TSG-SA WG4 Meeting #132S4-250861

Japan, Fukuoka, 19 – 23 May 2025 revision of S4aV250043

**Source: Xiaomi**

**Title: [VOPS] pCR on terminology for addressing the concept of video layer**

**Spec: 3GPP TS 26.265 1.1.0**

**Agenda item: 9.5**

**Document for: Decision**

**1. Introduction**

This contribution proposes to the functional concept of a coded video layer rather reusing the coding concept a (HEVC) layer and the related HEVC CLVS. This way this coded video layer concept can reused in the common system integration and video API regardless of the targeted codec.

**2. Reason for Change**

This changes derives from discussion on S4-250601 followed by discussion on S4aV250043.

3GPP TS 26.265 specifies stereoscopic video signal representation format as well as MV-HEVC operation points to encode it. With this operation point, the video bitstream contains the left and right views as two HEVC layers. Furthermore, TS 26.265 defines bitstream constraints for such operation points which means there is a need to define a concept which can correspond to a HEVC layer but possibly to other codec.

This contribution is a follow-up of S4aV250043 which proposes to define such a concept by focusing on the functional aspect rather an the coding aspect of HEVC CLVS.

**3. Proposal**

The proposal is to define the concept of Coded Video Layer (CVL) as well as fixing the definition of Bitstream to align on common definition of the term bitstream in video coding specification such as in HEVC.

****

Figure - Logical relationship between VOPS concepts as proposed in this contribution

It is proposed to agree the following changes to 3GPP TS 26.265 1.1.0.

\* \* \* First Change \* \* \* \*

# 3 Definitions of terms, symbols and abbrevitions

## 3.1 Terms

For the purposes of the present document, the terms given in TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in TR 21.905 [1].

**Access Unit:** Smallest individually accessible portion of data within a Bitstream to which unique timing information can be attributed.

**Bitstream:** A sequence of bits that forms the representation of any coded pictures and their associated data. This sequence of bits is formed by one or more coded video sequences (CVSs) where the CVS share identical metadata.

**Coded Video Sequence:** A sequence of access units that consists of a series of coded frames and any associated metadata (required for decoder and rendering initialization) and conforms to a specific video encoding format and aligns with a certain Operation Point, as defined in this document. Such coded video sequence (CVS) has no decoding dependency on any other prior CVS and consists, in decoding order, of information specifying the characteristics or format of the encoded video data, a single intra random access coded frame followed by zero or more dependent, on the intra random access coded frame, coded frames, and a series of associated coded metadata.

**Chroma:** a sample array or single sample representing one of the two colour difference signals related to the primary colours, represented by the symbols *Cb* and *Cr*.

**Coded Video Layer:** A sequence of coded frames within a Coded Video Sequence and any associated metadata which belong to the same vjdeo coding layer of the i-th order.

NOTE: A higher order CVL may depend on a lower order CVL for prediction purposes.

**Hero Eye**: The default eye in a stereo (stereoscopic) video pair, often determined by tags set by the cameras used to capture the video.

**Luma:** a sample array or single sample representing the monochrome signal related to the primary colours (denoted with the symbol *Y*),

**Operation Point:** A collection of discrete combinations of different video representation formats, including spatial and temporal resolutions, colour mapping, transfer functions, and the encoding format.

**Receiver:** A device capable of decoding and rendering any bitstream that is conforming to a certain Operation Point.

**[S4-250861\_r01] Sub-bitstream:** Bitstream that can be generated by extracting parts of a source *bitstream* and that is conforming to a certain Operation Point.

OR

**[S4-250797] Sub-bitstream:** Part of a *bitstream* that is conforming to a certain Operation Point.\

**[S4-250861\_r01] Video Layer sub-bitstream**: The *sub-bitstream* generated by extracting one or more CVLs from a source *bitstream*.

\* \* \* Second Change \* \* \* \*

3.3 Abbreviations

For the purposes of the present document, the abbreviations given in TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in TR 21.905 [1].

