|  |  |
| --- | --- |
| 3GPP TR 26.955 V0.2.0 (2020-05) | |
| Technical Report | |
| 3rd Generation Partnership Project;  Technical Specification Group TSG SA;  5G Video Codec Characteristics  (Release 17) | |
|  | |
|  |  |
|  | |
| The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP. The present document has not been subject to any approval process by the 3GPPOrganizational Partners and shall not be implemented. This Specification is provided for future development work within 3GPPonly. The Organizational Partners accept no liability for any use of this Specification. Specifications and Reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices. | |

|  |
| --- |
|  |
| ***3GPP***  Postal address  3GPP support office address  650 Route des Lucioles - Sophia Antipolis  Valbonne - FRANCE  Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16  Internet  http://www.3gpp.org |
| ***Copyright Notification***  No part may be reproduced except as authorized by written permission. The copyright and the foregoing restriction extend to reproduction in all media.  © 2020, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC).  All rights reserved.  UMTS™ is a Trade Mark of ETSI registered for the benefit of its members  3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners LTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners  GSM® and the GSM logo are registered and owned by the GSM Association |

Contents

Foreword 5

Introduction 6

1 Scope 7

2 References 7

3 Definitions of terms, symbols and abbreviations 9

3.1 Terms 9

3.2 Symbols 9

3.3 Abbreviations 10

4 Video Coding Capabilities in 3GPP Services 10

4.1 Introduction 10

4.2 TV Video Profiles 10

4.3 5G Media Streaming 11

4.3.1 Downlink Streaming 11

4.3.2 Uplink Streaming 11

4.4 Multimedia Telephony Services over IMS 11

4.5 VR Video Profiles 11

4.6 Messaging Services 11

4.7 Screen Content Coding 12

5 Common Test Conditions and Parameters 12

5.1 Introduction 12

5.2 Video Test Sequences 12

5.3 Key Performance Indicators and Metrics 12

5.4 Reference Software Tools 12

6 Relevant Scenarios 12

6.1 Introduction 12

6.2 Scenario 1: Full HD Streaming 13

6.2.1 Motivation 13

6.2.2 Description of the Anticipated Application 13

6.2.3 Source Format Properties 14

6.2.4 Encoding and Decoding Constraints 15

6.2.5 Performance Metrics 15

6.2.6 Interoperability Considerations 15

6.2.7 Test Sequences 16

6.2.7.1 Standard Dynamic Range 16

6.2.7.2 High Dynamic Range 16

6.2.8 Detailed Test Conditions 16

6.2.8.1 Overview 16

6.2.8.2 Reference Software AVC 1 16

6.2.8.3 Reference Software HEVC 1: HM16.20 16

6.2.9 External Performance Data 17

6.2.10 Additional Information 17

6.3 Scenario 2: 4K-TV 17

6.3.1 Motivation 17

6.3.2 Description of the Anticipated Application 17

6.3.3 Source Format Properties 19

6.3.5 Performance Metrics 20

6.3.6 Interoperability Considerations 20

6.3.7 Test Sequences 20

6.3.8 Detailed Test Conditions 20

6.3.8.1 Overview 20

6.3.8.2 Reference Software HEVC 1: HM16.20s 21

6.3.8.2 HEVC open-source implementation libx265 21

6.3.9 External Performance Data 21

6.4 Scenario 4: Online Gaming and Screen Content Scenario 21

6.4.1 Motivation 21

6.4.2 Description of the Anticipated Application 22

6.4.3 Source Format Properties 24

6.4.4 Encoding and Decoding Constraints 24

6.4.5 Performance Metrics 24

6.4.6 Interoperability Considerations 25

6.4.7 Test Sequences 25

6.4.8 Detailed Test Conditions 25

6.4.9 External Performance Data 25

6.4.10 Additional Information 25

6.5 Scenario 4: Messaging and Social Sharing 25

6.5.1 Motivation 25

6.5.2 Description of the Anticipated Application 26

6.5.3 Source Format Properties 28

6.5.4 Encoding and Decoding Constraints 28

6.5.5 Performance Metrics 29

6.5.6 Interoperability Considerations 29

6.5.7 Test Sequences 29

6.5.8 Detailed Test Conditions 29

6.5.9 External Performance Data 29

6.5.10 Additional Information 29

7 Characterization of Existing Codecs 29

8 Gaps and Optimization Potential 30

8.1 Identified Gaps and Deficiencies with Existing Codecs 30

8.2 Potential Requirements for New Codecs 30

9 Initial Information on new Codecs 30

10 Conclusions and Proposed Next Steps 30

Annex A Scenario Template 30

A.1 Introduction 30

A.2 Template 30

Annex B Details on Performance Metrics 32

Annex <X> (informative): Change history 33

For definitive guidance on drafting 3GPP TSs and TRs, see [3GPP TS 21.801](http://www.3gpp.org/DynaReport/21801.htm) supplemented by the 3GPP web page <http://www.3gpp.org/specifications-groups/delegates-corner/writing-a-new-spec>.

Ensure all blue guidance text is removed before submitting the TS/TR to the TSG for approval.

# 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, modal verbs have the following meanings:

**shall** indicates a mandatory requirement to do something

**shall not** indicates an interdiction (prohibition) to do something

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

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

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

The constructions "can" and "cannot" are not 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

The constructions "is" and "is not" do not indicate requirements.

# Introduction

Predictions on mobile video consumption are ever increasing. Different studies point to dominance of video traffic in 5G networks reaching from 65% total traffic in the short-term all the way to 90% by the end of the decade. Video is expected to be integral for services such as enhanced mobile media (such as mobile streaming services), home broadband and TV (for example in the context of 5G fixed wireless access services), immersive and interactive media in the context of eXtended Realities (XR) and cloud gaming as well as new media services from new verticals. This indicates that the user experience and efficiency of 5G networks will be heavily impacted by the quality of video compression technologies that are used with 5G services. Efficient video compression and decompression technologies required dedicated hardware for power and resource efficient real-time execution but are at the same time complex and costly in terms of implementation on integrated platforms. Hence, typically state-of-the-art video compression technologies last for several years and are used as generic service enablers for different applications and services, including traditional streaming and conversational services, but also new media services. This document analyzes the currently defined 3GPP-defined video compression technologies for their suitability for existing and emerging services in the context of 5G and identifies gaps and optimization potentials that would warrant the introduction of new video compression technologies.

# 1 Scope

The present document relevant interoperability requirements, performance characteristics and implementation constraints of video codecs in 5G services, and to characterize existing 3GPP video codecs, in particular H.264/AVC and H.265/HEVC in order to have a benchmark for the addition of potential future video codecs. For this purpose, the document:

* Collects a summary of the video coding capabilities in 3GPP services.
* Collects a subset of relevant scenarios for video codecs in 5G-based services and applications, including video formats (resolution, frame rates, color space, etc.), encoding and decoding requirements, adaptive streaming requirements.
* Collects relevant and exemplary test conditions and material for such scenarios, including test sequences.
* Defines performance metrics for such scenarios with focus on objective performance metrics.
* Collects relevant interoperability functionalities and enabling elements for video codecs in different 5G services such as MTSI and Telepresence (i.e. RTP based conversational communications), or 5G media streaming (e.g. based on DASH/CMAF) supporting the identified scenarios.
* Collects relevant criteria and key performance indicators for the integration of video codecs in 5G processing platforms, taking into account factors such as encoding and decoding complexity in the context of the defined scenarios.
* Characterizes the existing codecs H.264/AVC and H.265/HEVC in the context of the above scenarios and document the findings in a consistent manner.
* Identifies gaps and deficiencies of existing codecs in such use cases and derive requirements for potential new codecs.
* Collects initial information on how new codecs under development in ISO/IEC SC29 WG11 (MPEG)/JVET (in particular including VVC and EVC) may meet the above criteria.

