3GPP TSG-SA WG4 Meeting #131-bis-eS4- 250602

Online, 11 – 17 April 2025

**Source: Xiaomi**

**Title: Pseudo-CR on 26.265**

**Spec: 3GPP TS 26.265**

**Agenda item: 9.7**

**Document for: Decision**

**1. Introduction**

This contribution adds video operation points to support monochrome with the purpose of carrying depth and alpha maps in 26.265.

**2. Reason for Change**

First change introduces defines monochrome operation point to support carrage of depth and alpha maps.

Next change updates the table that lists the operation points to list the introduced monochrome OP.

Next change updates the HDTV video format, to support carriage of monochrome bitstreams.

Next change introduces a FullHDMono-Dec decoding capabilities profile to support decoding of monochrome bitstreams.

Next change introduces a FullHDMono-Enc encoding capabilities profile to support decoding of monochrome bitstreams.

**4. Proposal**

It is proposed to agree the following changes to 3GPP TS26.265 v1.0.0.

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

<ALL NEW TEXT>

6.3.7 3GPP HEVC Monochrome Operation Point

6.3.7.1 Introduction

The HEVC Monochrome Operation Point permits consistent distribution of monochrome video using HEVC. The remainder of this clause 6.3.7 defines the Bitstream and Receiver requirements for the 3GPP-HEVC-Monochrome receiver.

6.3.7.2 Bitstream Requirements

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

- the Bitstream shall be an **HEVC-Progressive Bitstream** as defined in clause 4.5.3.

Editor’s Note: The **HEVC-Format** Bitstream is not clearly defined in 4.5.3. Maybe to be added when clarified.

- the Representation Format included in the Bitstream shall conform to the 3GPP-HDTV Representation format as defined in clause 4.4.3.2.

Editor’s Note: We are proposing extending the 3GPP-HDTV Representation format – alternatively a new format can be created instead.

- the Bitstream shall be decodable by a decoder with **HEVC-FullHDMono-Dec** decoding capabilities.

Based on this, the following additional restrictions apply

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

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

- The value of chroma\_sample\_loc\_info\_present shall be set to 0.

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.

6.3.7.3 Receiver Requirements

Receivers conforming to the Operation Point 3GPP-HEVC-Monochrome shall support decoding and rendering Bitstreams with the restrictions defined in clause 6.3.7.2.

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

Receivers should ignore the content of all Video Parameter Sets (VPS) NAL units as defined in Recommendation ITU-T H.265 / ISO/IEC 23008-2 [h265].

NOTE 2: The VPS may be present to address requirements in other Operation Points, but the Bitstream also conforms to this Operation point.

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.

\* \* \* 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.

Editor’s Note: We are proposing extending the 3GPP-HDTV Representation format – alternatively a new format can be created instead.

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Video Format** | **Decoding Capabilities** | **Definition** |
| 3GPP-AVC-HD | 3GPP-HDTV (see clause 4.4.3.2) | AVC-FullHD-Dec (see clause 5.4) | 6.2.2 |
| 3GPP-HEVC-HD | 3GPP-HDTV (see clause 4.4.3.2) | HEVC-FullHD-Dec (see clause 5.4) | 6.3.2 |
| 3GPP-HEVC-HD-MONO | 3GPP-Mono (see clause 4.4.3.5) | HEVC-FullHDMono-Dec (see clause 5.4) | 6.3.7 |
| 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-3D | 3GPP-3DTV (see clause 4.4.3.4) | HEVC-UHD-Dec-2 (see clause 5.5) | 6.3.4 |
| 3GPP-MVHEVC-3D | 3GPP-3DTV (see clause 4.4.3.4) | MVHEVC-UHD-2 (see clause 5.5) | 6.3.6 |

