**3GPP TSG-SA WG4 Meeting #132 S4-250830**

**Japan, Fukuoka, 19 – 23 May 2025**

|  |
| --- |
| *CR-Form-v12.2* |
| **PSEUDO CHANGE REQUEST** |
|  |
|  |  | **CR** |  | **rev** |  | **Current version:** |  |  |
|  |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* |
|  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME | **X** | Radio Access Network |  | Core Network | **X** |

|  |
| --- |
|  |
| ***Title:***  |  |
|  |  |
| ***Source to WG:*** | Apple Inc. Tencent |
| ***Source to TSG:*** |  |
|  |  |
| ***Work item code:*** |  |  | ***Date:*** |  |
|  |  |  |  |  |
| ***Category:*** |  |  | ***Release:*** |  |
|  | *Use one of the following categories:****F*** *(correction)****A*** *(mirror corresponding to a change in an earlier release)****B*** *(addition of feature),* ***C*** *(functional modification of feature)****D*** *(editorial modification)*Detailed explanations of the above categories canbe found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | *Use one of the following releases:Rel-8 (Release 8)Rel-9 (Release 9)Rel-10 (Release 10)Rel-11 (Release 11)…Rel-16 (Release 16)Rel-17 (Release 17)Rel-18 (Release 18)Rel-19 (Release 19)* |
|  |  |
| ***Reason for change:*** | Several definitions of terminology and usage were missing related to layered HEVC. |
|  |  |
| ***Summary of change:*** | Provides definitions of terminology and its usage related to layered HEVC. A sketch of various HEVC concepts and their relation is depicted in following:**CVS:*** Consists of the following
	+ A CVS and layer relationships description unit (e.g. a VPS or an SPS)
		- May contain information about the overall combination of layers that could be useful to decoders to understand if they can consume the bitstream  (e.g. layer 0 is Main 10, layer 1 is Multiview Main 10)
		- Note: for single layers, an SPS may be used, which is fundamental for decoding
		- May contain info about the association of layers (e.g. layer 1 depends on layer 0 or layer 1 is alpha)
	+ Operating point information (In HEVC they are in the VPS, but could be in a separate unit)
		- For multiview stereo and alpha (4:2:0) there should be an OPS with only 2 layers, both being 4:2:0
		- Anything else (other than single layer 4:2:0) should be rejected
		- For multi-layer bitstreams there could be an operating point for 2 layers, but optionally there could be single layer operating points also (e.g. only texture A, only texture B (if not dependent), only alpha )
	+ One or more CLVSs
	+ Potentially metadata that apply to a set of layers

**CLVS:*** Consists of the following
	+ A Layer description unit (e.g. an SPS)
		- May contain the profile and level corresponding to just that CLVS e.g. Multiview Main 10 or Main 10 at level 5.2
	+ One or more random access points
	+ Other picture coding units and operational metadata (e.g. PPSs)
	+ Metadata for that layer (e.g. SEI messages)

**Video Layer sub-bitstream** * Contains one or more CLVS of a particular id
* May contain multiple layer description units and random access points
 |
|  |  |
| ***Consequences if not approved:*** | Key definitions for terminology and usage will remain missing related to layered HEVC. |
|  |  |
| ***Clauses affected:*** | 3.1, 3.3, 4.2, 5.3.2, 6.1, 6.3.6, 6.3.6.1, 6.3.6.2, 6.3.6.3 |
|  |  |
|  | **Y** | **N** |  |  |
| ***Other specs*** |  | **X** |  Other core specifications  | TS/TR ... CR ...  |
| ***affected:*** |  | **X** |  Test specifications | TS/TR ... CR ...  |
| ***(show related CRs)*** |  | **X** |  O&M Specifications | TS/TR ... CR ...  |
|  |  |
| ***Other comments:*** |  |
|  |  |
| ***This CR's revision history:*** |  |

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

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 associated data. This sequence of bits is formed by one or more coded video sequences (CVSs) where the CVS share identical metadata.

