



Title: **Joint Call for Proposals on Video Compression with Capability beyond HEVC**

Status: Output document of JVET

Purpose: Call for Proposals

Author(s) or Andrew Segall Email: asegall@sharplabs.com
Contact(s): Vittorio Baroncini Email: baroncini@gmx.com
 Jill Boyce Email: jill.boyce@intel.com
 Jianle Chen Email: cjianle@qti.qualcomm.com
 Teruhiko Suzuki Email: teruhikos@jp.sony.com

Source: Editors

1 Introduction

This document is a Call for Proposals on video coding technology with compression capabilities that significantly exceed those of the HEVC standard (Rec. ITU-T H.265 | ISO/IEC 23008-2) and its current extensions. This Call for Proposals (CfP) has been issued jointly by ITU-T SG16 Q6 (VCEG) and ISO/IEC JTC 1/SC 29/W G11 (MPEG), and the evaluation of submissions will be conducted in joint collaboration.

2 Purpose and procedure

A new generation of video compression technology that has substantially higher compression capability than the existing HEVC standard is targeted. More background information, as well as information about applications and requirements, is given in [1][2].

Companies and organizations that have developed video compression technology that they believe to be better than HEVC are invited to submit proposals in response to this Call.

To evaluate the proposed compression technologies, formal subjective tests will be performed. Results of these tests will be made public (although no direct identification of the proponents will be made in the report of the results unless it is specifically requested or authorized by a proponent to be explicitly identified and a consensus is established to do so). Prior to having evaluated the results of the tests, no commitment to any course of action regarding the proposed technology can be made.

Descriptions of proposals shall be registered as input documents to the proposal evaluation meeting of April 2018 (see the timeline in section 3). Proponents also need to attend this meeting to present their proposals. Further information about logistical steps to attend the meeting can be obtained from the listed contact persons (see section 9).

3 Timeline

The timeline of the Call for Proposals is as follows:

- 2017/10/31: Final Call for Proposals (public release)
- 2017/11/01: Formal registration period opens
- 2017/11/15: Anchor bitstreams will be available

- 2017/12/15: Formal registration period ends.
 Invoice for the testing fee (see section 6) sent by the test facility approximately one week after registration.
- 2018/02/15: Coded test material shall be received by the Test Coordinator¹. By this date, confirmation of the purchase order or payment shall be received.
- 2018/03/01: Subjective assessment starts
- 2018/04/02: Registration and submission of documents describing the proposals²
- 2018/04/10: Cross-checking of bitstreams and binary decoders (participation mandatory for proponents)
- 2018/04/10: Subjective test results available within standardization body
- 2018/04/11-20: Evaluation of proposals at standardization meeting³

Anticipated tentative timeline after CfP:

(referring to the first version of an anticipated new standard, which may be extended later in subsequent work):

- 2018/04 Test model selection process begins
- 2018/10 Test model selection established
- 2020/10 Final standard completed

4 Test categories, coding conditions, and anchors

Test categories for standard dynamic range (SDR), high dynamic range (HDR) and 360° content are defined in the three sub-sections below. Proponents are encouraged (but not required) to submit results for all test categories. However, submitters are required to provide results for all test cases in a given test category.

4.1 Standard Dynamic Range

4.1.1 Video test sequence formats and frame rates

All test material is progressively scanned and uses 4:2:0 colour sampling with either 8 or 10 bits per sample.

The classes of video sequences are:

- | | |
|--------------|--|
| Class SDR-A: | 3840×2160p 60 fps 10 bits: "FoodMarket4", "CatRobot1", "DaylightRoad2" |
| | 3840×2160p 50 fps 10 bits: "ParkRunning3" |
| | 3840×2160p 30 fps 10 bits: "Campfire" |
| Class SDR-B: | 1920×1080p 50fps 8 bits: "BasketballDrive", "Cactus" |
| | 1920×1080p 60fps 8 bits: "BQTerrace" |
| | 1920×1080p 60fps 10 bits: "RitualDance", "MarketPlace" |

¹ People who formally registered will receive instructions regarding how to submit the coded materials. If material is received later, the proposal may be excluded from testing.

² Contact persons will provide information about the document submission process and the document submission template. The draft document submission template may be found as JVET-H1003 [3]. Note that submitted documents will be made publicly available. Exceptions to public availability will be considered on a case by case basis upon request by the contributor.

³ Proponents are requested to attend this standardization meeting. The starting date has not been finalized; participants will be notified.

4.1.2 Coding conditions of submissions

Constraint cases are defined as follows:

- Constraint Set 1: not more than 16 frames of structural delay, e.g. 16 pictures “group of pictures (GOP)”, and random access intervals of 1.1 seconds or less. A random access interval of 1.1 seconds or less shall be defined as 32 pictures or less for a video sequence with a frame rate of 24, 25 or 30 frames per second, 48 pictures or less for a video sequence with a frame rate of 50 frames per second, 64 pictures or less for a video sequence with a frame rate of 60 frames per second, and 96 pictures or less for a video sequence with a frame rate of 100 frames per second.
- Constraint Set 2: no picture reordering between decoder processing and output, with bit rate fluctuation characteristics and any frame-level multi-pass encoding techniques to be described with the proposal. (A metric to measure bit rate fluctuation is implemented in the Excel file to be submitted for each proposal.)

Submissions shall include encodings for all video sequences in all classes, and each decoding shall produce the full specified number of pictures for the video sequence (no missing pictures).

Submissions shall be made for the test cases (combinations of classes and constraint sets) listed in Table 1.

Table 1 – Combinations of classes and constraint sets for the Standard Dynamic Range category

	Class SDR-A	Class SDR-B
Constraint set 1	X	X
Constraint set 2	–	X

- Submissions for Class SDR-A using Constraint Set 1, and submissions for class SDR-B using Constraint Set 1 and Constraint Set 2 will be evaluated by means of a formal subjective assessment and BD PSNR and rate [4][5] criteria.

For Class SDR-A using Constraint Set 1 defined above, submissions to this Call shall provide results for the target bit rate listed in Table 2. The submitted results shall not exceed the target bit rate points. The calculation of bit rate is further defined in Annex B.

Table 2 – Target bit rate points not to be exceeded for Class SDR-A⁴

Sequences	Target bit rates [kbit/s]			
	Rate 1	Rate 2	Rate 3	Rate 4
FoodMarket4	950	1400	2200	3800
CatRobot1	1500	2400	3800	6000
DaylightRoad2	1400	2200	3500	5500
ParkRunning3	2000	3300	6000	10000
Campfire	2000	3300	6000	10000

⁴ 1 kbit/s means 10³ bits per second, and 1 Mbit/s means 10⁶ bits per second.

For Class SDR-B using Constraint Set 1 and Constraint Set 2 defined above, submissions to this Call shall provide results for the target rate points listed in Table 3. The submitted results shall not exceed the target bit rate points. The calculation of bit rate is further defined in [Annex B](#).