\* \* \* Next 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. A component refers to an array or single sample from one of the three arrays (luma and two chroma) that compose a picture. 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** | **Service or Application 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 restrictions | Restrictions possible |
| Spatial Resolution height | The number of active lines per picture for the luma component.  Example values are 720 or 1080 for HD, and 2160 for UHD.  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 restrictions | Restrictions possible |
| 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 or 4:0:0 |  |
| 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  Luma and chroma components shall not differ |  |
| 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 the Rec. ITU-R BT.2100-2 perceptual quantization (PQ) system, or 18 to refer to the Rec. ITU-R BT.2100-2 hybrid log-gamma (HLG) system | BT.709, BT.2020 SDR, BT.2100 PQ, or BT.2100 HLG |  |
| Matrix Coefficients | Describes the matrix coefficients used in deriving the 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 in 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. | YCbCr BT.709, YCbCr BT.2020, or YCbCr 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 restrictions | 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. | Some 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. | Some 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 a value of 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 parameter VideoFullRangeFlag.  For video applications 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 (value set to 1) is commonly used. | No specific restrictions, but 0 is expected if not present. |  |
| Stereoscopic Video | Visual media may be stereoscopic, in which case 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).  NOTE: When distributing the signal, some systems may use different resolutions for one of the views.  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 that 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. |  |  |
| Monochrome Video | The video signal may monochrome, storing information only in the luma samples. Monochrome video can be used to store information for alpha and depth. |  |  |

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

<ALL NEW TEXT>

4.4.3.5 3GPP Monochrome Format

The 3GPP Monochrome (Mono) format uses a video signal to transmit depth and alpha information. It is based on the 3GPP HDTV, with the following changes:

- 4:0:0 colour subsampling is also considered.

- Full video range is used.

- Colour primaries are limited to those for depth and alpha.

An informative summary of the parameters of a 3GPP Monochrome format based on the parameters defined in Table 4.4.2-1 is provided in Table 4.4.3.5-1.

Table 4.4.3.5-1 Video Signal Parameters for 3GPP Mono format

|  |  |
| --- | --- |
| Parameter | Restrictions |
| Picture aspect ratio | 16:9 |
| Spatial Resolution width x height | 1920 × 1080  NOTE 1: Down-sampled resolutions may be created for distribution, for example in case of adaptive streaming.  NOTE 2: To accommodate the block coding structure of a given specification, quite often the encoded signal may be padded. In such cases, normative cropping is typically applied to remove spatial samples that are not intended to be 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 4:0:0. |
| Bit depth | The permitted values are 8 or 10 bit. The bit depth is the same for all samples. |
| Transfer Characteristics | Only the value 1, as defined in clause 8.1 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 the 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 | Full video range is used. |

\* \* \* Next 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* constraints as defined in clause 4.5.3.

- **HEVC-FullHDMono-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* 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* constraints as defined in clause 4.5.3.

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

5.4 Single-Instance Encoding Capabilities

The following encoding capabilities are defined:

**- AVC-FullHD-Enc:** the capability to encode a video signal to a bitstream that is decodable by a decoder that is *AVC-FullHD-Dec* capable as defined in clause 5.3 with the following additional constraints:

- up to 245,760 macroblocks per second;

- up to a frame size of 8,192 macroblocks;

- up to 240 frames per second;

- the chroma format being 4:2:0; and

- the bit depth being 8 bit;

NOTE 1: The 3GPP HDTV format if restricted to 8 bit as defined in clause 4.4.3.2 may be encoded with an **AVC-FullHD-Enc** capable encoder.

- **HEVC-HD-Enc**: the capability to encode a video signal with

- up to 33,177,600 luma samples per second;

- up to a luma picture size of 983,040 samples;

- up to 120 frames per second;

- the chroma format being 4:2:0; and

- the bit depth being 8 bit;

to a bitstream that is decodable by a decoder that is **HEVC-HD-Dec** capable as defined in clause 5.3.

NOTE 2: A restricted version of the 3GPP HDTV format as defined in clause 4.4.3.2 may be encoded with an HEVC-HD-Enc capable encoder.

**- HEVC-FullHD-Enc:** the capability to encode a video signal to a bitstream that is decodable by a decoder that is *HEVC-FullHD-Dec* capable as defined in clause 5.3 with the following additional constraints:

- up to 133,693,440 luma samples per second;

- up to a luma picture size of 2,228,224 samples;

- up to 240 frames per second;

- the chroma format being 4:2:0; and

- the bit depth being either 8 or 10 bit;

**- HEVC-FullHDMono-Enc:** the capability to encode a video signal to a bitstream that is decodable by a decoder that is *HEVC-FullHDMono-Dec* capable as defined in clause 5.3 with the following additional constraints:

- up to 133,693,440 luma samples per second;

- up to a luma picture size of 2,228,224 samples;

- up to 240 frames per second;

- the chroma format being 4:0:0; and

- the bit depth being either 8 or 10 bit;

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