# 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] 3GPP TS 26.114: "IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction".

[3] 3GPP TS 26.116: "Television (TV) over 3GPP services; Video profiles".

[4] 3GPP TS 26.118: "3GPP Virtual reality profiles for streaming applications".

[5] 3GPP TR 26.925: "Typical traffic characteristics of media services on 3GPP networks".

[6] 3GPP TR 26.928: "Extended Reality over 5G".

[7] Recommendation ITU-T H.264 (06/2017): "Advanced video coding for generic audiovisual services" | ISO/IEC 14496-10:2014/Amd.3 2016: "Information technology - Coding of audio-visual objects - Part 10: Advanced Video Coding".

[8] Recommendation ITU-T H.265 (11/2019): "High efficiency video coding" | ISO/IEC 23008-2:2020: "High Efficiency Coding and Media Delivery in Heterogeneous Environments - Part 2: High Efficiency Video Coding".

[9] 2020 Mobile Internet Phenomena Report, accessible here: <https://www.sandvine.com/download-report-mobile-internet-phenomena-report-2020-sandvine>, February 2020.

[10] 2019 Ericsson Mobility Report, accessible here: https://www.ericsson.com/4acd7e/assets/local/mobility-report/documents/2019/emr-november-2019.pdf, November 2019.

[11] T. Fautier, "New Codecs for 5G", DASH-IF Workshop on "Media Streaming meets 5G", December 2019, accessible here: <https://dashif.org/docs/workshop-2019/04-thierry%20fautier%20-%20Harmonic%20Codec%20Comparison%205G%20Media%20Workshop_Final%20v3.pdf>

[12] Bitmovin Video Developer Report, accessible here: <https://go.bitmovin.com/video-developer-report-2019>, September 2019.

[13] 3GPP TS 26.511: "5G Media Streaming (5GMS); Profiles, codecs and formats".

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

[15] Recommendation ITU-R BT.2020-2 (10/2015): "Parameter values for ultra-high definition television systems for production and international programme exchange".

[16] Recommendation ITU‑R BT.2100-1 (06/2017): "Image parameter values for high dynamic range television for use in production and international programme exchange".

[17] “You can watch Netflix on any screen you want, but you’re probably watching it on a TV”, https://www.vox.com/2018/3/7/17094610/netflix-70-percent-tv-viewing-statistics

[18] YouTube Revenue and Usage Statistics (2020), https://www.businessofapps.com/data/youtube-statistics/

[19] Ultra HD Forum service tracker B2C, https://ultrahdforum.org/uhd-service-tracker/“

[20] HIS Market “4K-TV and UHD: the whole picture”, <https://cdn.ihs.com/www/pdf/4ktv-uhd-ebook.pdf>

[21] A. Mercat, M. Viitanen, and J. Vanne, “UVG dataset: 50/120fps 4K sequences for video codec analysis and development,” Accepted to ACM Multimedia Syst. Conf., Istanbul, Turkey, June 2020.

[21] R. Jullian, Y. Chen, F. Galpin, E. François, M. Kerdranvat (InterDigital), “Extra results to JVET-N605 “Comparative study of video coding solutions VVC, AV1 and EVC versus HEVC””, document JVET-O0898, 15th Meeting: Gothenburg, SE, 3–12 July 2019.

[22] R. Jullian, Y. Chen, F. Galpin, E. François, M. Kerdranvat (InterDigital), “Comparative study of video coding solutions VVC, AV1 and EVC versus HEVC”, document JVET-N0605, 14th Meeting: Geneva, CH, 19–27 March 2019.

[23] S. Iwamura, S. Nemoto, A. Ichigaya, “[AHG13] Compression performance analysis for 4K and 8K HLG test sequences”, document JVET-N0828, 14th Meeting: Geneva, CH, 19–27 March 2019.

[24] A. Segall, E. François, W. Husak, S. Iwamura, D. Rusanovskyy, “JVET common test conditions and evaluation procedures for HDR/WCG video,” document JVET-P2011, 16th JVET meeting: Geneva, CH, 1–11 Oct. 2019.

[25] Nick Zarzycki, "The Complete Guide to Social Media Video Specs in 2018", July 23, 2018, <https://blog.hootsuite.com/social-media-video-specs/>

[26] YouTube Help, "Recommended upload encoding settings", <https://support.google.com/youtube/answer/1722171?hl=en>

[27] Facebook Help Center, "What are the video format guidelines for live streaming on Facebook?", <https://www.facebook.com/help/1534561009906955>

[28] Deepak Kumar, "All about TikTok video size [ Full Guide]-2020", February 2020, <https://tiktoktip.com/tiktok-size/>

[29] Snapchat business center, "Single Image or Video Specifications", <https://businesshelp.snapchat.com/en-US/article/top-snap-specs>

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

[31] GSMA White Paper, "Chinese operators make major RCS commitment: Whitepaper", April 9, 2020, <https://www.gsma.com/futurenetworks/latest-news/china-operators-make-major-rcs-commitment-whitepaper/>

[32] 3GPP TS 26.140, "Multimedia Messaging Service (MMS); Media formats and codecs".

[33] GSMA RCC.71, "RCS Universal Profile Service Definition Document", version 2.4, October 2019.

[34] GSMA NG.114, "IMS Profile for Voice, Video and Messaging over 5GS", February 2020.

[35] 3GPP TS 26.223, "Telepresence using the IP Multimedia Subsystem (IMS); Media handling and interaction"

[36] "NVIDIA VIDEO CODEC SDK", <https://developer.nvidia.com/nvidia-video-codec-sdk>

[37] Microsoft Online Documentation, "Real-time media calls and meetings with Microsoft Teams", <https://docs.microsoft.com/en-us/microsoftteams/platform/bots/calls-and-meetings/real-time-media-concepts>

# 3 Definitions of terms, symbols and abbreviations

This clause and its three subclauses are mandatory. The contents shall be shown as "void" if the TS/TR does not define any terms, symbols, or abbreviations.

## 3.1 Terms

For the purposes of the present document, the terms given in 3GPP 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 3GPP TR 21.905 [1].

Definition format (Normal)

**<defined term>:** <definition>.

**example:** text used to clarify abstract rules by applying them literally.

## 3.2 Symbols

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

Symbol format (EW)

<symbol> <Explanation>

## 3.3 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP 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 3GPP TR 21.905 [1].

AVC Advanced Video Coding

EVC Essential Video Coding

HEVC High-Efficiency Video Coding

VVC Versatile Video Coding

# 4 Video Coding Capabilities in 3GPP Services

## 4.1 Introduction

This clause summarizes the video coding capabilities in relevant existing 5G services.

As of today, two codecs are prominently referenced and available, namely H.264/AVC [7] and H.265/HEVC [8].

Both codecs are defined as part of the TV Video Profiles in 3GPP TS 26.116 [3] and are also the foundation of the VR Video Profiles in 3GPP TS 26.118 [4]. The highest defined profile/level combinations are:

* H.264/AVC Progressive High Profile Level 5.1 [7] with the following additional restrictions and requirements:
  + the maximum VCL Bit Rate is constrained to be 120Mbps with cpbBrVclFactor and cpbBrNalFactor being fixed to be 1250 and 1500, respectively.
  + the bitstream does not contain more than 10 slices per picture
* H.265/HEVC Main-10 Profile Main Tier Profile Level 5.1 [8] without any restrictions

More details on the codec capabilities and the the necessary interoperability requirements for different services are collected in the remainder of this clause.

