ETSI TR 126 966 V18.0.0 (2024-05)

TECHNICAL REPORT

5G;
Evaluation of new HEVC coding tools
(3GPP TR 26.966 version 18.0.0 Release 18)

Reference

DTR/TSGS-0426966vi00

Keywords

5G

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Important notice

The present document can be downloaded from:
https://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

If you find a security vulnerability in the present document, please report it through our Coordinated Vulnerability Disclosure Program:
https://www.etsi.org/standards/coordinated-vulnerability-disclosure

Notice of disclaimer & limitation of liability

The information provided in the present deliverable is directed solely to professionals who have the appropriate degree of experience to understand and interpret its content in accordance with generally accepted engineering or other professional standard and applicable regulations.
No recommendation as to products and services or vendors is made or should be implied.
No representation or warranty is made that this deliverable is technically accurate or sufficient or conforms to any law and/or governmental rule and/or regulation and further, no representation or warranty is made of merchantability or fitness for any particular purpose or against infringement of intellectual property rights.
In no event shall ETSI be held liable for loss of profits or any other incidental or consequential damages.

Any software contained in this deliverable is provided "AS IS" with no warranties, express or implied, including but not limited to, the warranties of merchantability, fitness for a particular purpose and non-infringement of intellectual property rights and ETSI shall not be held liable in any event for any damages whatsoever (including, without limitation, damages for loss of profits, business interruption, loss of information, or any other pecuniary loss) arising out of or related to the use of or inability to use the software.

Copyright Notification

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Legal Notice

This Technical Report (TR) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities. These shall be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between 3GPP and ETSI identities can be found under https://webapp.etsi.org/key/queryform.asp.

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

In the present document, certain modal verbs have the following meanings:

shall indicates a mandatory requirement to do something

shall not indicates an interdiction (prohibition) to do something

NOTE 1: The constructions "shall" and "shall not" are confined to the context of normative provisions, and do not appear in Technical Reports.

NOTE 2: The constructions "must" and "must not" are not used as substitutes for "shall" and "shall not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

should indicates a recommendation to do something

should not indicates a recommendation not to do something

may indicates permission to do something

need not indicates permission not to do something

NOTE 3: The construction "may not" is ambiguous and is not used in normative elements. The unambiguous constructions "might not" or "shall not" are used instead, depending upon the meaning intended.

can indicates that something is possible

cannot indicates that something is impossible

NOTE 4: The constructions "can" and "cannot" shall not to be used as substitutes for "may" and "need not".

will indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

will not indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

might indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

might not indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

is (or any other verb in the indicative mood) indicates a statement of fact

is not (or any other negative verb in the indicative mood) indicates a statement of fact

NOTE 5: The constructions "is" and "is not" do not indicate requirements.

1 Scope

This Technical Report gathers the opportunities for improving HEVC-based services. This includes documentation of motivating use cases and scenarios. Specifically, potential of improving on the following use cases are identified: the compression performance for stereoscopic 3D content, the network performance related to exploding adaptive streaming traffic, and the demands for very high-quality image applications. HEVC based solutions to address each opportunity are identified: HEVC Multiview profiles, HEVC Scalable profiles, and HEVC 4:4:4 (up to 10 bits) capable profiles. Methodologies to investigate and document the pros and cons of the proposed solutions for each use case are documented. Finally, conclusions are drawn on the relevancy of solutions and if any new normative specification work is to be done.

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".

[2] ISO/IEC 14496-10:2022: "Information technology — Coding of audio-visual objects — Part 10: Advanced video coding"

[3] ISO/IEC 23008-2:2015: "Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding"

[3] 3GPP TR 26.905: "Mobile stereoscopic 3D video".

[4] 3GPP TS 26.247: "Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH)".

[5] 3GPP TS 26.244: "Transparent end-to-end packet switched streaming service (PSS); 3GPP file format (3GP)".

[6] 3GPP TS 26.214: "IP Multimedia Subsystem (IMS); Multimedia Telephony; Media handling and interaction".

[7] 3GPP TS 26.218: "Virtual Reality (VR) profiles for streaming applications"

[8] 3GPP TS 26.347: "Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs"

[9] Vetro, Anthony. "Frame compatible formats for 3D video distribution." In 2010 IEEE International Conference on Image Processing, pp. 2405-2408. IEEE, 2010.

[10] Hannuksela, Miska M., Ye Yan, Xuehui Huang, and Houqiang Li. "Overview of the multiview high efficiency video coding (MV-HEVC) standard." In 2015 IEEE International Conference on Image Processing (ICIP), pp. 2154-2158. IEEE, 2015.

[11] ISO/IEC JTC1/SC29/WG11 MPEG2011 M22746, "AVC/MVC anchor coding for MFC", November 2011, Geneva, Switzerland.

[12] ISO/IEC JTC1/SC29/WG11 N16050, "MV-HEVC Verification Test Report", San Diego, US, Feb. 2016.

[13] ISO/IEC 14496-15:2022, "Information technology — Coding of audio-visual objects — Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format"

[14] "HTTP Live Streaming (HLS) authoring specification for Apple devices," https://developer.apple.com/documentation/http-live-streaming/hls-authoring-specification-for-apple-devices

[15] "ISO Base Media File Format and Apple HEVC Stereo Video Format additions," Version 0.9 (Beta) June 21, 2023

[16] "Apple HEVC Stereo Video," Interoperability Profile Version 0.9 (Beta) June 21, 2023

[17] Delbracio, Mauricio, Damien Kelly, Michael S. Brown, and Peyman Milanfar. "Mobile computational photography: A tour." Annual Review of Vision Science 7 (2021): 571-604.

[18] Camera & Imaging Products Association (CIPA) "Production, Shipment of Digital Still Camera January, January-January in 2017," 2016

[19] "Smartphones vs Cameras: Closing the gap on image quality," https://www.dxomark.com/smartphones-vs-cameras-closing-the-gap-on-image-quality/

[20] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG JVT-I018, "Color format downconversion for test sequence generation," 2003.

[21] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG JVT-I019, "Color format upconversion for video display," 2003.

[22] ISO/IEC 23008-12:2022: "Information technology - MPEG systems technologies - Part 12: Image File Format".

[23] ISO/IEC 14496-12:2022: "Information technology — Coding of audio-visual objects — Part 12: ISO base media file format".

[24] "Using HEIF or HEVC media on Apple devices," https://support.apple.com/en-us/HT207022

[25] "HEIF Imaging," https://source.android.com/docs/core/camera/heif

[26] ITU-T Recommendation T.81: "Information technology; Digital compression and coding of continuous-tone still images: Requirements and guidelines".

[27] 3GPP TR 26.948: "Study on video enhancements in 3GPP multimedia services"

[28] HTTP Live Streaming (HLS) Authoring Specification for Apple Devices, https://developer.apple.com/documentation/http_live_streaming/http_live_streaming_hls_authoring_specification_for_apple_devices

[29] Samira Afzal, Vanessa Testoni, Christian Esteve Rothenberg, Prakash Kolan, Imed Bouazizi, “A holistic survey of multipath wireless video streaming”, Journal of Network and Computer Applications, 212: 103581 (2023)

[30] ISO/IEC JTC1/SC29/WG11 N16051, "SHVC verification test report", February 2016, San Diego, USA.

[31] ISO/IEC JTC1/SC29/WG11 N16268, "Supplemental SHVC verification test report", June 2016, Geneva, CH.

[32] 3GPP TR 26.955: "Video codec characteristics for 5G-based services and applications"

[33] ISO/IEC 23000-19:2020, "Information technology — Multimedia application format (MPEG-A) — Part 19: Common media application format (CMAF) for segmented media"

[34] ISO/IEC JTC1/SC29/WG03 N01026, "Preliminary WD of ISO/IEC 23000-19 AMD New Structural CMAF Brand Profile", October 2023, Hannover, Germany.

[35] Recommendation ITU-R BT.2095-1 "Subjective assessment of video quality using expert viewing protocol (2016-2017) ", 06/2017.

[36] ISO/IEC JTC1/SC29/WG03 N01033, "Technology under consideration on CMAF", October 2023, Hannover, Germany.

[37] G. Tech, Y. Chen, K. Müller, J. -R. Ohm, A. Vetro and Y. -K. Wang, "Overview of the Multiview and 3D Extensions of High Efficiency Video Coding," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 35-49, Jan. 2016, doi: 10.1109/TCSVT.2015.2477935.

