|  |  |
| --- | --- |
| 3GPP TR 26.955 V0.4.0 (2020-11) | |
| 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 7

Introduction 8

1 Scope 9

2 References 9

3 Definitions of terms, symbols and abbreviations 12

3.1 Terms 12

3.2 Symbols 12

3.3 Abbreviations 12

4 Overview of video codec capabilities in 3GPP services in Release 16 14

4.1 Introduction 14

4.2 TV Video Profiles 14

[4.3 VR Video Profiles – To be kept only if interest raised on VR 360 video codec evaluation 15

4.4 5G Media Streaming 16

4.4.1 Introduction 16

4.4.2 5GMS Downlink Streaming default profile 16

4.4.2.1 H.264 (AVC) 16

4.4.2.2 H.265 (HEVC) 16

4.4.2.3 Encapsulation format 16

4.4.3 5GMS Uplink Streaming default profile 17

4.4.3.1 H.265 (HEVC) 17

4.4.3.2 Encapsulation format 17

4.4.4 5GMS Television (TV) profile 17

4.4.4.1 H.264 (AVC) 17

4.4.4.2 H.265 (HEVC) 17

4.4.4.3 Encapsulation format 17

[4.4.5 5GMS Downlink 360 Virtual Reality (VR) profile – to be removed in case no interested in evaluating 360 VR 17

4.4.5.1 H.264 (AVC) 17

4.4.5.2 H.265 (HEVC) 17

4.4.5.3 Encapsulation format 18

4.5 Multimedia Telephony Services over IMS 18

4.5.1 Introduction 18

4.5.2 H.264 (AVC) 18

4.5.3 H.265 (HEVC) 18

4.5.4 Encapsulation format 18

4.6 Messaging Services 18

4.7 Screen Content Coding 19

5 Test and Characterization Framework for Video Codecs 19

5.1 Overview 19

5.2 Original Sequences 20

5.3 Anchors 20

5.4 Reference Software Tools 21

5.4.1 General 21

5.4.2 H.264/AVC 21

5.4.3 H.265/HEVC 22

5.5 Metrics 22

5.5.1 General 22

5.5.2 SDR Metrics 22

5.5.3 HDR Metrics 22

5.6 Tests 22

5.7 Characterization 23

6 Relevant Scenarios 24

6.1 Introduction 24

6.2 Scenario 1: Full HD Streaming 24

6.2.1 Motivation 24

6.2.2 Description of the Anticipated Application 25

6.2.3 Source Format Properties 26

6.2.4 Encoding and Decoding Constraints 27

6.2.5 Performance Metrics 27

6.2.6 Interoperability Considerations 27

6.2.7 Test Sequences 28

6.2.7.1 Standard Dynamic Range 28

6.2.7.2 High Dynamic Range 28

6.2.8 Detailed Test Conditions 28

6.2.8.1 Overview 28

6.2.8.2 Reference Software AVC 1 28

6.2.8.3 Reference Software HEVC 1: HM16.20 28

6.2.9 External Performance Data 29

6.2.10 Additional Information 29

6.3 Scenario 2: 4K-TV 29

6.3.1 Motivation 29

6.3.2 Description of the Anticipated Application 29

6.3.3 Source Format Properties 31

6.3.4 Encoding and Decoding Constraints 32

6.3.5 Performance Metrics 32

6.3.6 Interoperability Considerations 32

6.3.7 Test Sequences 32

6.3.8 Detailed Test Conditions 32

6.3.8.1 Overview 32

6.3.8.2 Reference Software HEVC 1: HM16.20s 33

6.3.8.2 HEVC open-source implementation libx265 33

6.3.9 External Performance Data 33

6.4 Scenario 3: Screen Content Scenario 33

6.4.1 Motivation 33

6.4.2 Description of the Anticipated Application 34

6.4.3 Source Format Properties 34

6.4.4 Screen content codec tools 34

6.4.5 Encoding and Decoding Constraints 35

6.4.6 Performance Metrics 36

6.4.7 Interoperability Considerations 36

6.4.8 Test Sequences 36

6.4.9 Detailed Test Conditions 36

6.4.9.1 Overview 36

6.4.9.2 Reference Software AVC 36

6.4.9.3 Reference Software HEVC Main 10: HM16.22 36

6.4.9.4 Reference Software HEVC SCC: SCM 8.8 37

6.4.10 External Performance Data 37

6.4.11 Additional Information 37

6.5 Scenario 4: Messaging and Social Sharing 38

6.5.1 Motivation 38

6.5.2 Description of the Anticipated Application 39

6.5.3 Source Format Properties 41

6.5.4 Encoding and Decoding Constraints 41

6.5.5 Performance Metrics 42

6.5.6 Interoperability Considerations 42

6.5.7 Test Sequences 42

6.5.8 Detailed Test Conditions 42

6.5.9 External Performance Data 42

6.5.10 Additional Information 42

6.6 Scenario 5: Online Gaming 43

6.6.1 Motivation 43

6.6.2 Description of the Anticipated Application 44

6.6.3 Source Format Properties 44

6.6.3.1 Introduction 44

6.6.3.2 Category A: Low/medium dynamicity with low/medium complexity. 44

6.6.3.3 Category B: games with high dynamicity and low/medium complexity. 45

6.6.3.4 Category C: games with high dynamicity and high complexity. 45

6.6.3.5 Category D: photo-realistic games or games based on natural images/video. 45

6.6.3.6 Category E: XR game content 45

6.6.3.7 Summary 45

6.6.4 Codec and compression tools for games 46

6.6.5 Encoding and Decoding Constraints 46

6.6.6 Performance Metrics 47

6.6.7 Interoperability Considerations 47

6.6.8 Test Sequences 47

6.6.9 Detailed Test Conditions 47

6.6.9.1 Overview 47

6.6.9.2 Reference Software AVC 47

6.6.9.3 Reference Software HEVC Main 10: HM16.22 47

6.6.10 External Performance Data 48

6.6.11 Additional Information 48

7 Characterization of Existing Codecs 48

8 Gaps and Optimization Potential 48

8.1 Identified Gaps and Deficiencies with Existing Codecs 48

8.2 Potential Requirements for New Codecs 48

9 Initial Information on new Codecs 49

9.1 Introduction 49

9.2 Versatile Video Coding (VVC) 49

9.2.1 Overview 49

9.3 Essential Video Coding (EVC) 49

10 Conclusions and Proposed Next Steps 49

Annex A Scenario Template 50

A.1 Introduction 50

A.2 Template 50

Annex B Details on Performance Metrics 51

[B.1 Introduction 51

B.2 Objective metrics for evaluation of video coding efficiency experiments 51

B.3 Uncompressed Video in ISOBMFF 52

B.3.1 Overview 52

B.3.2. ISO/IEC 23001-17 – Background and Application for 5G Video 53

B.4 Sequence Quality Evaluation Tool 53

B.4.1 Overview 53

B.4.2 Usage of the QE tool 53

B.4.3 What does the tool 53

B.5 BD-rate numbers 54

B.5.1 Introduction 54

B.5.2 Per-sequence BD-rate 55

B.5.2 BD-Rate overall 55

B.6 Scripts and Tools 55

Annex C Candidate Test Sequences 56

C.1 Introduction 56

C.2 Gaming Test Sequences 56

C.2.1 Overview 56

C.2.2 AOV 56

C.2.2.1 Introduction 56

C.2.2.2 Source sequence properties 56

C.2.2.3 Copyright information 57

C.2.3 Baolei-Man 57

C.2.3.1 Introduction 57

C.2.3.2 Source sequence properties 57

C.2.3.3 Copyright information 57

C.2.4 Baolei-Woman 58

C.2.4.1 Introduction 58

C.2.4.2 Source sequence properties 58

C.2.4.3 Copyright information 58

C.2.5 Jianling-Temple 59

C.2.5.1 Introduction 59

C.2.5.2 Source sequence properties 59

C.2.5.3 Copyright information 59

C.2.6 Jiangling-Beach 60

C.2.6.1 Introduction 60

C.2.6.2 Source sequence properties 60

C.2.6.3 Copyright information 60

C.2.7 Heroes of the Storm 61

C.2.7.1 Introduction 61

C.2.7.2 Source sequence properties 61

C.2.7.3 Repository 61

C.2.7.4 Copyright information 61

C.2.8 Project CARS 62

C.2.8.1 Introduction 62

C.2.8.2 Source sequence properties 62

C.2.8.3 Repository 62

C.2.8.4 Copyright information 62

C.2.9 World of WarCraft: WoW 63

C.2.9.1 Introduction 63

C.2.9.2 Source sequence properties 63

C.2.9.3 Repository 63

C.2.9.4 Copyright information 63

C.2.10 Minecraft 64

C.2.10.1 Introduction 64

C.2.10.2 Source sequence properties 64

C.2.10.3 Repository 64

C.2.10.4 Copyright information 64

C.2.11 CS:GO 65

C.2.11.1 Introduction 65

C.2.11.2 Source sequence properties 65

C.2.11.3 Repository 65

C.2.11.4 Copyright information 65

C.2.12 StarCraft 66

C.2.12.1 Introduction 66

C.2.12.2 Source sequence properties 66

C.2.12.3 Repository 66

C.2.12.4 Copyright information 66

Annex <X> (informative): Change history 67

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>

[38] Recommendation ITU-T H.266 (07/2020): "Versatile Video Coding" | ISO/IEC 23090-3:2020: "Information technology — Coded representation of immersive media — Part 3: Versatile video coding".

[39] Recommendation ITU-T H.274 (07/2020): "Versatile supplemental enhancement information for coded video bitstreams" | ISO/IEC 23002-7:2020 "Information technology — MPEG video technologies — Part 7: Supplemental enhancement information messages for coded video bitstreams".

[40] [JCTVC-AA1006](http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=10735), “Verification test report for HEVC screen content coding extensions” V. Baroncini, H. Yu, R. Joshi, S. Liu, X. Xiu, J. Xu (editors)- 2017-09-08

[41] Intel Xe LP presented at Intel Architecture Day 2020. Slide 101 https://newsroom.intel.com/wp-content/uploads/sites/11/2020/08/Intel-Architecture-Day-2020-Presentation-Slides.pdf

[42] N. Barman, S. Zadtootaghaj, S. Schmidt, M. G. Martini and S. Möller, "GamingVideoSET: A Dataset for Gaming Video Streaming Applications," 2018 16th Annual Workshop on Network and Systems Support for Games (NetGames), Amsterdam, Netherlands, 2018, pp. 1-6. DOI: 10.1109/NetGames.2018.8463362

[43] N. Barman, S. Schmidt, S. Zadtootaghaj, M. G. Martini, and S. Möller. 2018. An Evaluation of Video Quality Assessment Metrics for Passive Gaming Video Streaming. In Proceedings of the 23rd Packet Video Workshop (PV '18). ACM, Amsterdam, Netherlands, 2018, pp. 7-12. DOI: https://doi.org/10.1145/3210424.3210434

[44] Technical Paper ITU-T HSTP-VID-WPOM (07/2020): "Working practices using objective metrics for evaluation of video coding efficiency experiments".

[45] 3GPP TR 26.902, "Video codec performance".

[46] 3GPP TS 26.244, "Transparent end-to-end packet switched streaming service (PSS); 3GPP file format (3GP)"

[47] ISO/IEC 23001-17, "Information technology — MPEG systems technologies — Part 17: Carriage of Uncompressed Video in ISOBMFF"

[48] Recommendation ITU-T P.910 (04/2008): "Subjective video quality assessment methods for multimedia applications".

[49] Ultra HD Forum Guidelines, <https://ultrahdforum.org/wp-content/uploads/UHD-Guidelines-V2.3-final.pdf>

[X] ISO/IEC 23094-1, "Information technology — General video coding — Part 1: Essential video coding"

[Y] ISO/IEC 23094-4, " Information technology — General video coding — Part 4: Conformance and Reference software for Essential Video Coding".

[W] ISO/IEC JTC 1/SC 29/WG 04 output document N0027, "Updated Verification Test Plan for Essential Video Coding for SDR Content", Online meeting, Oct. 2020.

[V] ISO/IEC JTC 1/SC 29/WG 04 output document N0030, "Report on Essential Video Coding compression performance verification testing for HDR/WCG content", Online meeting, Oct. 2020.

# 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].

AAC

ABR

ACM

AOV

ARIB

ATSC

AVC Advanced Video Coding

AVCHD

AVI

BMFF

CABAC

CAE

CBP

CBR

CGI

CHP

CMAF

CMP

CRA

CTA

CTU

DASH

DVB

ERP

EVC Essential Video Coding

FPS

GOP

HDMI

HDR

HDTV

HEVC High-Efficiency Video Coding

HFR

HLG

HLS

HMD

HRD

HTML

IDR

JVET

MKV

MMO

MMORPG

MMS

MOBA

MOV

MPD

MPEG

MSE

MTSI

OBS

PSNR

RAP

RCC

RCS

RDPCM

RGB

RPG

RTMP

RTP

RTS

RWP

SDK

SDR

SEI

SSIM

SVOD

UDP

UHD

UMTS

UVG

VBR

VVC Versatile Video Coding

# 4 Overview of video codec capabilities in 3GPP services in Release 16

## 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 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 |

Each of the Operation Point is associated to the support of a video codec and its Profile, Level and Tier (for HEVC) combined with additional bitstream constraints defined in TS 26.116 [3], clause 4. The Table 4.2-2 summarizes the video codec profile, tier and level associated to each operation point.

Table 4.2-2: Operation point video codec Profile/Tier/Level

| Operation Point name | Video Codec | Profile | Tier | Level |
| --- | --- | --- | --- | --- |
| H.264/AVC 720p HD | AVC | High | - | 3.1 |
| H.265/HEVC 720p HD | HEVC | Main | Main | 3.1 |
| H.264/AVC Full HD | AVC | High | - | 4.2 |
| H.265/HEVC Full HD | HEVC | Main-10 | Main | 4.1 |
| H.265/HEVC UHD | HEVC | Main-10 | Main | 5.1 |
| H.265/HEVC Full HD HDR | HEVC | Main-10 | Main | 4.1 |
| H.265/HEVC UHD HDR | HEVC | Main-10 | Main | 5.1 |
| H.265/HEVC Full HD HDR HLG | HEVC | Main-10 | Main | 4.1 |
| H.265/HEVC UHD HDR HLG | HEVC | Main-10 | Main | 5.1 |

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 [3], clause 4 and clause 5.

## [4.3 VR Video Profiles – To be kept only if interest raised on VR 360 video codec evaluation

The VR profiles for streaming applications defined in TS 26.118 [4] address the coded representation of 360 VR distribution signals. Table 4.X-1 provides an overview of the 360 VR relevant formats considered in the context of 3GPP VR profiles. The VR profiles follow the same logic as the TV Video profiles, they represent a list of interoperability points that are amended by the 3GPP services such as 5GMS.

Table 4.3-1: High-level Summary of Operation Points

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Operation Point name | Decoder | Bit depth | Typical  Original Spatial Resolution | Frame Rate | Colour space format | Transfer  Characteristics | Projection | Rotation | RWP | Stereo |
| Basic H.264/AVC | H.264/AVC HP@L5.1 | 8 | Up to 4k | Up to 60 Hz | BT.709 | BT.709 | ERP w/o padding | No | No | No |
| Main H.265/HEVC | H.265/HEVC MP10@L5.1 | 8, 10 | Up to 6k in mono and 3k in stereo | Up to 60 Hz | BT.709  BT.2020 | BT.709 | ERP w/o padding | No | Yes | Yes |
| Flexible H.265/HEVC | H.265/HEVC MP10@L5.1 | 8, 10 | Up to 8k in mono and 3k in stereo | Up to 120 Hz | BT.709  BT.2020 | BT.709,  BT.2100 PQ | ERP w/o padding CMP | No | Yes | Yes |

]

## 4.4 5G Media Streaming

### 4.4.1 Introduction

5G Media Streaming (5GMS) services are associated with a series of 5GMS profiles each of which containing a set of capability requirements associated to a service scenario. Default profiles (for downlink and uplink streaming) are also defined in case no other profile is claimed to be supported. The detailed requirements for 5G Media Streaming profiles and their associated codec requirements are defined in 3GPP TS 26.511 [13]. The following clauses focus on video codec requirements and recommendations in 5GMS specification as well as the encapsulation format.

NOTE: In some profiles the HD-HDR capability is used as the ability for a UE to present video signals with all the following features: at least Full-HD resolution, bit depth of at least 10, at least 60 frames per second, Wide Colour Gamut and High Dynamic Range.

### 4.4.2 5GMS Downlink Streaming default profile

#### 4.4.2.1 H.264 (AVC)

TS 26.511 [13] requires the support of H.264 (AVC) Progressive High Profile Level 3.1 [7] decoding, with the maximum VCL Bit Rate constrained to be 14 Mbps.

If the 5GMS client supports the reception of video and HD-HDR capabilities, then TS 26.511 [13] requires the support of H.264 (AVC) Progressive High Profile Level 4.0 [7] decoding.

#### 4.4.2.2 H.265 (HEVC)

TS 26.511 [13] recommends the support of H.265 (HEVC) Main Profile, Main Tier, Level 3.1 [8] decoding with no interlace support.

If the 5GMS client supports the reception of video and HD-HDR capabilities, then TS 26.511 [13] requires the support of H.265 (HEVC) Main10 Profile, Main Tier, Level 4.1 [8] decoding with no interlace support.

#### 4.4.2.3 Encapsulation format

TS 26.511 [13] requires the support of Common Media Application Format (CMAF) encapsulation as defined in ISO/IEC 23000-19 [30].

### 4.4.3 5GMS Uplink Streaming default profile

#### 4.4.3.1 H.265 (HEVC)

TS 26.511 [13] requires the support of H.265 (HEVC) Main10 Profile, Main Tier, Level 4.1[8] encoding with no interlace support.

#### 4.4.3.2 Encapsulation format

TS 26.511 [13] requires the support of Common Media Application Format (CMAF) as defined in ISO/IEC 23000-19 [30].

### 4.4.4 5GMS Television (TV) profile

#### 4.4.4.1 H.264 (AVC)

TS 26.511 [13] requires the support of *H.264/AVC 720p HD* as specified in TS 26.116 [3], clause 4.4.2.6.

TS 26.511 [13] recommends the support of *H.264/AVC Full HD* as specified in TS 26.116 [3], clause 4.4.3.6.

NOTE: Associated codec profiles and levels are listed in Table 4.2-2 of the present document.

#### 4.4.4.2 H.265 (HEVC)

TS 26.511 [13] recommends the support of *H.265/HEVC 720p HD*, *H.265/HEVC Full HD*, *H.265/HEVC, H.265/HEVC Full HD, H.265/HEVC UHD HDR*, *H.265/HEVC Full HD HDR HLG* and *H.265/HEVC UHD HDR HLG* as specified in TS 26.116 [3].

If the 5GMS client supports the Television (TV) profile and HD-HDR capabilities, TS 26.511 [13] requires the support of *H.265/HEVC Full HD* as specified in TS 26.116 [3].

NOTE: Associated codec profiles and levels are listed in Table 4.2-2 of the present document.

#### 4.4.4.3 Encapsulation format

TS 26.511 [13] requires the support of DASH as defined in 3GPP TS 26.116 [3], clause 5.

NOTE: Full alignment between DASH support in TS 26.116 [3] and CMAF support in TS 26.511 is still to be investigated.

### [4.4.5 5GMS Downlink 360 Virtual Reality (VR) profile – to be removed in case no interested in evaluating 360 VR

#### 4.4.5.1 H.264 (AVC)

TS 26.511 [13] requires the support of the *Basic H.264/AVC* Operation Point as defined in 3GPP TS 26.118 [4], clause 5.1.4.

NOTE: Associated codec profiles and levels are listed in Table 4.3-1 of the present document.

#### 4.4.5.2 H.265 (HEVC)

TS 26.511 [13] recommends the support of the *Main* *H.265/HEVC* Operation Point as defined in TS 26.118 [4], clause 5.1.5.

TS 26.511 [13] indicates that the *Flexible* *H.265/HEVC* Operation Point as defined in TS 26.118 [4], clause 5.1.6, may be supported.

NOTE: Associated codec profiles and levels are listed in Table 4.3-1 of the present document.

#### 4.4.5.3 Encapsulation format

TS 26.511 [13] requires the support of DASH as defined in 3GPP TS 26.118 [4], clause 5.2.

NOTE: Full alignment between DASH support in TS 26.116 [3] and CMAF support in TS 26.511 is still to be investigated.

]

## 4.5 Multimedia Telephony Services over IMS

### 4.5.1 Introduction

The Multimedia Telephony Services over IMS (MTSI) specification 3GPP TS 26.114 [2] covers the media aspects of 3GPP conversational services. For MTSI services the UE is associated with both encoder and decoder requirements. Therefore, the support for a given video codec with its profile and level is interpreted as the ability to:

- encode any video input which characteristics fit in the level limits and produce a bitstream conformant with the codec profile.

- decode any bitstream in the codec/profile/level configuration.

NOTE: Additional constraints related to conversational services such as decoding process, location of the parameter sets are defined in TS 26.114 [2], clause 5.2.2.

### 4.5.2 H.264 (AVC)

TS 26.114 [2], in clause 5.2.2 requires the support of H.264 (AVC) [7] Constrained Baseline Profile (CBP) Level 1.2.

TS 26.114 [2], in clause 5.2.2 also recommends H.264 (AVC) [7] Constrained High Profile (CHP) Level 3.1.

### 4.5.3 H.265 (HEVC)

TS 26.114 [2] requires the support of H.265 (HEVC) [8] Main Profile, Main Tier, Level 3.1 except for constrained terminals for which it is only recommended.

### 4.5.4 Encapsulation format

MTSI is built upon the Real Time Protocol (RTP) for which the payload formats are specified in TS 26.114 [2], clause 7.4.3.

## 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:

- TS 26.140 [32] requires the support of H.264 (AVC) [7] Constrained Baseline Profile (CBP) Level 1.3.

- TS 26.140 [32] recommends the support of H.264 (AVC) [7] High Profile Level 3.1 with frame\_mbs\_only\_flag=1 by MMS clients supporting HDTV video content at a resolution of 1280x720 (720p) with progressive scan at 30 frames per second.