Table 3 – Target bit rate points not to be exceeded for Class SDR-B

Sequences	Target bit rates [kbit/s]			
	Rate 1	Rate 2	Rate 3	Rate 4
BQTerrace	400	600	1000	1500
RitualDance	900	1500	2300	3800
MarketPlace	500	800	1200	2000
BasketballDrive	800	1200	2000	3500
Cactus	500	800	1200	2000

In all evaluations, it is anticipated that subjective evaluation will have primary importance. However, the committee will determine the importance of the different criteria through its normal consensus process.

Submissions to this Call shall obey the following additional constraints:

1. No use of pre-processing.
2. Only use post-processing if it is part of the decoding process, i.e. any processing that is applied to a picture prior to its use as a reference for inter prediction of other pictures. Such processing can also be applied to non-reference pictures.
3. Quantization settings should be kept static except that a one-time change of the quantization settings to meet the target bit rate is allowed. When any change of quantization is used, it shall be described. This includes a description of the one-time change.
4. Proponents are discouraged from optimizing encoding parameters using non-automatic means as well as optimizing encoding parameters on a per sequence basis within a Class. When any optimization is employed, it shall be described.
5. Proponents shall not use any part of the video coding test set as a training set for training large entropy coding tables, VQ codebooks, etc.

4.1.3 Anchors

Two anchors are generated by encoding the above video sequences. The first anchor uses the HM 16.16 software package⁵. This anchor uses static quantization parameter (QP) settings, though a one-time change of the quantization parameter from value QP to value QP+1 is used for some bitstreams in order to meet the defined target bit rates. These quantization parameter settings will be reported. The anchors obey the coding conditions in section 4.1.2. Additionally, a higher rate point bit-stream will be created for the purpose of calculating objective metrics.

The second anchor uses the Joint Exploration Test Model (JEM) software package. The Joint Video Exploration Team (JVET) maintains the JEM software package in order to study coding tools in a

⁵ Information for accessing the anchors can be obtained from the contact persons. For details, proponents may refer to the config files prepared for the anchors, which are available at the same location as the anchor bitstreams. A summary of the coding conditions used in these config files is provided below.

coordinated test model [7]. JEM anchor bitstreams are generated using this software package and obey the coding conditions in section 4.1.2. The JEM 7.0 software package is used to generate the anchors, though a later version may be used if needed.

The purpose of the anchors is to facilitate testing in accordance with BT.500 [6], by providing useful reference points demonstrating the behaviour of well-understood configurations of current technology while obeying the same constraints as imposed on the proposals. The anchors will be among the encodings used in the testing process; however, the purpose of the test is to compare the quality of video for proposals to each other rather than to the anchors.

4.2 High Dynamic Range

4.2.1 Video test sequence formats and frame rates

All test material is progressively scanned and uses 4:2:0 colour sampling with 10 bits per sample

The classes of video sequences are:

- | | |
|--------------|---|
| Class HDR-A: | 3840x2160p 60 fps HLG: "DayStreet", "PeopleInShoppingCenter", "SunsetBeach" |
| Class HDR-B: | 1920x1080p 50 fps PQ: "Market3", "Hurdles", "Starting" |
| | 1920x1080p 25 fps PQ: "ShowGirl2" |
| | 1920x1080p 24 fps PQ: "Cosmos1" |

4.2.2 Coding conditions of submissions

The coding conditions in section 4.1.2 shall apply to the HDR test category with the following exception:

- The quantization settings do not need to be kept static. A static temporal quantizer structure shall be used. However, the quantizer may then be adjusted within a frame as a function of the local, average luma value and/or the local, average chroma value. A one-time change of the temporal quantizer structure is also allowed to meet the target bit rate. If either a local adjustment or a one-time change is used, a description of the adjustment scheme shall be provided in the descriptive document submission.
- A colour volume transform within the YCbCr 4:2:0 sample domain may be performed on the input video signal prior to coding, and an inverse of the colour volume transform may be performed on the signal after decoding. A colour volume transform modifies a luma or chroma sample at a relative location in the picture based only on the value of the luma or chroma sample and on the value of other component samples located at the same or closest relative location. The colour volume transform should be kept static within a frame and sequence. The use of a colour volume transform for the purpose of rate allocation or denoising is discouraged. If a colour volume transform is used, a description of the transform shall be provided in the descriptive document submission. Respondents are additionally asked to provide descriptions of any multi-pass processing and any dynamic change in the colour volume transform.

Submissions shall be made for the test cases (combinations of classes and constraint sets) listed in Table 4.

Table 4 – Combinations of classes and constraint sets for the High Dynamic Range category

	Class HDR-A	Class HDR-B
Constraint set 1	X	X
Constraint set 2	–	–

- Submissions for Class HDR-A and Class HDR-B will be evaluated by means of a formal subjective assessment.
- Submissions for Class HDR-A and Class HDR-B will be further evaluated by the following metrics: PSNR, weighted PSNR, deltaE100 and PSNR-L100, as well as the Bjøntegaard Delta-Rate and Delta-Distortion for each metric. Metric definitions are provided in Annex D.

For Class HDR-A using Constraint Set 1 defined above, submissions to this Call shall, for each of the test cases defined above, submit results for the target bit rate points listed in Table 5. The submitted results shall not exceed the target bit rate points. The calculation of bit rate is further defined in [Annex B](#).

Table 5 – Target bit rate points not to be exceeded for the HDR-A category⁶

Sequences	Target bit rates [kbit/s]			
	Rate 1	Rate 2	Rate 3	Rate 4
DayStreet	680	1100	1750	2800
PeopleInShoppingCenter	640	960	1450	2150
SunsetBeach	1500	3000	6000	10000

For Class HDR-B using Constraint Set 1 defined above, submissions to this Call shall, for each of the test cases defined above, submit results for the target bit rate points listed in Table 6. The submitted results shall not exceed the target bit rate points. The calculation of bit rate is further defined in [Annex B](#).

Table 6 – Target bit rate points not to be exceeded for the HDR-B category⁷

Sequences	Target bit rates [kbit/s]			
	Rate 1	Rate 2	Rate 3	Rate 4
Market3	600	900	1500	2400
ShowGirl2	350	500	750	1200
Hurdles	450	800	1500	2000
Starting	450	600	850	1400
Cosmos1	500	900	1500	3000

4.2.3 Anchors

The anchor conditions and descriptions in section 4.1.3 shall apply to the HDR-A and HDR-B test categories with the following change for HDR-B:

1. The generation of the anchor does not use a static quantization parameter (QP) setting. Instead, the configuration allows the QP value to vary spatially, where the variation is an

⁶ 1 kbit/s means 10³ bits per second, and 1 Mbit/s means 10⁶ bits per second.

⁷ 1 kbit/s means 10³ bits per second, and 1 Mbit/s means 10⁶ bits per second.

explicit function of the average, local luma value. A one-time change of the quantization parameter from value QP to value QP+1 may also be applied in order to meet the defined target bit rates. The quantization parameter settings applied for the anchors will be reported.

4.3 360° Video

4.3.1 Video test sequence formats and frame rates

All test material is progressively scanned and uses 4:2:0 colour sampling with 8 or 10 bits per sample.

The classes of video sequences are:

Class 360: 8192×4096 30 fps 10 bits: "ChairliftRide"
 8192×4096 30 fps 8 bits: "KiteFlite", "Harbor", "Trolley"
 6144×3072 60 fps 8 bits: "Balboa"

4.3.2 Coding conditions of submissions

The coding conditions in section 4.1.2 shall apply to the 360° video test category, with the following exceptions:

Submissions shall be made for the test cases (combinations of classes and constraint sets) listed in Table 7.