[38] https://developer.apple.com/av-foundation/HEVC-Video-with-Alpha-Interoperability-Profile.pdf

[39] Fehn, Christoph. (2004). Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV. Proc SPIE. 5291.

[40] S. Shimizu and S. Sugimoto, ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Document JCT3V-G0151, “AHG 13: Results with quarter resolution depth map coding”, Jan. 2014.

[41] K. Wegner and O. Stankiewicz, ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Document JCT3V-B0151, “3D-HEVC with reduced resolution of depth”, Oct. 2012.

[42] Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 Document JCTVC-AA0039, "Additional methods for Luma Adjustment," April 2017.

[43] 3GPP TR 26.928: "Extended Reality (XR) in 5G"

3 Definitions of terms and abbreviations

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].

3.2 Symbols

For the purposes of the present document, the following symbols apply:

3.2 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].

4 Background

The video codec characteristics for 5G services are documented in TR 26.955 [32], and they demonstrate that the HEVC coding standard provides satisfactory performance to fulfil the needs of video service studied in the TR. It also recommended to consider upgrading specifications to support profiles, levels, and possibly features available in HEVC, including features that may include XR/AR type of services, as well as low and very low latency services. There is interest in the distribution, including streaming, of 3D movie content. Finally, the use of scalability could further enhance multi-bitrate systems such as video conferencing, or adaptive streaming, but may also provide additional benefits to end user devices, such as power adaptation. HEVC may be suitable to cater and enable such applications. This specification outlines these emerging applications for video coding, gather evidence whether specific new tools can provide advantage for specific services and applications, and conclude if normative specification work is needed on these aspects.

5 Scenarios

5.1 Scenario #1.1: Streaming of stereoscopic 3D content

5.1.1 Overview

There has been renewed interest in the distribution, including streaming, of 3D movie content, as evident by media coverage of recent 3D movie releases. Consumption of stereoscopic 3D video content is expected to rapidly grow given new AR related products beings launched.

5.1.2 Review of previous work

Evaluation of AVC based stereoscopic 3D coding techniques has been done in TR 26.905 [3] and its normative support has been added for 3GPP DASH in TS 26.247 [3], the 3GPP file format in TS 26.244 [5], IMS in TS 26.114 [6], VR profiles in TS 26.118 [7], and MBMS in TS 26.347 [8]. The work done in TR 26.905 [3] for Rel-11 focused mostly on stereoscopic viewing on TVs, while today's applications have grown far beyond these, given especially advancements in AR devices. Also, today's requirements on quality are much higher owing to higher quality displays and the available channel capacities.

Simulcast and frame packed HEVC video operating points are specified in TS 26.118 for VR streaming scenarios. With the established support for MV-AVC, simulcast and frame packed HEVC in 3GPP SA4 specifications, an assessment needs to be done to upgrade the support for multiview coding using MV-HEVC with its superior coding performance.

5.1.3 Evaluation criteria and metrics

The evaluation for the coding performance for stereoscopic 3D content needs to be performed based on the following evaluation criteria.

1. Assessment/discussion of hardware impact: there are two possibilities for this:

   a) There is existing hardware product-grade support for the tool. In that case, refer to the example hardware.

   b) There is no existing hardware support. In this case, a discussion/description with justifications on the expected impact on hardware implementation is provided, or reference to existing demos etc.

2. Codec performance evaluation:

   a) PSNR-based Rate-Distortion (RD) objective performance evaluation, where the RD performance is compared for various solutions with a fixed QP encoding setting to get the plotting data points. A better PSNR-based RD performance is preferred, keeping in view the expected hardware complexity impact.

   b) Subjective performance evaluation.

5.1.4 Evaluation methodology

5.1.4.1 Objective performance evaluation

For objective performance evaluation, suitable source test content is identified that is accepted by video experts as representative content. Some of the important parameters for the content are the resolution, framerate, bit depth, color subsampling, and duration, in addition to the number of views available. Reference software for a specific solution is to be used with fixed QP encoding settings to generate each plotting point on the PSNR RD curves. The encoding settings (e.g. prediction types IPP or IBP etc.) are decided by experts considering the complexity and latency needs for the scenario. The resulting curves can directly be used for comparison by plotting together or by comparing the Bjøntegaard Delta (BD) bitrate.

5.1.4.2 Subjective performance evaluation

Recommendation ITU-R BT.2095-1 Subjective assessment of video quality using expert viewing protocol [35] describes the method to subjectively assess video quality by means of the expert viewing protocol (EVP), with the participation of a reduced number of viewers, all selected among experts in the relevant video processing area. This methodology has been used in JVET for the assessment of multiview video codec performance. The EVP visual evaluation protocol is specified in detail in [35] with the following main features:

1. 9 experts participate as viewers in each EVP session,
2. The “unimpaired” Source video Clip (SRC) is shown once, followed by two Processed Video Sequences (PVSs),
3. Experts are required to compare the PVS with the SRC, and to rate them separately.

5.2 Scenario #1.2: Low delay applications of stereoscopic 3D video

5.2.1 Overview

While scenario #1.1 focuses the use case of streaming of stereoscopic 3D content, there are several other use cases for such content where the latency requirements are stricter compared to the lax latency requirements of the streaming use case. For example, with the advent of modern-era XR devices, video conversational applications exchange stereoscopic 3D content. Some of the other use cases may include the stereoscopic content exchange for split rendering over edge where a (partially) rendered stereoscopic view may be exchanged between the edge cloud server and the device. Such low latency applications will demand different source formats (resolutions, framerates etc.), coding settings, as well as transport considerations to cater for this lower latency requirement.

5.2.2 Review of previous work

The evaluation of AVC based stereoscopic 3D coding techniques done in TR 26.905 [3] was primarily focused on download and streaming scenarios. Similarly, most other normative aspects specified had been for download or streaming use cases e.g. in 3GPP DASH in TS 26.247 [3], the 3GPP file format in TS 26.244 [5], IMS in TS 26.114 [6], VR profiles in TS 26.118 [7], and MBMS in TS 26.347 [8]. Reduced resolution frame packing is not sufficient because of the detrimental impact on quality due to resampling, as noted in TR 26.905 [3].

TR 26.928 [43] (study on Extended Reality (XR) in 5G) has documented a video resolution of 2k x 1k per eye at 50/60 fps, 4-10 Mbps (viewport-dependent) in context of quality and bitrate considerations for omnidirectional visual formats, similarly in clause 6.3.8 (XR conversational application). Further traffic characteristics were not documented (noted as FFS).

Hence in addition to a study on the streaming applications of stereoscopic 3D video content, realtime delivery aspects also need to be studied.

5.2.3 Evaluation criteria and metrics

The evaluation for the performance of coding stereoscopic 3D content for low delay applications can be done in alignment with the evaluation for streaming applications. However, low delay configurations instead of random access ones, would need to be considered. Additional criteria include:

1. Assessment/discussion of hardware impact; there are two possibilities for this:
    a. There is existing hardware product-grade support for the tool. In that case, refer to the example hardware.
    b. There is no existing hardware support. In this case, a discussion/description with justifications on the expected impact on hardware implementations is provided, or reference to existing demos etc.
2. Codec performance evaluation:
    a. PSNR-based Rate-Distortion (RD) objective performance evaluation, where the RD performance is compared for various solutions with a fixed QP encoding setting to get the plotting data points. A better PSNR-based RD performance is preferred, keeping in view the expected hardware complexity impact.

    b. Subjective performance evaluation.

5.2.4 Evaluation methodology

5.2.4.1 Objective performance evaluation

For an objective performance evaluation, suitable source test content should be identified that is accepted by video experts as representative content. Some of the important parameters for the content include the resolution, framerate, bit depth, color subsampling, and the duration, of the content in addition to the number of views available. Reference software for a specific solution is to be used with fixed QP encoding settings to generate each plotting point on the PSNR RD curves. The encoding settings (e.g. prediction types IPP or IBB etc.) are to be decided by experts, considering the complexity and latency needs for the scenario. The resulting curves can directly be used for comparison by plotting them together with an anchor, i.e. simulcast encoding of both views, or by computing the Bjøntegaard Delta (BD) rate metric compared to the anchor.

5.2.4.2 Subjective performance evaluation

Same considerations are made as in clause 5.1.4.2, i.e. relying on previous strategy adopted by JVET for assessment of multiview video codec performance by using EVP [35].