**Coded layer-wise video sequence:** A coded video sequence (CVS) of a specific HEVC layer-wise video.

**Coded Video Sequence:** A sequence of bits 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*.

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

**HEVC Layer-wise** **Video**: a set of VCLs and the associated non-VCL NAL units of a sub-bitstream with the same value of nuh\_layer\_id.

Editor’s Note: Definitions for uncoded layer/video component e.g. alpha layer/component, texture layer/component may need to be specified.

**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.

**HEVC Layer-wise** **Sub-bitstream:** An HEVC Sub-bitstream is a part of an HEVC Bitstream corresponding to a specific HEVC layer-wise video.

\* \* \* Next 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

CLVS Coded layer-wise video sequence

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

NAL Network Abstraction Layer

RAP Random access point

SDR Standard Dynamic Range

UHD Ultra-High Definition

VCL Video Coding Layer

WCG Wide Colour Gamut

\* \* \* Next Change \* \* \* \*

4.2 Reference architectures and definitions

In order to define the normative aspects of this specification, reference architectures are defined. The core architecture is provided in Figure 4.2-1. The workflow addresses the generation of a *video bitstream* from a *video signal* using a *video encoder* as well as the decoding of a video bitstream by a *video decoder* and providing the resulting decoded video as well as associated metadata to a rendering and display process. The video signal can be composed of one or more video signal components, for example a video signal can include multiple views. Video signals follow certain representation formats and can be rendered in a device specific manner.

The video encoder as well as the video decoder may be configured to certain operations indicated by APIs in Figure 4.2-1. These APIs are not normatively specified but serve as an example reference to configure encoders and decoders as documented in Annex [A].

****

**Figure 4.2-1 Reference architecture for video operating points and capabilities**

Video encoders produce a sequence of *Coded Video Sequences,* as defined in clause 3.1, and the sequence of CVSs are referred to as *Bitstreams*.

An intra random access coded frame, together with the associated metadata, forms a Random Access Point (RAP) that permits to initialize decoding of the coded video sequence.

The decoder is provided with access units which correspond to pieces of the Bitstream that can be processed by the decoder to regenerate decoded video frames.

Figure 4.2-2 provides an overview of the data model and the definitions in this specification.



Editor’s Note: This figure is for illustrative purposes, informative and may be moved to an Annex.

**Figure 4.2-2 Informative Data model for illustration purposes**

In this case, configuration information is coded into metadata, that can be provided to the decoder to initialize the decoding of the CSVs included in the Bitstream.

A more system-centric architecture is provided in Figure 4.2-3. The workflow addresses the generation of a *transport stream* from a video signal using a *video encoder* and a *packager*. The package may include for example timing and metadata information. The de-packaging and decoding of the *transport stream* by a de-packager and a *video decoder*, respectively, allows for providing the resulting video signal as well as associated metadata to a rendering and display process. Again, the packager/encoder as well as the de-packager/decoder may be configured to certain operations indicated by APIs in Figure 4.2-2.



**Figure 4.2-3 Reference architecture for system operating points and capabilities**

Based on this introduction, the following terms are defined:

**- Operating Point:** A collection of different possible video formats including spatial and temporal resolutions, colour mapping, transfer functions, etc. and a video encoding format.

**- Bitstream**: A compressed media representation presented as a sequence of bits

- that forms the representation of any coded pictures and associated metadata data,

- this sequence of bits is formed by one or more CVSs and each CVS has identical metadata

- the sequence of bits conforms to a particular video coding specification/format and one or more Operating Points.

- comprised by access units that serve as units to be provided to decoders for regenerating frames.

**- Receiver**: A device that can ingest and decode any Bitstream that is conforming to a particular video coding specification and Operating Point, and optionally render it.