## 4.2 TV Video Profiles

The TV Video Profiles in TS 26.116 [3] address coded representations of TV distribution signals up to UHD-1 phase 2. Table 4.2-1 provides an overview of the TV relevant formats considered in the context of 3GPP TV Video Profiles.

In the context of TV Video Profiles, the following aspect are defined:

* **Bitstream:** A media bitstream that conforms to a video encoding format and certain Operation Point.
* **Operation Point:** A collection of discrete combinations of different content formats including spatial and temporal resolutions, colour mapping, transfer functions, etc. and the encoding format.
* **Receiver:** A receiver that can decode and render any bitstream that is conforming to a certain Operation Point.

Table 4.2-1: TV over 3GPP services Video Profile Operation Points (TS 26.116 [3])

| Operation Point name | Resolution format | Picture aspect ratio | Scan | Max. frame rate | Chroma format | Chroma sub-sampling | Bit depth | Colour space format | Transfer  Characteristics |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| H.264/AVC 720p HD | 1280 × 720 | 16:9 | Progressive | 30 | Y'CbCr | 4:2:0 | 8 | BT.709 | BT.709 |
| H.265/HEVC 720p HD | 1280 × 720 | 16:9 | Progressive | 30 | Y'CbCr | 4:2:0 | 8 | BT.709 | BT.709 |
| H.264/AVC Full HD | 1920 × 1080 | 16:9 | Progressive | 60 | Y'CbCr | 4:2:0 | 8 | BT.709 | BT.709 |
| H.265/HEVC Full HD | 1920 × 1080 | 16:9 | Progressive | 60 | Y'CbCr | 4:2:0 | 8; 10 | BT.709; BT.2020 | BT.709; BT.2020 |
| H.265/HEVC UHD | 3840 × 2160 | 16:9 | Progressive | 60 | Y'CbCr | 4:2:0 | 10 | BT.2020 | BT.2020 |
| H.265/HEVC Full HD HDR | 1920 x 1080 | 16:9 | Progressive | 60 | Y'CbCr | 4:2:0 | 10 | BT.2020 | BT.2100 PQ |
| H.265/HEVC UHD HDR | 3840 x 2160 | 16:9 | Progressive | 60 | Y'CbCr | 4:2:0 | 10 | BT.2020 | BT.2100 PQ |
| H.265/HEVC Full HD HDR HLG | 1920 x 1080 | 16:9 | Progressive | 60 | Y'CbCr | 4:2:0 | 10 | BT.2020 | BT.2100 HLG |
| H.265/HEVC UHD HDR HLG | 3840 x 2160 | 16:9 | Progressive | 60 | Y'CbCr | 4:2:0 | 10 | BT.2020 | BT.2100 HLG |

For TV Video profiles, interoperability with ISO BMFF based systems and the DASH Streaming is of most relevance. Hence, for a codec to be used in the context of TV Video Profiles, the following is defined in terms of interoperability:

1. The receiver requirements on elementary stream level
2. The encapsulation of an elementary stream into an ISO Base Media File Format track
3. The provisioning of the media as part of DASH Adaptation Set to support seamless switching
4. All MPD-level signalling for the codec to support capability discovery

For details, refer to TS 26.116, clause 4 and clause 5.

## 4.3 5G Media Streaming

### 4.3.1 Downlink Streaming

Ffs

### 4.3.2 Uplink Streaming

The detailed requirements for uplink streaming for 5G Media Streaming and the codec requirements are defined in 3GPP TS 26.511 [13]. For video encoding the support for **HEVC-FullHD-Enc** as defined in clause 4.2.2.2 of TS26.511 [13] is required together with the the sender requirements for **HEVC-FullHD-Enc** Operation Point, i.e. real-time encoding. Encapsulation is based on the Common Media Application Format (CMAF) as defined in ISO/IEC 23000-19 [30].

## 4.4 Multimedia Telephony Services over IMS

ffs

## 4.5 VR Video Profiles

Ffs

## 4.6 Messaging Services

3GPP TS 26.140 [32] specifies the media types, formats and codecs for the MMS within the 3GPP system. The document extends to codecs for speech, audio, video, still images, bitmap graphics, and other media in general, as well as scene description, multimedia integration and synchronization schemes.

Specifically, for video, the following capabilities are defined

H.264 (AVC) Constrained Baseline Profile (CBP) Level 1.3 is required to be supported.

- H.264 (AVC) High Profile Level 3.1 with frame\_mbs\_only\_flag=1 is recommended to be supported by MMS clients supporting HDTV video content at a resolution of 1280x720 (720p) with progressive scan at 30 frames per second.

- H.265 (HEVC) Main Profile, Main Tier, Level 3.1 decoder is recommended to be supported

The specification has not been updated since 2014.

## 4.7 Screen Content Coding

3GPP TS 26.223 [35] specifies a client for the IMS-based telepresence service supporting conversational speech, video and text transported over RTP. Telepresence is defined as a conference with interactive audio-visual communications experience between remote locations, where the users enjoy a strong sense of realism and presence between all participants (i.e. as if they are in same location) by optimizing a variety of attributes such as audio and video quality, eye contact, body language, spatial audio, coordinated environments and natural image size.

As specified in 3GPP TS 26.223 [35] clause 5.2, telepresence UEs are required to support:

- H.264 (AVC) [16] Constrained High Profile (CHP), Level 3.1

- H.264 (AVC) [16] Constrained Baseline Profile (CBP), Level 1.2 (for interworking with MTSI clients)

- H.265 (HEVC) [17] Main Profile, Main Tier, Level 4.1

Telepresence UEs are also recommended to support dedicated HEVC extensions defined specifically for such types of content:

- H.265 (HEVC) Screen-Extended Main, Main Tier, Level 4.1

- H.265 (HEVC) Screen-Extended Main 4:4:4, Main Tier, Level 4.1

# 5 Common Test Conditions and Parameters

## 5.1 Introduction

This clause defines common test conditions and parameters.

## 5.2 Video Test Sequences

## 5.3 Key Performance Indicators and Metrics

## 5.4 Reference Software Tools

# 6 Relevant Scenarios

## 6.1 Introduction

This clause collects relevant scenarios based on the template defined in Annex A.

## 6.2 Scenario 1: Full HD Streaming

### 6.2.1 Motivation

The 2020 Mobile Internet Phenomena Report from Sandvine [9] shows that mobile video downstream traffic accounts for more than 65% of the global application category traffic share.

According to Ericsson mobility report [10], video traffic in mobile networks is forecast to grow by around 30 percent annually through 2025 to account for three-quarters of mobile data traffic, from slightly more than 60 percent in 2019. The video traffic growth is driven by the increase of embedded video in many online applications, growth of video-on-demand (VoD) streaming services in terms of both subscribers and viewing time per subscriber, and the evolution toward higher screen resolutions on smart devices. All of these factors have been influenced by the increasing penetration of video-capable smart devices.

Furthermore, while UHD and 4K are trendy formats, the main application for mobile streaming is Full HD with 1080p at 50 or 60 frames per second is expected to be the format of choice for mobile streaming at scale. The distribution version may be downsampled to support adaptive bitrate streaming, possibly with High Dynamic Range (HDR) support. For detailed discussion please refer to the presentation at the DASH-IF Workshop Dec 2019 [11].

In terms of distribution, while in the past, streaming video was delivered primarily via RTMP or RTP, fewer and fewer devices support these aging protocols each year. Instead, the latest web standards support built-in video playback and HTML5 is now by far the preferred method for video playback. And adaptive bitrate protocols dominate the distribution. According to the developer report [12], adaptive bitrate streaming through HLS/DASH, using the CMAF/DASH based segment formats, provide vast majority for streaming video. The distribution is used for On-Demand and Live Streaming.