Table 7 – Combinations of classes and constraint sets for the 360° video category

	Class 360
Constraint set 1	X
Constraint set 2	–

- Submissions for Class 360 will be evaluated by means of a formal subjective assessment.
- Submissions for Class 360 will further be evaluated using the following objective metrics: E2E WS-PSNR and E2E S-PSNR-NN, as well as the Bjøntegaard Delta-Rate and Delta-Distortion for each metric. Metric definitions are provided in Annex E.

Submissions to this Call shall, for each of the test cases defined above, submit results for the target bit rate points (which are not to be exceeded) listed in Table 8. The calculation of bit rate is further defined in [Annex B](#).

Table 8 – Target bit rate points not to be exceeded for the 360° video category⁸

Sequences	Target bit rates [kbit/s]			
	Rate 1	Rate 2	Rate 3	Rate 4
Balboa	2000	3300	6000	10000
Chairlift	1500	2400	4000	7000
KiteFlite	1200	2400	4000	7000
Harbor	700	1200	2000	3300
Trolley	1500	2400	4000	7000

⁸ 1 kbit/s means 10³ bits per second, and 1 Mbit/s means 10⁶ bits per second.

Submissions to this Call shall obey the constraints in section 4.1.2 with the following change:

1. The quantization settings do not need to be kept static. Instead, the quantization settings may be adjusted within a picture as a function of the geometric position. If local adjustment is used, a description of the adjustment scheme shall be provided in the descriptive document submission.
2. Pre-processing may be used to perform a projection mapping operation, and post-filtering may be used to perform an inverse projection mapping operation. The projection mapping algorithms may allow dynamic changes within a sequence if an automatic selection algorithm is used. The same projection mapping operation and inverse projection mapping operation shall be used for all test sequences in the test case. If a projection mapping is used, a description of the projection mapping technique shall be provided in the descriptive document submission. Respondents are asked to provide information regarding at least: (i) the coded resolution of the projection map, (ii) the use of padding and blending, (iii) the use of global rotation, (iv) the use of multi-pass projection mapping, and (v) E2E WS-PSNR values comparing each test sequence to the result of applying the projection mapping algorithm and then converting this result back to the equirectangular projection format without compression. The coded resolution, defined as the number of samples provided as input to the encoder, shall be the same for all rate points in all sequences with the same original, spatial resolution.
3. Post-processing after decoding of seam artifacts near discontinuous edges is permitted in submissions but must be fully described. If post-processing of seam artifacts is performed, the post-processed pictures are used end-to-end objective metrics.

Decoder and/or projection format conversion binaries of submissions for Class 360 must decode and/or projection format convert each bitstream to a 4:2:0 YUV file of the same resolution as the input sequence, either 8192×4096 or 6144×3072. For subjective evaluation, a viewport using a previously undisclosed view path per sequence will be extracted from this decoded video using the 360Lib tool. The viewport will be generated by rectilinear projection using bilinear interpolation filters. The viewport will have a field of view of 78.1×49.1 degrees and a resolution of 1920×1080. The viewport will use progressive scanning and 4:2:0 colour sampling with 8 bits per sample.

4.3.3 Anchors

The anchor conditions and descriptions in section 4.1.3 shall apply to the 360° video category, with the following exceptions:

1. The equirectangular projection input pictures (at 8192×4096 or 6144×3072 resolution) will be spherically downsampled using the 360Lib software to 4096×2048 equirectangular projection format sequences representing the full 360°×180° sphere, and 8 luma samples of padding on each of the left and right sides on the picture will be added by copying samples, yielding 4112×2048 padded equirectangular projection pictures, in which the yaw range exceeds 360°, for encoding by the HM and JEM. After decoding, the 4112×2048 padded equirectangular projection pictures will use linear position-based weighted blending of the overlapping region to convert to 4096×2048 equirectangular projection format without padding, representing 360° of yaw range. 360Lib is used for spherical upsampling to the input resolution, 8K or 6K respectively. E2E objective metrics will use the 4096×2048 pictures with blending in their calculations.

5 Test sites and delivery of test material

The proposals submission material will be evaluated by means of a formal subjective assessment process. The tests will be conducted by the Test Coordinator and one or more sites. The sites used will be communicated to proponents upon request.

All proponents need to deliver, by the due date of 2018/02/15, an solid state drive (SSD) to the address of the Test Coordinator (see section 9). The disk shall contain the bitstreams, YUV files, and decoder executable⁹ used by the proponent to generate the YUV files from the bitstreams.

Reception of the disk will be confirmed by the Test Coordinator. Any inconvenience caused by unexpected delivery delay or a failure of the disk will be under the complete responsibility of the proponents. However, solutions will be negotiated to ensure that the data can still be included in the test if feasible, which means that correct and complete data needs to be available before the beginning of the test at the latest.

The MD5 checksums of all bitstreams, decoder executables, and YUV files shall be provided in a text file included on the disk in order to verify that the data can be read without corruption.

Further technical details on the delivery of the coded material are provided in [Annex B](#).

6 Testing fee

Proponents will be charged a fee per each category to which an algorithm proposal is submitted (but not multiple responses to a single category). The fee will be USD 6,000 per category. Such fee will cover the logistical costs without any profit. The fee is non-refundable after the formal registration is made.

7 Requirements on submissions

More information about file formats can be found in Annex B. Files of decoded sequences and bitstreams shall follow the naming conventions as specified in section C-8 of [Annex C](#).

Proponents shall provide the following; incomplete proposals will not be considered:

- A) Coded test material submission to be received on hard disk drive by 15 February 2018:
 - 1. Bitstreams for all test cases specified in Table 1, Table 4, and/or Table 7 and all bit rates as specified in Table 2 and Table 3, Table 5, Table 6, and/or Table 8.
 - 2. Decoded sequences (YUV files) for all test cases as specified in Table 1, Table 4, and/or Table 7 and all bit rates as specified in Table 2 and Table 3, Table 5, Table 6, and/or Table 8.
 - 3. Binary decoder executable.
 - 4. MD5 checksum files for A1-A3.
- B) Coded test material to be brought for the meeting in April 2018:
 - 1. Bitstreams for all test cases as specified in Table 1, Table 4, and/or Table 7 and all bit rates as specified in Table 2 and Table 3, Table 5, Table 6, and/or Table 8.
 - 2. Decoded sequences (YUV files) for all test cases as specified in Table 1, Table 4, and/or Table 7 and all bit rates as specified in Table 2 and Table 3, Table 5, Table 6, and/or Table 8.
 - 3. Binary decoder executable.
 - 4. MD5 checksum files for B1–B3.

⁹ The decoder executable must be able to perform the decoding operation on a computer with a Windows 10 operating system.