5.3 Scenario #2: High quality photography

5.3.1 Overview

The demand for high quality photography has been and continues to stay a dominating factor in cell phone market growth [17]. Reports such as [18] (processed and published by [19]) have shown in the past that smartphone shipments have been devouring not just point-and-shoot but also high-end DSLR cameras, by closing the gap in image quality. Additional encoding tools are needed to progress further in achieving even higher image quality.

5.3.2 Review of previous work

JPEG-based still image [26] support is provided in SA4 specifications, and suitable extensions to attain an even higher quality are explored in this scenario.

5.3.3 Evaluation criteria and metrics

The evaluation for high quality image encoding tools shall be done based on the following evaluation criteria.

1. Assessment/discussion of hardware impact: there are two possibilities for this:
    a. There is existing hardware product-grade support for the tool. In that case, refer to the example hardware.
    b. There is no existing hardware support. In this case, a discussion/description with justifications on the expected impact on hardware implementation is provided, or reference to existing demos etc.
2. Codec performance evaluation:
    a. Objective performance evaluation: e.g. PSNR-based Rate-Distortion (RD) performance evaluation, where the RD performance is compared for various solutions. A better PSNR-based RD performance is preferred, keeping in view the expected hardware complexity impact.

5.4 Scenario #3: Optimising multi-bitrate delivery

5.4.1 Overview

New video codecs have potential to assist further in optimising multi-bitrate delivery applications such as video conferencing, or adaptive streaming, and may also provide additional benefits to end user devices, such as power adaptation. One specific target of optimization is the storage space savings achieved by employing scalable video.

5.4.2 Review of previous work

SA4 has studied SHVC in TR 26.948 [27] in 2015, there are however possibility of exploring new scenarios since that time that will be pursued here.

5.4.3 Evaluation criteria and methodology

1. Assessment/discussion of hardware impact: there are two possibilities for this:
    a. There is existing hardware product-grade support for the tool. In that case, refer to the example hardware.
    b. There is no existing hardware support. In this case, a discussion/description with justifications on the expected impact on hardware implementation is provided, or reference to existing demos etc.
2. Codec performance evaluation:
    a. The performance evaluation of positive impact on streaming will be determined by the savings of storage space w.r.t. conventional streaming with similar quality. Calculations are to be done on representative scenario for adaptive streaming.

5.5 Scenario #4: Pose correction optimisation

5.5.1 Overview

This scenario deals with a split-rendering case where the device is running a pose correction method (e.g., using ATW). While pose correction is a good solution to cope with the latency introduced by the roundtrip communication and the rendering, it can introduce visual artifacts if only 2D projected images are used. As an example, a rendered scene may be composed by multiple elements having different sensitivity to time warping. For instance, the user-interface (UI) does not need to be corrected as its position won't change in the user's Field of View (FoV). A 3D object near the user may benefit from a time warping as the pose correction would address parallax differences. The far away background similarly to the UI does not need warping as parallax fall off in the distance. This is illustrated in the Figure 5.5.1-1 below.

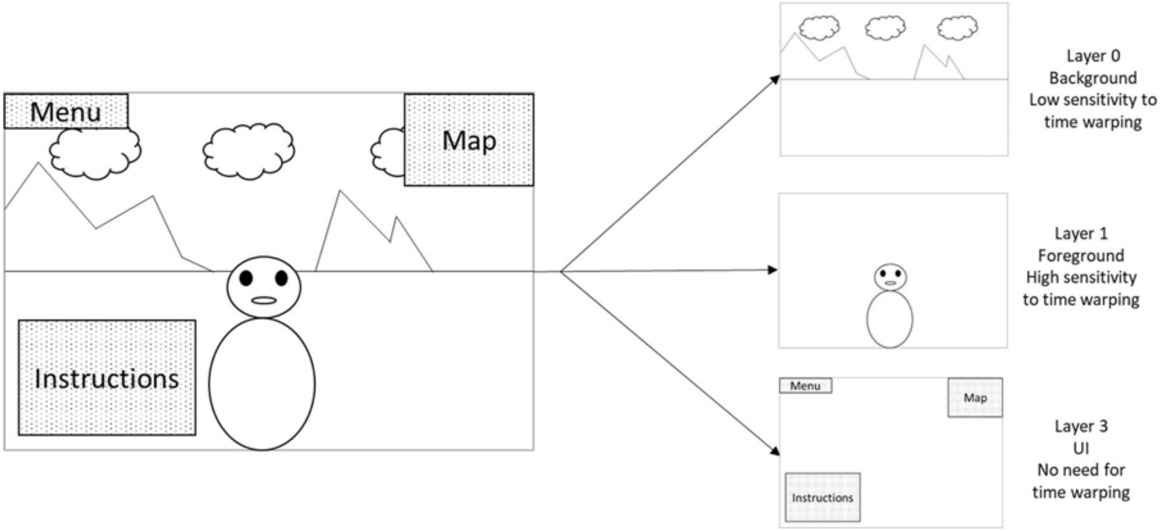

Figure 5.5.1-1: Illustration of layering in rendering

To maintain the effectiveness of pose correction, a rendering engine may apply segmentation and generate multiple layers of projected texture images that may be handled differently based on their time-warping sensitivity. Those different texture layers could be encoded and processed separately in multiple video streams but could also be encoded in a single stream with additional depths and alpha channels.

To drive the pose-correction and maximize the QoE, additional information may be provided to support segmentation into layers and to support the pose correction, indicating how the different texture layers should be handled by the pose correction engine. However, such optional metadata is currently not supported by OpenXR APIs.

Generally, the carriage of depth and alpha channels in the video bitstreams for proper scene and UI restitution allows to improve pose correction. New video codecs have the potential to address this scenario in a bandwidth efficient manner.

5.5.2 Review of previous work

The carriage of depth or alpha auxiliary channels has not been addressed until now.

5.5.3 Evaluation criteria and methodology

1. Assessment/discussion of hardware impact: there are two possibilities for this:
   a. There is existing hardware product-grade support for the tool. In that case, refer to the example hardware.
   b. There is no existing hardware support. In this case, a discussion/description with justifications on the expected impact on hardware implementation is provided, or reference to existing demos etc.
2. Codec performance evaluation can be evaluated in two possible ways:
   a. For single layer case, the performance evaluation of impact on bandwidth will be determined by the overhead introduced by adding additional channels to the video (alpha, depth, …) compared to traditional approach. It is expected that the additional cost is negligible.
   b. For multi-stream case, the performance evaluation of impact on bandwidth will be determined by measuring the overhead introduced by multiple encodings compared to a single-layer approach. It is expected that the additional cost is low.

6 Solutions

6.0 Mapping of Solutions to Scenarios

Table 6.0-1: Mapping of Solutions to Scenarios

Solution #	Solution Title	Scenario(s)
#1.1	HEVC simulcast	#1.1, #1.2
#1.2	HEVC Frame packing	#1.1, #1.2
#1.3	Multiview HEVC coding	#1.1, #1.2
#2.1	HEVC 4:2:0 coding	#2
#2.2	HEVC 4:2:2 coding	#2
#2.3	Native 4:4:4 coding - HEVC Main 4:4:4 profiles	#2
#2.4	Derived 4:4:4 coding - Layered use of HEVC 4:2:0 profiles	#2
#3.1	Scalable HEVC coding	#3
#4.1	MV-HEVC with auxiliary depth/alpha channels	#4

6.1 Solution #1.1: HEVC simulcast

6.1.1 Introduction

HEVC simulcast is considered as a baseline solution to addresses Scenario#1.

6.1.2 High-level Description

6.1.2.1 Overview MV-HEVC

This baseline solution uses two independent High Efficiency Video Coding (HEVC) [3] streams to transport the left- and right-eye view of the stereoscopic content. It represents a baseline or reference scenario that does not exploit any redundancy of the views during coding. Based on this fact that this simplistic solution does not optimize the performance, and due to its impacts that are noted in later in the evaluation, it is never practically used and is documented for reference/benchmark purpose only.

6.1.2.2 Transport of HEVC Simulcast

As noted in the overview, this solution is relevant for benchmark/reference purpose only and is not deployed, hence there is no existing support for its transport.

6.1.3 Evaluation

6.1.3.1 Assessment/discussion of hardware impact

This solution would require two independent video decoders, each to decode a given view, and hence it requires twice as much hardware for decoding as for a single 2D video stream.

6.1.3.2 Codec performance evaluation based on existing results

Subjective evaluation results using this technique as a reference to compare with 8-bit MV-HEVC are documented in [12].

6.2 Solution #1.2: HEVC frame packing

6.2.1 Introduction

HEVC frame packing is considered a solution that addresses Scenario#1.