### 6.2.2 Description of the Anticipated Application

In the context of 3GPP services, 5G Media Streaming [13] as well as the TV Video Profiles [3] are specifications addressing this streaming scenario. Both, 5G Media Streaming [13] and TV Video Profiles [3] builds on CMAF-based Segment formats and DASH distribution. From TS 26.116, the following operation points may be considered in scope of the Full HD Streaming Scenario (pending availability of appropriate test content):

- H.265/HEVC Full HD HDR, see TS26.116 [3] clause 4.5.3.

- H.264/AVC Full HD, see TS26.116 [3] clause 4.4.3.

- H.265/HEVC Full HD, see TS26.116 [3] clause 4.5.5.

- H.265/HEVC Full HD HDR HLG, see TS26.116 [3] clause 4.5.7.

These operation points are further informed by relevant operational experience with commercially available encoders and decoders.

The considered scenario is the distribution of content through DASH/CMAF based streaming. Important aspects that are expected to be considered when evaluating a codec in the context of this:

- Quality and Coding Efficiency:

- High and uninterrupted visual quality, taking into account the service constraints.

- Any savings can provide significant benefits due to the expected large volume of the traffic either in quality or network utilization.

- Adaptive Bitrate streaming:

- Multiple bit rates are provided, typically with a ladder of 30–50% to permit bandwidth adaptation. The use of constant bit rate (CBR) encoding maximises reuse of a common ladder of encoded representations across multiple distribution networks. The use of capped variable bit rate (VBR) encoding allows the bit rate to be varied according to the difficulty of the source material while maintaining the ability to distribute the encoded representations through distribution networks with fixed capacity. This also maximises reuse of a common ladder across multiple distribution networks.

- CMAF Fragments of size typically in the range of 1–6s to permit seamless switching for bit rate adaptation.

- Regular Random Access, typically every 1–2 seconds according to TS 26.116 [3]. To achieve clean switching in both sound and picture when moving between different encoded representations in the ladder, 3.84 seconds enables video segment boundaries to be aligned with an integer number of audio Access Units, if a 50fps video signal and 48kHz audio signal is used.

- Encoding in this scenario is typically done as

- Live and On-Demand distribution and encoding

- Server and Cloud-based Encoding

- No specific encoding latency constraints

### 6.2.3 Source Format Properties

Table 6.2-1 provides an overview of the different source signal properties following the information from TS26.116 [3]. This information is used to select proper test sequences.

Table 6.2-1 Source Format Properties for different operation point

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source Format Properties | H.264/AVC Full HD | H.265/HEVC Full HD | H.265/HEVC Full HD HDR | H.265/HEVC Full HD HLG |
| Spatial resolutions | 1920 × 1080 (Permitted Distribution formats: 1920 × 1080, 1600 × 900, 1280 × 720, 960 × 540, 854 × 480, 640 × 360,426 × 240) | | | |
| Chroma Format | Y'CbCr | | | |
| Chroma Subsampling | 4:2:0 | | | |
| Picture Aspect ratios | 16:9 | | | |
| Frame rates | 24; 25; 30; 50; 60; 24/1.001; 30/1.001; 60/1.001 Hz | | | |
| Bit Depth | 8 | 8, 10 | 10 | 10 |
| Colour space formats | BT.709 [14] | BT.709 [14]; BT.2020 [15] | BT.2020 [15] | BT.2020 [15] |
| Transfer Characteristics | BT.709 [14] | BT.709 [14]; BT.2020 [15] | BT.2100 [16] PQ | BT.2100 [16] HLG |

### 6.2.4 Encoding and Decoding Constraints

Table 6.2-2 provides an overview of encoding and decoding constraints for H.264/AVC Full HD and H.265/HEVC Full HD Profiles. This will support the definition of detailed test conditions.

Table 6.2-2 Encoding and Decoding Configurations

|  |  |  |
| --- | --- | --- |
| Encoding and Decoding Constraints | H.264/AVC Full HD | H.265/HEVC Full HD |
| Relevant Codec and Codec Profile/Levels according to TS26.116 and TS26.511. | H.264/AVC Progressive High Profile Level 4.2 [7] | HEVC/H.265 Main-10 Profile  Main Tier Level 4.1 [8] |
| Random access frequency | 1 second, 3.84 seconds [other numbers tbd] | 1 second, 3.84 seconds [other numbers tbd] |
| Error resiliency requirements | None | None |
| Bit rates and quality configuration | QP = [20, 23, 26, 29]  others | QP = [23[[1]](#footnote-2), 25, 28, 31, 34]  others |
| Bit rate parameters (CBR, VBR, CAE, HRD parameters) | Fixed QP  CBR 8–12 Mbit/s  VBR capped at 12 Mbit/s  others | Fixed QP  CBR 5–8 Mbit/s  VBR capped at 12 Mbit/s  others |
| ABR encoding requirements (switching frequency, etc.) | 1 second [other numbers tbd]  ABR through multiple QPs | 1 second [other numbers tbd]  ABR through multiple QPs |
| Latency requirements and specific encoding settings | No latency requirements beyond RAP so picture reordering allowed | No latency requirements beyond RAP so picture reordering allowed |
| Encoding complexity context | real-time encoding, cloud-based encoding, offline encoding, etc.  detailed parameters tbd | real-time encoding, cloud-based encoding, offline encoding, etc.  detailed parameters tbd |
| Required decoding capabilities | H.264/AVC Progressive High Profile Level 4.2 [7] | HEVC/H.265 Main-10 Profile  Main Tier Level 4.1 [8] |

### 6.2.5 Performance Metrics

The following performance metrics are considered for this scenario:

tbd

### 6.2.6 Interoperability Considerations

In order to use a codec in the context of 5G Media Streaming services in TS 26.511 and for TV Video profiles in TS 26.116, the following list provides a set of potentially relevant interoperability aspects for Full HD Streaming:

1. The receiver requirements on elementary stream level, in particular the profile/level and additional considerations.
2. The encapsulation of an elementary stream into an ISO Base Media File Format track
3. The definition of a CMAF media profile.
4. The static mapping of parameters to a DASH MPD, in particular to the MPD parameters, such as @mimeType, @codecs, etc
5. The dynamic mapping of parameters to a DASH MPD from a CMAF Master Header, in particular to the MPD parameters, such as @width, @height, etc.
6. All MPD-level signalling for the codec to support capability discovery
7. Encryption requirements and recommendations.
8. Capability discovery options, for example mapping to HTML-5, MSE and media capability APIs.
9. Source Buffer Initialization Requirements.
10. Playback Requirements, for example by referencing CTA WAVE Specifications
11. Relation to other specifications, such as in DVB, ATSC, MPEG, ETSI, etc.

For additional details, please refer to TS 26.116 and TS 26.511.