- C) Document to be submitted before the evaluation meeting in April 2018 shall contain the following:
1. A technical description of the proposal sufficient for full conceptual understanding and generation of equivalent performance results by experts and for conveying the degree of optimization required to replicate the performance. This description should include all data processing paths and individual data processing components used to generate the bitstreams. It does not need to include complete bitstream format or implementation details, although as much detail as possible is desired.
 2. An Excel sheet as will be provided to proponents by the contact persons (a draft contribution submission template may be found as JVET-H1003 [3]), with all white fields for the respective test cases filled. For objective metric values to be computed per picture, a precision of 2 digits after decimal point shall be used. BD measures [4][5] against the appropriate anchor will be automatically computed from the Excel sheets at the meeting where the evaluation is performed.
 3. The technical description shall also contain a statement about the programming language in which the software is written, e.g. C/C++ and platforms on which the binaries were compiled.⁹
 4. The technical description shall state how the proposed technology behaves in terms of random access to any picture within the sequence. For example, a description of the GOP structure and the maximum number of pictures that must be decoded to access any picture could be given.
 5. The technical description shall specify the expected encoding and decoding delay characteristics of the technology, including structural delay e.g. due to the amount of picture reordering and buffering, the degree of picture-level multi-pass decisions and the degree by which the delay can be minimized by parallel processing.
 6. The technical description shall contain information suitable to assess the complexity of the implementation of the technology, including the following but not limited to:
 - Encoding time¹⁰ (for each submitted bitstream) of the software implementation. Proponents shall provide a description of the platform and methodology used to determine the time. To help interpretation, a description of software and algorithm optimizations undertaken, if any, is welcome.
 - Decoding time¹¹ for each bitstream running the software implementation of the proposal, and for the corresponding constraint case anchor bitstream(s)¹² run on the same platform. Proponents shall provide a description of the platform and methodology used to determine the time. To help interpretation, a description of software optimisations undertaken, if any, is encouraged.
 - Expected memory usage of encoder and decoder.
 - Complexity characteristics of tools using inter picture redundancy, such as Motion Estimation (ME) / Motion Compensation (MC): E.g. number of reference pictures, sizes of picture (and associated decoder data) memories, sample value wordlength, block size, and motion compensation interpolation filter(s).

¹⁰ For example, using `ntimer` for Windows systems.

¹¹ For example, using `ntimer` for Windows systems.

¹² The decoder source code to be used to process the anchor bitstreams will be provided to proponents and must be compiled as-is, without modification of source code, compiler flags, or settings.

- Description of transform(s): use of integer/floating point precision, transform characteristics (such as length of the filter/block size), number of different transforms used, any adaptation of the transform(s).
- Degree of capability for parallel processing.

Furthermore, the technical description should point out any specific properties of the proposal (e.g. additional functionality such as benefit for 4:4:4 coding, error resilience, scalability, multi-view coding, screen content coding, or lightfield coding).

D) Optional information

Proponents are encouraged (but not required) to allow other committee participants to have access, on a temporary or permanent basis, to their encoded bitstreams and binary executables or source code.

8 Subsequent provision of source code and IPR considerations

Proponents are advised that, upon acceptance for further evaluation, it will be required that certain parts of any technology proposed be made available in source code format to participants in the core experiments process and for potential inclusion in the prospective standard as reference software. When a particular technology is a candidate for further evaluation, commitment to provide such software is a condition of participation. The software shall produce identical results to those submitted to the test. Additionally, submission of improvements (bug fixes, etc.) is certainly encouraged.

Furthermore, proponents are advised that this Call is being made subject to the common patent policy of ITU-T/ITU-R/ISO/IEC (see <http://www.itu.int/en/ITU-T/ipr/Pages/policy.aspx> or ISO/IEC Directives Part 1, Appendix I) and the other established policies of the standardization organizations.

9 Contacts

Prospective contributors of responses to the Call for Proposals should contact the following people:

Jens-Rainer Ohm
 RWTH Aachen University, Institute of Communications Engineering
 Melatener Str. 23, 52074 Aachen, Germany
 Tel. +49-241-8027671, Fax. +49-241-8022196, email ohm@ient.rwth-aachen.de

Gary Sullivan
 Microsoft Corporation
 One Microsoft Way, Redmond, WA 98052
 Tel. +1-425-703-5308, Fax. +1-425-936-7329, email garysull@microsoft.com

Vittorio Baroncini (Test Coordinator)
 Technical Director
 GBTech
 Viale Castello della Magliana, 38, 00148 – Rome, Italy
 Tel. +39-3335474643, email baroncini@gmx.com

10 References

- [1] “Requirements for Future Video Coding (FVC)”, ITU-T SG16/Q6 VCEG 56th meeting, Torino, Italy, July 2017, Doc. VCEG-BD03 (available at http://wftp3.itu.int/av-arch/video-site/1707_Tor/).
- [2] “Requirements for a Future Video Coding Standard v5”, ISO/IEC JTC1/SC29/WG11 MPEG, 119th meeting, Torino, Italy, July 2017, Doc. N17074 (available at <http://mpeg.chiariglione.org/meetings/119>).

- [3] “Draft template of proposal description document”, Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11), 8th Meeting, Macao, China, October 2017, Doc. JVET-H1003 (available at <http://phenix.it-sudparis.eu/jvet/>).
- [4] Gisle Bjontegaard, “Calculation of Average PSNR Differences between RD curves”, ITU-T SG16/Q6 VCEG 13th meeting, Austin, Texas, USA, April 2001, Doc. VCEG-M33 (available at http://wftp3.itu.int/av-arch/video-site/0104_Aus/).
- [5] Gisle Bjontegaard, “Improvements of the BD-PSNR model”, ITU-T SG16/Q6 VCEG 35th meeting, Berlin, Germany, 16–18 July, 2008, Doc. VCEG-AI11 (available at http://wftp3.itu.int/av-arch/video-site/0807_Ber/).
- [6] International Telecommunication Union – Radiocommunication Sector, *Methodology for the subjective assessment of the quality of television pictures*, Recommendation ITU-R BT.500-13 (available at <https://www.itu.int/rec/R-REC-BT.500>).
- [7] “Algorithm Description of Joint Exploration Test Model 7 (JEM7)”, Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11), 7th Meeting, Torino, Italy, July 2017, Doc. JVET-G1001 (available at <http://phenix.it-sudparis.eu/jvet/>).
- [8] “Algorithm descriptions of projection format conversion and video quality metrics in 360Lib Version 4”, Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11), 7th Meeting, Torino, Italy, July 2017, Doc. JVET-G1003 (available at <http://phenix.it-sudparis.eu/jvet/>).