In addition, on system level the following terms are defined:

**- System Operating Point:** A collection of different possible video formats including spatial and temporal resolutions, colour mapping, transfer functions, etc., a video encoding and a packaging format.

**- Transport Stream:** A packaged media bitstream that conforms to a particular video coding and packaging specification/format and one or more Operating Points.

**- System Receiver:** A receiver that can de-package and decode any system bitstream that is conforming to a particular System Operating Point, and optionally render it.

NOTE: A reference architecture for multiple decoders is for further study.

System Operating Points are not defined in this specification but are left for mappings to specific delivery protocols such as CMAF/DASH for 5G Media Streaming, or ISO BMFF for Messaging Services. However, this specification provides mapping principles to delivery protocols.

\* \* \* Next Change \* \* \* \*

5.3.2 HEVC Decoding Capabilities

The following decoding capabilities are defined:

- **HEVC-HD-Dec**: the capability to decode bitstreams containing [a single] or [at least one base] HEVC layer-wise sub-bitstream conforming to HEVC/ITU-T H.265 Main Profile, Main Tier, Level 3.1 [h265] with *progressive* constraints as defined in clause 4.5.3.

- **HEVC-FullHD-Dec**: the capability to decode bitstreams containing [a single] or [at least one base] HEVC layer-wise sub-bitstream conforming to HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 4.1 [h265] with *progressive* and *VUI* constraints as defined in clause 4.5.3.

- **HEVC-UHD-Dec**: the capability to decode bitstreams containing [a single] or [at least one base layer] HEVC layer-wise sub- bitstream with CLVSs conforming to HEVC/ITU-T H.265 Main 10 Profile, Main Tier, Level 5.1 [h265] with *progressive* and *VUI* constraints as defined in clause 4.5.3.

- **HEVC-8K-Dec**: the capability to decode bitstreams containing [a single] or [at least one base] HEVC layer-wise sub-bitstream with CLVSs conforming to HEVC/ITU-T H.265 Main10 Profile, Main Tier, Level 6.1 [h265] 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 containing two HEVC layer-wise video sub-bitstreams with the following constraints:

- The first HEVC layer-wise sub-bitstream having nuh\_layer\_id=0 with CLVSs conforming to the HEVC/ITU-T H.265 Main 10 profile [h265].

- A second HEVC layer-wise sub-bitstream with CLVSs conforming to the HEVC/ITU-T H.265 Multiview Main 10 [or Multiview Extended 10] profile [h265].

[NOTE: It is expected that the deployments are converging to using Multiview Extended 10 profile]

- All CLVSs in both HEVC layer-wise sub-bitstreams shall conform to Main Tier, Level 5.1.

- The device should be capable of supporting single HEVC layer-wise decoding of HEVC/ITU-T H.265 Main 10 Profile bitstreams at Main Tier, Level 5.2.

- All CLVSs in both HEVC layer-wise sub-bitstream shall follow the *progressive*. *VUI* constraints are specified by the operating point using this decoding capability.

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 containing [a single] or [at least one base] HEVC layer-wise sub-bitstream with CLVSs 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.

\* \* \* Next Change \* \* \* \*

6.1 Introduction

Video operation points define a restricted subset of representation signals and media capabilities. For each Video Operation Point, requirements for the Bitstream and for the Receiver are defined.

Table 6.1-1 provides an overview of defined video operation points.