6.2.2 High-level Description

6.2.2.1 Overview HEVC frame packing

Frame packing can be used as one of the options to deliver multiview (stereoscopic) video content. This solution is focused on reusing existing decoding HW and SW to deliver stereoscopic content and utilizes SEI messages to indicate how the content should be interpreted for viewing. For example, the frame packing arrangement SEI message is specified in the Advanced Video Coding (AVC) and High Efficiency Video Coding (HEVC) [3] cl D.3.16. specifications and could allow indicating a variety of frame packing arrangements, including spatial arrangements such as side-by-side or top-bottom, or temporal interleaving.

6.2.2.2 Transport of HEVC frame packing

The scheme for stereoscopic video arrangements ([23] cl 13.5.4) for restricted media tracks is one example of signalling that allows indicating the frame packing arrangement for a stereo pair.

6.2.3 Evaluation

6.2.3.1 Assessment/discussion of hardware impact

The use of frame packing allows the reuse of existing decoding HW and SW for the compression and delivery of stereoscopic content. SEI messages that identify the frame packing arrangement format used can be indicated in the bitstream to assist the decoding or display process to properly interpret, post-process, and/or display the decoded video data. However, frame packing can have an significant impact on the quality of the representation if full resolution is not used. If full resolution is used, the level requirements of a decoder may need to be increased. Such impact is noted in the following section. The increased sample rate needed for full resolution frame packing maybe the same as that for MV-HEVC.

6.2.3.2 Codec performance evaluation based on existing results

Though existing evaluations between simulcast, MVC, and MV-HEVC are available, as documented in clause 6.3.3.2, evaluations between frame packed HEVC and MV-HEVC are not.

Except for full-resolution spatial packing and temporal interleaving, retaining the same resolution for spatial frame packing with the same decoding level for the decoder would result in reduced video resolution for the views. This can have a considerable impact in visual quality. On the other hand, full resolution frame packing typically require higher level capability HEVC decoders, while also potentially being less efficient than MV-HEVC since it does not permit efficient exploitation of inter-layer redundancies. Spatial frame packing could also result in seam artifacts at the boundaries between two views.

Temporal interleaving would also require supporting double the frame rate and hence may increase the level requirements of the decoder. Although inter-layer prediction can be partially exploited, such is not supported for non-reference pictures in the base-layer, while constraints in the reference buffer specified by HEVC can negatively impact inter prediction.

In conclusion, compared to MV-HEVC, frame packed video:

- video commonly has reduced quality or increased bitrate requirements
- When stereoscopic MV-HEVC based content is used on a non-3D capable device, the content can be played back using only the base view for a 2D presentation. Frame-packed content require the interpretation of the frame packing arrangement SEI message, or analysis of the content to determine whether and, if yes, how the content would need to be processed (e.g. cropped) to extract and display a 2D representation from the decoded pictures.

6.3 Solution #1.2: Multiview HEVC coding

6.3.1 Introduction

This solution addresses Sceanrio#1.

6.3.2 High-level Description

6.3.2.1 Overview MV-HEVC

The Advanced Video Coding (AVC) (H.264) [2] and the High Efficiency Video Coding (HEVC) (H.265) [3] standards were initially intended for the compression of two-dimensional (2D) video. Multi-view extensions for HEVC were then developed, referred to as Multiview Video Coding (MVC) and Multiview HEVC (MV-HEVC) [3][10], respectively. The fundamental principle of both MVC and MV-HEVC is to re-use the coding tools of the underlying 2D AVC and HEVC coding respectively, so that implementations can be realized by software changes to high-level syntax in the slice header level and above [10]. For the case of HEVC, multiview profiles exist for coding both 8- and 10-bit content.

As a reference, MVC has been studied in detail in TR 26.905 [3] and its normative support has been added for 3GPP DASH in TS 26.247 [3], the 3GPP file format in TS 26.244 [5], IMS in TS 26.114 [6], VR profiles in TS 26.118 [7], and MBMS in TS 26.347. MVC does not currently support the encoding of 10-bit content.

6.3.2.2 Transport of MV-HEVC

6.3.2.2.1 Carriage in ISO BMFF

The carriage of MV-HEVC is specified in detail in [13] as one of the "Layered HEVC ((L-HEVC) extensions", including SHVC, MV-HEVC, and 3D-HEVC. Clause 9 of [13] specifies this L-HEVC elementary stream and sample definitions.

6.3.2.2.2 Adaptive Streaming

Encoding and encapsulation guidelines for MV-HEVC in HTTP Live Streaming (HLS) are documented in [14]. Currently, the following recommendations are currently provided for resolutions, bitrates and framerates for both SDR and HDR MV-HEVC content:

16:9 aspect ratio	**MV-HEVC SDR 30 fps**	**MV-HEVC HDR 30 fps**	**Frame rate**
640 x 360	246	272	≤ 30 fps
768 x 432	510	612	≤ 30 fps
960 x 540	1020	1241	≤ 30 fps
960 x 540	1530	1853	≤ 30 fps
960 x 540	2720	3281	Same as source
1280 x 720	4080	4930	Same as source
1280 x 720	5780	6936	Same as source
1920 x 1080	7650	9180	Same as source
1920 x 1080	9660	11900	Same as source
2560 x 1440	13770	16490	Same as source
3840 x 2160	19720	23630	Same as source
3840 x 2160	28560	34000	Same as source

6.3.2.2.3 Support in CMAF

CMAF (ISO/IEC 23000-19 [33]) signalling is required to convey the unique parameters for Multiview video encoded formats (e.g. how the views may be organized in switching sets, what possibilities are allowed or disallowed, which addressable units are relevant to application for stereoscopic vs. 2D displays, etc.).

As noted in clause 6.3.2.2.1, carriage of MV-HEVC is specified in ISO/IEC 14496-15 [13] (NAL-video file format) as one of the "Layered HEVC (L-HEVC) extensions", including SHVC, MV-HEVC, and 3D-HEVC. Clause 9 of 14496-15 specifies this L-HEVC elementary stream and sample definitions. Despite this, there is a need to enable CMAF-level functionality noted above. Based on this need, MPEG has started working on an MV-HEVC extension of CMAF in [34].

6.3.3 Evaluation

6.3.3.1 Assessment/discussion of hardware impact

Support for the multiview profiles of HEVC mostly involves SW level modifications since the support of multiview coding only involves high-level syntax signalling and coding tool considerations [10].

6.3.3.2 Codec performance evaluation based on existing results

The objective and subjective performance results comparing MVC and Simulcast HEVC (each view coded independently) with MV-HEVC are documented in [12]. The test sequences used for this evaluation are 1080p 8-bit 4:2:0 content either at 25 or 30 Hz. IPP encoding was used to generate the results. The objective results demonstrate significant performance improvements achieved by MV-HEVC against both MVC and simulcast HEVC, demonstrated by the Bjøntegaard Delta (BD) bitrates table reproduced here:

Test Sequence	BD-rate reduction of MV-HEVC [%] relative to	
	MVC	Simulcast HEVC
S03: Undo_Dancer	-45.7	-38.7
S04: GT_Fly	-52.9	-41.0
S13: Band06	-43.3	-31.7
S14: BMX	-60.6	-25.6
Average	-50.6	-34.2

Hence at least 30% performance gains were observed against simulcast HEVC. The corresponding subjective tests using the “Expert Viewing Protocol” (EVP) verified the objective gains via MOS for all the sequences above. For example, the results for the sequences "Undo Dancer" and "BMX" are copied in the following, other results in [12] follow these results similarly.