Annex A: Detailed description of test sequences

Table A-1 – SDR-A test sequence example pictures

 FoodMarket4	 CatRobot1	 DaylightRoad2	 ParkRunning3
 Campfire			

Table A-2 – SDR-B test sequence example pictures

 BQTerrace	 RitualDance	 MarketPlace	 BasketBallDrive
 Cactus			

Table A-3 – SDR test sequences

Sequence name	Resolution	Frame count	Frame rate	Chroma format	Bit depth
FoodMarket4	3840×2160	720	60	4:2:0	10
CatRobot1	3840×2160	600	60	4:2:0	10
DaylightRoad2	3840×2160	600	60	4:2:0	10
ParkRunning3	3840×2160	500	50	4:2:0	10
Campfire	3840×2160	300	30	4:2:0	10
BQTerrace	1920×1080	600	60	4:2:0	8
RitualDance	1920×1080	600	60	4:2:0	10
MarketPlace	1920×1080	600	60	4:2:0	10
BasketballDrive	1920×1080	500	50	4:2:0	8
Cactus	1920×1080	500	50	4:2:0	8

Table A-4 – SDR test sequence md5sums

Sequence name	Clip Identifier	MD5Sum	Source (or Owner)
FoodMarket4	S01	a378b34190f54f688d048a9a8b46a8ac	Netflix
CatRobot1	S02	03a89792693fd9ecfd72ef2590025e97	B<>COM
DaylightRoad2	S03	bf1d22643afb41b1521749d2749fb5f0	Huawei
ParkRunning3	S04	e7a1d1ebff269767ec4bffd2998d1154	Huawei
Campfire	S05	63d3d9f9e4e8b5c344e89840e84e6428	SJTU
BQTerrace	S10	efde9ce4197dd0b3e777ad32b24959cc	NTT DOCOMO Inc.
RitualDance	S11	a3cb399a7b92eb9c5ee0db340abc43e4	Netflix
MarketPlace	S12	dc668e7f28541e4370bdbdd078e61bba	Netflix
BasketballDrive	S13	d38951ad478b34cf988d55f9f1bf60ee	NTT DOCOMO Inc.
Cactus	S14	3fddb71486f209f1eb8020a0880ddf82	EBU/RAI

Note – Copyright information is included in the sequence ZIP container.

Table A-5 – HDR-A test sequence example pictures

 <p>DayStreet</p>	 <p>PeopleInShoppingCenter</p>	 <p>SunsetBeach</p>
--	---	--

Table A-6 – HDR-B test sequence example pictures

 <p>Market3</p>	 <p>ShowGirl2</p>	 <p>Hurdles</p>	 <p>Starting</p>
 <p>Cosmos1</p>			

Table A-7 – HDR test sequences

Sequence name	Resolution	Frame count	Frame rate	Chroma format	Bit depth
DayStreet	3840x2160	600	60	4:2:0	10
PeopleInShoppingCenter	3840x2160	600	60	4:2:0	10
SunsetBeach	3840x2160	600	60	4:2:0	10
Market3	1920x1080	400	50	4:2:0	10
ShowGirl2	1920x1080	339	25	4:2:0	10
Hurdles	1920x1080	500	50	4:2:0	10
Starting	1920x1080	500	50	4:2:0	10
Cosmos1	1920x856	240	24	4:2:0	10

Note: The capture frame rate of the Hurdles and Starting sequences was 100fps. However, these sequences are treated as 50fps sequences for the evaluation processes defined in this document.

Table A-8 – HDR test sequence md5sums

Sequence name	Clip Identifier	MD5Sum	Source (or Owner)
DayStreet	S31	2bd7bbcfb80ab329118a0a82101e2715	SONY
PeopleInShoppingCenter	S32	40395778572368438c0202fe3c017fa9	SONY
SunsetBeach	S33	7cc25ae4249203427c2f83db2d9f09f7	NHK
Market3	S41	c97abe47455fd12f6d6436cecfad7c7d	Technicolor
ShowGirl2	S42	44f1974d68f7799c71eea29fb72b245b	Stuttgart Media University
Hurdles	S43	bc3cba849d6f4ee74d39056600722aa5	EBU
Starting	S44	1cbc416696cb0dfcf4da9886eeb6a4a2	EBU
Cosmos1	S45	da4a2488c249720da0535f01c3693efa	Netflix

Note – Copyright information is included in the test sequence ZIP container.

Table A-9 – 360° test sequences

		
Balboa	ChairliftRide	KiteFlite
		
Harbor	Trolley	

Table A-10 – 360° video test sequences

Sequence name	Input resolution	Anchor resolution	Coded luma sample count of anchors	Frame count	Frame rate	Chroma format	Bit depth
Balboa	6144x3072	4096x2048	8388608	600	60	4:2:0	8
ChairliftRide	8192x4096	4096x2048	8388608	300	30	4:2:0	10
KiteFlite	8192x4096	4096x2048	8388608	300	30	4:2:0	8
Harbor	8192x4096	4096x2048	8388608	300	30	4:2:0	8
Trolley	8192x4096	4096x2048	8388608	300	30	4:2:0	8

Note – The sequences are omnidirectional 360° × 180° degree video and are stored in an equirectangular projection (ERP) format. The number of coded luma samples in the anchor is lower than the resolution of the input sequence.

Table A-11 – 360° video test sequence md5sums

Sequence	Clip Identifier	MD5Sum	Source (or Owner)
Balboa	S51	1457bb109ae0d47265f5a589cb3464d7	InterDigital
ChairliftRide	S52	9126f753bb216a73ec7573ecc4a280c3	GoPro
KiteFlite	S53	18c0ea199b143a2952cf5433e8199248	InterDigital
Harbor	S54	aa827fdd01a58d26904d1dbdbd91a105	InterDigital
Trolley	S55	25c1082d1e572421da2b16530718156d	InterDigital

Note – Copyright information is included in the sequence ZIP container.

Annex B:

Distribution formats for test sequences and decoded results, delivery of bitstreams and binary decoders, utilities and cross- check meeting day

Distribution of original video material files containing test sequences is done in YUV files with extension “.yuv”. A description of the YUV file format is available at <http://www.fourcc.org/> web site, designated as “yuv420p10le iyuv”.

HEVC Anchor bitstreams are provided with extension “.hevc”. JEM Anchor bitstreams are provided with extension “.jem”. Bitstream formats of proposals can be proprietary, but must contain all information necessary to decode the sequences at a given data rate (e.g. no additional parameter files). The file size of the bitstream will be used as a proof that the bit rate limitation from Table 2, Table 3, Table 5, Table 6, or Table 8 has been observed. The file extension of a proposal bitstream shall be “.bit”.

Decoded sequences shall be provided in the same “.yuv” format as originals, with the exception that the colour depth shall be 10 bits per component for all sequences.

All files delivered (bitstreams, decoded sequences and binary decoders) must be accompanied by an MD5 checksum file to enable identification of corrupted files. An MD5 checksum tool that may be used for that purpose is typically available as part of UNIX/LINUX operating systems; if this tool is used, it should be run with option “-b” (binary). For the Windows operating systems, a compatible tool can be obtained from <http://www.pc-tools.net/win32/md5sums/>. This tool should be run with additional option “-u” to generate the same output as under UNIX.

For the 360° video category, a binary projection format convertor may be provided, to be applied to the decoded YUV files. The output of the binary decoder, or the projection format convertor, if provided, must decode and/or convert each bitstream to a 4:2:0 YUV file of the same resolution as the input sequence.

The SSD should be shipped (for handling in customs) with a declaration “used hard disk drive for scientific purposes, to be returned to owner” and low value specification (e.g. EUR 20). The use of a SSD with substantially larger size than needed is discouraged. The hard disk drive should be a portable SSD with a USB 3.1 interface . NTFS file format shall be used.