- TS 26.140 [32] also recommends the support of H.265 (HEVC) [8] Main Profile, Main Tier, Level 3.1.

NOTE: 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) [7] Constrained High Profile (CHP), Level 3.1

- H.264 (AVC) [7] Constrained Baseline Profile (CBP), Level 1.2 (for interworking with MTSI clients)

- H.265 (HEVC) [8] Main Profile, Main Tier, Level 4.1

As specified in 3GPP TS 26.223 [35], Telepresence UEs are also recommended to support dedicated HEVC extensions defined specifically for such types of content:

- H.265 (HEVC) [8] Screen-Extended Main, Main Tier, Level 4.1

- H.265 (HEVC) [8] Screen-Extended Main 4:4:4, Main Tier, Level 4.1

# 5 Test and Characterization Framework for Video Codecs

## 5.1 Overview

This clause defines the characterization framework for video codecs for relevant 3GPP scenarios. For this purpose, the following is applied:

* A set of relevant scenarios are defined. The scenarios reflect a typical application for video codecs in 5G systems and networks. The scenarios are introduced in clause 6 and may be extended in future versions of this document.
* For each scenario, one or several reference sequences are defined that serve as the baseline for anchor generation.
* For each test scenario, one or several anchors are defined and generated. For details on anchors, please refer to clause 5.3. For anchor generation, reference software tools are used as introduced in clause 5.4.
* For each of the anchors, metrics are provided. Metrics are documented in detail in clause 5.5.
* Tests for new codecs can be developed and generated. They are equivalent to anchor generation, but possibly for other codecs. Tests are introduced in clause 5.6.
* Codecs are be characterized against anchors. Characterization is documented in terms of expected bitrate savings for a codec, and may include additional comparison parameters such as complexity increase, etc. The basic characterization is introduced in clause 5.7.

An overview of the anchor generation framework and the anchor metrics is provided in Figure 5.1-1. An integral part of this Technical Report is:

- Reference and anchor sequence formats to store raw video data

- One or multiple reference sequences for each of the defined scenarios

- Reference software encoders

- Anchor configuration files

- Anchor bitstreams in a well-defined anchor bitstream format

- A anchor metric computation based on a reference sequence, anchor bitstream and an anchor sequence.

- Conforming decoders to generate anchor sequences from an anchor bitstream

- Anchor metrics

NOTE: as the anchor sequences can be generated by conforming decoders, anchor sequences are not included in this technical report

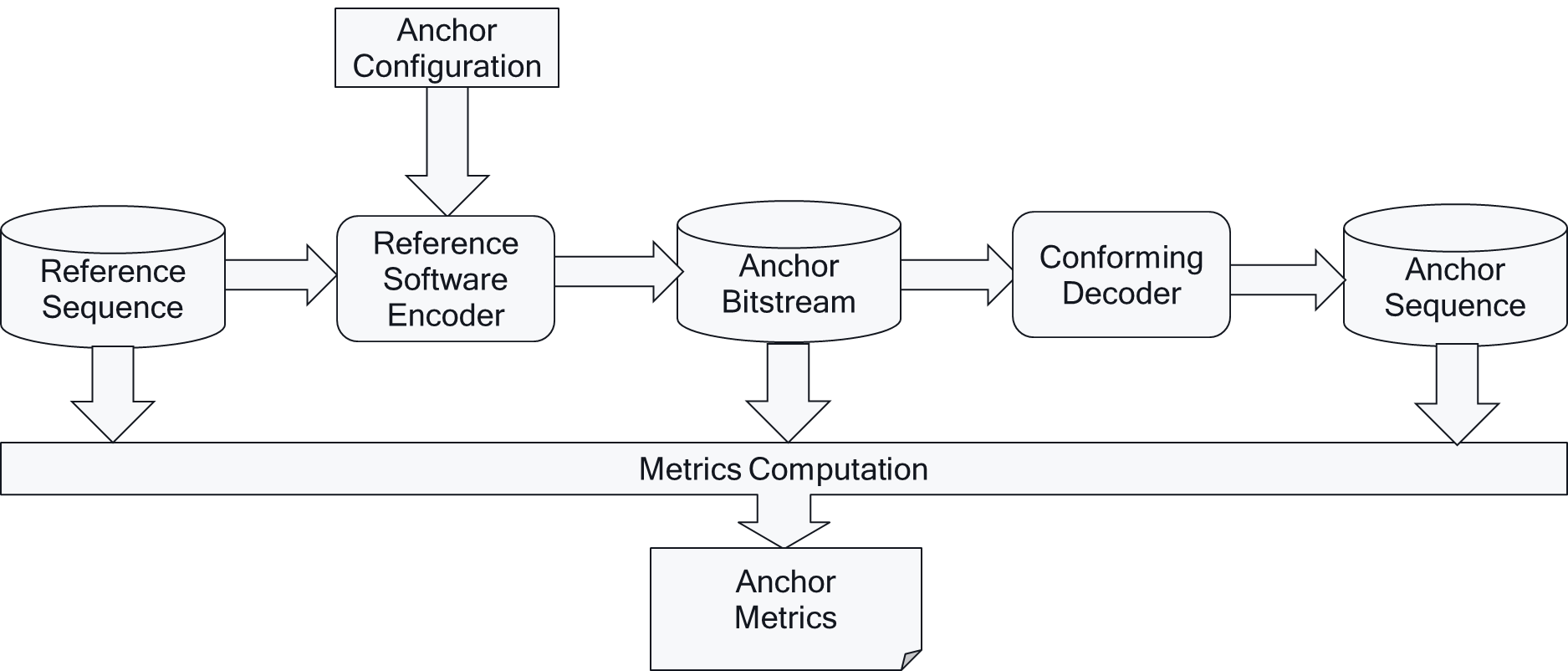


Figure 5.1-1 Anchor Generation Framework and Anchor Metrics Generation

## 5.2 Reference Sequences

This specification provides reference sequences that are used in order to provide anchors and are also made available in order to generate test results for other codecs. Reference sequences are selected to represent scenarios.

Reference sequences are collected in Annex C of this document among the properties and the license.

In order to address reference sequences with different raw formats and to maintain timing in these sequences, they are in a consistent format that maintains sequence properties as well as timing. The format follows the MP4 format for raw video sequences as defined in ISO/IEC 23001-17 [46]. More details on reference sequence format is provided in Annex B.3.

## 5.3 Anchors

Anchors provide a baseline that a tested method can be compared against. Anchors defined in this specification use a codec/profile/level that exists in an existing 3GPP specification as introduced in clause 4.

Anchors are collected in tuples to address different quality and bitrates that can then be used for evaluation over a larger set of operation points.

The following principle apply to anchor definitions:

* Each scenario typically has several well-defined anchors
* An anchor is a combination of:
  + Explanation what this anchor is important for
  + Reference sequence
  + Reference encoder
  + Encoder configuration
  + Encoding complexity estimation, if available
  + Variable encoder configuration to create multiple quality/bitrate variants (using for example QP variations or other bitrate/quality evaluation tools).
  + Anchor tuples creating multiple variants, each including
    - Anchor bitstream
    - Anchor Metrics
  + Additional recommended anchor information includes
    - MD5 check sum of the complete decoded yuv file
    - Output picture log
* Anchors and anchor tuples are an integral part of the Technical Report
* Anchors are preferably stored in an ISO BMFF-based container format.

Anchor tuples should be created over a wide range of parameters in order to provide sufficient data and overlap with expected test results to support the generation of characterization results (see clause 5.7).

The generation of anchor tuples is shown in Figure 5.3-1.

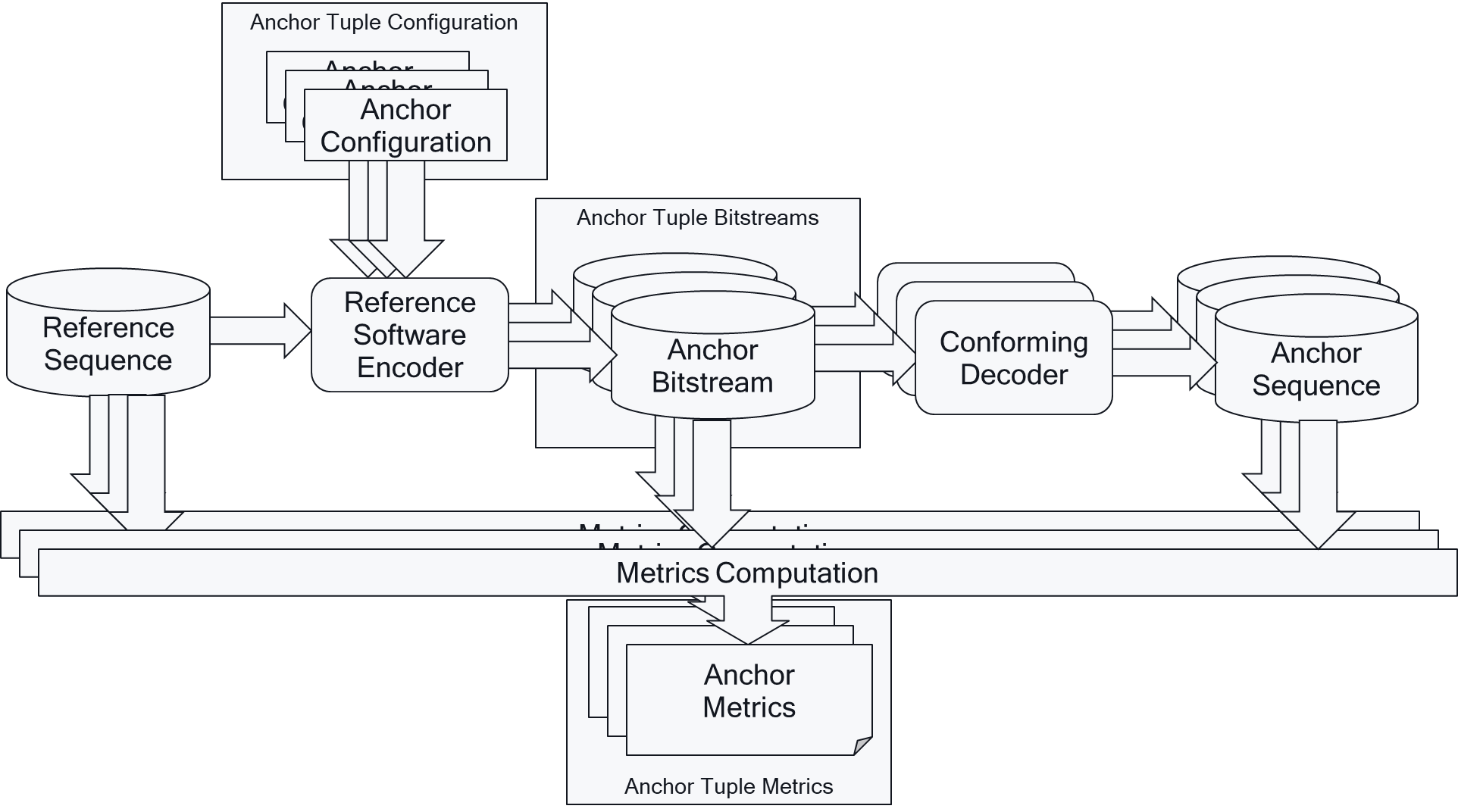


Figure 5.3-1: Anchor Tuple Generation Framework and Anchor Tuple Metrics Generation

## 5.4 Reference Software Tools

### 5.4.1 General

Anchors are generated with a dedicated reference software. The reference software is documented in order to permit repeatability of the anchor generation.

Reference configurations for reference encoders are provided in Annex D.

### 5.4.2 H.264/AVC

For H.264/AVC generated anchors, the following reference software has been used:

* *JM19.0* H.264/MPEG-4 AVC Reference Software: The reference software for H.264/MPEG-4 AVC is called JM (Joint Test Model). The JM software is maintained and can be downloaded from in the repository https://vcgit.hhi.fraunhofer.de/jct-vc/JM and development versions are available. For JM-based H.264/MPEG-4 AVC anchors, version 19.0 is used except specified otherwise. The reference software supports the following 3GPP profiles and levels

### 5.4.3 H.265/HEVC

For H.265/HEVC generated anchors, the following reference software has been used:

* *HM16.22* H.265/MPEG-H HEVC Reference Software: The reference software for H.265/MPEG-H HEVC is called HM (HEVC Test Model). The HM software is maintained and can be downloaded from in the repository https://vcgit.hhi.fraunhofer.de/jct-vc/HM and development versions are available. For HM-based H.265/MPEG-4 HEVC anchors, version 16.22 is used except specified otherwise. The reference software supports the following 3GPP profiles and levels
  + tbd

## 5.5 Metrics

### 5.5.1 General

Each anchor bitstream gets assigned multiple performance metrics, in particular:

* the bitrate of the sequence on ISO BMFF basis, i.e. the sum of the size of all samples divided by duration of the sequence.
* If Standard Dynamic Range (SDR) is used, then the metrics in clause 5.5.2 apply.
* If High Dynamic Range (HDR) is used, then the metrics in clause 5.5.3 apply.

These document does not specify metrics in the report, but the exact definition of metrics based on a provided computational script introduced in Annex B.4 and integral to this Technical Report.

Subjective evaluation of anchors streams is not excluded per se, but this specification does not define any recommended metric or method. Details are for further study.

### 5.5.2 SDR Metrics

For standard dynamic range (SDR) sequences, the following metrics are used:

* Peak-Signal to Noise Ratio (PSNRy) of luminance component
* Weighted PSNR over color components (wPSNR)
* Structural similarity metric (SSIM)
* Video Multimethod Assessment Fusion (VMAF)

For details refer to [44].

### 5.5.3 HDR Metrics

For high dynamic range (HDR) sequences, the following metrics are used:

* PSNRL100,
* wPSNR,
* DE100.

For details refer to [44].

## 5.6 Tests

Tests may be executed to compare against anchors defined in this specification. Tests, equivalently to anchors, are collected in tuples to address different quality and bitrates that can then be used for evaluation over a larger set of operation points.

A test is developed against an anchor and is a combination of:

* Anchor
* Reference Sequence
* Test encoder
* Test encoder configuration that matches the anchor configuration
* Encoding complexity estimation, if available
* Variable encoder configuration to create multiple quality/bitrate variants (using for example QP variations or other bitrate/quality evaluation tools).
* Test tuples creating multiple variants, each including
  + Test bitstream
  + Test Metrics
  + Additional recommended anchor information includes
    - MD5 check sum of the complete decoded yuv file
    - Output picture log
* Tests are an integral part of the Technical Report

The generation of test tuples is shown in Figure 5.6-1.

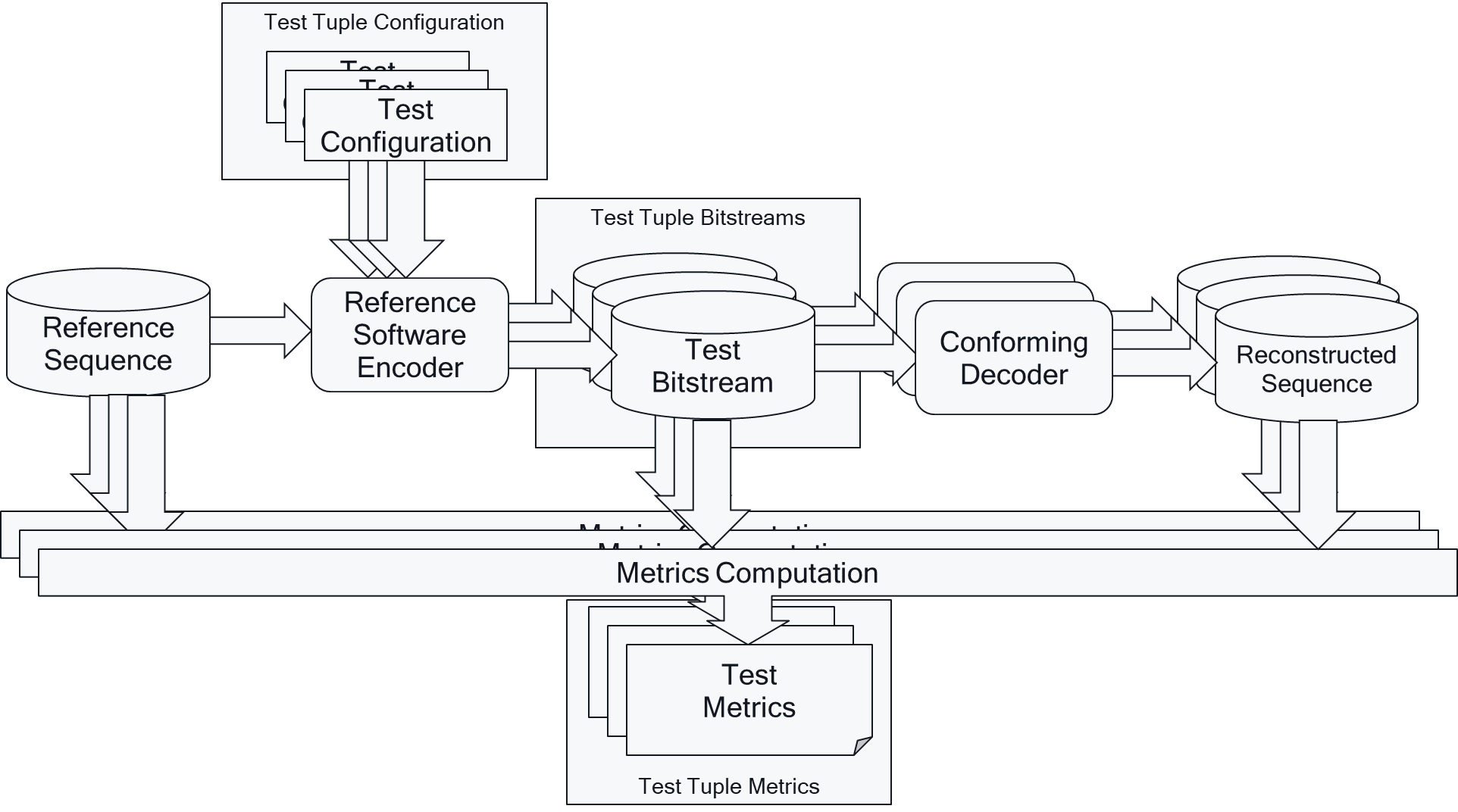


Figure 5.6-1: Test Tuple Generation Framework and Test Tuple Metrics Generation

## 5.7 Characterization

Characterization is the comparison of a codec under test with an anchor based on the framework introduced in this clause. Characterization in this report is based on Bjöntegard-Delta (BD)-rate information according to [44].

It is expected that for each of the scenarios, the following information is documented:

* The BD-rate for each anchor tuple and each metric.
* The average BD-rate for all anchors of the scenario against one codec for one metric.
* [The average BD-rate for all anchors of the scenario against one codec for all metrics]

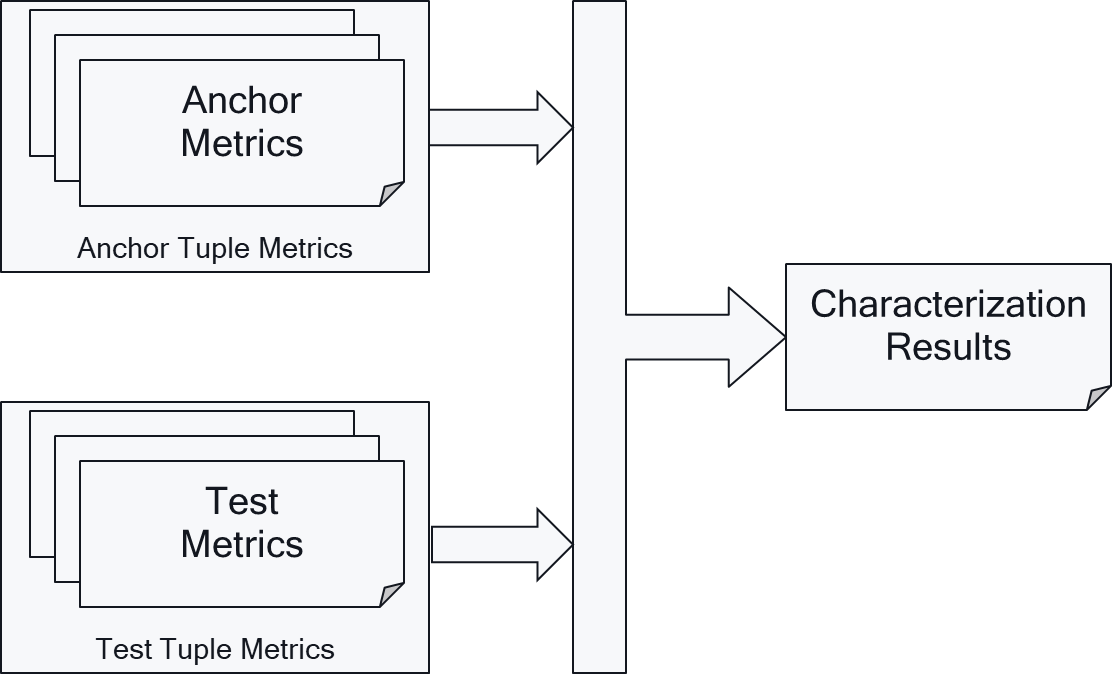


Figure 5.7-1: Characterization Framework

BD-Rate computation is supported by a script that uses anchor tuple and test tuple metrics to provide the characterization results as shown in Figure 5.7-1.

For details on BD-Rate computation, refer to Annex B.5.

# 6 Relevant Scenarios

## 6.1 Introduction

This clause collects relevant scenarios based on the template defined in Annex A. It also defines the anchors for each scenario based on the existing 3GPP codecs and profiles defined in clause 4.

## 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 | 1 second, 3.84 seconds |
| Error resiliency requirements | None | None |
| Bit rates and quality configuration | QP, for example [20, 23, 26, 29] | QP for example [23[[1]](#footnote-2), 25, 28, 31, 34] |
| 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  ABR through multiple QPs | 1 second  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. | real-time encoding, cloud-based encoding, offline encoding, etc. |
| 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

All metrics as defined in clause 5.5 apply.

