**3GPP TSG- Meeting #**

**, FL, United States, 19 -**

|  |
| --- |
| *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:*** | , 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:*** | Missing pieces in clause 4 |
|  |  |
| ***Summary of change:*** | Bug fixesAdds 3D TV signalAdds decoder API |
|  |  |
| ***Consequences if not approved:*** |  |
|  |  |
| ***Clauses affected:*** |  |
|  |  |
|  | **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:*** |

|  |  |  |  |
| --- | --- | --- | --- |
| [**S4-241893**](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/TSGS4_130_Orlando/Docs/S4-241893.zip) | [VOPS] Updates to Representation Formats | Qualcomm Incorporated, Tencent | Thomas Stockhammer |

 **Revisions**: none**Online Discussion**: * **Session 2: 16:00-18:00**
* Thomas presents.
* Alexis: The definition of hero eye could be refined.
* Thomas: Happy to update if there are proposals. A reference would be great.
* Waqar: Clause 4.4.3.4, 3D TV is a bit confusing since we want to focus on 3D movies. We need to differentiate sender and receiver ends. More flexibility is desirable. Also VoD vs live, there should be some variations.
* Gilles: having a guaranteed quality of experience is very useful for a service provider.
* Thomas: There is a misunderstanding. It is not the intent to restrict MV-HEVC for our intent. The idea is to define a source format to use for the generation of content, possibly using MV-HEVC but it can be used for another other encoded form of stereoscopic movies.
* Waqar: There can be stereo movies produced a long time ago which don’t follow those source format properties.
* Thomas: If there are more corner cases, let’s study. But this is just setting boundaries in which we can operate to produce content.
* Waqar: For example, why just 16:9.
* Thomas: Good I put it there because this is what I found. But open to hear that we need other aspect ratios and the same with bit depth. This is a starting point and we can extend, feel free to modify to loosen the boundaries.
* Alexis: 3D TV renamed to cinema would be better.
* Thomas: Yes, sounds good. I wanted to start with well defined signal properties.

**Decision**:* Session 2: Parked (revision expected) on hero eye and 3D TV renaming.

This also takes into account the comments received offline* It seems we are not consistent in the use of Rec. prior to some of the referenced specs. I also see the usage of Rec (no period).
* Reordering H.273 and H.274 in the references seems appropriate (ascending order). Also good to place them near H.264/H.265 instead of mixing other specs in between.
* There are still sections that should refer to H.273 but refer to H.274 (e.g. the chroma format and chroma sample location ). Since H.274 is mostly about the VUI of VVC and some SEI messages, we should likely avoid it as much as possible and only point to it if it is really needed. HEVC currently does not refer to it for anything essential.
* The definitions of the chroma sample locations for 0 and 2 are incorrect and in some way mixed up. 1 is also mentioned but not defined (which seems inconsistent if values 0 and 2 are defined).
* The expression “neither eye may be the hero eye” sounds odd to me. Would it be better to say “there is no requirement that either of the two eyes (or views?) is tagged as the hero eye”? Optionally adding “ in which case no hero eye tagging may be present in a bitstream” may also be helpful.
* My comment for 4.4.3.3 about the YCbCR and non constant  luminance lines from an earlier edit was not considered. Those imply somehow the same thing. Could they be joined into a single bullet?
* The same about chroma location 0 and 2 in the profiles. Those are wrong and I provided edits earlier. Were those shared?
* For the cinema format it says 4K is only supported but in the table you also see 1080p. Is it also useful to mention 4:2:0 YCbCr in the bullets? The text also still mentions this is based on the HDR profiles, which seems not to allow BT.709 or BT.2020. It is clarified later on but should that HDR mention in the text before the table be removed?
* Chroma location info is again (of course) wrong.
* The signal “has” not “have”.
* Shouldn’t we allow also 8 bit coding here? Sure, you can up convert but there may be advantages for SW encoders to use 8 bits (or SW decoders). Indeed there is zero diff for HW decoders.
 |

# ===== CHANGE =====

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non‑specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[bt709] Recommendation ITU-R BT.709-6 (06/2015): "Parameter values for the HDTV standards for production and international programme exchange"

[bt2100] Recommendation ITU-R BT.2100-2 (07/2018): "Image parameter values for high dynamic range television for use in production and international programme exchange"

[h264] Recommendation ITU-T H.264 (08/2021): "Advanced video coding for generic audiovisual services".