Before the evaluation meeting sessions, a one-day cross-check meeting will be held. Proponents shall bring another portable SSD with a USB 3.1 interface that can be connected to a Windows PC and contains original and decoded sequences in YUV format, bitstreams, binary decoder executables and all related checksum files. An adequate computer system shall also be brought to this meeting. Proponents shall specify the computing platform (hardware, OS version) on which the binary can be run. Should such a computing platform not be readily available, the proponent shall provide a computer adequate for decoder verification at this meeting. Further information will be exchanged with the proponents after the registration deadline.

Annex C:

Description of testing environment and methodology

The test procedure foreseen for the formal subjective evaluation will consider two main requirements:

- to be as much as possible reliable and effective in ranking the proposals in terms of subjective quality;
- to take into account the evolution of technology and laboratory set-up oriented to the adoption of FPD (Flat Panel Display) and video server as video recording and playing equipment.

Therefore, the test methods will be DSIS (Double Stimulus Impairment Scale; ITU-R BT.500) also known as DCR (Degradation Category Rating; ITU-T P.910)

C.1 DSIS

This test method is commonly adopted when the material to be evaluated shows a range of visual impairment that well distributes across all of the grading scale.

This method will be used under the schema of evaluation of the impairment, using an impairment scale made of 11 levels, ranging from "0" (lowest quality) to "10" (highest quality). The test will be held in multiple laboratories located in countries speaking different languages: This implies that it is better not to use categorical adjectives (e.g. excellent, good, fair etc.) to avoid any bias due to a possible different interpretation by naive subjects speaking different languages.

All the video material used for these tests will consist of video clips of 10 seconds duration.

The structure of the Basic Test Cell (BTC) of DSIS method is made by two consecutive presentations of the video clip under test; at first the original version of the video clip (SRC, Source Reference Content) is displayed, announced by the letter "A" on a mid gray screen (half a second), immediately afterwards the coded version of the video clip (PVS, Processed Video Sequence), announced by the letter "B" on a mid gray screen (half a second), is presented; then a message displays for 5 seconds asking the viewers to vote (see Figure C-1)



Figure C-1 – DSIS BTC

C.2 How to express the visual quality opinion with DSIS

The viewers will be asked to express their vote by putting a number in a box on a scoring sheet.

The scoring sheet for a DSIS test is made of a section for each BTC; each section is made of a numbered box (see Figure C-2).

The viewers will write a number in the box that corresponds to the number of the message "**Vote N**" that is shown on the screen. For example, writing "10" the subject will declare that he/she saw no difference between the SRC and the PVS, while writing "0" the subject will declare that he/she saw a significant and clearly visible amount of difference between the SRC and the PVS across the entire observed image.

The vote has to be written when the message "Vote N" appears on the screen. The number "N" is a numerical progressive indication on the screen aiming to help the viewing subjects to use the appropriate column of the scoring sheet.

Test name													
Session #													
1	2	3	4	5	6	7	8	9	10				
<input type="checkbox"/>													
11	12	13	14	15	16	17	18	19	20				
<input type="checkbox"/>													
21	22	23	24	25	26	27	28	29	30				
<input type="checkbox"/>													
31	32	33	34	35	36	37	38	39					
<input type="checkbox"/>													
								seat		subject			
1	2	3	4					<input type="checkbox"/>		<input type="checkbox"/>			
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					<input type="checkbox"/>		<input type="checkbox"/>			

Figure C-2 – Example of DSIS test method scoring sheet

C.3 Training and stabilization phase

The outcome of a test is highly dependent on a proper training of the test subjects.

For this purpose, each subject has to be trained by means of a short practice (training) session.

The video material used for the training session will be carefully selected from those of the test, taking care to represent as much as possible the extent of visual quality and the kind of impairment expected to appear during the test.

The stabilization phase uses the test material of a test session; three BTCs, containing one sample of best quality, one of the worst quality and one of medium quality, are duplicated at the beginning of the test session. By this way, the test subjects have an immediate impression of the quality range they are expected to evaluate during that session.

The scores of the stabilization phase are discarded. Consistency of the behaviour of the subjects will be checked inserting in the session a BTC in which original is compared to original.

C.4 The laboratory setup

The laboratory for a subjective assessment is planned to be set up following as much as possible the more relevant ITU Recommendations (e.g. ITU-R BT.500 and ITU-T P.910).

For the SDR and 360 tests, high quality TV sets will be used, taking care of disabling (as much as possible) all the internal local post processing of the images (e.g noise reduction etc.).

For the HDR tests, an appropriate display will be use (e.g. for the SONY BVM X300 for HDR-A, and the SIM2 for HDR-B)

The video play server, or the PC used to play video has to be able to support the display of both and UHD video formats, at 24, 30, 50 and 60 frames per second, without any limitation, or without introducing any additional temporal or visual artefacts.

The MUP player will be used for 360, SDR and HDR tests.

C.4.1 Viewing distance, seats and monitor size

For the SDR and 360 tests a 55” commercially available OLED TV set will be used (higher diagonal dimension is allowed).

For the HDR-A test a 30” monitor will be used; for the HDR-B test a 47” monitor will be used.

The viewing distance will be 2H, where H is equal to the height of the active part of the screen.

Four subjects will be seated in front of a 55” display; three subjects will be seated in front of a 47” display, and two subjects will be seated in front of a 30” display.

C.4.2 Viewing environment

The test laboratory has to be carefully protected from any external visual or audio pollution.

Internal general light has to be low (just enough to allow the viewing subjects to fill out the scoring sheets) and a uniform light has to be placed behind the monitor; this light will have an intensity of maximum 30 nits. No light source has to be directed to the screen or create reflections; ceiling, floor and walls of the laboratory have to be made of non-reflecting black or dark grey material (e.g. carpet or velvet).

C.5 Example of a test schedule for a day

This is an example of the schedule of the planned test activity for one day.

A time slot of one hour is dedicated every morning and afternoon to welcome, to screen and to train the viewers.

After the subjects’ screening for visual acuity and colour blindness, the subjects will be grouped in testing groups.

In the following example, four groups of subjects are created, according to the laboratory set-up¹³ and to the time constraints.

	DAY 1	DAY 2
9:00 – 10:00	Screening / training	Screening / training
10:00 -10:40	G1-S1	G1-S1
10:40 -11:20	G2-S1	G2-S1
11:20 -12:00	G1-S2	G1-S2
12:00 -12:40	G2-S2	G2-S2
13:00 – 14:00	Screening / training	Screening / training
14:00 -14:40	G3-S3	G3-S3
14:40 -15:20	G4-S3	G4-S3
15:20 -16:00	G3-S4	G3-S4
16:00 -16:40	G4-S4	G4-S4

Figure C-3 – Example of test schedule (one day)

C.6 Statistical analysis and presentation of the results

The data collected from the score sheets, filled out by the viewing subjects, will be stored in an Excel spreadsheet.

¹³ The viewing rooms of the laboratories could be different according to the test material and/or to the design of the laboratory. Large displays (e.g. monitor equal or wider than 50”) will allow to seat three (or more) subjects at the same time; a laboratory set-up, in which three wide monitors are available, will allow the creation of wider groups of viewers (three or more).

For each coding condition the Mean Opinion Score (MOS) and associated Confidence Interval (CI) values will be given in the spreadsheets.