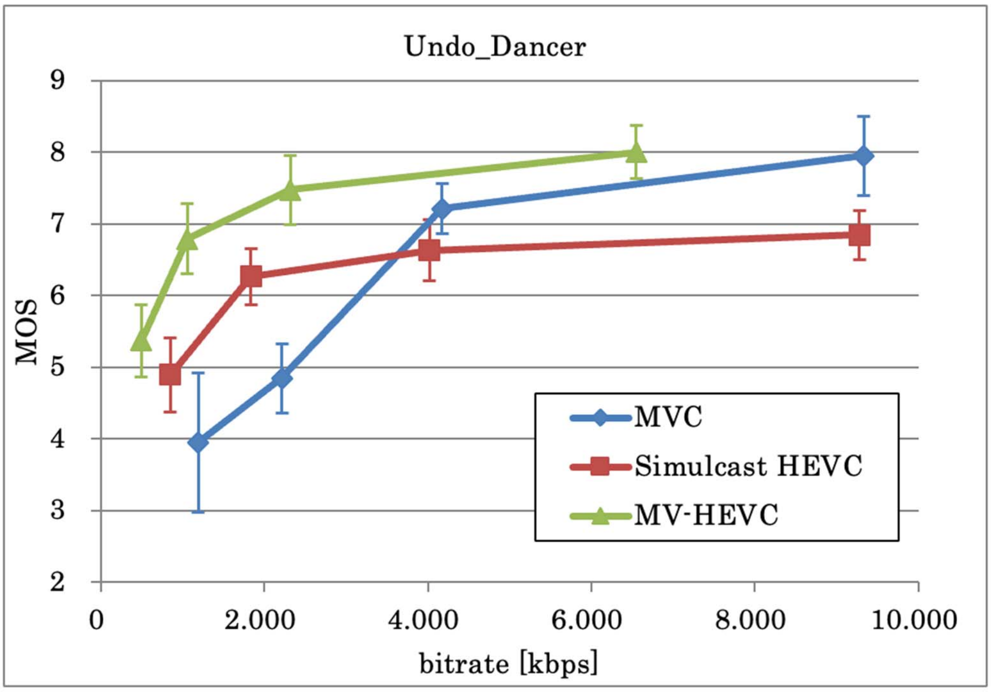

Figure 6.3.3.2-1: EVP results for sequence "Undo Dancer" [12]

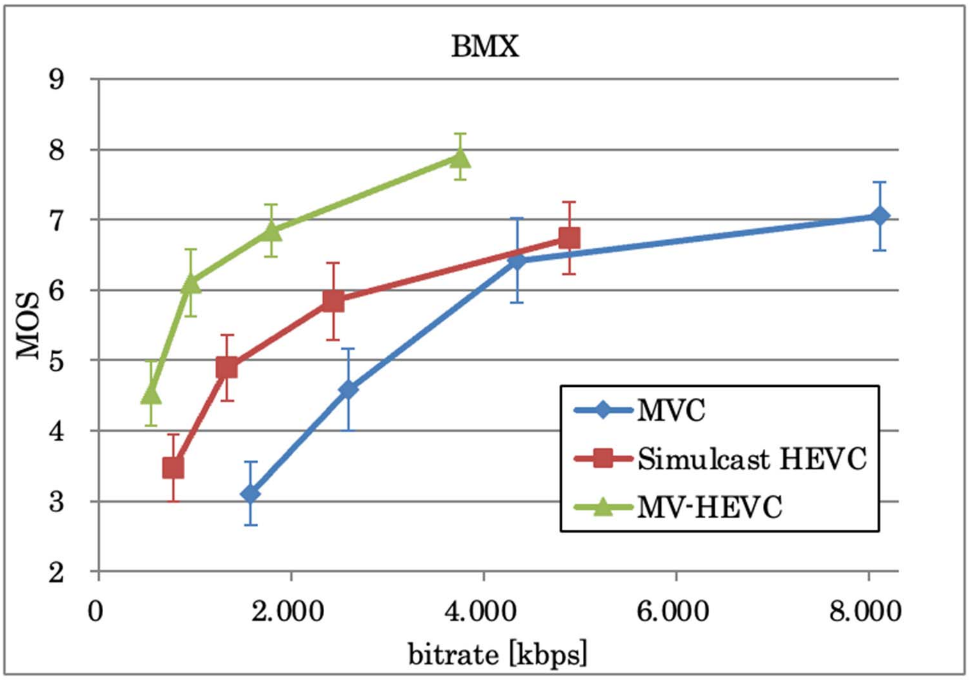

Figure 6.3.3.2-1: EVP results for sequence "BMX" [12]

Although no formal evaluation exists for the Multiview Main 10 profile of MV-HEVC, considering the large gains achieved as noted above, it is expected that it’s performance should be similar to what is demonstrated for 8-bit content, as reported for assessment of 2D video in [32].

6.4 Solution #2.1: HEVC 4:2:0 coding

6.4.1 Introduction

This solution of using 8 and 10 bit HEVC [3] 4:2:0 coding, depending on the source material, is the baseline solution for scenario#2. Such solution is already widely deployed, typically using the HEIF format [22]. 10-bits are also used to support High Dynamic Range (HDR) and Wide Colour Gamut (WCG) formats.

6.4.2 High-level Description

HEVC coding for still images using the HEIF file format [22] is widely deployed and supported by the current mobile ecosystem [24], [25]. This file format is designed to enable the interchange of images and image sequences, using the ISO base media file format as its basis [23]. When the requirements of the HEVC-specific brands are applied, the file format can be referred to as the HEVC Image File Format.

6.4.3 Evaluation

This is the baseline solution, i.e. baseline for evaluation of other solutions.

Assessment of all other solutions should be based on using this baseline technology, by taking 4:4:4 still image content, both in standard dynamic range (SDR) and high dynamic range (HDR) and first downconverting them to 4:2:0, while retaining the original bitdepth (i.e. 8 or 10 bits) using agreed downsampling methods (see JVT-I018[20]). Then such content can be coded with the appropriate HEVC 4:2:0 profile using the HEVC reference encoder (HM). Given the prevalence of the full range in still image content, full range signals should be generated across all conversion steps. For 8 bit material, it might also be desirable to explore the use of JPEG encoding for the same content. Chroma location of type 1, which is also prevalent in still image compression should be used for 8 bit material. For 10 bit content, including HDR, chroma location type 1 should be used.

After decoding, the content will be upconverted to 4:4:4 using a well agreed methodology (see JVT-I019 [21]). Afterwards, metrics will be computed for the upconverted content such as PSNR for the three colour components, Y, Cb, Cr in the 4:4:4 domain using the original content. The bits needed for coding these representations would also be considered.

6.5 Solution #2.2: HEVC 4:2:2 coding

6.5.1 Introduction

This solution uses 4:2:2 capable profiles that are already defined in HEVC for the coding of still images. Such images are then encapsulated in a file format based on the HEIF specification.

6.5.2 High-level Description

The HEVC video coding standard specifies profiles capable of coding images in a 4:2:2 coding format. This includes the HEVC Main 422 10, Main 422 12, Main 422 10 Intra, and Main 422 12 Intra profiles. These profiles are however not typically supported by mobile devices. Interest is primarily in applications limited to up to 10 bits of precision and therefore only profiles that satisfy this constrain should be evaluated.

6.5.3 Evaluation

6.5.3.1 Assessment/discussion of hardware impact

As noted above, there is a limited existing hardware support available for this solution and hence the hardware impact is potentially large.

6.5.3.2 Codec performance evaluation

Assessment should be based on taking the same 4:4:4 still image content as in baseline solution 2.1. The material can be then downconverted to 4:2:2, while retaining the original bitdepth (i.e. 8 or 10 bits) using an agreed horizontal downsampling method (see JVT-I018[20]). Then such content can be coded with the appropriate HEVC 4:2:2 profile using the HEVC reference encoder (HM). As in the previous solution, and given the prevalence of the full range in still image content, full range signals should be generated across all conversion steps.

After decoding, the content will be upconverted to 4:4:4 using a well agreed methodology (see JVT-I019 [21]). Afterwards, metrics will be computed for the upconverted content such as PSNR for the three colour components, Y, Cb, Cr in the 4:4:4 domain using the original content. The bits needed for coding these representations would also be considered. Although distortion is introduced in this process because of downconversion from 4:4:4 to 4:2:2 and the subsequent upconversion back to 4:4:4, this is likely to be smaller than what is observed and documented for 4:4:4 to 4:2:0 conversion [42].

Currently, there are no documented performance enhancements achieved by this solution.

6.6 Solution #2.3: Native 4:4:4 coding - HEVC Main 4:4:4 profiles

6.6.1 Introduction

This solution explores the use of the various 4:4:4 capable profiles that are already defined in HEVC for the coding of still images. Such images are then encapsulated in a file format based on the HEIF specification.

6.6.2 High-level Description

6.6.2.1 Overview

The HEVC video coding standard specifies the clear definition of several profiles capable of coding images in a 4:4:4 coding format. This includes the Main 4:4:4, Main 4:4:4 Still Picture, Main 4:4:4 10, Main 4:4:4 12, Main 4:4:4 10 Intra, and Main 4:4:4 12 Intra profiles, among others. Some of these profiles are already supported in some mobile devices but may not be widely available everywhere. These profiles are mostly targetting for the best coding performance, using the tools available in HEVC for the corresponding format(s) that they can support.