[h265] Recommendation ITU-T H.265 (09/2023): "High efficiency video coding".

[h273] Recommendation ITU-T H.273 (09/2023): "Coding-independent code points for video signal type identification".

[h274] Recommendation ITU-T H.274 (09/2023): "Versatile supplemental enhancement information messages for coded video bitstreams".

[CMAF] ISO/IEC 23000-19: "Information Technology Multimedia Application Format (MPEG-A) – Part 19: Common Media Application Format (CMAF) for segmented media".

[CENC] ISO/IEC 23001-7: "MPEG systems technologies - Part 7: Common encryption in ISO base media file format files".

[DPC] CTA-5003-B: "Web Application Video Ecosystem (WAVE): Device Playback Capabilities Specification", available at https://shop.cta.tech/products/web-application-video-ecosystem-device-playback-capabilities-cta-5003-b .

[6381] IETF RFC 6381: The 'Codecs' and 'Profiles' Parameters for "Bucket" Media Types.

[MSE] 3GPP TR 26.857, "5G Media Service Enablers"

[3dtv] A. Quested and B. Zegel, "3D-TV production standards - first report of the ITU-R Rapporteurs", EBU Technical Review, 2011 Q2, https://tech.ebu.ch/publications/trev\_2011-Q2\_3dtv\_quested

# ===== 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].

**Bitstream:** A sequence of bits that conforms to a specific video encoding format and aligns with a certain Operation Point.

**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.

**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 content 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.

# ===== CHANGE =====

### 4.4.2 Video signal parameters

Video signals considered in this specification are represented by a sequence of pictures, where a *picture* can represent either an array of *luma* samples in a monochrome format or an array of luma samples and two corresponding arrays of *chroma* samples in a 4:2:0, 4:2:2, or 4:4:4 colour format. Only *progressive* signals are considered. The Luma component represents a sample array or single sample representing the monochrome signal related to the primary colours (denoted with the symbol *Y*), and a chroma component represents 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*.

Video signals are typically described by a set of parameters that are required for the proper rendering of the decoded signal. Table 4.4.2-1 documents typical video signal parameters and provides a definition and/or reference.