### 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 Selected Reference Sequences

Table 6.2.7-1 provides the selected reference sequences for this scenario. Keys are identified to refer to the sequences in the context of the scenario. The sequences are named and a reference to the details of the sequence is provided. A justification is provided, why this sequence is selected.

Table 6.2.7-1 Selected Reference Sequences for Full HD Scenario

|  |  |  |  |
| --- | --- | --- | --- |
| Key | Name | Reference | Justification/Comment |
| S1-R1 | WAVE-ToS01 | Annex C.3.2.2 | Good long test sequence and full resolution of 1080p. |
| S1-R2 | WAVE-ToS02 | Annex C.3.2.2 | Good long test sequence at subsampled resolution of 900p. |
| S1-R3 | WAVE-ToS03 | Annex C.3.2.2 | Good long test sequence at subsampled resolution of 720p. |
| S1-R4 | WAVE-CRO01 | Annex C.3.3.2 | Good HDR Test sequence at full resolution of 1080p. |
| S1-R4 | WAVE-CRO02 | Annex C.3.3.2 | Good HDR Test sequence at subsampled resolution of 900p. |
| S1-R4 | WAVE-CRO03 | Annex C.3.3.2 | Good HDR Test sequence at subsampled resolution of 720p. |

### 6.2.8 Anchor Generation Details

#### 6.2.8.1 Overview

This clause provides details on how to generate the anchors for the Full HD Scenario.

#### 6.2.8.2 H.264/AVC Anchors

##### 6.2.8.2.1 Overview

Table 6.2.8.2.1-1 provides an overview of the H.264/AVC anchor tuples. Keys are identified to refer to the anchors in the context of the scenario.

Table 6.2.8.2.1-1 Anchor Tuple generation with H.264/AVC for Full HD Scenario

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Key | Clause | Reference Sequence | Reference Encoder | Configuration | Variations |
| S1-A1-264 | 6.2.8.2.2 | S1-R1 | JM19.0 | JM-01 | QP = [17, 20, 23, 26, 29, 32] |
| S1-A2-264 | 6.2.8.2.2 | S1-R2 | JM19.0 | JM-02 | QP = [17, 20, 23, 26, 29, 32] |
| S1-A3-264 | 6.2.8.2.2 | S1-R3 | JM19.0 | JM-03 | QP = [17, 20, 23, 26, 29, 32] |
|  |  |  |  |  |  |

##### S1-A1-264: One second intra

This anchor tuple produces an anchor over 60 seconds with a typical configuration of H.264/AVC:

- Profile H.264/AVC Progressive High-Profile Level 4.2 [7]

- Random access and switching at 1 second interval

- Fixed QP settings

- IBBP GOP structures

Different sequences are used with resolutions 1080p, 900p and 720p.

#### H.265/HEVC Anchors

##### 6.2.8.2.1 Overview

Table 6.2.8.2.1-1 provides an overview of the H.265/HEVC anchor tuples. Keys are identified to refer to the anchors in the context of the scenario.

Table 6.2.8.2.1-1 Anchor Tuple generation with H.265/HEVC for Full HD Scenario

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Key | Clause | Reference Sequence | Reference Encoder | Configuration | Variations |
| S1-A1-265 | 6.2.8.3.2 | S1-R1 | JM19.0 | HM-01 | QP |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

##### S1-A1-265

<Explanation what this anchor is important for>

### 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 = {10,20,30,40} Mbps [49]  CBR and capped-VBR | B = {10,20,30,40} Mbps [49]  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

The Annex C.3 describes in details the selection process conducted to build the test sequences considered for this scenario, the outcome of this process is reported in the table below.

Table 6.3-3 : Selected test material for 4K-TV scenario

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Name | Resolution | Frame-rate | ColorGamut | Number of Frames / Duration | Scene Cut |
| 4EVER\_Brest\_Sedof | 3840 x 2160 | 60 | BT.2020 | 600 | 0 |
| Rain\_fruits | 3840 x 2160 | 50 | BT.709 | 600 | 0 |
| ParkJoy | 3840 x 2160 | 50 | BT.709 | 500 | 0 |
| Indoor Soccer | 3840 x 2160 | 23.98 | BT.709 | 17s | 4 |
| Netflix\_ToddlerFountain | 4096 x 2160 | 59.94 | BT.709 | 1199 | 0 |
| Netflix\_TunnelFlag | 4096 x 2160 | 59.94 | BT.709 | 600 | 0 |
| RiverBank | 3840 x 2160 | 50 | BT.709 | 12s | 0 |
| Netflix\_Boat | 4096 x 2160 | 59.94 | BT.709 | 300 | 0 |

### 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 3: 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 rely on screen content types such as videoconferencing and webinars with presentation slides displayed, or remote desktop applications for demos or remote assistance. These services include massive online videoconferencing systems and telepresence systems both being particularly popular with remote collaboration.

For telepresence with 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.

As an example of hardware implementation, the Intel Xe LP™ chipset supports HEVC screen content profile (see slide 101 of [41]).

### 6.4.2 Description of the Anticipated Application

3GPP until now has very restricted set of services but based on the considerations in clause 6.4.1, 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

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

The considered scenario is low-latency streaming or conversational. 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. Typically, it means the ability to have non perceptible intra refreshes and the ability to maintain stability on low frequency areas (such as uniform backgrounds) as well as maintaining details on high frequencies (particularly for text)

- 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

### 6.4.3 Source Format Properties

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

Table 6.4-1 Screen Content Sharing source properties

|  |  |
| --- | --- |
| Source format properties | Screen Content |
| Spatial resolution | 1920 x 1080 |
| Chroma format | Y’CbCr, RGB |
| Chroma subsampling | 4:2:0, 4:4:4 |
| Picture aspect ratio | 16:9 |
| Frame rates | 25, 30, 50, 60 Hz |
| Bit depth | 8 |
| Colour space formats | BT.709 |
| Transfer characteristics | BT.709 |

### 6.4.4 Screen content codec tools

Table 6.4-2 provides an overview of the different codec tools per profile that may be suitable for coding screen content sequences.

Table 6.4-1 Screen Content Tools per Profile

|  |  |  |
| --- | --- | --- |
| **Screen content tools** | **AVC** | **HEVC** |
| main profile | Not applicable | Transform skip |
| range extension profile | not applicable | Residual Differential Pulse Code Modulation (RDPCM) (implicit intra/explicit inter), |
| screen content profile | Not applicable | Intra Block Copy (full frame or less), Palette, Adaptive Color Transform |

Note 1: HEVC specifies nested profiles. The screen content profile contains all the screen content tools from the main and range extension profiles.

Note 2: Tools specific results could be gathered from JCT-VC.

### 6.4.5 Encoding and Decoding Constraints

Table 6.4-3 provides an overview of encoding and decoding constraints for Screen Content scenario using AVC and HEVC codecs operating points. This will support the definition of detailed test conditions.

Table 6.4-3 Encoding and Decoding Configurations for 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  H.265/HEVC Screen Content profile  Level 4.1, Level 5.1, Level 6.1 |
| **RAP period** | 1 second, infinite | 1 second, infinite |
| **Bit rate parameters (CBR, VBR, CAE, HRD parameters)** | Fixed QP | Fixed QP |
| **Latency requirements and specific encoding settings** | Low-latency requirements, Low-delay configuration | Low-latency requirements,  Low-delay B, low-delay-P |
| **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  H.265/HEVC Screen Content profile  Level 4.1, 5.1, 6.1 [8] |

### 6.4.6 Performance Metrics

Performance is assessed using BD-Rate computation, with PSNR metrics as objective quality criterion.

Note: other metrics suitable for SCC content are TBD.

### 6.4.7 Interoperability Considerations

tbd

### 6.4.8 Test Sequences

### 6.4.8 Test Sequences

The test sequences illustrating the screen content scenario are described in Annex C. They contain either synthetic content from a presentation such as a slide deck with text and graphics.

### 6.4.9 Detailed Test Conditions

#### 6.4.9.1 Overview

The screen content configuration relies on low delay encoding modes with the optional use of Screen Content encoding tools when possible (foe HEVC).

#### 6.4.9.2 Reference Software H.264 (AVC)

The bitstreams are generated with the H.264 (AVC) reference software called the JM (Joint model) in version 19.0.

- <http://iphome.hhi.de/suehring/tml/download/>

The encoder configuration file is:

- <https://vc-git.hhi.fraunhofer.de/jct-vc/JM/-/blob/master/cfg/encoder_main.cfg>

With the following adaptations:

- ProfileIDC = 77 (Main Profile)

- LevelIDC = 42 (for 1080p60) : 52 (for 4Kp60)

- IntraPeriod = 0 (only first) or 64 (approaching 1s for 60fps)

- IDRPeriod = [Same as IntraPeriod]

- QPISlice = QPPSlice = [17, 22, 27, 32, 37, 42]

- SearchRange = 64 for fullHD ; 128 for 4K

- NumberOfReferenceFrames = 2

- NumberBFrames = 0 [TBC]

- SearchMode = 0 (fast full search)

#### 6.4.9.3 Reference Software H.265 (HEVC) Main 10 Profile

To generate the anchor bitstream, the following HEVC reference software was used:

- <https://vcgit.hhi.fraunhofer.de/jct-vc/HM/-/tags/HM-16.22>

The settings are defined as follow:

- https://vcgit.hhi.fraunhofer.de/jct-vc/HM/-/blob/HM-16.22/cfg/encoder\_lowdelay\_P\_main10.cfg

Each source sequence is encoded with the following changes:

- QP: [22, 27, 32, 37]

Editor’s Note: QP values may need to be refined

- For the intra period, two configurations are tested

- IntraPeriod with no fix interval

- GOPSize is equal to 8. Each P picture refers to up to 4 preceding pictures in display order

- IntraQPOffset is -1. P picture QP offsets are alternatively equal to 4, 5, and are set to 1 every 8 pictures.

- IntraPeriod such that 1 second is achieved

- DecodingRefreshType: (2) IDR

- IntraQPOffset and QPoffsets are set equal to 0

- Each P picture refers to immediately preceding pictures in display order.

#### 6.4.9.4 Reference Software H.265 (HEVC) SCC Profile

To generate the anchor bitstream, the following reference software for HEVC Screen Content profile was used:

SCM-8.8: https://vcgit.hhi.fraunhofer.de/jct-vc/HM/-/tags/HM-16.21+SCM-8.8

The settings are defined as follow:

- <https://vcgit.hhi.fraunhofer.de/jct-vc/HM/-/blob/HM-16.21+SCM-8.8/cfg/encoder_lowdelay_main_scc.cfg>

Each source sequence is encoded with

- QP: [22, 27, 32, 37]

Editor’s Note: QP values may need to be refined

- For the intra period, two configurations are tested:

- IntraPeriod with no fix interval

- The settings are defined in the attached file “encoder\_lowdelay\_P\_main\_scc\_gop8”:

- GOPSize is equal to 8. Each P picture refers to up to 4 preceding pictures in display order

- IntraQPOffset is -1. P picture QP offsets are alternatively equal to 4, 5, and are set to 1 every 8 pictures.

- IntraPeriod such that 1 second is achieved

- The settings are defined in the [soon to be provided] attached file “encoder\_lowdelay\_P\_main\_constant\_qp\_scc\_gop8”.

- DecodingRefreshType: (2) IDR

- IntraQPOffset and P pictures QPoffsets are set equal to 0

- Each P picture refers to immediately preceding pictures in display order.

### 6.4.8 Test Sequences

The test sequences illustrating the screen content scenario are described in Annex C. They contain either synthetic content from a presentation such as a slide deck with text and graphics.

### 6.4.9 Detailed Test Conditions

#### 6.4.9.1 Overview

The screen content configuration relies on low delay encoding modes with the optional use of Screen Content encoding tools when possible (foe HEVC).

#### 6.4.9.2 Reference Software H.264 (AVC)

The bitstreams are generated with the H.264 (AVC) reference software called the JM (Joint model) in version 19.0.

- <http://iphome.hhi.de/suehring/tml/download/>

The encoder configuration file is:

- <https://vc-git.hhi.fraunhofer.de/jct-vc/JM/-/blob/master/cfg/encoder_main.cfg>

With the following adaptations:

- ProfileIDC = 77 (Main Profile)

- LevelIDC = 42 (for 1080p60) : 52 (for 4Kp60)

- IntraPeriod = 0 (only first) or 64 (approaching 1s for 60fps)

- IDRPeriod = [Same as IntraPeriod]

- QPISlice = QPPSlice = [17, 22, 27, 32, 37, 42]

- SearchRange = 64 for fullHD ; 128 for 4K

- NumberOfReferenceFrames = 2

- NumberBFrames = 0 [TBC]

- SearchMode = 0 (fast full search)

#### 6.4.9.3 Reference Software H.265 (HEVC) Main 10 Profile

To generate the anchor bitstream, the following HEVC reference software was used:

- <https://vcgit.hhi.fraunhofer.de/jct-vc/HM/-/tags/HM-16.22>

The settings are defined as follow:

- https://vcgit.hhi.fraunhofer.de/jct-vc/HM/-/blob/HM-16.22/cfg/encoder\_lowdelay\_P\_main10.cfg

Each source sequence is encoded with the following changes:

- QP: [22, 27, 32, 37]

Editor’s Note: QP values may need to be refined

- For the intra period, two configurations are tested

- IntraPeriod with no fix interval

- GOPSize is equal to 8. Each P picture refers to up to 4 preceding pictures in display order

- IntraQPOffset is -1. P picture QP offsets are alternatively equal to 4, 5, and are set to 1 every 8 pictures.

- IntraPeriod such that 1 second is achieved

- DecodingRefreshType: (2) IDR

- IntraQPOffset and QPoffsets are set equal to 0

- Each P picture refers to immediately preceding pictures in display order.

#### 6.4.9.4 Reference Software H.265 (HEVC) SCC Profile

To generate the anchor bitstream, the following reference software for HEVC Screen Content profile was used:

SCM-8.8: https://vcgit.hhi.fraunhofer.de/jct-vc/HM/-/tags/HM-16.21+SCM-8.8

The settings are defined as follow:

- <https://vcgit.hhi.fraunhofer.de/jct-vc/HM/-/blob/HM-16.21+SCM-8.8/cfg/encoder_lowdelay_main_scc.cfg>

Each source sequence is encoded with

- QP: [22, 27, 32, 37]

Editor’s Note: QP values may need to be refined

- For the intra period, two configurations are tested:

- IntraPeriod with no fix interval

- The settings are defined in the attached file “encoder\_lowdelay\_P\_main\_scc\_gop8”:

- GOPSize is equal to 8. Each P picture refers to up to 4 preceding pictures in display order

- IntraQPOffset is -1. P picture QP offsets are alternatively equal to 4, 5, and are set to 1 every 8 pictures.

- IntraPeriod such that 1 second is achieved

- The settings are defined in the [soon to be provided] attached file “encoder\_lowdelay\_P\_main\_constant\_qp\_scc\_gop8”.

- DecodingRefreshType: (2) IDR

- IntraQPOffset and P pictures QPoffsets are set equal to 0

- Each P picture refers to immediately preceding pictures in display order.

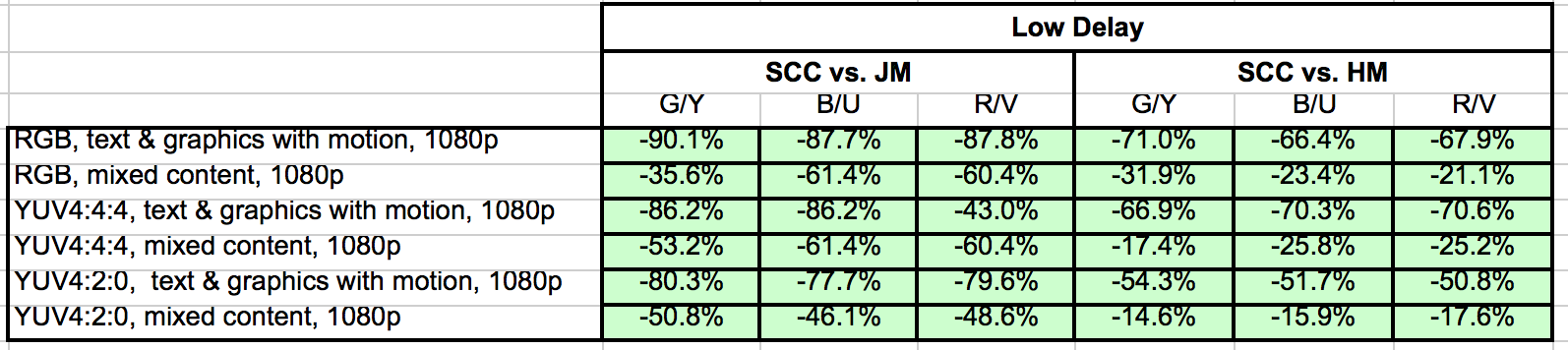
### 6.4.10 External Performance Data

In JCTVC-AA1006 [40], verification test report for HEVC screen content coding extensions, “the coding performance of HEVC screen content model (SCM) reference software with the SCC extensions enabled is compared with that of HEVC test model (HM) without the SCC extensions, as well as with the AVC joint model (JM) reference software in both lossy and mathematically lossless compression modes using All-Intra (AI), Random Access (RA), and Low-delay B (LB) encoding structures and using similar encoding rate-distortion optimization techniques.

BD-rate savings of SCM over JM and SCM-w/o-SCC for RGB, YUV 4:4:4, and YUV 4:2:0 colour sampling formats were calculated by using the actual 648 test points that were used in the subjective testing across 6 JCT-VC sequences.

The document contains details on sequences, test parameters and configuration files, as well as subjective results. In the context of this study, only the objective results for lossy Low-delay B are reported in Table 6.4.10-1.

Table 6.4.10-1: BD-rate savings of SCM over JM and HM for LB coding configuration



### 6.4.11 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

All metrics as defined in clause 5.5 apply.

### 6.5.6 Interoperability Considerations

tbd

### 6.5.7 Selected Reference Sequences

Table 6.5.7-1 provides the selected reference sequences for this scenario. Keys are identified to refer to the sequences in the context of the scenario. The sequences are named and a reference to the details of the sequence is provided. A justification is provided, why this sequence is selected.

Table 6.5.7-1 Selected Reference Sequences for Social sharing and messaging

|  |  |  |  |
| --- | --- | --- | --- |
| Key | Name | Reference | Justification/Comment |
| S4-R1 | tbd | Annex C.X |  |
| S4-R2 | tbd | Annex C.X |  |
| S4-R3 | tbd | Annex C.X |  |
| S4-R4 | tbd | Annex C.X |  |
| S4-R4 | tbd | Annex C.X |  |
| S4-R4 | tbd | Annex C.X |  |

### 6.5.8 Anchor Generation Details

#### 6.5.8.1 Overview

This clause provides details on how to generate the anchors for the Social sharing and messaging Scenario.

#### 6.5.8.2 H.264/AVC Anchors

##### 6.5.8.2.1 Overview

Table 6.5.8.2.1-1 provides an overview of the H.264/AVC anchor tuples. Keys are identified to refer to the anchors in the context of the scenario.

Table 6.5.8.2.1-1 Anchor Tuple generation with H.264/AVC for Social sharing and messaging

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Key | Clause | Reference Sequence | Reference Encoder | Configuration | Variations |
| S4-A1-264 | 6.5.8.2.2 | S4-R1 | JM19.0 | JM-01 | QP = [17, 20, 23, 26, 29, 32] |
| S4-A2-264 | 6.5.8.2.2 | S4-R2 | JM19.0 | JM-02 | QP = [17, 20, 23, 26, 29, 32] |
| S4-A3-264 | 6.5.8.2.2 | S4-R3 | JM19.0 | JM-03 | QP = [17, 20, 23, 26, 29, 32] |
|  |  |  |  |  |  |

##### S4-A1-264 One second intra

This anchor tuple produces an anchor over 60 seconds with a typical configuration of H.264/AVC:

* Profile H.264/AVC Progressive High-Profile Level 4.0 [7]
* Random access at 1 second interval
* Fixed QP settings
* IPPPP structures

#### H.265/HEVC Anchors

##### 6.5.8.3.1 Overview

Table 6.5.8.3.1-1 provides an overview of the H.265/HEVC anchor tuples. Keys are identified to refer to the anchors in the context of the scenario.

Table 6.5.8.3.1-1 Anchor Tuple generation with H.265/HEVC for Social sharing and messaging Scenario

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Key | Clause | Reference Sequence | Reference Encoder | Configuration | Variations |
| S1-A1-265 | 6.2.8.3.2 | S1-R1 | JM19.0 | HM-01 | QP |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

##### 6.5.8.3.2 S4-A1-265

<Explanation what this anchor is important for>

### 6.5.9 External Performance Data

tbd

### 6.5.10 Additional Information

Tbd

# 6.6 Scenario 5: Online Gaming

### 6.6.1 Motivation

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 for HMD-based consumption was discussed and introduced in detail in TR 26.928 [6]. 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 APIs 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]. As another 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;

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.

For games rendered on 2D displays such as smart phones, 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

### 6.6.2 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 2D displays:

- 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 online gaming scenario is low-latency streaming, typically using UDP/IP based distribution to minimize protocol latencies. 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 traditional computer-generated content.

- The ability to compress photorealistic computer-generated content.

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

- Considered settings for encoding:

- Ultra low latency and Low-latency settings

- No specific error resilience mechanisms

- Encoding in this scenario is typically done as

- Real-time encoding

- Cloud-based encoding

### 6.6.3 Source Format Properties

#### 6.6.3.1 Introduction

Video games have different characteristics that are important to take into account when encoding the rasterized frames produced by the game engine. In TR 26.928 [6], clause 4.2.2, a few different types of games and their interaction delay tolerance are documented. However, TR 26.928 [6] does not differentiate the characteristics of the content. This aspect is addressed in the following.

In particular, the following characteristics are important:

- Dynamicity of content: how frequent rasterized frames change when compared to previous frame

- Complexity of content: how much content changes between frames and how complex such changes are

- Type of content: traditional CGI, photo-realistic CGI or natural images/video

Depending on these characteristics as well as the interaction delay tolerance, video games can be organized into different categories as document in the remainder of this clause.