### 6.2.7 Test Sequences

#### 6.2.7.1 Standard Dynamic Range

Example sequences are here https://media.xiph.org/video/derf/

#### 6.2.7.2 High Dynamic Range

tbd

### 6.2.8 Detailed Test Conditions

#### 6.2.8.1 Overview

tbd

#### 6.2.8.2 Reference Software AVC 1

tbd

#### 6.2.8.3 Reference Software HEVC 1: HM16.20

As reference software for HEVC, the following was used

- <https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.20/>

Example setting: <https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.20/cfg/encoder_randomaccess_main10.cfg> with following proposed changes

- IntraPeriod: Intra Period such that 1 second is achieved

- DecodingRefreshType: 1 (CRA) 🡺 2 (IDR)

- GOPSize: adjusted to Intra

- QP: [25, 28, 31, 34]

### 6.2.9 External Performance Data

tbd

### 6.2.10 Additional Information

Tbd

## 6.3 Scenario 2: 4K-TV

### 6.3.1 Motivation

Streaming towards mobile devices is undoubtly the first natural use-case expected for 5G-media streaming. However, consumption of video services on fixed receivers (e.g. TV sets) remains a preferable way of experiencing the high-quality content, whether it is for on-demand (e.g., blockbuster movies) or live services (e.g., sport events). Recent reports from largest VOD platforms such as Netflix confirm that assumption and show that the primary way of watching content remains fixed TV screens, covering 70% of devices 6 months after subscription [17]. In the same way, YouTube indicates that service usage on fixed TV set remains an inevitable way of accessing the content, with 250M of hours viewed per day on TV screens [18]. As 5G media streaming targets a wide range of connected devices and should be able to deliver video streams to many compatible high-resolution receivers, (e.g. 5G-HDMI-sticks, 5G-StB/5G-MediaGateway or even 5G-TV sets) the inclusion of 4K TV Scenarios for 5G Video codec evaluation is important.

First, the 4K-TV set is currently the most established way of displaying premium quality services using latest technology improvements for video content, including High-Dynamic-Range (HDR) and Wide-Colour-Gamut (WCG). Latest statistics from the Ultra-HD forum indicate that 148 UHD services are currently on-air, 74% being linear, 45% of those using HDR [19]. In addition, a large number of SVOD operators propose 4K access in their subscription packages (e.g. Netflix™ and Amazon Prime™). All these services may eventually take advantage of 5G-network capabilities to increase the device reach and enlarge audiences. This scenario is also endorsed by strong shipment forecasts, as indicated in the latest IHS 4K-TV UHD bluebook [20].

### 6.3.2 Description of the Anticipated Application

In the context of 3GPP services, 5G Media Streaming [13] as well as the TV Video Profiles [3] are specifications addressing this 4K-TV scenario. Both, 5G Media Streaming [13] and TV Video Profiles [3] build on CMAF-based Segment formats and DASH distribution. From 3GPP TS 26.116, the following operation points may be considered in scope of the 4K-TV Streaming Scenario (pending availability of appropriate test content):

- H.265/HEVC UHD, see 3GPP TS 26.116 [3] clause 4.5.4.

- H.265/HEVC UHD HDR, see 3GPP TS 26.116 [3] clause 4.5.6.

- H.265/HEVC UHD HDR HLG, see 3GPP TS 26.116 [3] clause 4.5.8.

This scenario is based on CMAF (including LL-DASH and HLS-LL) distribution of UHD-TV video services over 5G networks to 5G/non-5G capable devices. This includes 5G-equipped devices (e.g. smartphone, tablets, …) but also other devices gateway (e.g. TV sets, HDMI-Stick…) accessing services through a “5G-gateway” which can be a mobile phone or a home gateway. As multiple linear services will be delivered in parallel (news, sport, talk show…) in a similar manner as traditional TV services in a multiplex (potentially using multicast/broadcast delivery over 5G). In certain environments, High Frame Rate (HFR) beyond 60 fps is considered, e.g. in DVB and ATSC broadcast specifications. 3GPP does not have any HFR TV video profiles yet.

Important aspects that are expected to be considered when evaluating a codec in the context of this 4K-TV scenario are:

- Quality and Coding Efficiency:

- High and uninterrupted visual quality, considering the service constraints.

- Any savings can provide significant benefits due to the expected large volume of the traffic either in quality or network utilization.

- Adaptive Bitrate streaming:

- Multiple bit rates are provided, typically with a ladder of 30–50% to permit bandwidth adaptation. The use of constant bit rate (CBR) encoding maximises reuse of a common ladder of encoded representations across multiple distribution networks. The use of capped variable bit rate (VBR) encoding allows the bit rate to be varied according to the difficulty of the source material while maintaining the ability to distribute the encoded representations through distribution networks with fixed capacity. This also maximises the usage of a common ladder across multiple distribution networks.

- CMAF Fragments of size typically in the range of 1–6s to permit seamless switching for bit rate adaptation.

- Regular Random Access, typically every 1–2 seconds according to 3GPP TS 26.116 [3]. To achieve clean switching in both sound and picture when moving between different encoded representations in the ladder, 3.84 seconds enables video segment boundaries to be aligned with an integer number of audio Access Units, if a 50fps video signal and 48kHz audio signal is used.

- Encoding in this scenario is typically done as

- Live and On-Demand distribution and encoding

- Server and Cloud-based Encoding

- Capable of encoding multiple services at variable bitrate, inside a fixed dedicated resource (statistical multiplexing).

### 6.3.3 Source Format Properties

Table 6.3-1 provides an overview of the different source signal properties for 4K-TV. This information is used to select proper test sequences.

Table 6.3-1 4K-TV source format properties

|  |  |
| --- | --- |
| Source format properties | 4K-TV |
| Spatial resolution | 3840 x 2160  (Permitted distribution formats: 2560 × 1440, 1920 × 1080, 1600 × 900, 1280 × 720) |
| Chroma format | Y’CbCr |
| Chroma subsampling | 4:2:0 |
| Picture aspec ratio | 16:9 |
| Frame rates | 24; 50; 60; 24/1.001; 60/1.001; [100; 120] Hz |
| Bit depth | 10 |
| Colour space formats | BT.2020 [15] |
| Transfer characteristics | BT.2020 [15], BT.2100 [16] (PQ & HLG) |

NOTE: High Frame Rate (HFR) is not supported by 3GPP TV Video profiles defined in 3GPP TS 26.116 [3] in release 16. However, HFR is introduced in this clause for consideration on the video codec performances.

6.3.4 Encoding and Decoding Constraints

Table 6.3-2 provides an overview of encoding and decoding constraints for 4K-TV category using legacy codec HEVC. This will support the definition of detailed test conditions. It is noted that no relevant profiles exist in TS26.116 and TS26.511 for HFR 4K-TV content.

Table 6.3-2 Encoding and Decoding Configurations for 4K-TV with legacy HEVC codec

|  |  |  |
| --- | --- | --- |
| Encoding and Decoding Constraints | H.265/HEVC 4K-TV | H.265/HEVC 4K-TV HFR |
| Relevant Codec and Codec Profile/Levels according to TS26.116 and TS26.511. | H.265/HEVC Main-10 Profile  Level 5.1 [8] | No relevant 3GPP profiles, should be aligned with H.265/HEVC Main-10 Profile Level 5.2 [8] |
| RAP period | 3.84sec, 1sec | 3.84sec, 1sec |
| Bit rate parameters (CBR, VBR, CAE, HRD parameters) | B = {10,15,20,25} Mbps  CBR and capped-VBR | B = {10,15,20,25} Mbps  CBR and capped-VBR |
| Latency requirements and specific encoding settings | No latency requirements beyond RAP so picture reordering allowed | No latency requirements beyond RAP so picture reordering allowed |
| Encoding complexity context | real-time encoding, cloud-based encoding, offline encoding, etc. | real-time encoding, cloud-based encoding, offline encoding, etc. |
| Required decoding capabilities | H.265/HEVC Main-10 Profile  Level 5.1 [8] | H.265/HEVC Main-10 Profile  Level 5.2 [8] |

### 6.3.5 Performance Metrics

Performance is assessed using BD-Rate computation, with PSNR, SSIM and VMAF metrics as objective quality criterion. Regarding complexity considerations, encoding/decoding runtime is provided.