The MOS and CI values will be used to draw graphs. The graphs will be drawn grouping the results for each video test sequence. No graph grouping results from different video sequences will be considered.

C.8 Proponents identification and file names

Each Proponent submitting to the CfP will be identified with a two digit code preceded by the letter “P” (e.g. P01 P02 ... Pnn).

Each coded video file provided for a submission will be identified by a name formed by the below listed combination of letters and numbers:

PnnSxxCzRy.<filetype>

where:

- Pnn identifies the Proponent
- Sxx identifies the original video clip used to produce the coded video, as identified in the tables of [Annex A](#)
- Cz identifies the constraint set z (z=1 or z=2), as identified in Table 1, Table 4 and Table 7
- Ry identifies the bit rate y, as identified in Table 2, Table 3, Table 5, Table 6, or Table 8
- <filetype> identifies the kind of file:
 - o .bit = bitstream
 - o .yuv = decoded video clip in YUV format

Annex D: Description of HDR video category test metrics

In this Annex, further information about the wPSNR and deltaE100, and PSNR-L100 metrics are provided. The metrics deltaE100 and PSNR-L100 shall be calculated using the HDRMetrics v0.15 utility available at <https://gitlab.com/standards/HDRTools>.

D.1 wPSNR

The wPSNR, or weighted PSNR, metric is calculated from the weighted mean squared error of the pixel values. An implementation of the metric is provided in the JEM software, and that implementation shall be used for calculation of the metric using the weighting functions provided below. The metric shall be computed in the PQ transform domain, which means an HLG sequence should be converted to the PQ domain for the calculation of the metric. Configuration files for this conversion will be provided at the same locations as the anchor bitstreams.

As an informative description of the metric follows:

The wPSNR metric is calculated as:

$$wPSNR = 10 * \log \frac{X^2}{wMSE},$$

where X is the maximum pixel value for the specific bit depth and $wMSE$ is given as

$$wMSE = \sum_{all\ pixels\ i\ in\ block} w_i \left(luma(x_{orig,i}) \right) * (x_{orig,i} - x_{dec,i})^2,$$

where $w_i \left(luma(x_{orig,i}) \right)$ is a weight that is a function of the luma value corresponding to the pixel $x_{orig,i}$, $x_{orig,i}$ is the original value at location i , and $x_{dec,i}$ is the reconstructed value at location i .

The calculation of the weight is computed as:

$$\begin{aligned} y_i &= 0.015 * luma(x_{orig,i}) - 1.5 - 6; \\ y_i &= y_i < -3 ? -3 : (y_i > 6 ? 6 : y_i); \\ w_i(luma(x_{orig,i})) &= pow(2.0, y_i \div 3.0); \end{aligned}$$

In all cases, the metric is calculated individually for a single luma or chroma channel and then used to compute a Bjøntegaard Delta-Rate and Delta-PSNR rate.

D.2 deltaE100-based metric

An implementation of the deltaE100 distortion metric is provided in the HDRTools software, and that implementation shall be used for calculation of the metric using the HDRMetric utility and configuration provided with the HDRTools software. Configuration files for computing the deltaE100 metric for HDR-A content will be provided at the same locations as the anchor bitstreams.

An informative description of the metric follows.

The original and test material must first be converted to linear-light 4:4:4 RGB EXR. For example, if the material is in the YCbCr BT.2100 4:2:0 PQ 10 bit format, it must be converted to a 4:4:4 RGB BT.2100 OpenEXR file using configuration provided with the HDRTools software.

Subsequently, the following steps should be applied for each (R,G,B) sample, within the content to be compared, i.e. original source (content 1) and test material (content 2):

- Convert the content to the XYZ colour space
- Convert the content from the XYZ to Lab space using the following equations. Computations are performed using double floating point precision.

- o $\text{invY}_n = 1.0 \div Y_n$, with $Y_n = 100$
- o $\text{invX}_n = \text{invY}_n \div 0.95047$
- o $\text{invZ}_n = \text{invY}_n \div 1.08883$
- o $y_{\text{Lab}} = \text{convToLab}(y * \text{invY}_n)$
- o $L = 116.0 * y_{\text{Lab}} - 16.0$
- o $a = 500.0 * (\text{convToLab}(x * \text{invX}_n) - y_{\text{Lab}})$
- o $b = 200.0 * (y_{\text{Lab}} - \text{convToLab}(z * \text{invZ}_n))$

with $\text{convToLab}(x)$ is defined as

if $x \geq 0.008856$

$$\text{convToLab}(x) = x^{(1/3)}$$

else

$$\text{convToLab}(x) = 7.78704 * x + 0.137931$$

- The deltaE100 distance DE between two samples (L1,a1,b1) and (L2,a2,b2) is then computed as follows:

$$c_{\text{Ref}} = \sqrt{a_1 * a_1 + b_1 * b_1}$$

$$c_{\text{In}} = \sqrt{a_2 * a_2 + b_2 * b_2}$$

$$cm = (c_{\text{Ref}} + c_{\text{In}}) \div 2.0$$

$$g = 0.5 * (1.0 - \sqrt{cm^{7.0} \div (cm^{7.0} + 25^{7.0})})$$

$$a_{\text{PRef}} = (1.0 + g) * a_1$$

$$a_{\text{PIn}} = (1.0 + g) * a_2$$

$$c_{\text{PRef}} = \sqrt{a_{\text{PRef}} * a_{\text{PRef}} + b_1 * b_1}$$

$$c_{\text{PIn}} = \sqrt{a_{\text{PIn}} * a_{\text{PIn}} + b_2 * b_2}$$

$$h_{\text{PRef}} = \arctan(b_1, a_{\text{PRef}})$$

$$h_{\text{PIn}} = \arctan(b_2, a_{\text{PIn}})$$

$$\text{deltaLp} = L_1 - L_2$$

$$\text{deltaCp} = c_{\text{PRef}} - c_{\text{PIn}}$$

$$\text{deltaHp} = 2.0 * \sqrt{c_{\text{PRef}} * c_{\text{PIn}}} * \sin((h_{\text{PRef}} - h_{\text{PIn}}) \div 2.0)$$

$$l_{\text{pm}} = (L_1 + L_2) \div 2.0$$

$$\text{cpm} = (\text{cPRef} + \text{cPIn}) \div 2.0$$

$$\text{hpm} = (\text{hPRef} + \text{hPIn}) \div 2.0$$

$$\text{rC} = 2.0 * \text{sqrt}(\text{cpm}^{7.0} \div (\text{cpm}^{7.0} + 25^{7.0}))$$

$$\text{dTheta} = \text{DEG30} * \text{exp}(-((\text{hpm} - \text{DEG275}) \div \text{DEG25}) * ((\text{hpm} - \text{DEG275}) \div \text{DEG25}))$$

$$\text{rT} = -\text{sin}(2.0 * \text{dTheta}) * \text{rC}$$

$$\text{t} = 1.0 - 0.17 * \text{cos}(\text{hpm} - \text{DEG30}) + 0.24 * \text{cos}(2.0 * \text{hpm}) +$$

$$0.32 * \text{cos}(3.0 * \text{hpm} + \text{DEG6}) - 0.20 * \text{cos}(4.0 * \text{hpm} - \text{DEG63})$$