#### 6.6.3.2 Category A: Low/medium dynamicity with low/medium complexity.

This category includes games such as board games, turn-by-turn strategy games, management/simulation games or non-realtime role-playing games (RPG) in which content may not change over several consecutive frames and changes are typically limited.

This category also includes games such as adventure games, casual games, or platform games in which although content may change at every single frame, changes are limited to animation of sprites or simple global movements of the content.

The common characteristics of the games in this category is that their playability can support longer interaction delay tolerance (500 – 1000ms according to TR 26.928 [6]) and their content is typically considered to video encode.

#### 6.6.3.3 Category B: games with high dynamicity and low/medium complexity.

This category includes games such as fighting games, racing games, real-time strategy (RTS) games or real-time RPGs in which content is very dynamic but changes are either limited or simple transforms.

The common characteristics of the games in this category is that their playability requires shorter interaction delay tolerances (100ms according to TR 26.928 [6]) while their content is still considered simple to video encode (with high benefits from prediction coding).

#### 6.6.3.4 Category C: games with high dynamicity and high complexity.

This category includes games such as first-person shooters (FPS), Massive Multiplayer Online (MMO) games and racing games in which content is very dynamic with possibly very significant changes regularly in the content.

The common characteristics of the games in this category is that their playability requires shorter interaction delay tolerances (100ms according to TR 26.928) and their content is typically considered as complex content to video encode.

#### 6.6.3.5 Category D: photo-realistic games or games based on natural images/video.

The main characteristics of the games in this category is that their content is typically considered as more complex content to video encode.

#### 6.6.3.6 Category E: XR game content

For a detailed analysis, please refer to clause 6.6.1 and TR26.928, clause 4.4.

#### 6.6.3.7 Summary

Table 6.6.3.7-1 provides an overview of the different source signal properties for Online Gaming. This information is used to select proper test sequences.

Table 6.6.3.7-1 Online Gaming source properties

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source format properties | Category A  low/med dynamicity & low/med complexity games | Category B  high dynamicity & low/med complexity games | Category C  high dynamicity & high complexity games | Category D  photo-realistic or natural video games | Category E  XR games |
| Spatial resolution | 1280x720, 1920x1080, 3840x2160 | 1280x720, 1920x1080, 3840x2160 | 1280x720, 1920x1080, 3840x2160 | 1280x720, 1920x1080, 3840x2160 | 1920x1080, 2048x1024, 3840x2160, 4096x2048, 8192x4096 |
| Chroma format | Y’CbCr | Y’CbCr | Y’CbCr | Y’CbCr | Y’CbCr |
| Chroma subsampling | 4:2:0, 4:4:4 | 4:2:0, 4:4:4 | 4:2:0, 4:4:4 | 4:2:0, 4:4:4 | 4:2:0, 4:4:4 |
| Picture aspect ratio | 16:9 | 16:9 | 16:9 | 16:9 | 16:9; 2:1 |
| Frame Buffers | 1 | 1 | 1 | 1 | 2 (1 per eye) |
| Frame rates | 30, 50, 60 Hz | 30, 50, 60, 90, 120 Hz | 30, 50, 60, 90, 120 Hz | 30, 50, 60, 90, 120 Hz | 30, 50, 60, 90, 120 Hz |
| Bit depth | 8 | 8 | 8, 10 | 8, 10 | 8, 10 |
| Colour space formats | BT.709, BT.2020 | BT.709, BT.2020 | BT.709, BT.2020 | BT.709, BT.2020 | BT.709, BT.2020 |
| Transfer characteristics | N/A | N/A | BT.2100 (HDR) | BT.2100 (HDR) | BT.2100 (HDR) |

### 6.6.4 Codec and compression tools for games

Table 6.6.4-1 provides an overview of the different coding tools per profile that may be suitable for coding different types of gaming sequences.

Table 6.6.4-1 Gaming Codec Tools per Profile

|  |  |  |
| --- | --- | --- |
| **Profile** | **AVC** | **HEVC** |
| **main profile** | **NA** | TBD |
| **range extension profile** | NA | TBD |
| **screen content profile** | NA | Palette, TBD |

Note: Palette being heavily used in the design of games, this tool should be particularly tested for game content.

For photo-realistic games (Category D), certain "screen content" tools may not provide much benefits for this type of game content and would need to be tested.

### 6.6.5 Encoding and Decoding Constraints

Table 6.6.5-1 provides an overview of encoding and decoding configurations for the online gaming scenario using AVC and HEVC codecs as operation points. This will support the definition of detailed test conditions.

Table 6.6.5-1 Encoding and Decoding Configurations for Online Gaming

|  |  |  |
| --- | --- | --- |
| **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, infinite | 1 second, infinite |
| **Bit rate parameters (CBR, VBR, CAE, HRD parameters)** | Fixed QP | Fixed QP |
| **Latency requirements and specific encoding settings** | Low-latency requirements  Low-delay configuration  With and without use of future reference frames | Low-latency requirements,  Low-delay-P  Low-delay-B  With (moderate latency to non-critical latency games) and without (ultra low and low latency game) use of future reference frames. |
| **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 4.1, 5.1, 6.1 [8] |

NOTE: Depending on the type of game (First person, Massive Multi-Player Online Role Play Game, Strategy) the latency requirements differs, usually ranging from less than 50ms to more than 200ms, thus leading to different encoding constraints and configuration.

[Table 6.6.5-2 provides an overview of latency constraints for each game category and identifies which coding tools are applicable and potentially useful for each category.

**Table 6.6.5-2 Latency constraints tools for different game categories**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Latency Constraints** | **Category A** | **Category B** | **Category C** | **Category D** | **Category E** |
| **Low latency requirement** | optional | Yes | Yes | Yes/No | Yes |
| **Low delay P** | optional | required | required | optional | required |
| **Low delay B** | optional | required | required | optional | required |
| **Use of future reference frames** | optional | No | No | optional | No |

]

### 6.6.6 Performance Metrics

tbd

### 6.6.7 Interoperability Considerations

tbd

### 6.6.8 Test Sequences

[Test sequences illustrating the Online Gaming scenario should ideally contain synthetic content with detailed textures and realistic movements. This should include FPS, RPG and Strategy games.]

For candidate test sequences, please refer to Annex C.2.

### 6.6.9 Anchors generation

#### 6.6.9.1 Overview

#### 6.6.9.2 Reference Software AVC

#### 6.6.9.3 Reference Software HEVC Main 10: HM16.22

To generate the anchor bitstreams, the following HEVC reference software is used:

- https://vcgit.hhi.fraunhofer.de/jct-vc/HM/-/tags/HM-16.22

The settings are defined as follow:

- https://vcgit.hhi.fraunhofer.de/jct-vc/HM/-/blob/HM-16.22/cfg/encoder\_lowdelay\_main10.cfg

Each source sequence is encoded with the following changes:

- QP: [22, 27, 32, 37]

Editor’s Note: QP values may need to be refined

- For the intra period, two configurations are tested

- IntraPeriod with no fix interval

- GOPSize is equal to 8. Each B picture refers to up to 4 preceding pictures in display order

- IntraQPOffset is -1. B picture QP offsets are alternatively equal to 4, 5, and are set to 1 every 8 pictures.

- IntraPeriod such that 1 second is achieved

- DecodingRefreshType: (2) IDR

- IntraQPOffset and B pictures QPoffsets are set equal to 0

- Each B picture refers to immediately preceding pictures in display order.

### 6.6.10 External Performance Data

tbd

### 6.6.11 Additional Information

tbd

# 7 Metrics and Characterization for Existing Codecs

## 7.1 Introduction

## 7.2 H.264/AVC Anchor Metrics

Document the anchors and results for all scenarios for H.264/AVC.

## 7.3 H.265/HEVC Anchor Metrics

Document the anchors and results for all scenarios for H.265/HEVC.

## 7.4 H.265/HEVC Characterization against H.264/AVC

Do a first characterization, also to show the how the results are preferably presented.

# 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

## 9.1 Introduction

This clause 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 based on the characterization results provided for example by ISO/IEC SC29 WG11 (MPEG)/JVET.

Editor’s Note: It is expected that the preliminary coding performances documented in the following clauses for VCC and EVC will be updated by the latest results from JVET and MPEG when made available.

## 9.2 Versatile Video Coding (VVC)

### 9.2.1 Overview

ITU-T SG 16 informed that the development of the new Versatile Video Coding (VVC) standard has been completed for Consent at the June-July 2020 meeting of ITU-T SG16 and is also thereafter undergoing final approval in ISO/IEC JTC 1/SC 29/WG 11 (MPEG). After final approval, the dual reference to the jointly developed VVC standard will be Rec. ITU-T H.266 | ISO/IEC 23090-3 [38].

VVC is the latest in a series of such jointly developed ITU-T Recommendations and International Standards for video coding, and is the direct successor to HEVC (Rec. ITU-T H.265 | ISO/IEC 23008-2) [8] and AVC (Rec. ITU-T H.264 | ISO/IEC 14496-10) [7]. VVC provides significant coding efficiency improvements over HEVC. Plans are underway to conduct a verification test with formal subjective testing to confirm that VVC achieves about a 50% bit rate reduction vs. HEVC for equal subjective video quality. Test results demonstrate that VVC provides about a 40% bit rate reduction for 4K/UHD test sequences using objective metrics. Application areas especially targeted for the use of VVC include ultra-high definition 4K and 8K video, video with a high dynamic range and wide colour gamut, and video for immersive media applications such as 360° omnidirectional video, as well as conventional standard-definition and high-definition video content. In addition to improving coding efficiency, VVC also provides highly flexible syntax supporting such use cases as subpicture bitstream extraction, bitstream merging, and layered coding scalability.

The VVC standard includes the specification of a "Main 10" profile that supports 8- and 10-bit 4:2:0 video, a "Main 10 4:4:4" profile with 4:4:4 and 4:2:2 format support, corresponding "Multilayer Main 10" and "Multilayer Main 10 4:4:4" profiles with support for layered coding, and "Main 10 Still Picture" and "Main 10 4:4:4 Still Picture" profiles for still image coding employing the same coding tools as in the corresponding video profiles.

ITU-T also announced AAP Consent for Rec. ITU-T H.274 “Versatile supplemental enhancement information for coded video bitstreams” (VSEI), developed jointly as twin text with ISO/IEC JTC 1/‌SC 29/‌WG 11 as ISO/IEC 23002-7 [39]. VSEI specifies the syntax and semantics of video usability information (VUI) parameters and supplemental enhancement information (SEI) messages for use with coded video bitstreams. VSEI is particularly intended for use with VVC, although it is drafted in a manner intended to be sufficiently generic that it may also be used with other types of coded video bitstreams.

## 9.3 Essential Video Coding (EVC)

### 9.3.1 Overview

The development of the MPEG-5 Essential Video Coding (EVC) standard is completed and its specification has been published in October 2020 as ISO/IEC 23094-1 [X].

The main goal of the EVC standard is to provide significantly improved compression capability over previous generation of video coding standards with timely publication of commercial terms. The EVC standard has been developed to provide a video codec for emerging delivery protocols and networks, such as 5G, enabling the delivery of high-quality video services to an ever-growing audience by providing improved coding performance.

The MPEG-5 EVC defines two important profiles, "Baseline" profile and "Main" profile. It was the design objective that the "Baseline" profile contains only technologies that are older than 20 years. The "Main" profile includes additional tools, each of which can be either cleanly disabled or switched to the corresponding baseline tool on an individual basis. Additionally, for still image coding, "Main Still Picture" and "Baseline Still Picture" profiles, which employ the same coding tools as in the corresponding video profiles, are defined.

MPEG is currently conducting verification tests [W] with formal subjective testing to confirm that the EVC Main profile achieves significant bit-rate reductions vs. HEVC for equal subjective video quality for SDR content and that the EVC Baseline profile achieves similar bit-rate reductions vs. AVC for equal subjective video quality for SDR content. The verification tests have been completed for HDR content and the results are reported in [V]. They show that the EVC Main profile provides around 36% of bitrate reduction for HDR content at UHD resolution and 35% for HDR content at HD resolution for equal subjective quality.

Editors’note: This will be updated when full subjective test results are released by MPEG and will be documented as external reference.

Application areas especially targeted for the use of EVC include ultra-high definition 4K and 8K video, video with a high dynamic range and wide colour gamut, and video for immersive media applications such as 360° omnidirectional video, as well as conventional standard-definition and high-definition video content.

# 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

## [B.1 Introduction

This Annex provides a detailed overview on performance metrics that are potentially relevant for the evaluation of video codecs.

## B.2 Objective metrics for evaluation of video coding efficiency experiments

The ITU-T Technical Paper [40] provides a description of Bjøntegaard Delta rate (BD-rate) measurement practices for video coding experiments. The document provides a concept-level overview of recent practices and provides references to technical papers that describe further details. In addition, the document provides comments on why some of the choices were made and identifies situations where caution must be taken when interpreting the results.

To calculate the Bjøntegaard Delta bit rate, a distortion metric needs to be used. For standard-dynamic range video, the distortion metric primarily used has been the Peak Signal to Noise Ratio (PSNR). There are certainly some weaknesses to the PSNR-based BD-rate measure in terms of its correspondence with human perception of fidelity, but the use of PSNR-based BD-rate measures is the most prevalent in the video coding standardization community.

For high-dynamic range (HDR) material additional aspects that influence the usability of BD-rate calculations. For screen content material (i.e., material that has not been captured by a camera), no need for a special metric has been seen and PSNR-based BD-rate is used for this category.

According to [40], there are several steps in the BD-rate calculation process, where the result in each step is calculated from the result obtained in the previous step:

1. Calculation of PSNR for individual frames, see clause B.4 for details.
2. Calculation of per-sequence PSNR and bit rate values for each quantization parameter (QP) value. The QP value influences the resulting bit rate. Hence, compressing the sequence several times with different QPs ensures that the final BD-rate measurement will reflect the performance at many different bit rates. See clause B.4 for details.
3. Calculation of BD-rate values, for details see clause B.5.

In clause B.3, a reference file format to store raw elementary streams are introduced. These formats can be used together with the sequence quality evaluation tool defined in clause B.4.

An explanation to the collected tools are provided in Annex B.6.

<Editor’s Note: Additional metrics such as VMAF or SSIM may be calculated and also used for BD-rate evaluation>

## B.3 Uncompressed Video in ISOBMFF

### B.3.1 Overview

Raw video formats follow the concepts developed in TR 26.902 [45]. The raw video sequences are stored in a format such that display timing is maintained. The 3GP file format as defined in TS 26.244 [46] was viewed as most appropriate.

Attached to this document is a package *ISOFileFormatConverter\_711.zip* which allows conversion of yuv video files in different formats to and from 3GPP/ISO BMFF file format. This package contains the relevant library as well as sample applications.

Raw video is stored as samples in a video track in ISO Base Media Format family files (such as MP4, 3GP and so on). That is, it uses the same video handler, video media header, etc., as a video track containing, for example, H.264/AVC. Raw video can take various formats - based on choice of colour model, sub-sampling, and so on. As is usual in ISO files, the format of the video (the 'decoder' needed) is declared by the sample entry 4-character-code.

This encapsulation format adds boxes to the sample entry to parameterize the video.

Based on the requirements in this Technical Report, the following aspects of the raw data need to be provided:

1) Spatial Resolution width

2) Spatial Resolution height

3) Chroma Format

4) Chroma Sub-Sampling

5) Picture Aspect ratio

6) Frame rate

7) Bit depth

8) Color space format

9) Transfer characteristics

10) Duration of the sequence

In addition, a sample needs to be defined along with the storage format.

### B.3.2. ISO/IEC 23001-17 – Background and Application for 5G Video

<tbd>

## B.4 Sequence Quality Evaluation Tool

### B.4.1 Overview

The quality evaluation tool is an extension to the tool provided TR 26.902 [45]. However, channel impairment aspects described in [45] are not studied in this TR. However, channel impairment aspects described in [45] are not studied in this TR.

For quality evaluation, an archive *QualityEvaluation\_711.zip* is attached to this document. Note that some libraries in the package *ISOFileFormatConverter\_711.zip* are required.

Metrics are defined based on a script.

### B.4.2 Usage of the QE tool

The usage of the QE tool is as follows:

QualEval [OrigSeq in ISOBMFF format] [AnchorBitstream in ISOBMFF format] [EncodedSeq in CMAF format]

whereby

- OrigSeq corresponds to one the original sequence in ISO BMFF format as defined in clause B.3.

- Anchor Bitstream corresponds to the encoded bitstream in ISO BMFF format.

- EncodedSeq corresponds corresponds to the reconstructed sequence without transmission in ISO BMFF format as defined in clause B.3.

The output of the program are the following six values:

- The number of video frames in OrigSeq.

- The number of video frames in ReconSeq.

- The average PSNR\_Y, PSNR\_U and PSNR\_V as well as the weighted PSNR in dB.

- The PSNR of the average MSE\_Y, of the average MSE\_U and the average MSE\_V in dB as well as the PSNR of the weighted average MSE.

- The bitrate value.

- <other values should be added, SSIM and VMAF>

### B.4.3 What does the tool

Following the description in [44], the tool does the following for Standard-Dynamic Range (SDR) and Hybrid Log-Gamma (HLG):

- For each sample of the original sequence

- checks for the presented frame in the reconstructed sequence.

- If a frame with the same presentation time exists in the reconstructed sequence, this one is taken for comparison.

- If no frame with the same presentation time exists in the reconstructed sequence, the one which is closest in the past is taken.

- If the frame in the reconstructed frame does not have the same resolution as the original one, the reconstructed frame is up-sampled to match the original frame size.

- The mean square error, MSE\_Y, is calculated between the luma channel 𝑑𝑒𝑐𝑌 of the reconstructed output image and the luma channel 𝑜𝑟𝑖𝑔𝑌 of the original image according to [44], equation (7-1) with H the height and W the width.

- The luma PSNR value, PSNR\_Y, for the frame is then calculated according to [44], equation (7-2) with bitDepth identical to the signal bit depth.

- In addition, the MSE numbers for the two chroma, MSE\_U and MSE\_V, are calculated accordingly.

- In addition, the PSNR numbers for the two chroma, PSNR\_U and PSNR\_V, are calculated accordingly.

- For each sample of the encoded entire sequence, the following values are computed

- the sample size according to the sample table box for encoded value is extracted.

- For the entire sequence, the following values are computed

- The aggregate PSNR\_Y, PSNR\_U and PSNR\_V for a sequence are calculated as the average of the PSNR values for the individual frames according to [44], equation (7-3).

- The aggregate PSNR as a weighted average for each frame according to [40], equation (7-5) is calculated.

- The aggregate MSE\_Y, MSE\_U and MSE\_V for a sequence is calculated as the average of the MSE values. The resulting PSNRs of the aggregated MSE\_Y, MSE\_U and MSE\_V is calculated using the equation in [44], equation (7-2).

- The PSNR of the averaged MSE is calculated.

- The aggregate sample size in bytes is used and divided by the duration of the sequence.

For HDR sequences with PQ transfer characteristics, the following additional metrics are created as defined in [44], clause 8: PSNRL100, wPSNR and DE100.

## B.5 BD-rate numbers

### B.5.1 Introduction

The previous sections provide PSNR values and bit rate values for a single sequence encoding (using a single encoder configuration). For the BD-rate computation, two streams are needed

1) Anchor: the anchor refers to the baseline that is used to check the improvements for a tested configuration.

2) Test: test is the tested method under investigation.

BD-Delta provides the aggregated bitrate savings over a range of bitrates at the same quality. Any quality measure Q may be used as defined in B.4.

The following assumptions are taken:

- Several experiments have been carried out (for example for a different QP) for the anchor and the test, each resulting in a different bitrate results Ra[i], i=1,…, Na and Rt[j], j=1,…, Nt, respectively, as well as quality results Qa[i], i=1,…, Na and Qt[j], j=1,…, Nt,, respectively. Such results can be mapped to a single bitrate saving value.

- The results may be carried out with the same resolution for all encodings or for different resolutions (applying a bitrate ladder). The tool in B.4.3 takes care of appropriate scaling to the highest resolution.

### B.5.2 Per-sequence BD-rate

The quality-rate function for anchor and test results in a curve. The idea for a sequence is to estimate the area between these two curves to compute an average bit rate savings for equal measured quality. A detailed explanation of the how to do this is provided in [44], clause 7.3.

def BD\_Q(RA, QA, RT, QT, piecewise=0):

lRA = np.log(RA)

lRT = np.log(RT)

QA = np.array(QA)

QT = np.array(QT)

p1 = np.polyfit(lRA, QA, 3)

p2 = np.polyfit(lRT, QT, 3)

# integration interval

min\_int = max(min(lRA), min(lRT))

max\_int = min(max(lRA), max(lRT))

# find integral

if piecewise == 0:

p\_int1 = np.polyint(p1)

p\_int2 = np.polyint(p2)

int1 = np.polyval(p\_int1, max\_int) - np.polyval(p\_int1, min\_int)

int2 = np.polyval(p\_int2, max\_int) - np.polyval(p\_int2, min\_int)

else:

# See https://chromium.googlesource.com/webm/contributor-guide/+/master/scripts/visual\_metrics.py

lin = np.linspace(min\_int, max\_int, num=100, retstep=True)

interval = lin[1]

samples = lin[0]

v1 = scipy.interpolate.pchip\_interpolate(np.sort(lRA), QA[np.argsort(lRA)], samples)

v2 = scipy.interpolate.pchip\_interpolate(np.sort(lRT), QT[np.argsort(lRT)], samples)

# Calculate the integral using the trapezoid method on the samples.

int1 = np.trapz(v1, dx=interval)

int2 = np.trapz(v2, dx=interval)

# find avg diff

avg\_diff = (int2-int1)/(max\_int-min\_int)

return avg\_diff

### B.5.2 BD-Rate overall

Once the BD-rate values for a set of test sequences have been determined, they are typically combined using an arithmetic average. For details refer to [44], clause 7.4.