### 6.3.6 Interoperability Considerations

In order to use a codec in the context of 5G Media Streaming services in 3GPP TS 26.511 [13] and for TV Video profiles in 3GPP TS 26.116 [3], the same considerations for interoperability as for FullHD according to clause 6.2.6 apply.

For additional details, please refer to 3GPP TS 26.116 [3] and 3GPP TS 26.511 [13].

### 6.3.7 Test Sequences

Following sources exist and may be used:

- 4K sequences from Ultra Video Group [21],

- Derf test material available at https://media.xiph.org/video/derf/

### 6.3.8 Detailed Test Conditions

#### 6.3.8.1 Overview

First, the legacy codec HEVC is tested to assess the relevance of what’s already in the 3GPP specification for this particular 4K-TV scenario. For this first test, two HEVC implementations are compared (x265 and reference software HM), according to encoding constraints derived from Table 6.3-2. For x265, encoding presets are selected to cover the desired encoding complexity contexts, for live and offline encoding. The tested rate-control modes are CBR and capped-VBR to fit with the possible delivery methods (single or multiple services inside a resource). In this first test, the HEVC reference implementation HM is evaluated with CBR rate-control on, at coding-tree-unit (CTU) granularity. In addition, fixed QP encoding is also carried out to provide additional anchor points aligned with formal MPEG/JVET common test conditions (potentially for future comparison with other codecs, if needed).

#### 6.3.8.2 Reference Software HEVC 1: HM16.20s

As reference software for HEVC, the following is used

- <https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.20/>

Example setting: <https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.20/cfg/encoder_randomaccess_main10.cfg> with following proposed changes

- IntraPeriod: Intra Period such that 1 second is achieved

- DecodingRefreshType: 1 (CRA) 🡺 2 (IDR)

- GOPSize: adjusted to Intra

- QP: [25, 28, 31, 34]

Tbd

#### 6.3.8.2 HEVC open-source implementation libx265

Tbd

### 6.3.9 External Performance Data

Tbd (with [21] [22] [23] as references ).

## 6.4 Scenario 4: Online Gaming and Screen Content Scenario

### 6.4.1 Motivation

This scenario mostly motivates cases for which content goes beyond videographic content, an in particular includes computer generated imagery (CGI). Several application spaces are obvious and serve as motivation and reference, in particular gaming, 3D content and telepresence including screen and slide sharing.

According to the 2020 Mobile Internet Phenomena Report from Sandvine [9] gaming is continuing to grow on mobile network. The improved performance of 4G and the coming promise of 5G will continue to drive at least casual gamers to mobile networks.

Online gaming was discussed and introduced in detail in TR 26.928 [X]. At least the following use cases are in context of Online gaming:

- Use Case 5: Untethered Immersive Online Gaming

- Use Case 6: Immersive Game Spectator Mode

For raster-based split rendering, according to TR 26.928, clause 4.4, rasterized 3D scenes available in frame buffers are provided by the XR engine and need to be encoded, distributed, and decoded. According to clause 4.2.1, relevant formats for frame buffers are 2k by 2k per eye, potentially even higher. Frame rates are expected to be at least 60fps, potentially higher up to 90 fps. The formats of frame buffers are regular texture video signals that are then directly rendered. As the processing is graphics centric, formats beyond commonly used 4:2:0 signals and YUV signals may be considered. It is known from experiments that with H.264/AVC the bitrates are in the order of 50 Mbps per eye buffer. It is expected that this can be reduced to lower bitrates with improved compression tools as for example available for H.265/HEVC. For use case 5 from above and split rendering, encoding is required to be done in low-latency based on the considerations in TR 26.928. For the spectator mode, higher latency may be acceptable.

As an example, a comprehensive set of API including high-performance tools, samples and documentation for hardware accelerated video encode and decode on Windows and Linux for NVIDIA™ Video Codec SDK is available [N]. For example, in a game recording and streaming scenario like streaming to Twitch.tv using Open Broadcaster Software (OBS), encoding being completely offloaded to NVENC makes the graphics engine bandwidth fully available for game rendering. As of May 2020, the following formats are supported for hardware-based encoding as documented on the high-end Turing encoding:

- H.264 (AVCHD) YUV 4:2:0, YUV 4:4:4, and Lossless, all 8 bit, Max Resolution 4096 x 4096;

- H.265 (HEVC) YUV 4:2:0, YUV 4:4:4, and Lossless, all 10 bit, Max Resolution 8192 x 8192;

[ Comment from Xiaomi

In typical cloud gaming environments, the game server produces rasterized frames at a fixed resolution, framerate and colour bit depth which are negotiated with the player client. Negotiation takes into account game capabilities, player choices and eventually bandwidth constraints.

Typical characteristics of rasterized frames produced by the game engine are:

* Resolution of 720p, 1080p or 4K
* Framerate of 30fps, 60fps or 120 fps
* Typical color bit depth of 8bits (RGB frames) but higher bit depth may be offered for HDR compatible games

Rasterized frames are directly passed to a video encoder (typically H.264 but H.265 may be used in a few environments) and content is live encoded to fit target quality. As an example, the following quality categorization may be done:

* High Quality: 4k at 60/120fps with an average throughput of 60/100 Mbps
* Main Quality: 1080p at 60/120fps with an average throughput of 30/40 Mbps
* Low Quality: 720p/1080p at 30fps with an average throughput of 10/12 Mbps

]

For telepresence and screen-sharing applications, some information related to video is collected in the following:

* MS Teams™ [26] as of end of 2019.
  + There are several formats supported for video. Two key properties of a video format are its frame size and color format. Supported frame sizes include 640x360 ("360p"), 1280x720 ("720p"), and 1920x1080 ("1080p"). Supported color formats include NV12 (12 bits per pixel) and RGB24 (24 bits per pixel).
  + A "720p" video frame contains 921,600 pixels (1280 times 720). In the RGB24 color format, each pixel is represented as 3 bytes (24-bits) comprised of one byte each of red, green, and blue color components. Therefore, a single 720p RGB24 video frame requires 2,764,800 bytes of data (921,600 pixels times 3 bytes/pixel). At a frame rate of 30fps, sending 720p RGB24 video frames means processing approximately 80 MB/s of content (which is substantially compressed by the H.264 video codec before network transmission).
* Other tools are for further study.

### 6.4.2 Description of the Anticipated Application

3GPP until now has very restricted set of services, but based on the considerations in clause 6.Y.1, the following encoding benchmark capabilities are considered for decoding:

- H.264 (AVCHD) YUV 4:2:0, YUV 4:4:4, 8 bit, Max Resolution 1920x1080 and 4096 x 2048

- H.265 (HEVC) YUV 4:2:0, YUV 4:4:4, 10 bit, Max Resolutions 4096 x 2048, 8192 x 4096

The considered scenario is low-latency streaming, possibly using UDP/IP based distribution. Important aspects that are expected to be considered when evaluating a codec in the context of this:

- Quality and Coding Efficiency:

- The ability to compress computer-generated content.

- The ability compress YUV 4:2:0 and 4:4:4 content

- Considered settings for encoding:

- Low-latency settings

- No specific error resilience mechanisms

- Encoding in this scenario is typically done as

- Real-time encoding

- Cloud-based encoding

### 6.4.3 Source Format Properties

Table 6.4-1 provides an overview of the different source signal properties for Online Gaming and Screen Content Sharing. This information is used to select proper test sequences.