**Table 6.1-1 Overview of Video Operation Points**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Video Format** | **Decoding Capabilities** | **Definition** |
| 3GPP-AVC-HD | 3GPP-HD (see clause 4.4.3.2) | AVC-FullHD-Dec (see clause 5.4) | 6.2.2 |
| 3GPP-HEVC-HD | 3GPP-HD (see clause 4.4.3.2) | HEVC-FullHD-Dec (see clause 5.4) | 6.3.2 |
| 3GPP-HEVC-HD-HDR | 3GPP-HDR (see clause 4.4.3.3) | HEVC-FullHD-Dec (see clause 5.4) | 6.3.3 |
| 3GPP-HEVC-UHD-HDR | 3GPP-HDR (see clause 4.4.3.3) | HEVC-UHD-Dec (see clause 5.4) | 6.3.4 |
| 3GPP-HEVC-Stereo | 3GPP-Stereo (see clause 4.4.3.4) | HEVC-Frame-Packed-Stereo-Dec (see clause 5.5) | 6.3.5 |
| 3GPP-MV-HEVC-Stereo | 3GPP-Stereo (see clause 4.4.3.4) | MV-HEVC-UHD-Dec, HEVC-UHD-Dec (see clause 5.3.2) | 6.3.6 |

\* \* \* Next Change \* \* \* \*

6.3.6 3GPP MV-HEVC Stereo

6.3.6.1 Introduction

The MV-HEVC Stereo Operation Point permits consistent distribution of stereoscopic content using MV-HEVC. The remainder of this clause 6.3.6 defines the Bitstream and Receiver requirements for the 3GPP-MV-HEVC-Stereo receiver.

6.3.6.2 Bitstream Requirements

Editor’s Note: this needs additional signaling:

* Layer dependency is possible, but not needed. Can be two independent layers, i
* 3D reference displays information SEI message

A 3GPP-MV-HEVC-Stereo Bitstream shall conform to the following requirements

- the Representation Format included in the Bitstream shall conform to the 3GPP Stereoscopic format as defined in clause 4.4.3.4.

- The bitstream shall conform to the constraints specified in the **MV-HEVC-UHD** decoding capabilities as defined in clause 5.3.2.

- the Bitstream shall be decodable by

- a decoder with **HEVC-UHD-Dec** decoding capabilities as defined in clause 5.3.2.

- a decoder with **MV-HEVC-UHD** decoding capabilities as defined in clause 5.3.2.

Based on this, the following additional restrictions apply

- The chroma sub-sampling shall be 4:2:0 and the value of chroma\_format\_idc shall be set to 1.

- AuxId[ lId ] shall be equal to 0 in the VPS extension for sub-bitstream with nuh\_layer\_id != 0.

Editor’s Note: this should refer to the bitstream element and not the variable AuxId.

- The aspect\_ratio\_idc value shall be set to 1, indicating a square pixel format.

- In the VUI, either

- the values of colour\_primaries, transfer\_characteristics and matrix\_coeffs each shall be set to 1.

- The value of chroma\_sample\_loc\_type\_top\_field shall be set to 0.

- or

- the values of colour\_primaries and matrix\_coeffs each shall be set to 9, and the value of transfer\_characteristics shall be set to one of the following values: 14 (for SDR with WCG), 16 (for PQ) and 18 (for HLG).

- The value of the chroma\_sample\_loc\_type\_top\_field shall be set to 2.

The timing information may be present.

- If the timing information is present, i.e. the value of vui\_timing\_info\_present\_flag is set to 1, then the values of vui\_num\_units\_in\_tick and vui\_time\_scale shall be set according to the frame rates allowed for each operation point. The timing information present in the video Bitstream should be consistent with the timing information signalled at the system level.

- The frame rate shall not change between two RAPs. fixed\_frame\_rate\_flag value, if present, shall be set to 1.

Bitstreams not required to be associated with frame packing information for all coded video sequences. It is also possible that such information, when present, may differ from one coded video sequence to another.

6.3.6.3 Receiver Requirements

Receivers conforming to this Operation Point 3GPP-MV-HEVC-Stereo shall support decoding and rendering Bitstreams with the restrictions defined in clause 6.3.6.2.

NOTE 1: Rendering includes adherence to the parameters signalled in the bitstream to characterize the distributed Representation format.

There are no requirements on output timing conformance for H.265/HEVC decoding (Annex C of [6]). The Hypothetical Reference Decoder (HRD) parameters, if present, should be ignored by the Receiver.

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