## B.6 Scripts and Tools

[Some baseline scripts are provided here: <https://bitmovin.com/vp9-vs-hevc-h265/>. A full tool will be made available here on github <https://github.com/haudiobe/5GVideo> and the ported to 3GPP repository]

]

Annex C  
Candidate Test Sequences

## C.1 Introduction

tbd

## C.2 Gaming Test Sequences

### C.2.1 Overview

A set of gaming sequences is described in this sub-clause C.2.

Table C.2.1-1 proposes a summary of the gaming sequences according to the categories defined in section 6.6.3.

Table C.2.1-1, Gaming sequences categorization.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Sequence description** | | | **Categories** | | | | |
| Game sequence title | Game type | Sequences Characteristics | A Low/medium dynamicity and complexity | B:  High dynamicity Low/medium complexity | C:  High dynamicity and complexity | D:  photo-realistic games games based on natural images/video | E:  XR game content  Note: no sequences in that category |
| Arena of Valor (AOV)  (Tencent) | multiplayer online battle arena | fix +  Smooth travelling. Small incrustation. Stable at 90%. One pop-up incrustation. | X |  |  |  |  |
| Baolei-Man  (Tencent) | online multiplayer first person shooter | Pursuit. Some irregular camera traveling. Small incrustation.  Stable at 99% |  | X |  |  |  |
| Baloei man 4k  (Tencent) | online multiplayer first person shooter | Pursuit. Some irregular camera traveling. Small incrustation. Stable at 99% |  | X |  |  |  |
| Baolei-Woman  (Tencent) | online multiplayer  first person shooter | Pursuit. Lots of irregular camera traveling. Small incrustation. Stable at 90% |  |  | X |  |  |
| Baolei-woman 4K  (Tencent) | online multiplayer  first person shooter | Pursuit. Lots of irregular camera traveling. Small incrustation. Stable at 90% |  |  | X |  |  |
| Jianling-Temple  (Tencent) | massive multiplayer online role-playing | Battle/pursuit. Lots of irregular camera traveling. Stable at 90% One small pop-up incrustation |  |  | X |  |  |
| Jianling-Beach  (Tencent) | massive multiplayer online role-playing | Battle/pursuit. Dark scene. Irregular camera traveling. Small incrustation. Stable at 99% |  | X |  |  |  |
| Heroes of the Storm part 1  (kingston) | a crossover multiplayer online battle arena | Fix, plus 2 smooth small travelling | X |  |  |  |  |
| Project CARS part 1  (kingston) | Car racing game | Realistic rendering. Constant smooth movement. Small incrustation. Stable at 90% |  |  |  | X |  |
| WoW part 2  (kingston) | multiplayer online role play | Dark scene, irregular camera traveling. Small incrustation. Stable at 90% | X |  |  |  |  |
| Minecraft  (twitch) | Casual game sandbox, brick construction | Graphic : Lots of irregular camera traveling. Very small incrustation. Stable at 100% | X |  |  |  |  |
| CS:GO  (twitch) | a multiplayer first-person shooter | 1/ first half of sequence is dark (300pict)  2/ fade (15 pict)  3/ action : first-person shooter | X (first part) |  | X (3rd part) |  |  |
| StarCraft  (twitch) | Real Time Strategy game | fix +  smooth travelling Small incrustation stable at 90%. |  | X |  |  |  |

### C.2.2 AOV

#### C.2.2.1 Introduction

Arena of Valor (AOV) ©Tencent Games – is a 3D MOBA (multiplayer online battle arena) game. Within a limited size map 2 groups of 5 vs 5 or 3 vs 3 players take part of a battle for which each player chooses a hero with particular power and abilities. The figure C.2.2-1 illustrates the gameplay of AOV with a screenshot of the AOV test sequence.



Figure C.2.2-1: Screenshot of AOV test sequence

#### C.2.2.2 Source sequence properties

The AOV test sequence has the following properties:

- Resolution: 1920 x1080

- Scan: Progressive

- Frame rate: 60 fps

- Bit depth: 8-bit

- Length: 600 frames (10s)

- YUV format: YUV 4:2:0

- Colour Components: Y’CbCr

- Colour space: ITU-R BT.709

#### C.2.2.3 Copyright information

The AOV video sequence is made available under the following copyright disclaimer.

*Copyright © 1998 - 2018 Tencent. All Rights Reserved.*

*This sequence and all intellectual property rights therein remain the property of Shenzhen Tencent Computer System Co. Ltd. and Tencent America LLC. This sequence may only be used for the purpose of developing, testing and promulgating technology standards (the Purpose). Shenzhen Tencent Computer System Co. Ltd. and Tencent America LLC. make no warranties with respect to this sequence and expressly disclaim any warranties regarding its fitness for any purpose. This disclaimer shall remain attached to this sequence except when this sequence is used for the Purpose.*

### C.2.3 Baolei-Man

#### C.2.3.1 Introduction

The *Baolei-Man* test sequence is an extract from a famous online multiplayer game in which many players play in a battle royale mode meaning that in survival mode, the last standing one is the winner. The figure C.2.3-1 shows a screenshot of the *Baolei-Man* test sequence.

Une image contenant herbe, intérieur, homme, table

Description générée automatiquement

Figure C.2.3-1: Screenshot of Baolei-Man test sequence

#### C.2.3.2 Source sequence properties

The Baolei-Man test sequence is available in two original versions (fullHD and 4K) with the following properties specified hereafter in Table C.2.3-1:

Table C.2.3-1: Baolei-Man sequences properties

|  |  |  |
| --- | --- | --- |
|  | 1080p | 2160p |
| Resolution | 1920x1080 | 3840x2160 |
| Scan | Progressive | Progressive |
| Frame Rate | 60 fps | 60 fps |
| Bit Depth | 8 | 10 |
| Length | 600 frames | 600 frames |
| YUV format | YUV 4:2:0 | YUV 4:2:0 |
| Color components | Y’CbCr | Y’CbCr |
| Colour space | BT.709 | BT.709 |

#### C.2.3.3 Copyright information

The Baolei-Man video sequences (in FullHD and 4K resolutions) are made available under the following copyright disclaimer.

*Copyright © 1998 - 2020 Tencent. All Rights Reserved.*

*This sequence and all intellectual property rights therein remain the property of Shenzhen Tencent Computer System Co. Ltd.. This sequence may only be used for the purpose of developing, testing and promulgating technology standards (the Purpose). Shenzhen Tencent Computer System Co. Ltd. make no warranties with respect to this sequence and expressly disclaim any warranties regarding its fitness for any purpose. This disclaimer shall remain attached to this sequence except when this sequence is used for the Purpose.*

### C.2.4 Baolei-Woman

#### C.2.4.1 Introduction

The *Baolei-Woman* test sequence is also an extract from the same game as for *Balei-Man*. The figure C.2.4-1 shows a screenshot of the *Baolei-Woman* test sequence.

Une image contenant télévision, vidéo, moniteur, pièce

Description générée automatiquement

Figure C.2.4-1: Screenshot of Baolei-Woman test sequence

#### C.2.4.2 Source sequence properties

Similarly to the *Baolei-Man* test sequence, *Baolei-Woman* is also available in two original versions (fullHD and 4K) with the following properties specified hereafter in Table C.2.4-1:

Table C.2.4-1: Baolei-Woman sequences properties

|  |  |  |
| --- | --- | --- |
|  | 1080p | 2160p |
| Resolution | 1920x1080 | 3840x2160 |
| Scan | Progressive | Progressive |
| Frame Rate | 60 fps | 60 fps |
| Bit Depth | 8 | 10 |
| Length | 600 frames | 600 frames |
| YUV format | YUV 4:2:0 | YUV 4:2:0 |
| Color components | Y’CbCr | Y’CbCr |
| Colour space | BT.709 | BT.709 |

#### C.2.4.3 Copyright information

The Baolei-Woman video sequences (in FullHD and 4K resolutions) are made available under the following copyright disclaimer.

*Copyright © 1998 - 2020 Tencent. All Rights Reserved.*

*This sequence and all intellectual property rights therein remain the property of Shenzhen Tencent Computer System Co. Ltd.. This sequence may only be used for the purpose of developing, testing and promulgating technology standards (the Purpose). Shenzhen Tencent Computer System Co. Ltd. make no warranties with respect to this sequence and expressly disclaim any warranties regarding its fitness for any purpose. This disclaimer shall remain attached to this sequence except when this sequence is used for the Purpose.*

### C.2.5 Jianling-Temple

#### C.2.5.1 Introduction

The *Jianling-Temple* test sequence is an extract from another famous massive multiplayer online role-playing game (MMORPG) having two gameplay modes Player-versus-Environment (PvE) and Player-versus-player (PvP). The figure C.2.5-1 is a screenshot of the *Jianling-Temple* test sequence.

Une image contenant herbe, extérieur, équitation, homme

Description générée automatiquement

Figure C.2.5-1: Screenshot of Jiangli-Temple test sequence

#### C.2.5.2 Source sequence properties

The *Jiangli-Temple* test sequence has the following properties:

- Resolution: 1920 x1080

- Scan: Progressive

- Frame rate: 60 fps

- Bit depth: 8-bit

- Length: 600 frames (10s)

- YUV format: YUV 4:2:0

- Colour Components: Y’CbCr

- Colour space: ITU-R BT.709

#### C.2.5.3 Copyright information

The *Jiangli-Temple* video sequence is made available under the following copyright disclaimer.

*Copyright © 1998 - 2020 Tencent & NCSOFT. All Rights Reserved.*

*This sequence and all intellectual property rights therein remain the property of Shenzhen Tencent Computer System Co. Ltd. and NCSOFT Co.Ltd.. This sequence may only be used for the purpose of developing, testing and promulgating technology standards (the Purpose). Shenzhen Tencent Computer System Co. Ltd. and NCSOFT Co.Ltd. make no warranties with respect to this sequence and expressly disclaim any warranties regarding its fitness for any purpose. This disclaimer shall remain attached to this sequence except when this sequence is used for the Purpose.*

### C.2.6 Jiangling-Beach

#### C.2.6.1 Introduction

The *Jianling-Beach* test sequence is another extract from the same MMORPG game as the *Jiangli-Temple* sequence. The figure C.2.6-1 is a screenshot of the *Jianling-Beach* test sequence.

Une image contenant ordinateur, portable, différent, assis

Description générée automatiquement

Figure C.2.6-1: Screenshot of Jiangli-Beach test sequence

#### C.2.6.2 Source sequence properties

The *Jiangli-Beach* test sequence has the following properties:

- Resolution: 1920 x1080

- Scan: Progressive

- Frame rate: 60 fps

- Bit depth: 8-bit

- Length: 600 frames (10s)

- YUV format: YUV 4:2:0

- Colour Components: Y’CbCr

- Colour space: ITU-R BT.709

#### C.2.6.3 Copyright information

The *Jiangli-Beach* video sequence is made available under the following copyright disclaimer.

*Copyright © 1998 - 2020 Tencent & NCSOFT. All Rights Reserved.*

*This sequence and all intellectual property rights therein remain the property of Shenzhen Tencent Computer System Co. Ltd. and NCSOFT Co.Ltd.. This sequence may only be used for the purpose of developing, testing and promulgating technology standards (the Purpose). Shenzhen Tencent Computer System Co. Ltd. and NCSOFT Co.Ltd. make no warranties with respect to this sequence and expressly disclaim any warranties regarding its fitness for any purpose. This disclaimer shall remain attached to this sequence except when this sequence is used for the Purpose.*

## C.2.7 Heroes of the Storm

### C.2.7.1 Introduction

Heroes of the Storm is a crossover multiplayer online battle arena video game developed and published by Blizzard Entertainment. The figure C.2.7.1-1 illustrates the gameplay of Heroes of the Storm with a screenshot of the test sequence provided by Kingston University, London and Technical University, Berlin. [42] [43]



Figure C.2.7.1-1: Screenshot of Heroes of the Storm test sequence

### C.2.7.2 Source sequence properties

The Heroes of the Storm part 1 test sequence provided by Kingston University, London and Technical University, Berlin has the following properties:

* Resolution: 1920 x1080
* Scan: Progressive
* Frame rate: 30 fps
* Bit depth: 8-bit
* Length: 900 frames. 300 first frames to be used.
* YUV format: YUV 4:2:0
* Colour Components: Y’CbCr
* Colour space: ITU-R BT.709

### C.2.7.3 Repository

The Heroes of the Storm video sequence part 1 provided by Kingston University, London and Technical University, Berlin is available at this location:

<https://kingston.app.box.com/v/GamingVideoSET>. Part 2

### C.2.7.4 Copyright information

The Heroes of the Storm video sequence provided by Kingston University, London and Technical University, Berlin is made available under the following copyright disclaimer:

<https://kingston.app.box.com/v/GamingVideoSET/file/289578680253>

## C.2.8 Project CARS

### C.2.8.1 Introduction

Project CARS is a racing video game series developed by Slightly Mad Studios and published by Bandai Namco Entertainment. The figure C.2.8.1-1 illustrates the gameplay of Project CARS with a screenshot of the test sequence provided by Kingston University, London and Technical University, Berlin. [42] [43]

****

Figure C.2.8.1-1: Screenshot of Project CARS test sequence

### C.2.8.2 Source sequence properties

The Project CARS sequence part 1 provided by Kingston University, London and Technical University, Berlin, has the following properties:

* Resolution: 1920 x1080
* Scan: Progressive
* Frame rate: 30 fps
* Bit depth: 8-bit
* Length: 900 frames. 300 first frames to be used.
* YUV format: YUV 4:2:0
* Colour Components: Y’CbCr
* Colour space: ITU-R BT.709

### C.2.8.3 Repository

The Project CARS video sequence part 1 provided by Kingston University, London and Technical University, Berlin is available at this location:

<https://kingston.app.box.com/v/GamingVideoSET>. Part 2

### C.2.8.4 Copyright information

The Project CARS video sequence provided by Kingston University, London and Technical University, Berlin is made available under the following copyright disclaimer:

<https://kingston.app.box.com/v/GamingVideoSET/file/289578680253>

## C.2.9 World of WarCraft: WoW

### C.2.9.1 Introduction

WoW is an multiplayer online role play developed and published by Blizzard Entertainment where players control a character avatar within a game world in third- or first-person view, exploring the landscape, fighting various monsters, completing quests, and interacting with non-player characters or other players. The figure C.2.9.1-1 illustrates the gameplay of WoW with a screenshot of the test sequence, which is a dark scene, provided by Kingston University, London and Technical University, Berlin. [42] [43]



Figure C.2.9.1-1: Screenshot of WoW test sequence

### C.2.9.2 Source sequence properties

The WoW test sequence part 2, provided by Kingston University, London and Technical University, Berlin, has the following properties:

* Resolution: 1920 x1080
* Scan: Progressive
* Frame rate: 30 fps
* Bit depth: 8-bit
* Length: 900 frames. 300 first frames to be used.
* YUV format: YUV 4:2:0
* Colour Components: Y’CbCr
* Colour space: ITU-R BT.709

### C.2.9.3 Repository

The WoW video sequence part 2 provided by Kingston University, London and Technical University, Berlin is available at this location:

<https://kingston.app.box.com/v/GamingVideoSET>. Part 2

### C.2.9.4 Copyright information

The WoW video sequence provided by Kingston University, London and Technical University, Berlin is made available under the following copyright disclaimer:

<https://kingston.app.box.com/v/GamingVideoSET/file/289578680253>

## C.2.10 Minecraft

### C.2.10.1 Introduction

The Minecraft test sequence is an extract from a famous a sandbox video game developed by Mojang Studios in which players explore a blocky, procedurally-generated 3D world with infinite terrain, and may discover and extract raw materials, craft tools and items, and build structures or earthworks. The figure C.2.10.1-1 shows a screenshot of the Minecrafttest sequence provided by DERF\_TWITCH.



Figure C.2.10.1-1: Screenshot of Minecraft test sequence

### C.2.10.2 Source sequence properties

The Minecraft test sequence provided by DERF\_TWITCH has the following properties:

* Resolution: 1920 x1080
* Scan: Progressive
* Frame rate: 60 fps
* Bit depth: 8-bit
* Length: 600 frames (10s)
* Color format: 4:2:0
* Colour Components: Y’CbCr
* Colour space: ITU-R BT.709

### C.2.10.3 Repository

The Minecraft sequence provided by DERTF-Twitch can be downloaded in y4m format at following URL: <https://media.xiph.org/video/derf/twitch/y4m/>

and can be converted into YUV format with ffmpeg by using the following exemplary command line:

ffmpeg -i CSGO.y4m CSGO.yuv

### C.2.10.4 Copyright information

The copyright under which the Minecraft video sequence from DERF\_Twitch is made available on that link:

*<https://media.xiph.org/video/derf/twitch/copyright.txt>*

## C.2.11 CS:GO

### C.2.11.1 Introduction

The CS:GO test sequence from Twitch is an extract from a is a multiplayer first-person shooter video game developed by Valve and Hidden Path Entertainment where two opposing teams, known as the Terrorists and the Counter-Terrorists, compete in game modes to complete objectives, such as securing a location to plant or defuse a bomb and rescuing or guarding hostages. The figure C.2.11.1-1 shows a screenshot of the CS:GOtest sequence provided by DERF-TWITCH.



Figure C.2.11.1-1: Screenshot of CS:GO test sequence

### C.2.11.2 Source sequence properties

The CS:GO test sequence provided by DERF-Twitch has the following properties:

* Resolution: 1920 x1080
* Scan: Progressive
* Frame rate: 60 fps
* Bit depth: 8-bit
* Length: 600 frames (10s)
* Color format: 4:2:0
* Colour Components: Y’CbCr
* Colour space: ITU-R BT.709

### C.2.11.3 Repository

The CS:GO sequence provided by DERF-Twitch can be downloaded in y4m format at following URL: <https://media.xiph.org/video/derf/twitch/y4m/>

and can be converted into YUV format with ffmpeg by using the following exemplary command line:

ffmpeg -i CSGO.y4m CSGO.yuv

### C.2.11.4 Copyright information

The copyright under which the CS:GO video sequence provided by DERF-Twitch is made available can be found on that link:

*<https://media.xiph.org/video/derf/twitch/copyright.txt>*

## C.2.12 StarCraft

### C.2.12.1 Introduction

The StarCraft test sequence from Twitch is an extract from a famous Real Time Strategy game developed and published by Blizzard Entertainment where players enter a tournament of 5 games and chose to play one of three alien categories with distinct characteristic and abilities. The figure C.2.12.1-1 shows a screenshot of the StarCrafttest sequence provided by DERF-Twitch.



Figure C.2.12.1-1: Screenshot of StarCraft test sequence

### C.2.12.2 Source sequence properties

The StarCraft test sequence provided by DERF-Twitch has the following properties:

* Resolution: 1920 x1080
* Scan: Progressive
* Frame rate: 60 fps
* Bit depth: 8-bit
* Length: 600 frames (10s)
* Color format: 4:2:0
* Colour Components: Y’CbCr
* Colour space: ITU-R BT.709

### C.2.12.3 Repository

The StarCraft sequence provided by Derf-Twitch can be downloaded in y4m format at following URL: <https://media.xiph.org/video/derf/twitch/y4m/>

and can be converted into YUV format with ffmpeg by using the following exemplary command line:

ffmpeg -i CSGO.y4m CSGO.yuv

### C.2.12.4 Copyright information

The copyright under which the StarCraft video sequence provided by DERF-Twitch is made available can be found on that link:

*<https://media.xiph.org/video/derf/twitch/copyright.txt>*

## C.3 4K-TV Test Sequences

### C.3.1 SDR Category

#### C.3.1.1 Candidate Test Sequences

In this Annex, a wide range of SDR test material is reported. The candidate sequences are extracted from test material used in video processing community. The following sections describe the test material in terms of format characteristics and copyright.

#### C.3.1.1.1 Ultra Video Group Sequences

The Ultra Video Group (UVG) sequences are publicly available, and can be used under CC BY-NC 3.0 license. The sequences are available for download at <https://github.com/ultravideo/UVG-4K-Dataset> and an overview is provided in the table below.

Table C.3-1 : UVG test material description

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Resolution | Frame-rate | ColorGamut | duration |
| CityAlley | 3840 x 2160 | 50 | BT.709 | 12s |
| FlowerFocus | 3840 x 2160 | 50 | BT.709 | 12s |
| FlowerKids | 3840 x 2160 | 50 | BT.709 | 12s |
| FlowerPan | 3840 x 2160 | 50 | BT.709 | 12s |
| RaceNight | 3840 x 2160 | 50 | BT.709 | 12s |
| RiverBank | 3840 x 2160 | 50 | BT.709 | 12s |
| SunBath | 3840 x 2160 | 50 | BT.709 | 6s |
| Twilight | 3840 x 2160 | 50 | BT.709 | 12s |

#### C.3.1.1.2 JVET Sequences

The JVET test material for UHD SDR sequences is composed by six sequences described in the table below. The licensing information referenced in the table is described below:

1. CC BY-NC-ND 4.0
2. Our dataset is available for free and academic research only, no commercial use. Please cite our paper (L. Song, X. Tang, W. Zhang, X. Yang, P. Xia, The SJTU 4K Video Sequence Dataset, the Fifth International Workshop on Quality of Multimedia Experience (QoMEX2013), Klagenfurt, Austria, July 3rd-5th, 2013) in any published work if you use those video sequences.
3. Please note that this test material remains the property of the Institute of Research & Technology b<>com and is solely provided “as is" to ITU-T SG16 Q6 and ISO/IEC JTC1/SC29/WG11, on a non-transferable and non-exclusive basis, and solely for the purpose of development, testing and promulgation of video coding standards. Any disclosure to third parties, exhibition, broadcast, distribution, public display or other exploitation of the test material is strictly prohibited.
4. The proposed test sequences in this document and all intellectual property rights therein remain the property of Huawei Technologies Co., Ltd., and National Engineering Research Center.