Table 4.4.2-1 Video Signal Parameters

|  |  |  |
| --- | --- | --- |
| Parameter | Definition | 3GPP restrictions |
| Spatial Resolution width | the number of active samples per line for the luma component.Example values are 1280 or 1920 for HD, and 3840 for UHD.NOTE: The width does not restrict the encoding resolution to fixed values. Cropping parameters can be indicated that prescribe decoders the need to remove spatial video samples in a partially filled coding block that are not intended for presentation. | no direct restrictions, but services may provide subsets. |
| Spatial Resolution height | the number of active lines per picture for the luma component.Typical values are 720 or 1080.NOTE: The height does not restrict the encoding resolution to fixed values. Cropping parameters can be indicated that prescribe decoders the need to remove spatial video samples in a partially filled coding block that are not intended for presentation. | no direct restrictions, but services may provide subsets. |
| Scan Type | indicates the source scan type of the pictures as defined in clause 7.3 of Rec. ITU-T H.273.Typical value is progressive | progressive only |
| Chroma format indicator | indicates whether the picture has only a luma component or that the picture has three colour components that consist of a luma component and two associated chroma components, such that the width and height of each chroma component are the width and height of the luma component divided by a factor defined by the chroma format as defined in Rec. ITU-T H.274, clause 7.3.  | 4:2:0 |
| Bit depth | Indicates the bit depth for the samples of the luma component and the samples of the two associated chroma components.Note that in general, the bit depth of the luma component and of the two associated chroma components may differ.Typical values are 8 or 10 bits. | 8 or 10 bits |
| Colour primaries  | indicates the chromaticity coordinates of the source colour primaries as specified in clause 8.1 of Rec. ITU-T H.273.Typical values are 1 to refer to Rec. ITU-R BT.709-6 [bt709] or 9 to refer to Rec. ITU-R BT.2020-2 and Rec. ITU-R BT.2100-2.  | BT.709 or BT.2020/BT.2100 |
| Transfer Characteristics | either indicates the reference opto-electronic transfer characteristic function of the source picture as a function of a source input linear optical intensity input or indicates the inverse of the reference electro-optical transfer characteristic function as a function of an output linear optical intensity as defined in clause 8.2 of Rec. ITU-T H.273.Typical values are 1 to refer to Rec. ITU-R BT.709-6, 14 to refer to Rec. ITU-R BT.2020-2 (10 bit), 16 to refer to Rec. ITU-R BT.2100-2 perceptual quantization (PQ) system, or 18 to refer to Rec. ITU-R BT.2100-2 hybrid log-gamma (HLG) system | BT.709 SDR, BT.2020 SDR, BT.2100 PQ, or BT.2100 HLG |
| Matrix Coefficients | describes the matrix coefficients used in deriving luma and chroma signals from the green, blue, and red primaries. A video full range flag may be supplied with this parameter specifying the scaling and offset values applied in association with the Matrix coefficients. For detailed definition refer to clause 8.2 of Rec. ITU-T H.273.Typical values are 1 to refer to the non constant luminance YCbCr representation Rec. ITU-R BT.709-6 or 9 to refer to the non constant luminance YCbCr representations in Rec. ITU-R BT.2020-2 and Rec. ITU-R BT.2100-2. | BT.709 or BT.2020/BT.2100 |
| Frame rate | Typical values, using frames per second, are: 120, 120/1.001, 100, 60, 60/1.001, 50, 30, 30/1.001, 25, 24, 24/1.001 | no direct restrictions, but services may only permit a restricted subset. |
| Frame packing | indicates a frame packing arrangement, if present, as defined in clause 8.4 of Rec. ITU-T H.273. | Typically restricted to no frame packing, but applications may use frame packing. |
| Projection | indicates a projection, if present, as defined in Rec. ITU-T H.274, clause 7.3, and typically refers to packing arrangements in clause 8.6 of Rec. ITU-T H.274. | Typically restricted to no projection, but applications may use projections. |
| Sample aspect ratio | indicates width-to-height aspect ratio of the luma samples of the associated pictures as defined in clause 7.3 of Rec. ITU-T H.273.Typical value is 1 | No specific restrictions, but 1 is expected. |
| Chroma sample location type | specifies the location of the chroma samples relative to the luma samples for frames as defined in Rec. ITU-T H.273, clause 8.7.Typical values are 0 (chroma samples are horizontally co-sited with and vertically centered between the first luma sample at the top-left corner and the first two luma samples at the top-left corner, respectively) or 2 (chroma samples are co-sited with the luma sample at the top-left corner). Note that 1 is common for still images. | No specific restrictions, but 0 is expected if not present. For HDR the value is typically set to 2. |
| Range | Specifies how luma and chroma samples are represented in digital video as defined in Rec. ITU-T H.273, clause 8.3 using the VideoFullRangeFlag. Only the value set to 0 is used, i.e. the video range or restricted range is applied where the luma values range from 16 to 235 in an 8-bit system, and chroma values range from 16 to 240. For 10-bit systems, the values are multiplied by 4. Note that for still images full range is commonly used. | No specific restrictions, but 0 is expected if not present. |
| Stereoscopic Video | Visual media may be stereoscopic, in which a view is available to be presented to the left eye and another view is available to be presented simultaneously to the right eye. The presentation of both the left and right views allows for an effect known as stereopsis, which can be defined as "the perception of depth produced by the reception in the brain of visual stimuli from both eyes in combination; binocular vision." For signal representations, [3dtv] recommends that the Left and Right eyes comply to regular image formats such as Rec. ITU-R BT.709 and any necessary 3D-specific metadata is incorporated with the data. Hence, for stereoscopic video, two synchronized video signals are available, each with identical format parameters (such as the ones defined in this table). Additional metadata that may be added with stereoscopic video:- hero eye is the default eye in a stereo (stereoscopic) video pair, often determined by tags set by the cameras used to capture the video. If so signaled, this indicates the other stereo eye view is derived from the specified stereo eye and may be useful when choosing which eye to use in a monoscopic viewing environment. There is no requirement that either of the two eyes (or views) is tagged as the hero eye in which case no hero eye tagging may be present.  |  |

# ===== CHANGE =====

#### 4.4.3.1 Introduction

While a variety of formats may be used based on the video signal parameters defined in clause 4.4.2, for consistent programs and signals, several video formats are defined by a set of restrictions using the video signal parameters in clause 4.4.2. These signals are primarily used to distribute TV and movie content.

New 3GPP Video formats may be defined.

# ===== CHANGE =====

#### 4.4.3.2 High-Definition TV

3GPP High-Definition TV (HDTV) formats are defined based on Rec. ITU-R BT-709-6 [bt709]. 3GPP HDTV formats shall conform to Rec. ITU-R BT-709-6 [bt709] with the following restrictions:

- Only the following formats are included 24/P, 25/P, 30P, 50/P and 60/P. Interlace and progressive segmented frame signals are excluded.