AVC Advanced Video Coding

CENC Common ENCryption

CMAF Common Media Application Format

CVL Coded Video Layer

CVS Coded Video Sequence

DPC Device Playback Capabilities

FFS For Further Study

HDR High Dynamic Range

HDTV High-Definition TeleVision

HEVC High Efficiency Video Coding

HLG Hybrid Log-Gamma

MSE Media Source Extension

MV-HEVC MultiView extensions of HEVC

RAP Random access point

SDR Standard Dynamic Range

UHD Ultra-High Definition

WCG Wide Colour Gamut

\* \* \* Third Change \* \* \* \*

5.3.2 HEVC Decoding Capabilities

The following decoding capabilities are defined:

- **HEVC-HD-Dec**: the capability to decode bitstreams conforming to both, HEVC/ITU-T H.265 Main Profile, Main Tier, Level 3.1 [h265] bitstreams with *progressive* constraints as defined in clause 4.5.3.

- **HEVC-FullHD-Dec**: the capability to decode bitstreams conforming to HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 4.1 [h265] bitstreams with *progressive* and *VUI* constraints as defined in clause 4.5.3.

- **HEVC-UHD-Dec**: the capability to decode bitstreams conforming to HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 5.1 [h265] bitstreams with *progressive* and *VUI* constraints as defined in clause 4.5.3.

- **HEVC-8K-Dec**: the capability to decode bitstreams conforming to HEVC/ITU-T H.265 Main10 Profile, Main Tier, Level 6.1 [h265] bitstreams with *progressive* and *VUI* constraints as defined in clause 4.5.3 and further constraints:

- the bitstream does not exceed the maximum luma picture size in samples of 33,554,432,

- the maximum VCL Bit Rate is constrained to be 80 Mbps with CpbVclFactor and CpbNalFactor being fixed to be 1000 and 1100, respectively.

* **MV-HEVC-UHD-Dec**: the capability to decode bitstreams with an HEVC/ITU-T H.265 Main 10 Profile base layer (layer\_id=0), and a single HEVC/ITU-T H.265 Multiview Main 10 [or Multiview Extended 10] layer (layer\_id=1) [h265]. Each layer shall conform to Main Tier, Level 5.1, while the device should be capable of supporting single layer decoding of HEVC/ITU-T H.265 Main 10 Profile bitstreams at Main Tier, Level 5.2. All layers shall follow the *progressive* and *VUI* constraints as defined in clause 4.5.3.

Editor’s Note: The removal of brackets for Extended 10 is subject to verification that we can playback such content on receivers. For this purpose, we recommend check using the VET-AM1008-v1 with direct http link to the test streams: <https://www.itu.int/wftp3/av-arch/jvet-site/bitstream_exchange/HEVCMultiview/under_test/>.

- **HEVC-Frame-Packed-Stereo-Dec**: the capability to decode bitstreams conforming to HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 6.0 [h265] with *frame-packing* and *VUI* *constraints* as defined in clause 4.5.3

NOTE: The increase from Level 5.2 for MV-HEVC-UHD-Dec to Level 6.0 in HEVC-Frame-Packed-Stereo-Dec is only to handle larger buffers per frame. There is no increase in the pixels/second between the two capabilities.

\* \* \* Fourth Change \* \* \* \*

# 7 Common System Integration

7.1 Introduction

This clause documents general functionalities that are relevant for integration of video codecs into delivery systems to support common APIs on encoders and decoders.

7.2 Functional Definitions

### 7.2.1 General

##### 7.2.1.1 Summary

This clause defines functional definitions for system integration in Table 7.2.1.1-1. The remainder of this

Table 7.2.1.1-1 Functional Definitions

|  |  |  |
| --- | --- | --- |
| Term | Summary | Details |
| Codec String | A single value identifying the codec indicated to render the content in the Bitstream as defined in IETF RFC 6381. | 7.2.1.2 |
| Decoder Configuration | a data structure storing essential parameters needed for decoding and rendering a video stream. | 7.2.1.3 |
| Random Access Point | A byte position in the Bitstream, for which in combination with the Decoder Configuration, the Bitstream can be randomly accessed, i.e. in decoding order the Bitstream carries sufficient information to access the media in the stream. | 7.2.1.4 |
| Access Unit (AU) | See Clause 3.1 |  |
| Coded Video Layer (CVL) | See Clause 3.1 | 7.2.1.7 |
| Coded access unit (CAU) | bits corresponding to an Access Unit | 7.2.1.5 |
| Random Access CAU | A CAU that starts with a random access point | 7.2.1.6 |