$$\text{sH} = 1.0 + (0.015 * \text{cpm} * \text{t})$$

$$\text{sC} = 1.0 + (0.045 * \text{cpm})$$

$$\text{sL} = 1.0 + (0.015 * (\text{lpm} - 50) * (\text{lpm} - 50) \div \text{sqrt}(20 + (\text{lpm} - 50) * (\text{lpm} - 50)))$$

$$\text{deltaLpSL} = \text{deltaLp} \div \text{sL}$$

$$\text{deltaCpSC} = \text{deltaCp} \div \text{sC}$$

$$\text{deltaHpSH} = \text{deltaHp} \div \text{sH}$$

$$\text{deltaE100} = \text{sqrt}(\text{deltaLpSL} * \text{deltaLpSL} + \text{deltaCpSC} * \text{deltaCpSC} + \text{deltaHpSH} * \text{deltaHpSH} + \text{rT} * \text{deltaCpSC} * \text{deltaHpSH})$$

with

$$\text{DEG275} = 4.7996554429844$$

$$\text{DEG30} = 0.523598775598299$$

$$\text{DEG6} = 0.1047197551196598$$

$$\text{DEG63} = 1.099557428756428$$

$$\text{DEG25} = 0.436332$$

After this process, the deltaE100 values for each frame are averaged within the “Distortion” specified window. Finally, a PSNR based value is derived as:

$$\text{PSNR_DE} = 10 * \text{log}_{10}(\text{PeakValue} \div \text{deltaE100})$$

where PeakValue is set to 10,000.

D.3 PSNR-L100

An implementation of the PSNR-L100 distortion metric is provided in the HDRTools software, and that implementation shall be used for calculation of the metric using the HDRMetric utility and configuration provided with the HDRTools software. Configuration files for computing the PSNR-L100 metric for HDR-A content will be provided at the same locations as the anchor bitstreams.

An informative description of the metric follows:

PSNR-L100 represents the distortion in the lightness domain of the CIELab colour space. The derivation of Lab values from the linear representation of the signal is similar as given in the description of deltaE100. The mean absolute error (MAE) in the L domain is used to compute the PSNR-L100 as follows:

$$\text{PSNR-L100} = 10 * \log_{10}(\text{PeakValue} \div \text{MAE})$$

where PeakValue is set to 10,000.

Annex E: Description of 360° video objective test metrics

The following objective metrics will be provided for sequences in the 360° video category: E2E WS-PSNR and E2E S-PSNR-NN. Figure E-1 illustrates the processing chain for computation of objective metrics.

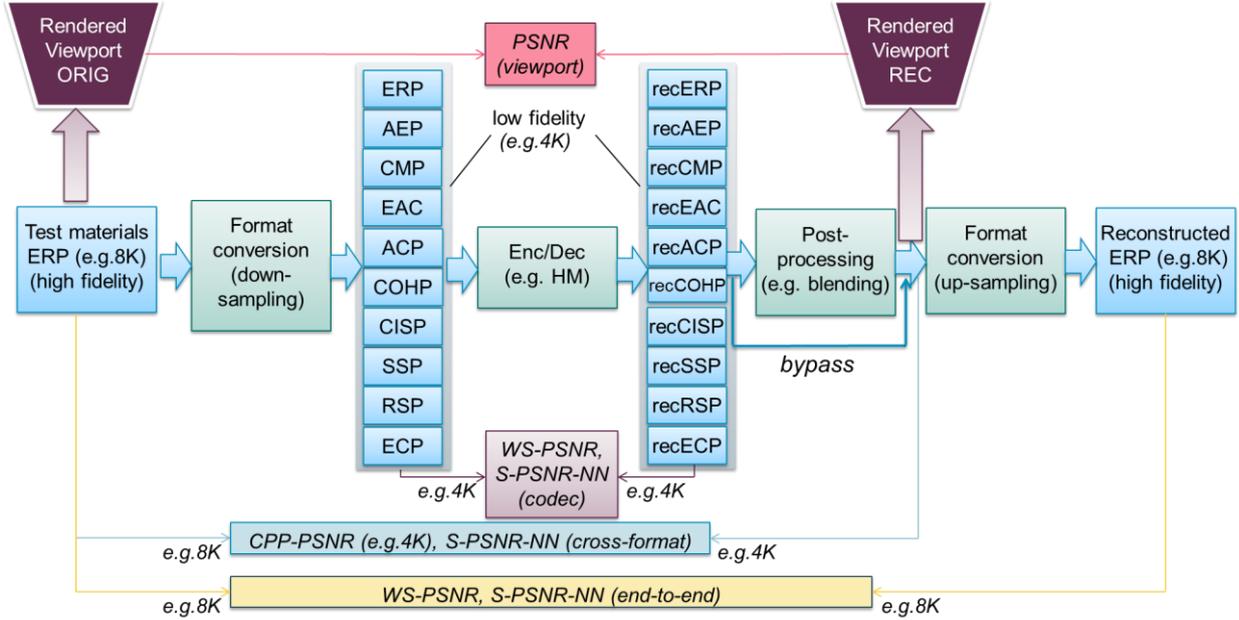


Figure E-1 – Processing chain for 360° video objective metrics

E.1 WS-PSNR

WS-PSNR calculates PSNR using all image samples on the 2D projection plane. The distortion at each position (i, j) is weighted by the spherical area covered by that sample position. For each position (i, j) of an $M \times N$ image on the 2D projection plane, denote the sample values on the reference and test images as $y(i, j)$ and $y'(i, j)$, respectively, and denote the spherical area covered by the sample as $w(i, j)$. The weighted mean squared error (WMSE) is first calculated as:

$$WMSE = \frac{1}{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} w(i, j)} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (y(i, j) - y'(i, j))^2 * w(i, j) \quad (E-1)$$

The WS-PSNR is then calculated as:

$$WS_PSNR = 10 \log \left(\frac{MAX_I^2}{WMSE} \right) \quad (E-2)$$

where MAX_I is the maximum intensity level of the images.

Since WS-PSNR is calculated based on the projection plane, different weights are derived for different projection formats. A description of the WS-PSNR weight derivation is to be provided for the projection formats in the submissions.

E2E WS-PSNR metric will use the ERP format at the same resolution as the input sequence.

For an $M \times N$ image in the ERP format, the weight $w(i, j)$ at position (i, j) is calculated as:

$$w(i, j)_{ERP} = \cos \frac{(j + 0.5 - N/2)\pi}{N} \quad (\text{E-3})$$

E.2 S-PSNR-NN

S-PSNR-NN uses a set of 655,362 uniformly sampled positions on the sphere [8] to measure 360-degree video quality. S-PSNR-NN is calculated using the following steps:

1. For a point “s” on the sphere, apply 3D-to-2D coordinate mapping to find the corresponding positions “r” and “t” in the reference and test video projection planes, respectively;
 2. Calculate the distortion between the samples values at “r” and “t”. When “r” and/or “t” are at fractional sample positions, use nearest neighbor rounding;
 3. Calculate the overall distortion by accumulating the distortion at each “s” in the set of points, and calculate PSNR from distortion
-