Interest is primarily in applications limited to up to 10 bits of precision and therefore only profiles that satisfy this constrain should be evaluated.

6.6.3 Evaluation

6.6.3.1 Assessment/discussion of hardware impact

As noted above, there is a limited existing hardware support available for this solution and hence the hardware impact is potentially large.

6.6.3.2 Codec performance evaluation

Assessment should be based on taking the same 4:4:4 still image content as in baseline solution 2.1 and coding them with the appropriate HEVC 4:4:4 profile using the HEVC reference encoder (HM). No bitdepth or format conversion needs to be performed. For such content then metrics such as PSNR for the three colour components, Y, Cb, Cr in the 4:4:4 domain should be computed using the original, 4:4:4, content. Unlike baseline solution 2.1, no upconversion or downconversion needs to be performed. The bits needed for coding these representations would also be considered.

6.7 Solution #2.4: Derived 4:4:4 coding- Layered use of HEVC 4:2:0 profiles

6.7.1 Introduction

This solution explores the use of derived 4:4:4 coding, where a base layer image, that is coded in 4:2:0 mode, is augmented using auxiliary images, to derive the 4:4:4 chroma format representation. Such capabilities can be achieved, for example, in HEIF, and are currently used for other applications. This permits decoders that are not capable of native 4:4:4 HEVC coding to still be able to encode and decode 4:4:4 content through simple software support.

6.7.2 High-level Description

6.7.2.1 Overview

The HEIF specification permits a concept called derived images, which permits the signaling of instructions to the decoder on how to combine a set of images together to generate an alternative representation of that same image. The concept could easily be used also for the support of 4:4:4 images. In this scenario a derived image can be based on a base, 4:2:0, image and one or two more images that contain the chroma information in the 4:4:4 format. Additional instructions would exist that provide information to the decoder on how to extract this chroma information and how to apply them onto the base image to achieve the desired, 4:4:4, output.

As one approach, a single enhancement image may be used that contains both Cb and Cr components stacked together, e.g. in a side by side or over-under representation. Such data are placed in the “luma” plane of that image and dummy data, e.g. a value of 128 for 8 bit data, is added in the “chroma” planes of that same image. This new image is then coded independently from the base layer image. During decoding, a decoder may select to discard the 4:2:0 version of the chroma information and instead replace that information from the information provided in this enhancement image.

As a different implementation, the enhancement image may contain predicted residuals for the Cb and Cr components given upscaled versions of the chroma values in the 4:2:0 representation. However, we do not advocate for this approach, even if it may appear more efficient in terms of coding efficiency, since that creates reconstruction dependencies of the 4:4:4 chroma values with the coding and upscaling of the 4:2:0 chroma values. There is no guarantee, for example, that all implementations could use a particular chroma upscaler while any further transcoding of the 4:2:0 representation could have an adverse effect in the reconstruction of the 4:4:4 representation.

The two chroma planes could also be coded in separate enhancement images if that is desired. A decoder can select to decode one of both enhancement images and augment either one or both components.

HEIF is also capable in achieving region of interest enhancement if that is desired.

As in the previous cases, interest is primarily in applications limited to up to 10 bits of precision and therefore only profiles that satisfy this constrain should be evaluated.

6.7.3 Evaluation

6.7.3.1 Assessment/discussion of hardware impact

Unlike solution 2.2, this approach allows existing HW, that support HEVC 4:2:0 profiles, to be used for the delivery of 4:4:4 content. The only requirement would be to perform the reconstruction in SW, after decoding of the multiple layers.

In this scenario additional images over scenario 1 should be coded that only contain the chroma planes. These chroma planes could either be coded as two separate images or stacked together in either a side by side or over under representation. The bit-depth of the original content will be retained also for the chroma planes. Metrics will be computed using the decoded chroma data from these additional coded images, while the bits of scenario one will be augmented by the bits also needed for coding these additional representations.

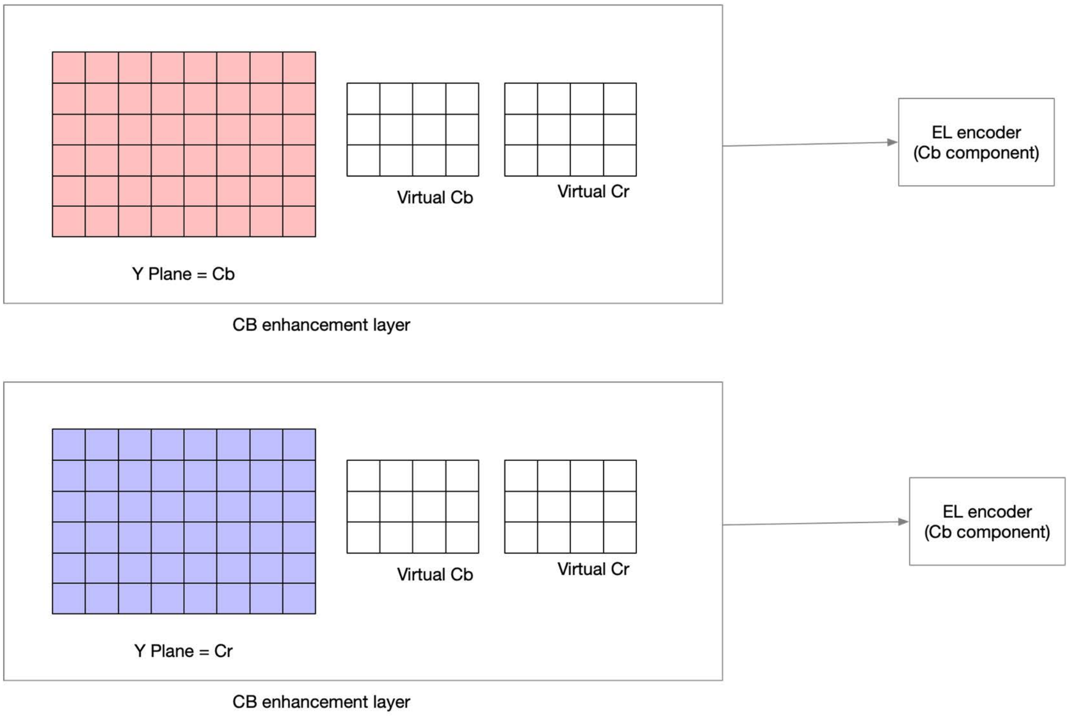

Figure 6.7.3.1-1: Enhancement layers for the creation of a 4:4:4 derived representation

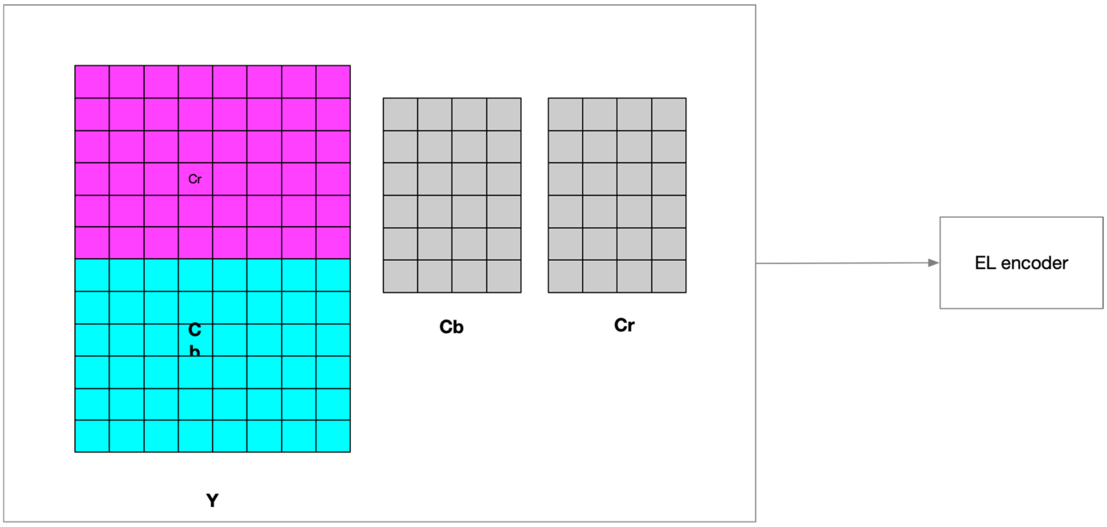

Figure 6.7.3.1-2: Single enhancement layer using stacking for the creation of a 4:4:4 derived representation

6.7.3.2 Codec performance evaluation

In this scenario, in addition to the bistreams used for solution 2.1, the chroma planes would also have to be coded in full resolution, either by packing the two chroma planes together and coding them as a single image or by coding each chroma plane independently. After decoding, the PSNR for these two chroma planes would have to be computed compared to the original 4:4:4 chroma planes and that value should be used in place of the Cb/Cr PSNR values of solution 2.1. In addition, the extra bit overhead of coding the full resolution chroma planes needs to be included in the evaluation and when comparing with either solution 2.1 or solution 2.2.