Table 6.4-1 Screen Content and Online Gaming source properties

|  |  |  |
| --- | --- | --- |
| Source format properties | Screen Content | Online Gaming |
| Spatial resolution | 1920 x 1080 | 1920x1080, 2048 x 1024, 4096 x 2048, 8192 x 4096 |
| Chroma format | Y’CbCr | Y’CbCr |
| Chroma subsampling | 4:2:0, 4:4:4 | 4:2:0, 4:4:4 |
| Picture aspec ratio | 16:9 | 16:9; 2:1 |
| Frame rates | 25, 30, 50, 60 Hz | 30, 50, 60, 90, 120 Hz |
| Bit depth | 8 | 8, 10 |
| Colour space formats | BT.709 | BT.709, BT.2020 |
| Transfer characteristics | BT.709 | BT.2100 (HDR) |

### 6.4.4 Encoding and Decoding Constraints

Table 6.4-2 provides an overview of encoding and decoding constraints Online Gaming and Screen Content scenario using AVC and HEVC codecs. This will support the definition of detailed test conditions.

Table 6.4-2 Encoding and Decoding Configurations for Online Gaming and Screen Content

|  |  |  |
| --- | --- | --- |
| Encoding and Decoding Constraints | AVC | HEVC |
| Relevant Codec and Codec Profile/Levels according to TS26.116 and TS26.511. | H.264/AVC Main Profile  Level 4.0 [X] | H.265/HEVC Main-10 Profile  Level 4.1, Level 5.1, Level 6.1 |
| RAP period | 1 second, no intra | 1 second, no intra |
| Bit rate parameters (CBR, VBR, CAE, HRD parameters) | Fixed QP | Fixed QP |
| Latency requirements and specific encoding settings | Low-latency requirements, no backward-compatible prediction | Low-latency requirements, no backward-compatible prediction |
| Encoding complexity context | real-time encoding | real-time encoding. |
| Required decoding capabilities | H.264/AVC Main Profile  Level 4.0 [X] | H.265/HEVC Main-10 Profile  Level 4.2, 5.2, 6.2 [8] |

### 6.4.5 Performance Metrics

### 6.4.6 Interoperability Considerations

### 6.4.7 Test Sequences

[Test sequences illustrating the Online Gaming scenario should ideally contain synthetic content with detailed textures and realistic movements.

Tests sequences illustrating the screen content scenario should ideally contain either snyhtetic content from a presentation such as a slide deck with text and graphics or natural content presenting participants to a video conference call, or a mix of the two]

### 6.4.8 Detailed Test Conditions

[For online gaming, would like to test at least AVC and HEVC in 480p/720p/1080p with bitrates ranging from 5-20Mbps for AVC and 2 to 15Mbps for HEVC, with AVC Main Profile and HEVC Main10 profile, both codecs in low delay mode

For screen content, would like to at least test HEVC, 720p and 1080p, Main and Screen Extended Main profiles]

### 6.4.9 External Performance Data

tbd

### 6.4.10 Additional Information

tbd

## 6.5 Scenario 4: Messaging and Social Sharing

### 6.5.1 Motivation

According to the 2020 Mobile Internet Phenomena Report from Sandvine [9] Video traffic continues to grow worldwide, and the increasing popularity of mobile consumers sharing video has not only caused growth in downstream traffic, but also in upstream traffic as well. Instagram ™ grew in the upstream as more consumers share images and videos. TikTok ™, Snapchat ™ (video), FaceTime ™, and even Facebook ™ Live were all in the top 50 applications worldwide on the upstream that are video-sharing-centric. Messaging applications, especially on the upstream, continue to become a critical part of the mobile experience, replacing old style text messaging, and increasingly are video-based. Four of the top 20 applications on the upstream are messaging apps.

Some typical examples and restrictions in April 2020 are provided in the following:

1. WhatApp™ [25]
   1. The maximum size of the video that you can share is 16 MB.
   2. Various container formats that are supported by include MP4, MKV, AVI, 3GP, and MOV. H.264/AVC video codec and AAC audio codec are needed today.
2. YouTube™ [26]
   1. H.264/AVC is the recommended codec with the following settings
      1. Progressive scan
      2. High Profile
      3. 2 consecutive B frames
      4. Closed GOP. GOP of half the frame rate.
      5. CABAC
      6. Variable bitrate. No bitrate limit required, though we offer recommended bit rates below for reference
      7. Chroma subsampling: 4:2:0
   2. Resolution Formats: 360p, 480p, 720p, 1080p, 1440p, 2160p
   3. Both SDR and HDR are possible
   4. The standard aspect ratio is 16:9
3. Facebook Live ™ [27]. To live stream on Facebook™, these video format guidelines are provided:
   1. Recommended max bit rate is 4000 Kbps (4 mbps).
   2. Max: 1080p (1920x1080) resolution, at 60 frames per second.
   3. An I-frame (keyframe) must be sent at least every 2 seconds throughout the stream.
   4. H264 encoded video.
4. TikTok™ [28], some video restrictions
   1. Upload from Android, up to 72 MB at most. upload videos from iOS up to 287.6 MB.
   2. Video dimensions should be 1080 x 1920
   3. MP4 or MOV file format. Video should be H.264/AVC encoded
5. Snapchat™ [25][2)], The latest information from 2018
   1. Recommended size: 1080 by 1920 pixels (9:16 aspect ratio)
   2. Recommended specs: .MP4 or MOV, H.264 encoded, maximum file size 1GB

According to Sandvine's report [9], sharing and uploading content as part of social sharing is predominantly pictures and videos that uploaded directly into a cloud and uploaded to one or many social networks, and then discussed (or shared again) over messaging networks. The relevant quality-of-experience factors include the quality of shared content, the time it takes to upload, the costs associated with the upload and also the processing and battery consumption requirements on the device to prepare the content for upload.

In another activity, three large operators released a 5G messaging white paper [31] to promote enhanced messaging services based on Rich Communication Services (RCS). This also shows the relevance of operator-based messaging services. Generally, uplink resources are even more precious and costly in 5G network operation and hence efficient technologies are vital for mass-scale services. This aspect is also considered by GSMA RCS Universal Profile specification, promoted as the industry standard for RCS Business Messaging, ensuring the telecoms industry remains at the centre of digital communications [34].

### 6.5.2 Description of the Anticipated Application

In the context of 3GPP services, 5G Media Streaming [13] provides the following encoding benchmark capabilities:

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

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

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

- up to 240 frames per second, and

- the Chroma format being 4:2:0, and

- the bit depth being either 8 or 10 bit,

to a bitstream that is decodable by a decoder that is **HEVC-FullHD-Dec** capable as defined in clause 4.2.2.1 of TS26.511 and defined as the capability to decode H.265 (HEVC) Main10 Profile, Main Tier, Level 4.1 [3] bitstreams that have general\_progressive\_source\_flag equal to 1, general interlaced\_source\_flag equal to 0, general\_non\_packed\_constraint\_flag equal to 1, and general\_frame\_only\_constraint\_flag equal to 1.

Based on the considerations in clause 6.X.1, it is also recommened to take into account the AVC-FullHD-Enc capabilities as defined in TS26.511 [13]:

**AVC-FullHD-Enc**: the capability to encode a video signal with

- up to 245,760 macroblocks per second, and

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

- up to 240 frames per second, and

- 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 **AVC-HD-Dec** capable as defined in clause 4.2.1.1 of TS26.511 and defined as the capability to decode H.264 (AVC) Progressive High Profile Level 4.0 [2] bitstreams.