Table C.3-2 : JVET test material description

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Name | Resolution | Frame-rate | ColorGamut | Number of frames | License |
| Tango2 | 3840 x 2160 | 60 | BT.709 | 294 | See 1. |
| FoodMarket4 | 3840 x 2160 | 60 | BT.709 | 300 | See 1. |
| Campfire | 3840 x 2160 | 30 | BT.709 | 300 | See 2. |
| Catrobots1 | 3840 x 2160 | 60 | BT.709 | 300 | See 3. |
| Daylightroad2 | 3840 x 2160 | 60 | BT.709 | 300 | See 4. |
| Parkrunning3 | 3840 x 2160 | 50 | BT.709 | 300 | See 4. |

#### C.3.1.1.3 Netflix Sequences

The Netflix dataset is composed by eighteen 4K sequences described in the table below. The sequences can be used under the CC BY-NC-ND 4.0 license restriction.

Table C.3-3 : Netflix test material description

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Resolution | Frame-rate | ColorGamut | Number of frames |
| Netflix\_Aerial | 4096 x 2160 | 59.94 | BT.709 | 1199 |
| Netflix\_DinnerScene | 4096 x 2160 | 59.94 | BT.709 | 1199 |
| Netflix\_RollerCoaster | 4096 x 2160 | 59.94 | BT.709 | 1199 |
| Netflix\_DrivingPOV | 4096 x 2160 | 59.94 | BT.709 | 1199 |
| Netflix\_BarScene | 4096 x 2160 | 59.94 | BT.709 | 1199 |
| Netflix\_ToddlerFountain | 4096 x 2160 | 59.94 | BT.709 | 1199 |
| Netflix\_Dancers | 4096 x 2160 | 59.94 | BT.709 | 1199 |
| Netflix\_PierSeaside | 4096 x 2160 | 59.94 | BT.709 | 1199 |
| Netflix\_WindAndNature | 4096 x 2160 | 59.94 | BT.709 | 1199 |
| Netflix\_Boat | 4096 x 2160 | 59.94 | BT.709 | 300 |
| Netflix\_FoodMarket | 4096 x 2160 | 59.94 | BT.709 | 600 |
| Netflix\_Tango | 4096 x 2160 | 59.94 | BT.709 | 294 |
| Netflix\_BoxingPractice | 4096 x 2160 | 59.94 | BT.709 | 254 |
| Netflix\_Narrator | 4096 x 2160 | 59.94 | BT.709 | 300 |
| Netflix\_TunnelFlag | 4096 x 2160 | 59.94 | BT.709 | 600 |
| Netflix\_RitualDance | 4096 x 2160 | 59.94 | BT.709 | 600 |
| Netflix\_FoodMarket2 | 4096 x 2160 | 59.94 | BT.709 | 300 |
| Netflix\_SquareAndTimelapse | 4096 x 2160 | 59.94 | BT.709 | 600 |
| Netflix\_Aerial | 4096 x 2160 | 59.94 | BT.709 | 1199 |
| Netflix\_DinnerScene | 4096 x 2160 | 59.94 | BT.709 | 1199 |
| Netflix\_RollerCoaster | 4096 x 2160 | 59.94 | BT.709 | 1199 |

#### C.3.1.1.4 EBU Sequences

The EBU dataset is composed by two sequences described in the table below. The sequences can be used under the CC BY-NC-ND 4.0 license restriction.

Table C.3-4 : EBU test material description

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Resolution | Frame-rate | ColorGamut | Number of frames |
| Rain\_fruits | 3840 x 2160 | 50 | BT.709 | 600 |
| Lupo\_candlelight | 3840 x 2160 | 50 | BT.709 | 600 |

#### C.3.1.1.5 SVT Sequences

The SVT dataset is composed by five sequences described in the table below. The sequences can be used under the license available at <https://tech.ebu.ch/webdav/site/tech/shared/hdtv/svt-multiformat-conditions-v10.pdf> .

Table C.3-5 : SVT test material description

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Resolution | Frame-rate | ColorGamut | Number of frames |
| CrowdRun | 3840 x 2160 | 50 | BT.709 | 500 |
| ParkJoy | 3840 x 2160 | 50 | BT.709 | 500 |
| DucksTakeOff | 3840 x 2160 | 50 | BT.709 | 500 |
| IntoTree | 3840 x 2160 | 50 | BT.709 | 500 |
| OldTownCross | 3840 x 2160 | 50 | BT.709 | 500 |

#### C.3.1.1.6 4ever Sequences

The 4ever dataset is composed by two sequences described in the table below. The sequences can be used under the license described below:

*The proposed test sequences Brest\_Sedof\_3840x2160p60\_10b.yuv and Paris\_Montmartre\_3840x2160p60\_10b, and all intellectual property rights therein remain the property of 4EVER consortium partners: Orange, AMP Visual TV, ATEME, France Télévisions, GlobeCast, InterDigital, Highlands Technologies Solutions, INSA Rennes, TeamCast and Télécom ParisTech*

* + *The following uses are allowed for the proposed sequences:* 
    - *Can be published in technical papers, played at technology research and development events.*
    - *Can be used in 3GPP SA4.*
  + *The following uses are not allowed for the proposed sequences:* 
    - *Do not publish snapshots in product brochures.*
    - *Do not redistribute video with a commercial product.*

Table C.3-6 : 4ever test material description

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Resolution | Frame-rate | ColorGamut | Number of frames |
| 4EVER\_Brest\_Sedof | 3840 x 2160 | 60 | BT.2020 | 600 |
| 4EVER\_Paris\_Montmartre | 3840 x 2160 | 60 | BT.2020 | 600 |

#### C.3.1.1.7 CableLabs Sequences

The CableLabs dataset is composed by eight sequences described in the table below. The sequences can be used under the CC BY-NC-ND 3.0 license restriction.

Table C.3-7 : CableLabs test material description

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Name | Resolution | Frame-rate | ColorGamut | duration | Comment on selected intervals |
| Lifting Off | 3840 x 2160 | 23.98 | BT.709 | 3:21 | Nice colored pictures, som burned in text, scene cuts (2) |
| Seconds that count | 3840 x 2160 | 23.98 | BT.709 | 5:22 | Not critical |
| Unspoken Friend | 3840 x 2160 | 23.98 | BT.709 | 4:17 | Not critical |
| Indoor Soccer | 3840 x 2160 | 23.98 | BT.709 | 4:12 | Fast motion, scene cuts (4) |
| Skateboarding | 3840 x 2160 | 23.98 | BT.709 | 4:36 | Not critical |
| Eldorado | 3840 x 2160 | 23.98 | BT.709 | 3:02 | Panning up/down on a tree (oak), no scene cut |
| Moment of Intensity | 3840 x 2160 | 59.94 | BT.709 | 3:10 | Skate board, no scene cuts |
| Ancient Thought | 3840 x 2160 | 23.98 | BT.709 | 3:08 | Not critical |

All of the sequences are not of interest for compression evaluation. A selection of some critical time intervals have been conducted based on experience on compression. Especially intervals with fast or complex motion, some including scene cuts, representative of TV contents have been selected. Static pictures, or pictures with artistic intent with few spatial details have been discarded. To remain in the scope of the scenario, the following sub sequences have been selected as final candidates:

* Lifting off : from 00:05 to 00:22
* Indoor Soccer : from 01:06 to 01:23
* Eldorado : from 01:33 to 01:50
* Life Untouched : from 00:00 to 00:22

The extraction of subpart from original ProRes sequences is realized with FFmpeg release 4.3.1 using the following commandline:

./ffmpeg.exe -i $IN.mov -vcodec rawvideo -ss $START\_TIME -t $DURATION -pix\_fmt yuv420p10le $OUT.yuv

Where $IN and $OUT are respectively the input and output sequences, $START\_TIME and $DURATION respectively the start time and duration in hh:mm:ss format.

#### C.3.1.2 Selected Test Sequences

In order to decide wich sequences should be used in this specification, the following process has been used:

1. Run informative encoding using x265 on the candidate sequences
2. Run SI-TI metrics computation on the candidate sequences
3. Select two diverse datasets based on R-D and SI-TI information.

#### C.3.1.2.1 x265 Encoding

To represent candidate sequences in the R-D plane, fixed QP encodings are carried out for QPs [12, 17,22,27,32,37,42] using ffmpeg release 4.3.1, with the following command line:

./ffmpeg.exe -pix\_fmt yuv420p10le -s $Wx$H -framerate $FR -i $IN -vcodec libx265 -qp $QP -preset fast -x265-params “psnr=1:ssim=1” $OUT.mp4

Where $W and $W are respectively the sequence width and height, $FR the sequence frame-rate, $QP the selected quantization parameter. $IN and $OUT are respectively the input and output files.

From the obtained log, the overall bitrate and PSNR are extracted to be used for plotting. The encoding result are presented below, all sequences being represented in the targeted [10-40] Mbps range (red box).