- Only the Non-Constant Luminance Y'C'BC'R signal format is considered.

An informative summary of the parameters of a 3GPP HDTV format based on the parameters defined in Table 4.4.2-1 is provided in Table 4.4.3.2-1.

Table 4.4.3.2-1 Video Signal Parameters for 3GPP HDTV format

|  |  |
| --- | --- |
| Parameter | Restrictions |
| Spatial Resolution width | the number of active samples per line is 1920. |
| Spatial Resolution height | the number of active lines per picture for the luma component is 1080.NOTE: Typically the encoded signal has 1088 lines and cropping is applied to remove spatial samples that are not presented. |
| Scan Type | the source scan type of the pictures as defined in clause 7.3 of Rec. ITU-T H.273 is progressive |
| Chroma format indicator | The chroma format indicator is 4:2:0.  |
| Bit depth | The permitted values are 8 or 10 bit. |
| Colour primaries  | Only the value 1 as defined in clause 8.2 of Rec. ITU-T H.273 is permitted. |
| Transfer Characteristics | Only the value 1 as defined in clause 8.2 of Rec. ITU-T H.273 is permitted. |
| Matrix Coefficients | Only the value 1 as defined in clause 8.2 of Rec. ITU-T H.273 is permitted. |
| Frame rates | The permitted values are 60, 60/1.001, 50, 30, 30/1.001, 25, 24, 24/1.001 fps. |
| Frame packing | No frame packing is applied. |
| Projection | No projection is used. |
| Sample aspect ratio | The pixel aspect ratio is 1 (square pixel), i.e. only the value 1 as defined in clause 7.3 of Rec. ITU-T H.273 is permitted. |
| Chroma sample location type | the location of chroma samples relative to the luma samples for progressive frames as defined in Rec. ITU-T H.273, clause 8.7 is set to 0 (chroma samples are horizontally co-sited with and vertically centered between the first luma sample at the top-left corner and the first two luma samples at the top-left corner, respectively). |
| Range | The restricted video range is used.  |

# ===== CHANGE =====

#### 4.4.3.3 High Dynamic Range TV

3GPP High Dynamic Range (HDR) TV formats are defined based on Rec. ITU-R BT-2100-2 [bt2100]. 3GPP HDR TV formats shall conform to ITU-R BT-2100-2 [bt2100] with the following restrictions:

- Only 4:2:0 colour subsampling is considered

- Only the Non-Constant Luminance Y'C'BC'R signal format is considered

- Only 10-bit representations are considered

An informative summary of the parameters of a 3GPP HDR TV format based on the parameters defined in Table 4.4.2-1 is provided in Table 4.4.3.3-1.

Table 4.4.3.3-1 Video Signal Parameters for 3GPP HDR TV format

|  |  |
| --- | --- |
| Parameter | Restrictions |
| Picture aspect ratio | 16:9 |
| Spatial Resolution width x height | 7680 × 4320, 3840 × 2160, 1920 × 1080NOTE: For 1080, typically the encoded signal has 1088 lines and cropping is applied to remove spatial samples that are not presented. |
| Scan Type | the source scan type of the pictures as defined in clause 7.3 of Rec. ITU-T H.273 is progressive |
| Chroma format indicator | The chroma format indicator is 4:2:0.  |
| Bit depth | The permitted value is 10 bit. |
| Colour primaries  | Only the value 9 as defined in clause 8.2 of Rec. ITU-T H.273 is permitted. |
| Transfer Characteristics | Only the values 16 (for PQ) and 18 (for HLG) as defined in clause 8.2 of Rec. ITU-T H.273 are permitted. |
| Matrix Coefficients | Only the value 9 as defined in clause 8.2 of Rec. ITU-T H.273 is permitted. |
| Frame rates | The permitted values are 120, 120/1.001,100, 60, 60/1.001, 50, 30, 30/1.001, 25, 24, 24/1.001 fps. |
| Frame packing | No frame packing is applied. |
| Projection | No projection is used. |
| Sample aspect ratio | The pixel aspect ratio is 1 (square pixel), i.e. only the value 1 as defined in clause 7.3 of Rec. ITU-T H.273 is permitted. |
| Chroma sample location type | the location of chroma samples relative to the luma samples for progressive frames as defined in Rec. ITU-T H.273, clause 8.7 is set to 2 (chroma samples are co-sited with the luma samples at the top-left corner). |
| Range | The restricted video range is used.  |

# ===== CHANGE =====

#### 4.4.3.4 3GPP Stereoscopic Cinema Format

The stereoscopic 3D TV format uses two signals, one for the left eye and another view for the right eye as defined in Table 4.4.2-1. The components for each eye closely follow the specifications of the 3GPP HDR signals, but there are some restrictions and extensions, namely:

- Only 4:2:0 colour subsampling is considered

- Frame rates include high frame rate for movies, namely 48 fps.