6.8 Solution #3.1: Scalable HEVC coding

6.8.1 Introduction

Several video coding standards and technologies, such as AVC and HEVC, include scalable extensions, which enable these technologies to provide "flexible" experiences to end users, such as allowing spatial, SNR, or bitdepth scalability. It is claimed that such functionalities can reduce the bitrate/storage needed by certain applications that may require multiple instances of the same video to be available to the end-user, e.g., in a multi-conferencing scenario simultaneously supporting multiple heterogeneous devices and networks. It has been argued, however, that such solutions have little benefits, if any, while adding a lot in terms of complexity, compared to existing solutions for adaptive streaming, such as Dynamic Adaptive Streaming over HTTP (DASH) and HTTP Live Streaming (HLS).

Such statements seem to be mostly based on the assumption that scalable coding would completely replace the existing adaptive streaming solutions. Instead, a more plausible alternative could be the use of scalability as a way of augmenting adaptive streaming systems by still using a solution with multiple independent bitstreams encoded at different bitrates and resolutions [28], while augmenting some or all of these bitstreams with 1 (preferably) or more enhancement layers.

Looking further in the future, in recent years new network protocols [29] are being discussed for the delivery of media and other services, such as QUIC and Multipath QUIC (MP-QUIC). Scalability can even better fit within such new protocols since it could better enable prioritization and delivery of different packets (i.e., the protocol could handle differently the base layer versus the enhancement layer or layers) with less waste in bandwidth.

Other benefits of scalability include power adaptation, simultaneous support of multiple screens with different capabilities (e.g., resolution, SDR vs HDR etc.). Scalability can be especially useful for multi-conferencing applications. On the other hand, the implementation cost of supporting scalable systems based on the Scalable HEVC profiles can be considered as minimal since that mostly involves SW level modifications in end devices because of its design.

6.8.2 High-level Description

6.8.2.1 Overview using scalable HEVC for adaptive streaming

An example is shown in Table 6.8.2-1, where a scalable layer is introduced when a change of resolution occurs from one stream to the next.

Table 6.8.2-1: Example Bitrate ladder for a Scalable Adaptive Streaming solution

Streams	16:9 aspect ratio	HEVC (base layer)	Enhancement layer	Frame rate
R1	640 x 360	145	77.5 at 768 x 432	≤ 30 fps
R2	768 x 432	300	150 at 960 x 540	≤ 30 fps
R3	960 x 540	600		≤ 30 fps
R4	960 x 540	900		≤ 30 fps
R5	960 x 540	1600	400 at 1280 x 720	Same as source
R6	1280 x 720	2400		Same as source
R7	1280 x 720	3400	550 at 1920 x 1080	Same as source
R8	1920 x 1080	4500		Same as source
R9	1920 x 1080	5800	1150 at 2560 x 1440	Same as source
R10	2560 x 1440	8100	1750 at 3840 x 2160	Same as source
R11	3840 x 2160	11600		Same as source
R12	3840 x 2160	16800		Same as source

An advantage that this could introduce is that this could considerably reduce the storage required to support the additional intermediate bitrates that the enhancement layers could result in. In the above example, if additional streams would be introduced, that would increase bitrate requirements by 23.4Mbps, an increase of ~30% in storage compared to the current number of streams, while scalability would only require ~4Mbps, an increase in storage of only ~7%. Alternatively, a service may decide to convert some of the existing bitstreams to enhancement layers and save on storage, while retaining the content instead of phasing them out from their service a bit too early. Even if storage is becoming cheaper, deploying new storage systems can be quite expensive while such storage is preferred to be used to store new content.

In addition to storage savings, encryption/decryption complexity may also be reduced. It would be sufficient to only encrypt the base layer signals and not the enhancement layers, which would reduce the overall complexity of decrypting the video on the client.

6.8.2.2 Transport of Scalable HEVC

6.8.2.2.1 Carriage in ISO BMFF

The carriage of scalable HEVC is specified in detail in [13] as one of the "Layered HEVC ((L-HEVC) extensions", including SHVC, MV-HEVC, and 3D-HEVC. Clause 9 of [13] specifies the L-HEVC elementary stream and sample definitions.

6.8.2.2.3 Support in CMAF

Carriage of scalable HEVC is specified by ISO/IEC 23000-19 (CMAF) [33] Annex H.

Currently however, the CMAF specification restricts the spatial resolution of the enhancement layer be to be either 1.5, 2, or 3 times that of the base layer both horizontally and vertically in Annex H.4.2.2 (General constraints). This raises some issues:

1. It omits the spatial resolution ratio of 1.0 for the enhancement layer that can be used for purposes beyond spatial resolution scalability, e.g., to provide bit-depth scalability.

2. These 3 ratios omit several other possible ratios, e.g., going beyond the ratio value of 3, or using some other typical ratios such as 1.25.

Based on this, MPEG has started studying this issue in [36] to ensure if such limitations can be addressed without creating any backward compatibility issues.

6.8.3 Evaluation

6.8.3.1 Assessment/discussion of hardware impact

The difference of HEVC and SHVC implementation is a high-level employing same low level coding tools, hence the hardware impact on implementations is manageable.

6.8.3.2 Performance evaluation

Based on the representative scenario evaluation, using the scalable streams save 23% of the otherwise required additional storage. Finally, some information about the performance of SHVC in different application scenarios is documented in [30] and [31].

6.9 Solution #4.1: MV-HEVC with auxiliary depth/alpha channels

6.9.1 Introduction

This solution explores the use of auxiliary alpha or depth channels, complementary to an HEVC bitstream to enable rendering optimization based on the auxiliary alpha/depth channels. This can be done in two ways:

- Solution 4.1-A: An MV-HEVC bitstream carrying a single video layer and alpha/depth video channels.

- Solution 4.1-B: Multiple MV-HEVC bitstreams, each carrying a texture layer and with alpha/depth channels.

6.9.2 High-level Description

6.9.2.1 Introduction

This solution explores the usage of MV-HEVC to carry the alpha and depth information as auxiliary channels. The carriage of such data is described in clause 6.9.2.2.

Additional information on possible SEI messaging transmitted to drive pose-correction is also documented for information in clause 6.9.2.3 but is not supported at this stage by OpenXR APIs and thus is excluded from this evaluation.

6.9.2.2 Carriage of alpha and depth auxiliary channels with MV-HEVC

The usage of auxiliary pictures in HEVC is part of the multi-layer extensions. The carriage of auxiliary data such as depth or alpha channels is defined by the ScalabilityId signalled through the scalability_mask_flag in the Video Parameter Set (VPS). This is possible by configuring the scalability mask index to ‘3’, the value reserved for enabling “Auxiliary” as scalability dimension, as highlighted in yellow in Table 6.9.2-1.

Table 6.9.2-1: Mapping of ScalabilityId to scalability dimensions, as specified in HEVC (see Table F.1)

Scalability mask index	Scalability dimension	ScalabilityId mapping
0	Texture or depth	DepthLayerFlag
1	Multiview	ViewOrderIdx
2	Spatial/Quality scability	DependencyId
3	Auxiliary	AuxId
4-15	Reserved	

The selection of alpha/depth auxiliary pictures is then set by the AuxId which can be configured as defined in the Table 6.9.2-2 below. Setting value '1' would signal the auxiliary picture is an Alpha plane while '2' would indicate a depth picture. Additional information about how to interpret and process those channels can be carried in SEI messages, through the Alpha channel and depth representation information SEI messages.

Table 6.9.2-2: Mapping of AuxId to the type of auxiliary pictures, as specified in HEVC (see Table F.2)

AuxId	Name of AuxId	Type of auxiliary pictures	SEI message describing interpretation of auxiliary pictures
1	AUX_ALPHA	Alpha plane	Alpha channel information
2	AUX_DEPTH	Depth picture	Depth representation information
3..127		Reserved	
128..159		Unspecified	
160..255		Reserved	

6.9.2.3 Additional information on SEI messages

Additionally, alternative SEI messages can be carried to indicate how the picture texture should be rendered and processed in the device, based on information carried through the alpha or depth channel. In the case of pose-correction parameters, the GUI can be isolated from the rest of the picture through specific depth ranges, or alpha values. The strength or sensibility to the pose-correction can be also indicated for each depth or alpha range value.