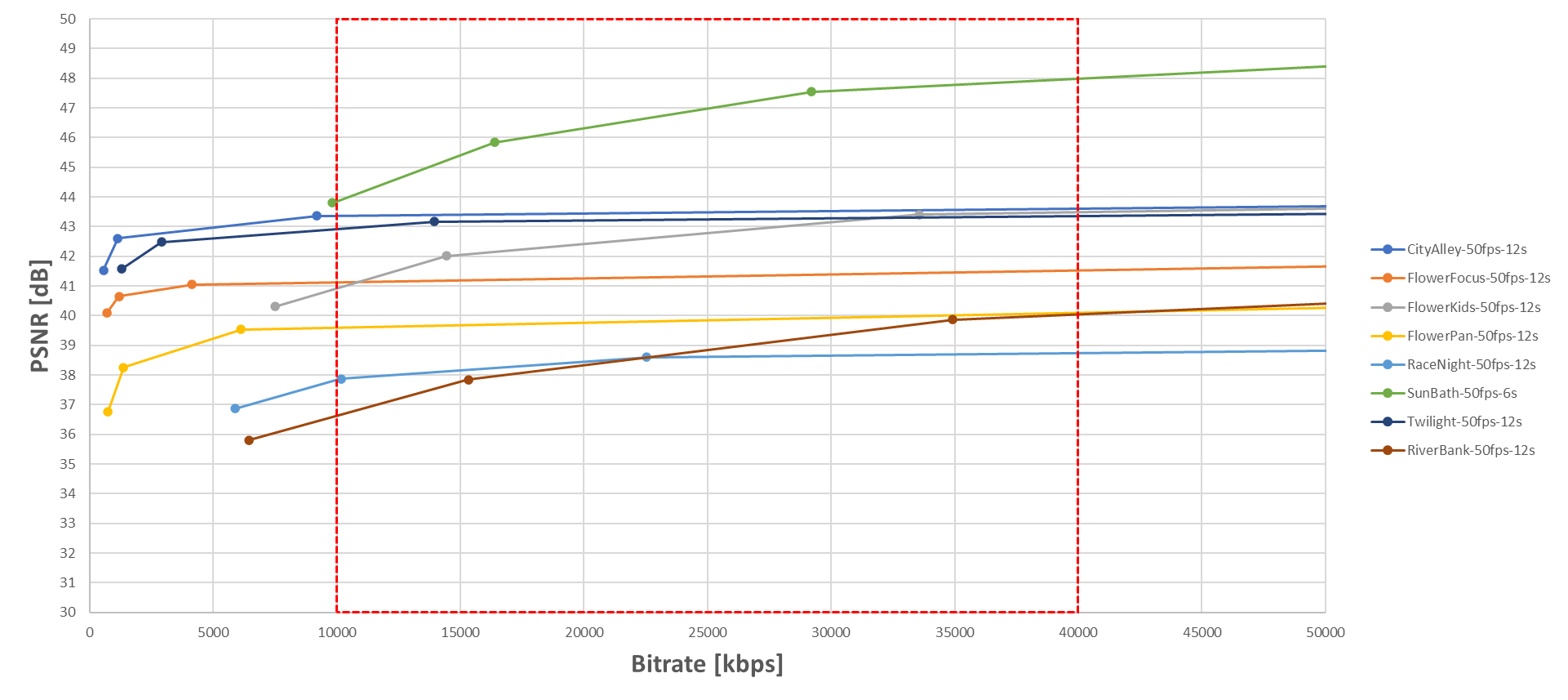


Figure C.3-1 : R-D curves for the UVG dataset

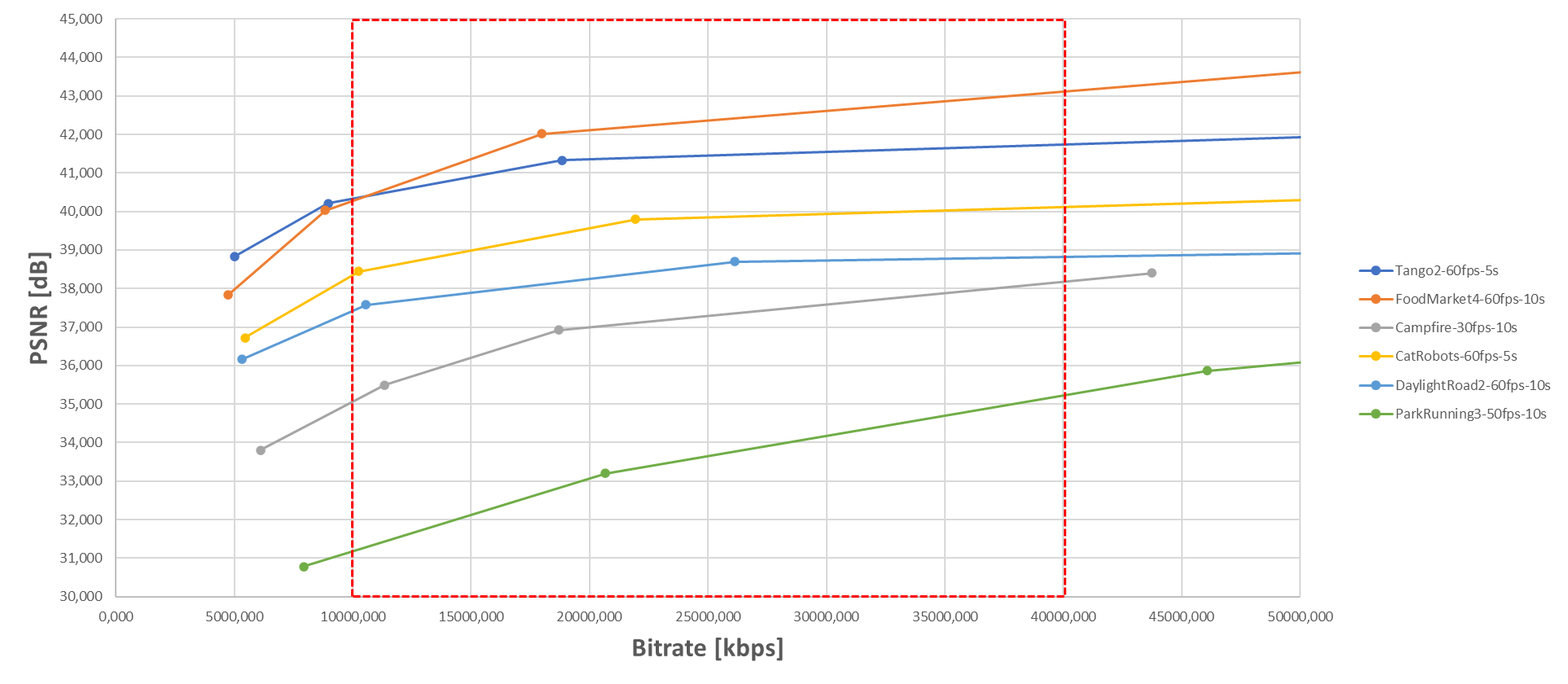


Figure C.3-2 : R-D curves for the JVET dataset

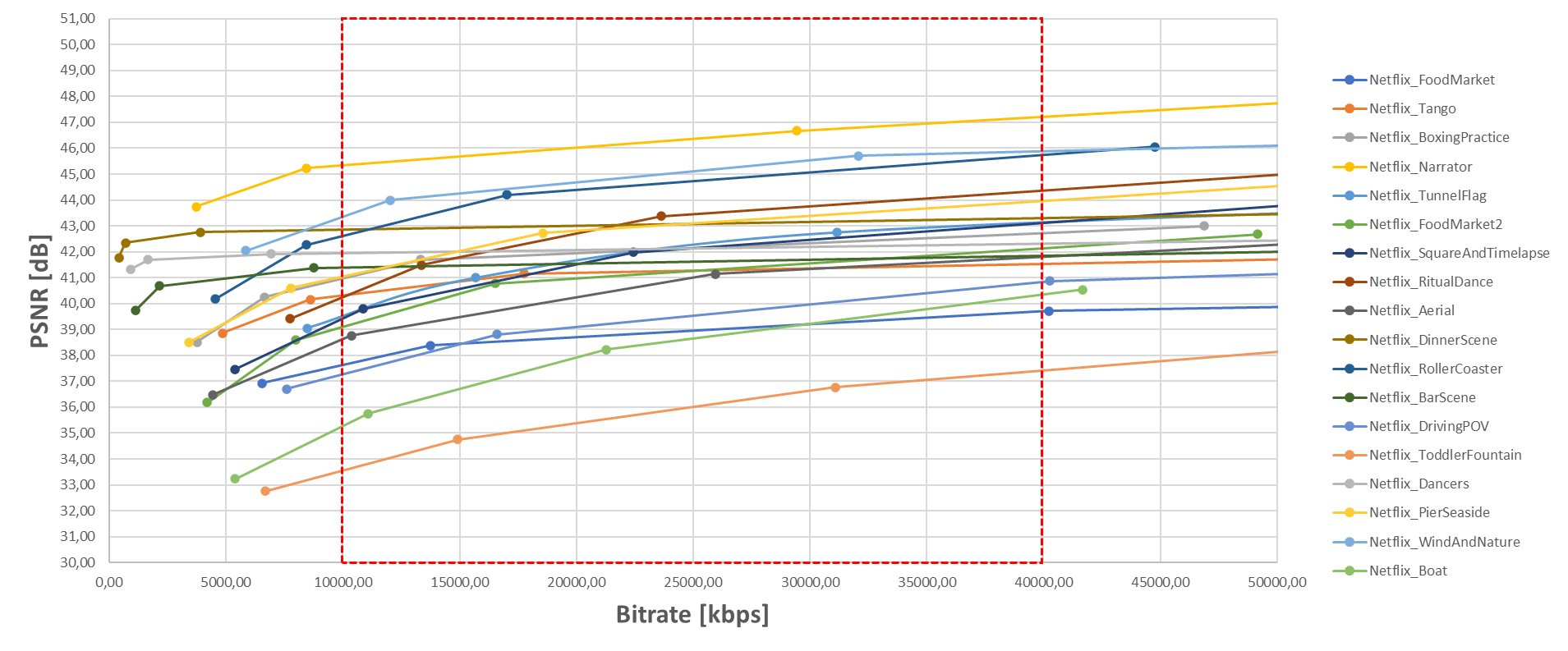


Figure C.3-3 : R-D curves for the Netflix dataset

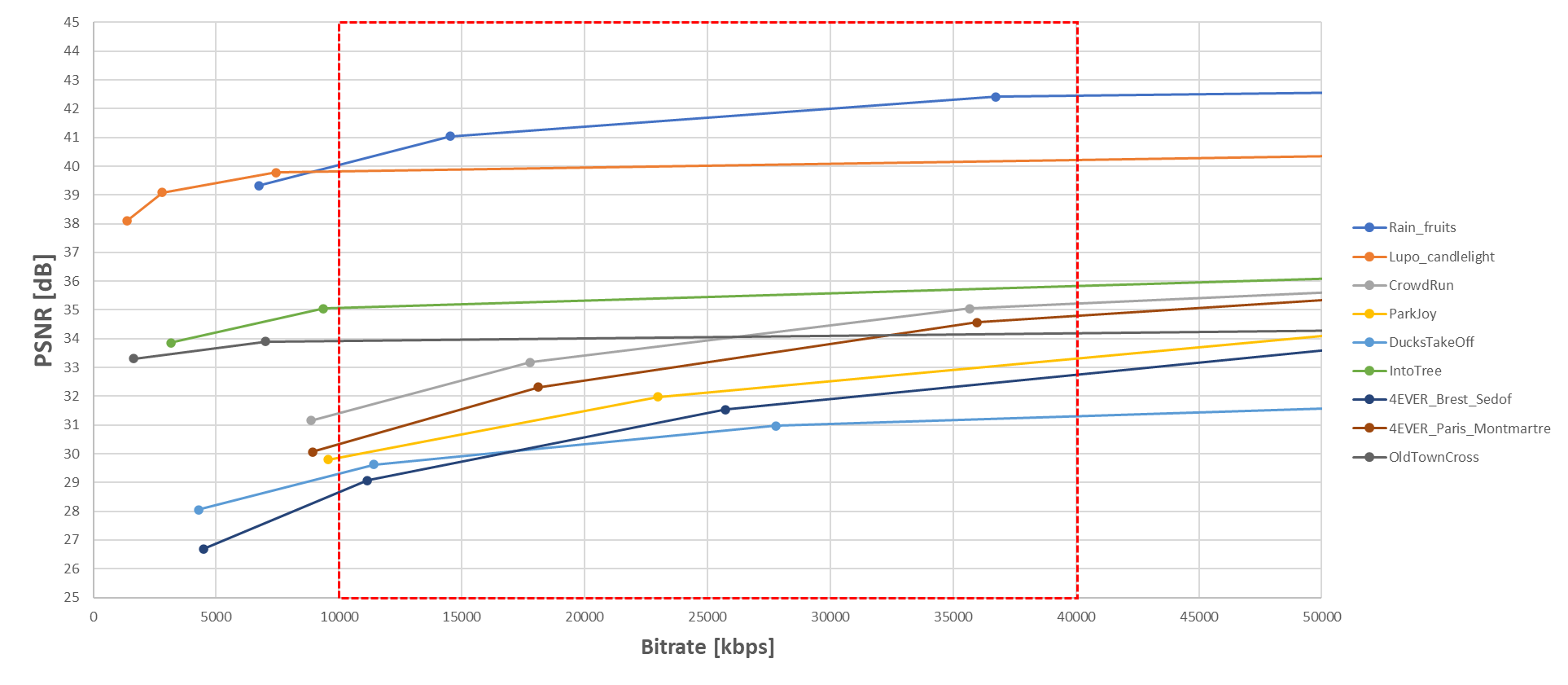


Figure C.3-4 : R-D curves for the EBU, SVT and 4ever datasets

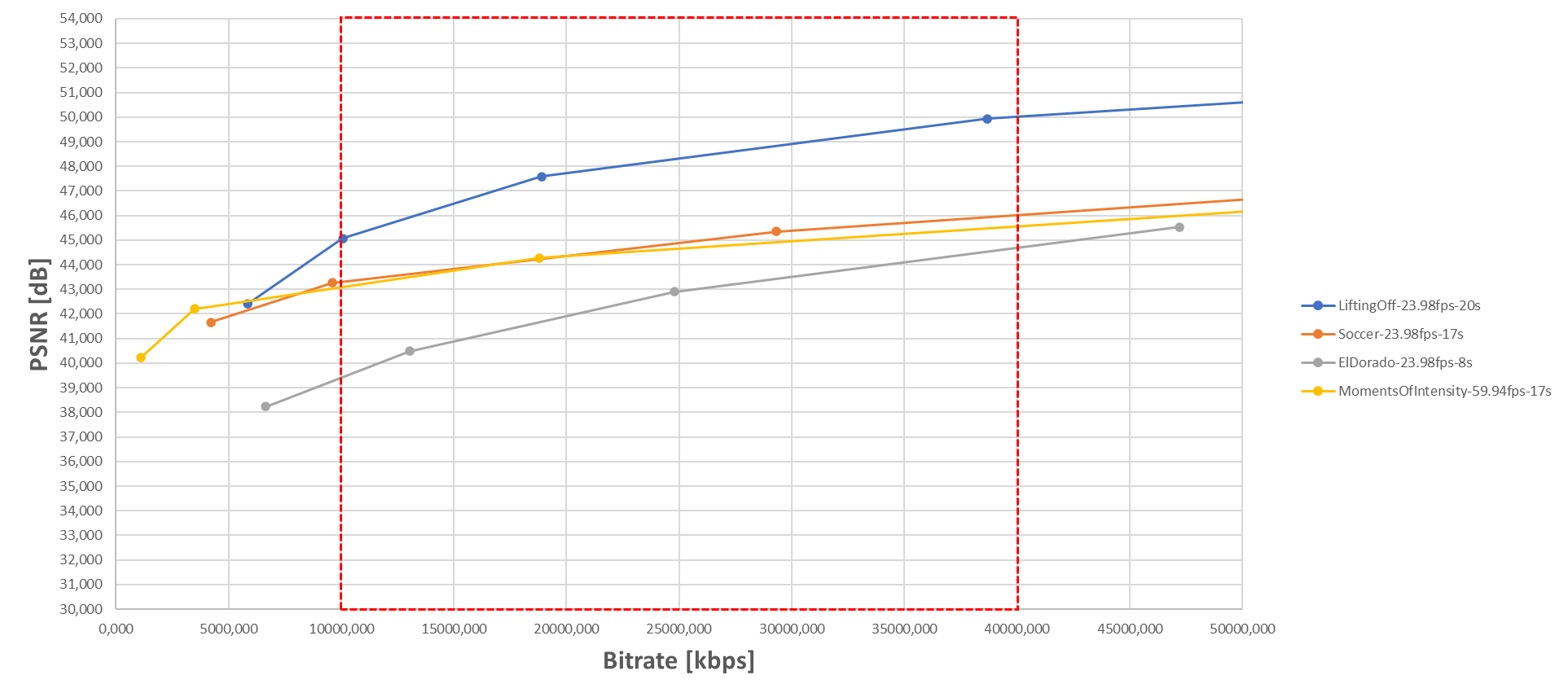


Figure C.3-5 : R-D curves for the CableLabs dataset

#### C.3.1.2.2 SI-TI Metrics

The analysis is based on spatial perceptual information (SI) and temporal perceptual information (TI) metrics calculated as defined in ITU-P.910 recommendation [40]. In order to reduce the bias introduced by scene-cuts in TI metric computation, the TI at scene-cut boundaries are removed from computation.

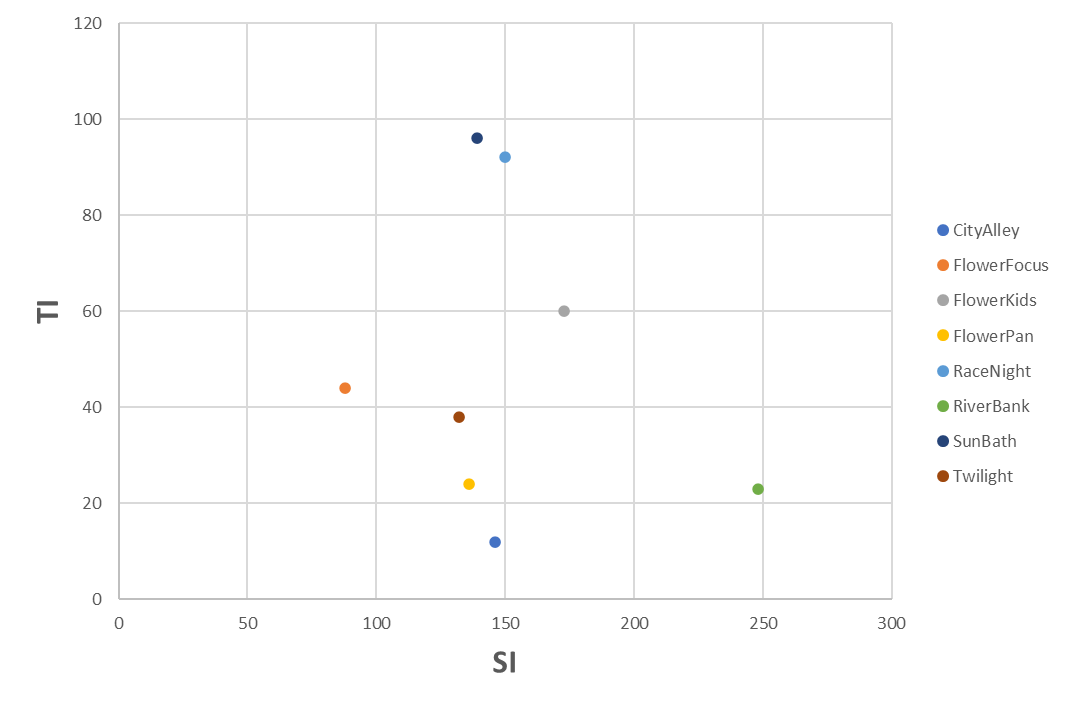


Figure C.3-6 : SI-TI for the UVG dataset

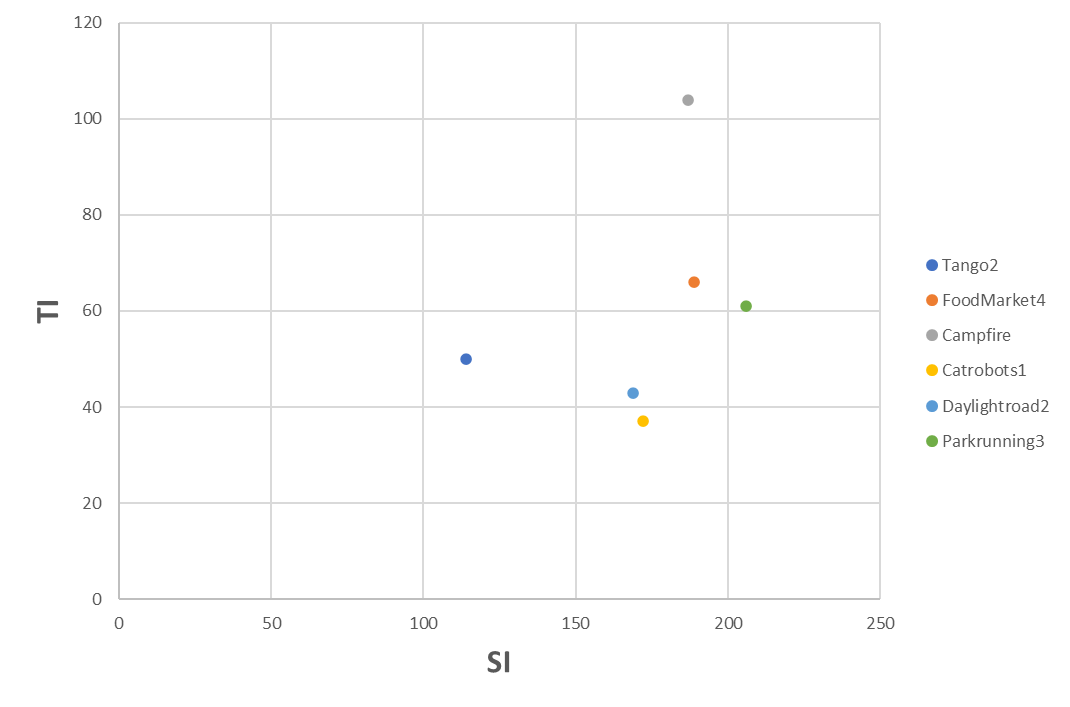


Figure C.3-7 : SI-TI for the JVET dataset

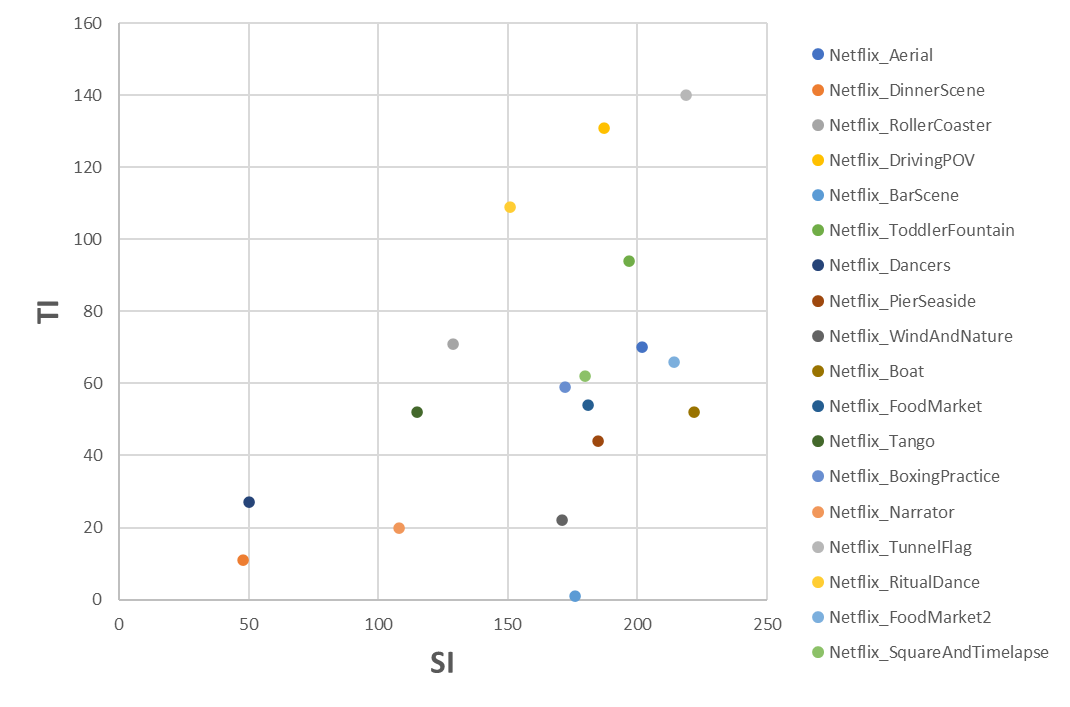


Figure C.3-8 : SI-TI for the Netflix dataset

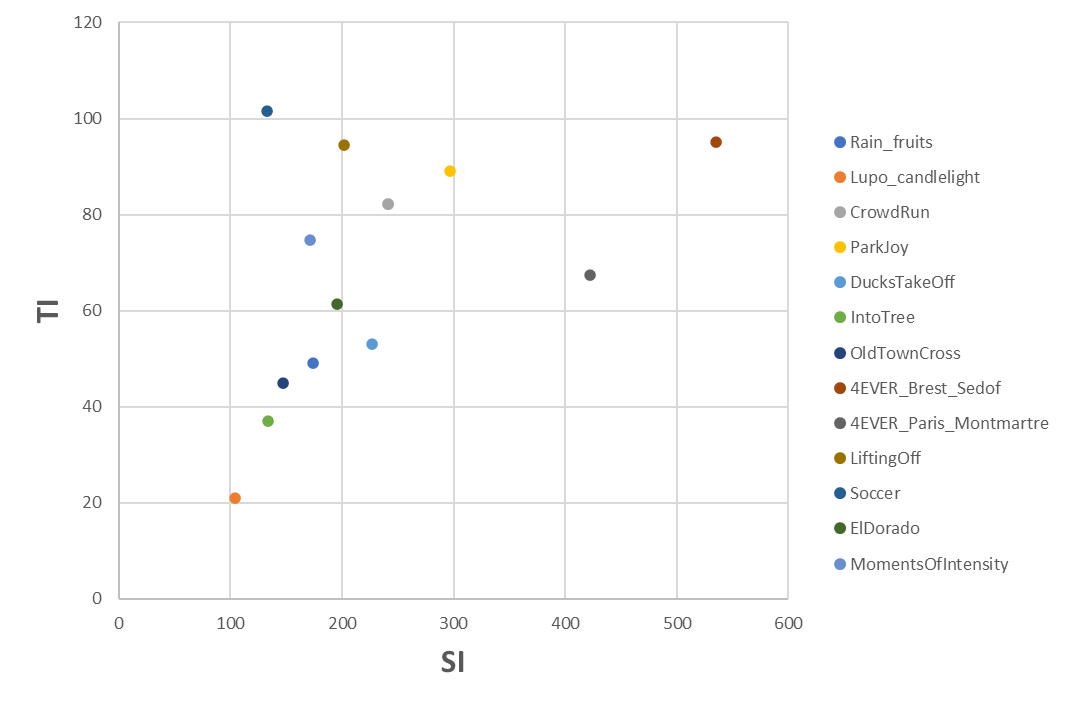


Figure C.3-9 : SI-TI for the EBU, SVT, 4ever and CableLabs dataset

To be added later: reference to the SI-TI scripts will be available (ETSI git)

#### C.3.1.2.3 Sequence Selection

The first step consists in classifying the sequences in two categories : easy and challenging. This reflects the complexity of the sequences from an encoding perspective. Based on the R-D data, the QP points covering the application range [10-40]Mbps are identified. For the lower (10Mbps) and higher (40Mbps) boundaries, a linear interpolation is achieved between the two closest QP points to estimate the distortion at the boundaries: D\_10 and D\_40. If the distortion difference D\_DIFF=(D\_40-D\_10) is less than or equal to a threshold Thr=1.5, then the sequence is classified as follows:

* Low-dynamic: the PSNR dynamic on the application bitrate range is limited
* High-dynamic: the PSNR dynamic on the application bitrate range is high

The results of this first classification step is reported in the table below.

**Table C.3-8 : Sequence classification based on R-D i**nformation in the [10-40] Mbps range, Thr=1.5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Name | D\_10 | D\_40 | D\_DIFF | Dynamic |
| JVET | Tango2 | 40,33 | 41,44 | 1,12 | Low |
| FoodMarket4 | 40,28 | 42,37 | 2,09 | High |
| Campfire | 35,06 | 37,29 | 2,23 | High |
| Catrobots1 | 38,36 | 39,85 | 1,50 | Low |
| Daylightroad2 | 37,42 | 38,61 | 1,19 | Low |
| Parkrunning3 | 31,17 | 33,65 | 2,48 | High |
| UVG | CityAlley | 43,37 | 43,49 | 0,12 | Low |
| FlowerFocus | 41,13 | 41,33 | 0,20 | Low |
| FlowerKids | 40,92 | 42,78 | 1,86 | High |
| FlowerPan | 39,60 | 39,85 | 0,25 | Low |
| RaceNight | 37,82 | 38,62 | 0,79 | Low |
| RiverBank | 36,62 | 38,84 | 2,22 | High |
| SunBath | 43,85 | 46,98 | 3,13 | High |
| Twilight | 42,92 | 43,24 | 0,33 | Low |
| Netflix | Netflix\_Aerial | 38,62 | 41,00 | 2,37 | High |
| Netflix\_DinnerScene | 42,85 | 43,08 | 0,23 | Low |
| Netflix\_RollerCoaster | 42,61 | 44,73 | 2,11 | High |
| Netflix\_DrivingPOV | 37,27 | 39,54 | 2,27 | High |
| Netflix\_BarScene | 41,39 | 41,62 | 0,23 | Low |
| Netflix\_ToddlerFountain | 33,56 | 36,01 | 2,45 | High |
| Netflix\_Dancers | 41,96 | 42,14 | 0,18 | Low |
| Netflix\_PierSeaside | 41,05 | 43,10 | 2,06 | High |
| Netflix\_WindAndNature | 43,35 | 45,09 | 1,75 | High |
| Netflix\_Boat | 35,27 | 38,64 | 3,37 | High |
| Netflix\_FoodMarket | 37,62 | 38,95 | 1,32 | Low |
| Netflix\_Tango | 40,31 | 41,27 | 0,97 | Low |
| Netflix\_BoxingPractice | 40,97 | 42,14 | 1,17 | Low |
| Netflix\_Narrator | 45,33 | 46,36 | 1,03 | Low |
| Netflix\_TunnelFlag | 39,47 | 42,06 | 2,59 | High |
| Netflix\_RitualDance | 40,25 | 43,46 | 3,21 | High |
| Netflix\_FoodMarket2 | 39,11 | 41,27 | 2,16 | High |
| Netflix\_SquareAndTimelapse | 39,43 | 42,15 | 2,72 | High |
| EBU | Rain\_fruits | 40,04 | 41,69 | 1,65 | High |
| Lupo\_candlelight | 39,81 | 40,01 | 0,20 | Low |
| SVT | CrowdRun | 31,41 | 33,94 | 2,53 | High |
| ParkJoy | 29,86 | 32,13 | 2,27 | High |
| DucksTakeOff | 29,31 | 30,75 | 1,44 | Low |
| IntoTree | 35,06 | 35,45 | 0,39 | Low |
| OldTownCross | 33,93 | 34,06 | 0,13 | Low |
| 4ever | 4EVER\_Brest\_Sedof | 28,66 | 31,42 | 2,76 | High |
| 4EVER\_Paris\_Montmartre | 30,33 | 33,19 | 2,86 | High |
| CableLabs | LiftingOff | 45,01 | 48,31 | 3,30 | High |
| Soccer | 43,31 | 44,89 | 1,58 | High |
| ElDorado | 39,41 | 44,29 | 4,88 | High |
| MomentsOfIntensity | 43,09 | 44,66 | 1,57 | High |

It is observed that 26 sequences are classified as “High” while 19 are “Low”. In order to keep sequences that are relevant for a codec characterization, the low-dynamic sequences are removed. Then, the next step consists in removing sequences having some similarities, using SI-TI information computed in Section C.3.1.2.2. The sequences are represented below in the SI-TI plane with a k-means clustering to help the decision.

Chart

Description automatically generated

Figure C.3-9 : SI-TI clustering for the Challenging sequences (10 clusters)

In order to remain representative in the sense of sources origin, format, and SI-TI values, the following approach was taken:

* 1 sequence (out of 6) from the C2 cluster (EBU: Rain-fruits)
* 1 sequence (out of 4) from the C3 cluster (CableLabs: Soccer)
* 2 sequences (out of 8) from the C4 cluster (Netfllix: ToddlerFountain, Boat)
* 1 sequence (out of 1) from the C5 cluster (UVG: Riverbank)
* 1 sequence (out of 2) from the C6 cluster (Netflix: TunnelFlag)
* 1 sequence (out of 1) from the C8 cluster (SVT: ParkJoy)
* 1 sequence (out of 1) from the C10 cluster (4ever: Brest-Sedof)

The following eight sequences are selected:

* 1 From 4ever: Brest-Sedof (3840x2160p 60fps)
* 1 From EBU: RainFruits (3840x2160p 50fps)
* 1 From SVT: ParkJoy (3840x2160p 50fps)
* 1 From CableLabs: Soccer (3840x2160p 23.98fps)
* 3 From Netflix:, TunnelFlag, Boat and Toddlerfountain (4096x2160p 59.94fps)
* 1 From UVG: RiverBank (3840x2160p 50fps)

### C.3.2 HDR Category

### C.3.3 HFR Category

Tbd

## C.4 HD Test Sequences

### C.4.1 Overview

### [C.4.2 CTA WAVE Tears of Steel

#### C.4.2.1 Introduction

CTA WAVE has produced a derivate of Tears of Steel with dedicated overlays for testing purposes. The content has been subsampled and different frame rates are produced. Details can be found here: http://dash.akamaized.net/WAVE/index.html.