Based on future expectations of higher quality uploads, it is also recommened to take into account the HEVC-UHD-Enc capabilities as defined in TS26.511 [13]:

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

- up to 534,773,760 luma samples per second, and

- up to a luma picture size of 8,912,896 samples, and

- up to 480 frames per second, and

- the Chroma format being 4:2:0, and

- the bit depth being either 8 or 10 bit,

to a bitstream that is decodable by a decoder that is **HEVC-UHD-Dec** capable as defined in clause 4.2.2.1 of TS26.511 and defined as the capability the capability to decode H.265 (HEVC) Main10 Profile, Main Tier, Level 5.1[3] bitstreams that have general\_progressive\_source\_flag equal to 1, general interlaced\_source\_flag equal to 0, general\_non\_packed\_constraint\_flag equal to 1, and general\_frame\_only\_constraint\_flag equal to 1.

The considered scenario is the uploading and uplink streaming into the ISO/BMFF and CMAF container formats. Important aspects that are expected to be considered when evaluating a codec in the context of this:

- Quality and Coding Efficiency:

- The ability to compress a video sequence targeting the maximum file size and maintaining high quality.

- The ability to compress a video stream in real time to the available uplink streaming resources.

- Considered settings for encoding:

- Regular random access at least every 2 seconds, preferably more often

- No specific encoding latency constraints are applicable

- Encoding in this scenario is typically done as

- Real-time encoding for social sharing

- Offline encoding for messaging

- UE-based Encoding

### 6.5.3 Source Format Properties

Table 6.5-1 provides an overview of the different source signal properties for Social Sharing and Messaging. This information is used to select proper test sequences.

Table 6.5-1 Social sharing and source format properties

|  |  |
| --- | --- |
| Source format properties | Social Sharing |
| Spatial resolution | 1920 x 1080  (Permitted encoding formats: 1920 × 1080, 1280 × 720, 854 × 480) |
| Chroma format | Y’CbCr |
| Chroma subsampling | 4:2:0 |
| Picture aspec ratio | 16:9 |
| Frame rates | 25, 30, 50, 60 Hz |
| Bit depth | 8, 10 |
| Colour space formats | BT.709, BT.2020 |
| Transfer characteristics | BT.709, BT.2100 (HDR) |

### 6.5.4 Encoding and Decoding Constraints

Table 6.5-2 provides an overview of encoding and decoding constraints for Social sharing and messaging category using AVC and HEVC codecs. This will support the definition of detailed test conditions.

Table 6.5-2 Encoding and Decoding Configurations for Social sharing and messaging

|  |  |  |
| --- | --- | --- |
| Encoding and Decoding Constraints | AVC HD | HEVC HD |
| Relevant Codec and Codec Profile/Levels according to TS26.116 and TS26.511. | H.264/AVC Main Profile  Level 4.0 [X] | H.265/HEVC Main-10 Profile  Level 4.0 [8] |
| RAP period | 1 and 5 seconds | 1 and 5 seconds |
| Bit rate parameters (CBR, VBR, CAE, HRD parameters) | B = {0.5, 1, 2, 5} Mbps  Capped-VBR (social sharing) and VBR (messaging)  Fixed QP | B = {0.5, 1, 2, 5} Mbps  Capped-VBR (social sharing) and VBR (messaging)  Fixed QP |
| Latency requirements and specific encoding settings | No latency requirements beyond RAP so picture reordering allowed | No latency requirements beyond RAP so picture reordering allowed |
| Encoding complexity context | real-time encoding (social sharing), offline encoding (messaging) | real-time encoding (social sharing), offline encoding (messaging). |
| Required decoding capabilities | H.264/AVC Main Profile  Level 4.0 [X] | H.265/HEVC Main-10 Profile  Level 4.0 [8] |

### 6.5.5 Performance Metrics

tbd

### 6.5.6 Interoperability Considerations

tbd

### 6.5.7 Test Sequences

Tbd

https://photos.app.goo.gl/6QrmTTMoizVtxAoQA

### 6.5.8 Detailed Test Conditions

tbd

### 6.5.9 External Performance Data

tbd

### 6.5.10 Additional Information

tbd

# 7 Characterization of Existing Codecs

# 8 Gaps and Optimization Potential

## 8.1 Identified Gaps and Deficiencies with Existing Codecs

ffs

## 8.2 Potential Requirements for New Codecs

Ffs

Reference to interop requirements in clause 4.6.

# 9 Initial Information on new Codecs

ffs

# 10 Conclusions and Proposed Next Steps

ffs

Annex A  
Scenario Template

## A.1 Introduction

This annex provides a proposed template to introduce a Scenario for 5G Video. This template has been used to collect the scenarios in this report.

## A.2 Template

The following aspects are considered to be important for a scenario

1. Scenario name <give the scenario a catchy name>
2. Motivation for the scenario: Why is the scenario relevant for 5G and video? What is the expected traffic?
3. Description of the scenario: This provides a description of the scenario addressing potentially the relation to a service *5G-based services and applications*, including video formats (resolution, frame rates, color space, etc.), encoding and decoding requirements, adaptive streaming requirements, predominantly based on scenarios defined for 5G media streaming as well as for TR 26.925 and TR 26.928
4. Supporting companies and 3GPP members:
   1. This documents the 3GPP members that support this scenario in terms of providing the information, test material, test requirements and the characterization for the tests. For each of the identified necessities, a tick box is is created in the template.
   2. Preferably several 3GPP members are included in the support, and in addition a video service provider may be included (not necessarily a 3GPP member).
   3. Cross-verification is preferably done by the supporters of the scenario
5. Source format properties: This defines a clear range of the considered and relevant source formats, including the signal properties, but also the characteristics of the content. As an example, the source formats as defined in TS26.116 may be used which include:
   1. Spatial resolutions
   2. Chroma Format
   3. Chroma Subsampling
   4. Aspect ratios
   5. Frame rates
   6. Colour space formats
   7. Transfer Characteristics
   8. Bit depth
   9. Other signal properties
6. Encoding and decoding constraints and settings: Typical encoding constraints and settings such as
   1. Relevant Codec and Codec Profile/Levels according to TS26.116 and TS26.511.
   2. Random access frequency
   3. Error resiliency requirements
   4. Bitrates and quality requirements
   5. Bitrate parameters (CBR, VBR, CAE, HRD parameters)
   6. ABR encoding requirements (switching frequency, etc.)
   7. Latency requirements and specific encoding settings
   8. Encoding context: real-time encoding, on device encoding, cloud-based encoding, offline encoding, etc.
   9. Required decoding capabilities
7. Performance Metrics and Requirements:
   1. A clear definition on how the performance needs to be evaluated including metrics, etc addressing the main KPIs of the scenario.
   2. Objective measures such as PSNR, VMAF, etc, may be used.
   3. Subjective evaluation is not excluded and may be done, but needs commitment
8. Interoperability Considerations for the application
   1. Streaming with DASH/HLS/CMAF
   2. RTP based delivery
9. Test Sequences
   1. A set of selected test sequences that are provided by the proponents in order to do the evaluation. They should cover a set of source format properties
10. Detailed test conditions:
    1. Provides a proposal for detailed test conditions, for example based on a reference software together with the sequences and configuration parameters.
11. External Performance data
    1. References to external performance data that can be added, for example other SDOs, public documents and so on.
12. Additional Information

Annex B  
Details on Performance Metrics

Annex <X> (informative):  
Change history

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Change history** | | | | | | | |
| **Date** | **Meeting** | **TDoc** | **CR** | **Rev** | **Cat** | **Subject/Comment** | **New version** |
| 2020-04 | SA4#108 | S4-200661 |  |  |  | Initial Version | 0.0.1 |
| 2020-04 | SA4#108 | S4-200666 |  |  |  | Version agreed at SA4#108-e | 0.1.0 |
| 2020-06 | SA4#109 | S4-200891 |  |  |  | Version agreed at SA4#109-e | 0.2.0 |

1. Achieves the objective of high quality for more difficult source material. [↑](#footnote-ref-2)