##### 7.2.1.2 Codec String

The *Codec String* provides means to identify the codec needed to decode and render the content in the Bitstream. The codec parameters shall also include the profile and level information where applicable. The content of this parameter shall conform to the id-simple production of IETF RFC 6381:2011, subclause 3.2, without the enclosing DQUOTE characters. The codec identifier for the media format, mapped into the name space for codecs as specified in IETF RFC 6381:2011, subclause 3.3, shall be used.

##### 7.2.1.3 Decoder Configuration

The *Decoder Configuration* provides parameters about the Bitstream and shall follow the format defined in ISO/IEC 14496-15 including:

- profile, tier, level

- constraints flags

- chroma format

- bit depth chroma and luma

- frame rates, average or constant

- layering structure

- NAL units

- VPS (Video Parameter Set): Contains parameters that apply to the entire video sequence.

- SPS (Sequence Parameter Set): Contains parameters that apply to a sequence of pictures.

- PPS (Picture Parameter Set): Contains parameters that apply to individual pictures.

- declarative SEI NAL unit, as specified in ISO/IEC 23008-2. 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 as instances of nalUnit.

##### 7.2.1.4 Random Access Point

Different types of Random Access Points are defined as follows:

**- Closed loop RAP (CL-RAP)** is an intra coded picture that can identify a RAP in a bitstream. It can be the first coded picture or can appear later in a bitstream. Each CL-RAP is the first picture in decoding order of a coded video sequence (CVS) but does not need to be an output picture or be the first picture in display order. All frames that follow a CL-RAP in decoding order and belong in the same coded video sequence are decodable and can potentially be all output by the decoder depending on their coding parameters.

**- Open loop RAP (OL-RAP)** is an intra coded frame that can identify a RAP in a bitstream. It can be the first frame in the bitstream in decoding order or can appear later in the bitstream. An OL-RAP does not need to be an output picture or be the first picture in display order. Other pictures that follow the OL-RAP in coding order can refer to an OL-RAP for prediction. However, an OL-RAP, if it is the first picture in the bitstream in decoding order, may also be followed in coding order by some pictures that can refer to pictures that are not present in the bitstream. In that case, these pictures cannot be decoded. These pictures can be referred to as leading pictures. Subsequently, when those pictures are detected, they are not decoded and can be discarded by the decoder.

**- Gradual decoder refresh (GDR) access point** identifies a RAP in a bitstream from where decoding operations can start by a decoder. However, unlike other RAP types, decoding may not be instantaneous and may initially result in decoding errors in the decoded and reconstructed pictures. Nevertheless, these decoding errors are expected to disappear after a certain maximum period, from which point decoding can continue without any further decoding errors.

##### 7.2.1.5 Coded Access Unit

Editor’s Note: This needs to be completed.

##### 7.2.1.6 Random Access CAU

Editor’s Note: This needs to be completed.

##### 7.2.1.7 Coded Video Layer

Different types of Coded Video Layer are defined as follows:

* **Independent CVL** is a CVL which does not depend on any other CVL in the CVS for prediction purposes.
* **Output CVL** is a CVL whose coded pictures are meant to be output after decoding.
* **Base CVL** is an Independent CVL and Output CVL and it is the first CVL in the CVS.
* **Dependent CVL** is a CVL that depends on a CVL of a lower order for prediction purposes.
* **Auxiliary CVL** is an Independent CVL and Output CVL that is not the first CVL of the CVS.

Editor’s Note: This needs to be completed.

### 7.2.2 AVC

Editor’s Note: This needs to be completed.

### 7.2.3 HEVC

Editor’s Note: This needs to be completed.

\* \* \* End of Changes \* \* \* \*