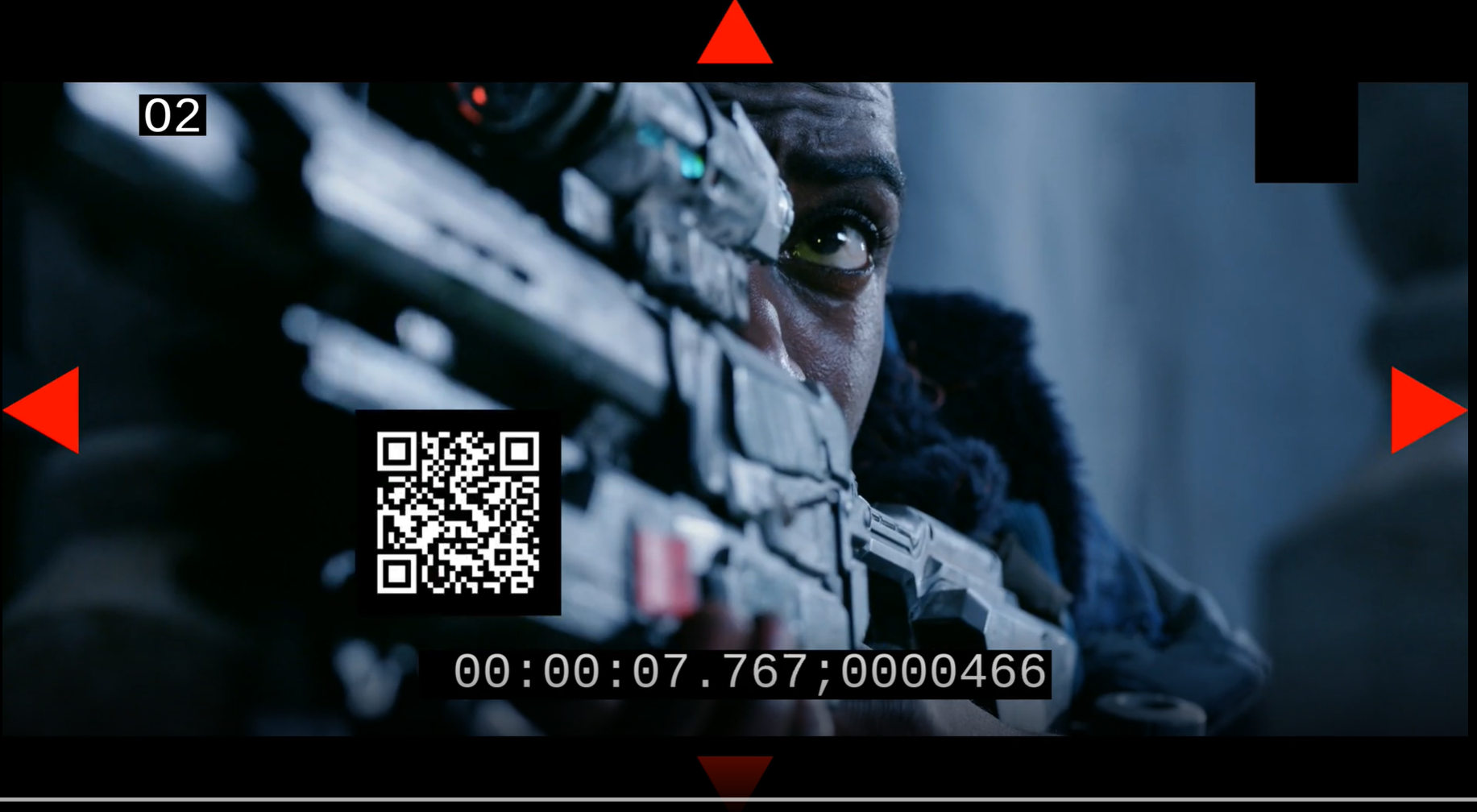


Figure C.3.2-1: Screenshot of CTA WAVE’s ToS test sequence

#### C.4.2.2 Reference sequence properties

The Tears of Steel sequence has the following properties:

- Resolution: 1920 x1080

- Scan: Progressive

- Frame rate: 60 fps

- Bit depth: 8-bit

- Length: 3600 frames (60s)

- YUV format: YUV 4:2:0

- Colour Components: Y’CbCr

- Colour space: ITU-R BT.709

The original version as well as subsampled ones are available

* WAVE-ToS-01: <http://dash.akamaized.net/WAVE/Mezzanine/tos_L1_1920x1080@60_60.mp4>
* WAVE-ToS-02: <http://dash.akamaized.net/WAVE/Mezzanine/tos_K1_1600x900@60_60.mp4>
* WAVE-ToS-03: http://dash.akamaized.net/WAVE/Mezzanine/tos\_J1\_1280x720@60\_60.mp4

#### C.4.2.3 Copyright information

The Teers of Steel video sequence is made available under the following copyright disclaimer.

*The proceedings and results of the Mango Open Movie project are being licensed under the Creative Commons Attribution 3.0 license. This material is a result of the Mango Open Movie project, who owns the copyright. The Mango Open Movie project (licensor) has licensed it under the Creative Commons Attribution 3.0 license (<https://creativecommons.org/licenses/by/3.0/>). In short, this means you can freely reuse and distribute this content, also commercially, for as long you provide a proper attribution.*

*The attribution is: (CC) Blender Foundation | mango.blender.org*

### C.4.3 CTA WAVE Croatia

#### C.4.3.1 Introduction

CTA WAVE has produced a derivate of EBU’s Croatia sequence with dedicated overlays for testing purposes. The content has been subsampled and different frame rates are produced. This sequence is an HDR sequence. Details can be found here: http://dash.akamaized.net/WAVE/index.html.



Figure C.3.3-1: Screenshot of Croatia test sequence

#### C.4.3.2 Reference sequence properties

The Tears of Steel sequence has the following properties:

- Resolution: 1920 x1080

- Scan: Progressive

- Frame rate: 50 fps

- Bit depth: 10-bit

- Length: 3000 frames (60s)

- YUV format: YUV 4:2:0

- Colour Components: Y’CbCr

- Colour space: ITU-R BT.2100 PQ

The original version as well as subsampled ones are available

* WAVE-CRO-01: http://dash.akamaized.net/WAVE/Mezzanine/croatia\_L1\_1920x1080@50\_60.mp4
* WAVE-CRO-02: http://dash.akamaized.net/WAVE/Mezzanine/croatia\_K1\_1600x900@50\_60.mp4
* WAVE-CRO-03: http://dash.akamaized.net/WAVE/Mezzanine/croatia\_J3\_1280x720@50\_60.mp4

#### C.4.3.3 Copyright information

The Croatia video sequence is made available under the following copyright disclaimer.

*This material has been recorded by the EBU, who owns the copyright. The EBU (licensor) has licensed it under the Creative Commons Attribution 4.0 license (https://creativecommons.org/licenses/by/4.0/). In short, this means you can freely reuse and distribute this content, also commercially, for as long you provide a proper attribution.*

#### ]

## C.5 Messaging Sequences

### [

### C.5.1 Overview

### C.5.2 Powder

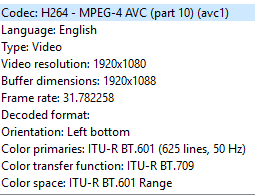
#### C.5.2.1 Introduction

A donated sequence with outdoor filming.



Figure C.X.2-1: Screenshot of Powder sequence

#### C.5.2.2 Reference sequence properties



#### C.5.2.3 Copyright information

*This material has been recorded by the Thomas Stockhammer, who owns the copyright. Thomas Stockhammer (licensor) has licensed it under the Creative Commons Attribution 4.0 license (https://creativecommons.org/licenses/by/4.0/). In short, this means you can freely reuse and distribute this content, also commercially, for as long you provide a proper attribution.*

]

## C.6 Screen Content Test Sequences

### [

### C.6.1 Overview

Screen content test sequences have been generated for the purpose of this study with the objective to illustrate commonly encountered situations during videoconferences, particularly when sharing presentation slides. Some critical configurations are part of this test set such as moving text at different speeds and various transitions between the slides.

### C.6.2 Moving Text 2

#### C.6.2.1 Introduction

The sequence *MovingText2* is composed of several lines of text horizontally scrolling at different speeds over a gradation of background colours. The sequence has been generated in 4K resolution at 60 frames per second with an authoring tool and exported into Apple ProRes 422 before being converted into YUV4:2:0. The figure C.X.2-1 is a screenshot of the *MovingText2* test sequence.

Une image contenant texte

Description générée automatiquement

Figure C.X.2-1: Screenshot of MovingText2 test sequence

#### C.6.2.2 Source sequence properties

The *MovingText2* test sequence has the following properties:

- Resolution: 3840 x 2160 and 1920 x1080

- Scan: Progressive

- Frame rate: 60 fps

- Bit depth: 10-bit and 8-bit

- Length: 600 frames (10s)

- YUV format: YUV 4:2:0

- Colour Components: Y’CbCr

- Colour space: ITU-R BT.709

#### C.6.2.3 Copyright information

The *MovingText2* video sequence has been produced for the purpose of this study on the characterization of video codecs in 3GPP.

### C.6.3 Text Mix Transitions

#### C.6.3.1 Introduction

The sequence *TextMixTransitions* is composed of 2 slides containing fixed text at different sizes. Those 2 slides appear twice each during 2.5s. 1s long transitions are inserted in between (fade, slide, turning page). The sequence has been generated in 4K resolution at 60 frames per second with an authoring tool and exported into Apple ProRes 422 before being converted into YUV4:2:0. The figure C.X.3-1 is a screenshot of the *TextMixTransitions* test sequence.

Une image contenant texte

Description générée automatiquement

Figure C.X.3-1: Screenshot of TextMixTransitions test sequence

#### C.6.3.2 Source sequence properties

The *TextMixTransitions* test sequence has the following properties:

- Resolution: 3840 x 2160 and 1920 x1080

- Scan: Progressive

- Frame rate: 60 fps

- Bit depth: 10-bit and 8-bit

- Length: 600 frames (10s)

- YUV format: YUV 4:2:0

- Colour Components: Y’CbCr

- Colour space: ITU-R BT.709

#### C.6.3.3 Copyright information

The *TextMixTransitions* video sequence has been produced for the purpose of this study on the characterization of video codecs in 3GPP.

### C.6.4 Graphics Mix Simple

#### C.6.4.1 Introduction

The sequence *GraphicsMixSimple* is composed of 3 slides containing different types of graphics. Those 3 slides appear once each during 3.33s. Transitions are simple scene cut in between. The sequence has been generated in 4K resolution at 60 frames per second with an authoring tool and exported into Apple ProRes 422 before being converted into YUV4:2:0. The figure C.X.4-1 is a screenshot of each of the 3 scenes contained in the *GraphicsMixSimple* test sequence.

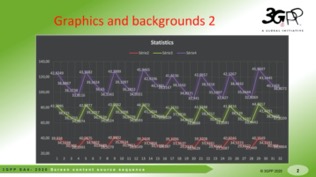


Figure C.X.4-1: Screenshot of the 3 scenes in GraphicsMixSimple test sequence

#### C.6.4.2 Source sequence properties

The *GraphicsMixSimple* test sequence has the following properties:

- Resolution: 3840 x 2160 and 1920 x1080

- Scan: Progressive

- Frame rate: 60 fps

- Bit depth: 10-bit and 8-bit

- Length: 600 frames (10s)

- YUV format: YUV 4:2:0

- Colour Components: Y’CbCr

- Colour space: ITU-R BT.709

#### C.5.4.3 Copyright information

The *GraphicsMixSimple* video sequence has been produced for the purpose of this study on the characterization of video codecs in 3GPP.

### C.6.5 Graphics Mix Transition

#### C.6.5.1 Introduction

The sequence *GraphicsMixTransition* is the same sequence as *GraphicsMixSimple* described in clause C.X.4, except the fact that 1s long transitions have been inserted composed of 3 slides containing different types of graphics. The 3 slides appear once each during 3.33s. Transitions are fading and page turning in between. The sequence has been generated in 4K resolution at 60 frames per second with an authoring tool and exported into Apple ProRes 422 before being converted into YUV4:2:0. The figure C.X.4-1 is a screenshot of each of the 3 scenes contained in the *GraphicsMixSimple* test sequence.

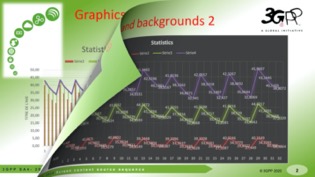
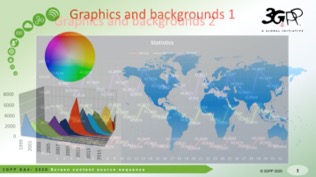


Figure C.X.5-1: Screenshot of the 2 transitions in GraphicsMixTransitions test sequence

#### C.6.5.2 Source sequence properties

The *GraphicsMixTransitions* test sequence has the following properties:

- Resolution: 3840 x 2160 and 1920 x1080

- Scan: Progressive

- Frame rate: 60 fps

- Bit depth: 10-bit and 8-bit

- Length: 600 frames (10s)

- YUV format: YUV 4:2:0

- Colour Components: Y’CbCr

- Colour space: ITU-R BT.709

#### C.6.5.3 Copyright information

The *GraphicsMixTransitions* video sequence has been produced for the purpose of this study on the characterization of video codecs in 3GPP.

]

Annex D  
Reference Configurations

## D.1 Introduction

## D.2 H.264/AVC JM19.0 Reference Configurations

### D.2.1 JM-01: Low-latency

Add details

## D.3 H.265/HEVC HM16.22 Reference Configurations

### D.3.1 HM-01: Low-latency

Add details

Variations QP

Annex E  
Data Management and Hosting

## E.1 Introduction

All reference sequences and all anchors are hosted on a publically accessible web-server here: [https://dash-large-files.akamaized.net/WAVE/3GPP/5GVideo](https://urldefense.com/v3/__https:/dash-large-files.akamaized.net/WAVE/3GPP/5GVideo__;!!GjvTz_vk!H9Rjd2jziclpofrTZRCQ7sdADrTFH6e2aYwPlUfHgT8MYKLWV_Qijd_wFmgc$" \t "_blank).

A frontend is provided here: [https://dash-large-files.akamaized.net/WAVE/3GPP/5GVideo](https://urldefense.com/v3/__https:/dash-large-files.akamaized.net/WAVE/3GPP/5GVideo__;!!GjvTz_vk!H9Rjd2jziclpofrTZRCQ7sdADrTFH6e2aYwPlUfHgT8MYKLWV_Qijd_wFmgc$" \t "_blank)/index.html.

## E.2 Reference Sequences

### E.2.1 Hosting

All reference sequences are hosted at <https://dash-large-files.akamaized.net/WAVE/3GPP/5GVideo/ReferenceSequence>. Below this folder, for each reference sequence a decicated folder is created. The folder includes at least the following information:

* Raw video file
* MD5 hash
* Metadata file
* Copyright and license file

### E.2.2 Uploading

For uploading a new reference sequence, please create a new issue here <https://github.com/haudiobe/5G-Video-Content/issues/new?assignees=&labels=request&template=reference_sequence.md&title=>

Please specify the following information:

* **Contact Person** Max Mustermann, [mustermann@max.com](mailto:mustermann@max.com)
* **Related Scenario** Specify the related scenario
* **Accessibility of Sequence** Where can the sequence be accessed. Provide URL and instructions.
* **Copyright/License** Provide a copyright/license statement with the sequence
* **Status** Is the sequence proposed to 3GPP or already agreed as reference sequence
* **Metadata** Provide all relevant metadata for the sequence
* **MD5 Hash** Provide and MD5 hash for the sequence

### E.2.3 Downloading

Sequence can be downloaded for free from the above folders. Licensing terms must be obeyed. Downloaded files should be MD5 verified

## E.3 Anchors

### E.3.1 Hosting

All anchors are hosted at <https://dash-large-files.akamaized.net/WAVE/3GPP/5GVideo/Anchors>. Below this folder, a hierarchy is created

* Scenario
  + Codec
    - Anchor Tuple
      * Anchor

a dedicated folder is created. The folder includes at least the following information:

* Binary video bitstream
* MD5 hash of video bitstream
* MD5 hash of reconstructed video sequence
* Reference to anchor generation script
* Creation data
* Anchor Results

### E.3.2 Uploading

For uploading anchors, please create a new issue here <https://github.com/haudiobe/5G-Video-Content/issues/new?assignees=&labels=request&template=anchor_tuple.md&title=>

Please specify the following information:

### E.3.3 Downloading

Anchors can be downloaded for free from the above folders. Licensing terms must be obeyed. Downloaded files should be MD5 verified.

Annex F  
Software Packages

## F.1 Introduction

tbd

## F.2 Readme

# Quickstart

Given an anchor definition, the tool performs encoding, reconstruction and runs additional metrics computation for all variants.

1. add the samples/references/yuv420\_1280x720\_8bit\_rec709.yuv sequence
2. set the environment variable: HM\_ENCODER=/path/to/HM/bin/TAppEncoderStatic
3. generate the sample HM anchor

./cmd.py ./samples/anchors/sample\_hm.json encode decode`

## Using docker

a sample [Dockerfile](https://docs.docker.com/get-docker/) is provided to build an image containing all the dependencies : HM, JM, and VTM

git clone https://github.com/haudiobe/5GVideo.git

cd 5GVideo

docker build -t anchortools -f ./docker/Dockerfile .

# add the missing `./samples/references/yuv420\_1280x720\_8bit\_rec709.yuv` sequence

docker run --mount type=bind,source=./samples,target=/samples -it anchortools cmd.py /samples/anchors/sample\_hm.json

the --mount option mounts the directory source path, and makes it available as target path in the running container.

# Usage

./cmd.py ./anchor.json [encode] [decode] [metrics]

* the tool assumes you have reference encoders compiled, with environment variables pointing to the executables, see below.

## encode

./cmd.py ./anchor.json encode decode

runs the **reference encoder** to generate both bitstream, and reconstructed sequence for all the anchors, with the reconstructed chroma format as specified in the encoder config.

./cmd.py ./anchor.json encode

runs the reference encoder to generate bitstream only.

## decode

./cmd.py ./anchor.json decode

runs the **reference decoder** to reconstruct the bitstream, with the output chroma format matching the bitstream.

## metrics

./cmd.py ./anchor.json metrics

generates some metrics for each variant defined in the anchor

the metrics options uses third party tools, see below.

# Reference encoders

the following environment variables are needed depending on the encoder/decoder you want to use:

[JM](https://vcgit.hhi.fraunhofer.de/jct-vc/JM)

JM\_ENCODER=/path/to/JM/bin/lencod\_static

JM\_DECODER=/path/to/JM/bin/ldecod\_static

[HM](https://vcgit.hhi.fraunhofer.de/jct-vc/HM)

HM\_ENCODER=/path/to/HM/bin/TAppEncoderStatic

HM\_DECODER=/path/to/bin/TAppDecoderStatic

[VTM](https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM)

VTM\_ENCODER=/path/to/bin/EncoderAppStatic

VTM\_DECODER=/path/to/bin/DecoderAppStatic

### Adding a custom encoder

implement the ReferenceEncoder interface (encoder\_id, encode\_variant, decode\_variant) and decorate your class (@register\_encoder)

# Anchor definition

example ***anchor.json*** :

{

"description": "human readable description, use case, settings overview ...",

"reference": "path/to/reference/sample.yuv",

"encoder": "HM",

"encoder\_cfg": "path/to/anchor/encoder\_cfg.cfg",

"variants": {

"variant\_id0": {

"OptionKey": "OptionValue0",

"OptionKey2": "{ANCHOR\_DIR}/sub0.cfg",

[...]

},

"variant\_id1": {

"OptionKey": "OptionValue1",

"OptionKey2": "{ANCHOR\_DIR}/sub1.cfg",

[...]

},

[...]

}

}

### notes :

* if test\_sequence or encoder\_cfg is a relative path, it is interpreted as **relative to anchor.json**.
* when used in the variant options {ANCHOR\_DIR} is expanded to the anchor's encoder config directory: **path/to/anchor**.

## Variant options to encoder CLI args mapping

### **HM**, **VTM**

{

"encoder\_cfg": "/encoder.cfg",

"variants": {

"variant\_id0": {

"-k": "v",

"--Key": "value"

}

}

}

maps to -c /encoder.cfg -k v --Key=value

### **JM**

{

"encoder\_cfg": "/encoder.cfg",

"variants": {

"variant\_id0": {

"-f": "file.cfg",

"Key": "value"

}

}

}

maps to -d /encoder.cfg -f file.cfg -p Key=value

## **Output**

the above configuration would generate the following :

# variant bitstream

path/to/anchor/encoder\_cfg.variant\_id0.bit

# encoder log (VTM has additional .opl file)

path/to/anchor/encoder\_cfg.variant\_id0.enc.log

# reconstructed variant

path/to/anchor/encoder\_cfg.variant\_id0.yuv

# variant metrics

path/to/anchor/encoder\_cfg.variant\_id0.csv

[...]

path/to/anchor/encoder\_cfg.variant\_id1.bit

path/to/anchor/encoder\_cfg.variant\_id1.yuv

path/to/anchor/encoder\_cfg.variant\_id1.enc.log

path/to/anchor/encoder\_cfg.variant\_id1.csv

[...]

# averaged metrics, one variant per row

path/to/anchor/encoder\_cfg.csv

# Raw video sequence descritpion

YUV sequences are currently described through a sidecar file.

eg. for the above path/to/reference/sample.yuv, add the following path/to/reference/sample.json

{

"width": 1280,

"height": 720,

"chroma\_format": "yuv",

"chroma\_subsampling": "420",

"bitdepth": 8,

"fps": 30,

"color\_space": "rec709",

"transfer": null,

"framecount": 30

}

# Current limitations

* fps is converted to integer
* only planar YUV reference sequences are supported
* the **transfer** and **color\_space** properties are currently ignored, however color space conversions can be configured through the encoder.cfg file and through variant options for each encoder
* metrics computation assumes reference sequence and reconstructed sequences share the same chroma format, frame packing and bitdepth

# Dependencies [metrics]

The open-source gpac application is needed for metrics computation.

Detailed build instruction please refer to : <https://github.com/gpac/gpac/wiki/Build-Introduction>

the path to the gpac executable can be configured through environment variable, eg. :

GPAC\_APP=/path/to/bin/gpac

Annex G  
Reference Configurations

## G.1 Introduction

tbd

## G.2 H.264/AVC JM19.0 Reference Configurations

### G.2.1 JM-01: 1080p Streaming, 1 second

tbd

### G.2.2 JM-02: 900p Streaming, 1 second

tbd

### G.2.3 JM-03: 720p Streaming, 1 second

## G.3 H.265/HEVC HM16.22 Reference Configurations

### G.3.1 HM-01: 1080p Streaming, 1 second

Add details

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

- Variations QP

### G.3.2 HM-02: 900p Streaming, 1 second

Add details

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

- Variations QP

### G.3.3 HM-03: 720p Streaming, 1 second

Add details

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

- Variations QP

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 |
| 2020-05 | SA4#109 | S4-200891 |  |  |  | Version agreed at SA4#109-e | 0.2.0 |
| 2020-08 | SA4#110 | S4-201243 |  |  |  | Version agreed at SA4#110-e | 0.3.0 |
| 2020-11 | SA4#111 | S4-201515 |  |  |  | Version agreed at SA4#111-e | 0.4.0 |

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