- the spatial resolution is restricted to a maximum value of 4K

- Only the Non-Constant Luminance Y'C'BC'R signal format is considered

An informative summary of the parameters of a 3GPP Stereoscopic 3D TV format based on the parameters defined in Table 4.4.2-1 is provided in Table 4.4.3.4-1.

Table 4.4.3.4-1 Video Signal Parameters for 3GPP Stereoscopic 3D Cinema format

|  |  |
| --- | --- |
| Parameter | Restrictions |
| Picture aspect ratio | 16:9 |
| Spatial Resolution width x height | 3840 × 2160, 1920 × 1080NOTE: For 1080, typically the encoded signal has 1088 lines and cropping is applied to remove spatial samples that are not presented. |
| Scan Type | the source scan type of the pictures as defined in clause 7.3 of Rec. ITU-T H.273 is progressive |
| Chroma format indicator | The chroma format indicator is 4:2:0.  |
| Bit depth | The permitted values are 8 or 10 bit. 8 bit is only permitted for SDR. |

|  |  |
| --- | --- |
| Colour primariesTransfer CharacteristicsMatrix Coefficients | Only the following value combinations are permitted: (1, 1, 1), (9,14, 9), (9, 16, 9), and (9, 18, 9) for SDR HD, SDR UHD, HDR PQ, and HDR HLG, respectively. |
|  |  |
|  |  |
| Frame rates | The permitted values are 60, 60/1.001, 48, 48/1.001, 50, 30, 30/1.001, 25, 24, 24/1.001 fps. |
| Frame packing | No frame packing is applied. |
| Projection | No projection is used. |
| Sample aspect ratio | The pixel aspect ratio is 1 (square pixel), i.e. only the value 1 as defined in clause 7.3 of Rec. ITU-T H.273 is permitted. |
| Chroma sample location type | For SDR HD, the location of chroma samples relative to the luma samples for progressive frames as defined in Rec. ITU-T H.273, clause 8.7 is set to 0.For SDR UHD, HDR PQ, and HLG, the location of chroma samples relative to the luma samples for progressive frames as defined in Rec. ITU-T H.273, clause 8.7 is set to 2. |
| Range | The restricted video range is used.  |
| Stereoscopic Video | A signal for the Left and for the Right Eye is provided whereby the signals have the identical parameters as above and are timely synchronized. |

# ===== CHANGE =====

## 4.5 Reference API parameters

### 4.5.1 Introduction

When media is played back, the decoder and the playback pipeline need to be initialized. For this purpose, certain parameters are required. In CTA-5003 [DPC], a media playback model is described that is aligned with with HTML 5.1 and the <video> element, as well as the Media Source Extensions.

### 4.5.2 Video Decoder API Parameters

Based on CTA-5003 [DPC], Table 4.5.2-1 provide relevant parameters that need to be attached to the content, in order to establish media playback properly, and serve as an API. The parameters are used for the following purposes:

- to identify the capability of the device in order to check of the signal can be played back

- to initialize the decoding and playback platform to allocate the resources for decoding and rendering

Table 4.4.3.4-1 Video Signal Parameters for 3GPP Stereoscopic 3D TV format

|  |  |  |
| --- | --- | --- |
| Parameter | Restrictions | Status |
| width | specifies the width of a video player, in pixels | required |
| height | specifies the width of a video player, in pixels. | required |
| media type | specifies the media type of the component, in this case video | required |
| format | specifies the format of the media, for example mp4 | required |
| profiles | specifies the profile of the format, for example 'cmfc' | optional |
| codecs | specifies through a well-defined string the codec used for the signal  | required |
| Video format parameters | specifies additional video format parameters as defined in Table 4.4.2.1 to describe the signal and to initialize the encoder. | optional |

### 4.5.3 Video Encoder API Parameters

Video encoder API parameters are for further study.