A specific SEI message is needed to carry out this information, which can be done for example through a private ITU-T 35 message, or by defining a new one in MPEG. A message carrying the desired information is provided in the Table 6.9.2-3. The provided SEI handles all possible scenarios.

Table 6.9.2-3: Possible payload for pose-correction parameters SEI

pose_correction_parameters(payloadSize) {	**Descriptor**
pcp_metric	u(1)
pcp_n_intervals	u(16)
for (i=0; i<pcp_n_intervals; i++){	
pcp_interval_upper_bound[i]	u(16)
pcp_interval_correction_sensitivity[i]	u(16)
}	u(1)
}	

With the following semantic:

- pcp_metric indicates what metric is used to extract different layers from video texture. 0 means alpha ranges are used, 1 means depth ranges are used.
- pcp_n_intervals indicates in how many intervals the layering is described for the selected metric.
- pcp_interval_upper_bound[i] indicates the upper bound value of the i-th interval.
- pcp_interval_correction_sensitivity[i] indicates the intensity of pose correction that should be applied on the i-th interval in the received frame. 0 means no pose correction should be applied, other values describes different degrees of pose-correction sensitivity.

The SEI messaging driving the pose-correction is depicted here for information but is currently not supported by OpenXR APIs and is then not included in the performance evaluation.

6.9.3 Evaluation

6.9.3.1 Assessment/discussion of hardware impact

This potential solution requires the device to decode the auxiliary channels and forward them to the XR runtime. This task is expected to be straightforward and light in terms of processing. As documented in [37], the carriage of auxiliary pictures does not impact the decoding of the primary layers for which the complexity remains unchanged. The decoding of the added auxiliary channels can be done by reusing single-layer HEVC decoding instances. A demuxer software update on top of an existing 4:2:0 decoder is expected to be sufficient to enable the feature with minimal complexity and power consumption overhead.

6.9.3.2 Codec performance evaluation

In this scenario, additional data is carried, through auxiliary pictures. As multiple solutions are possible, the performance should be evaluated as follows:

- For stereoscopic content:
  - A 2-views MV-HEVC bitstream with up to two auxiliary pictures for depth and alpha channels.
  - A 3D-HEVC+depth bitstream with one auxiliary channel for alpha channel.
- For regular 2D content:
  - A 2D MV-HEVC bitstream with up to two auxiliary pictures for depth and alpha channels.

As the coding of the auxiliary pictures themselves would not change between those configurations, it is needed to identify what would be the impact on distribution when adding those auxiliary pictures to a regular 2D or stereo HEVC encoded bitstream to enable pose correction optimization. The performance of alpha channel coding with HEVC is supported in the industry, at distribution friendly data rate [38], and is then not subject to particular concerns in terms of performance.

Regarding the depth channel coding, the 5:1 fixed ratio has been established as typically a good value to be used when it comes to static bitrate allocation between texture and depths [39] for older codecs. However, the solution #4.1 focuses on HEVC, for which the topic was addressed during MV-HEVC standard development. From [40], it is estimated that the ratio between texture and rate can be lowered to reach an overhead in the range of 8%, which can be further reduced when adjusting the depth resolution [41]. Thus, it is assessed that the coding and distribution of depth channel can be done at a reasonable and acceptable additional data rate.

7 Conclusions and proposed next steps

7.1 Conclusions for scenario #1.1, #1.2:

Comparing solution #1.1 (HEVC simulcast), solution #1.2 (HEVC frame packing) and solution#1.3 (Multiview HEVC coding), the following conclusions can be drawn for the stereoscopic content delivery scenarios:

- HEVC simulcast:
  - This is the most basic solution to address the stereoscopic HEVC delivery scenario.
  - It adds no new signalling.

  - Uses 2x HEVC encode/decode chains to provide stereoscopic video.
  - Does not exploit inter-view redundancy.
  - Application addresses the needed signalling aspects to realize immersive viewing.

- HEVC frame packing:
  - Reuses existing decoding hardware, albeit to achieve full resolution of the two views, a higher profile/level may be needed.
  - Addresses signalling via SEI messages.
  - For temporally interleaved frame packing, it could exploit inter-view redundancies for referenced frames, but not for non-referenced ones. However, the same frame packing scheme also results in a reduction of the available reference frames for each view given specified reference buffer constraints in the specification, which can impact coding performance.

- MV-HEVC:
  - Reuses the same low-level decoding tools as single layer HEVC decoding.
  - Better exploits inter-view redundancies by even allowing inter-view prediction from non-reference frames, without also additionally limiting the size of the reference buffer.
  - When used on a non-3D capable device, the content can be played back using only the base view for a 2D presentation.
  - Has better coding efficiency compared to either HEVC simulcast and HEVC frame packing.

Based on the assessment, MV-HEVC and HEVC frame packing are suitable solutions for addressing scenario#1.1 and #1.2 for stereoscopic content delivery, where MV-HEVC represents a more versatile tool. With HEVC simulcast and HEVC frame packing already included in SA4 specifications, and given the coding benefits it provides compared to alternative solutions, it is recommended to add support for stereoscopic MV-HEVC to the related specifications.

7.2 Conclusions for scenario #2:

Solution #2.3 (native 4:4:4 coding) and solution #2.4 (derived 4:4:4 coding) can achieve better visual quality than the baseline solution #2.1 (HEVC 4:2:0 coding). Solution #2.4 (derived 4:4:4 coding) however can achieve this improvement by reusing existing hardware support, without a need for a specialised hardware (as is needed for solution #2.3). However, a higher level may be needed for Solution #2.4.

At the time of drawing the conclusions, MPEG has agreed to include the technology for the solution #2.4 (derived 4:4:4 coding) into the HEIF amendment.

7.3 Conclusions for scenario #3:

Solution #3.1 (scalable HEVC coding) shows improvement potential for enhancing the adaptive streaming experience by allowing more switchable representations to be made available, while optimising storage overhead for this purpose. Scalable HEVC is also supported by MPEG specifications such as CMAF. The need to do normative work will be driven by industry interest in this direction.

7.4 Conclusions for scenario #4:

Solution #4.1 “MV-HEVC with depth/alpha channels” shows feasibility of the combination of MV-HEVC with depth and/or alpha channels, both in terms of additional data rate and complexity. The support and usage of depth and alpha channels in the industry is relevant for various application, potentially including closed caption insertion and pose-correction. Based on such usage cases and their potential benefits, it may be desirable to add support for carriage of these channels in 3GPP specifications.

Annex A: Change history

Change history							
Date	**Meeting**	**TDoc**	**CR**	**Rev**	**Cat**	**Subject/Comment**	**New version**
2023-08	SA4#125	S4-231295	-	-	-	Skeleton	0.0.1
2023-08	SA4#125	S4-231294	-	-	-	Implements agreements in: S4aV230053 (On HEVC Multiview coding)	0.0.2
2023-08	SA4#125	S4-231550	-	-	-	Implements agreements in: S4-231289 (On HEVC Multiview coding), S4-231536 (On HEVC 4:4:4 coding), S4-231291 (On HEVC scalable coding)	0.1.0
2023-11	SA4#126	S4-232006	-	-	-	Implements agreements in: S4-231818 (Updates on MV-HEVC), S4-231819 (Latency sensitive multiview), S4-231820 (Updates on scalable HEVC coding), S4-232040 (Pose correction optimisation), S4-232036 (Scope and background)	0.2.0
2023-12	SA#102	SP-231304				Version 1.0.0 created by MCC	1.0.0
2024-02	SA4#127	S4-240470	-	-	-	Implements agreements in: S4-240172 (Miscellaneous corrections), S4-240173 (Updates on HEVC Evaluations), S4-240174 (Updates on LD MV-HEVC), S4-240455 (Evaluation of Solution #4.1), S4-240471 (On framepacking), S4-240472 (Updates to HEVC 4:4:4 solutions), S4-240473 (Updated conclusions)	1.1.0
2024-03	SA#103	SP-240033				Version 1.0.0 created by MCC	2.0.0
2024-03						Version 18.0.0 created by MCC	18.0.0

History

Document history		
V18.0.0	May 2024	Publication