Joint Video Experts Team (JVET)
of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29
22nd Meeting, by teleconference, 20–28 Apr. 2021

Document: JVET-V2020

Title:	**VVC Verification Test Report for High Definition (HD) and 360° Standard Dynamic Range (SDR) Video Content**		
Status:	Output document approved by JVET		
Purpose:	Report		
Author(s) or Contact(s):	Mathias Wien Vittorio Baroncini	Email:	wien@lfb.rwth-aachen.de baroncini@gmx.com
Source:	Verification Test Coordinators		

Executive Summary

This document reports verification test results comparing VVC to its predecessor HEVC on high definition (HD, 1920×1080) with random access (RA) and low delay (LD) configurations and 360° video content with random access configuration using formal subjective visual quality assessment testing. All content in this testing was of standard dynamic range (SDR). The RA configuration was chosen to represent streaming or broadcast applications. The LD configuration was chosen to represent conversational and gaming-type applications. In the 360° video category, two types of projection formats were used to demonstrate the versatility of VVC: padded equirectangular projection (PERP) and padded or generalized cubemap projection (PCMP/GCMP). While PERP is available for both, HEVC and VVC, the GCMP format was only available for VVC and the PCMP format was used for HEVC. The purpose of the verification test was to confirm that the coding efficiency objective for the VVC standard has been met: achieving a substantial bit-rate reduction for the same level of subjective visual quality relative to the HEVC Main Profile. The compression performance of the HEVC reference software HM-16.22, the VVC reference software VTM-11.0 were used for the comparison in all categories. For HD RA, the open-source VVC encoder implementation VVenC-0.3 was added to the comparison. The HM reference software encoder for HEVC and the VTM reference software encoder for VVC used essentially the same rate-distortion optimization encoding techniques. VVenC was used to represent an example of practical encoding as may be found in product implementations and is reported to be more than 100 times faster than the VTM encoder.

The testing used the degradation category rating (DCR) test method (as in ITU-T P.910) with an 11-point impairment scale (as in Rec. ITU-R BT.500). The results of a visual assessment of VVC compared to HEVC by naïve test subjects are reported. The assessment included four test sequences each in the HD RA and 360° video categories. They were encoded in a random-access configuration with a random-access interval of 1.07 seconds. In the HD LD category, six sequences were assessed, with three sequences showing conversational and three sequences representing gaming-type content. The measured mean opinion score (MOS) figures indicate a significant improvement of VVC relative to HEVC in all categories resulting in overall average bit-rate savings estimates of 49% and 51% for VTM-11.0 and VVenC-0.3 in the HD RA category as well as 37% for VTM-11.0 in the HD LD category. For 360° video with PERP and GCMP/PCMP, overall average bit-rate savings estimates of 50% and 56% are reported for VTM-11.0. The reported results document the substantial compression benefit provided by the VVC over its predecessor HEVC in the tested categories. VVC provides several new compression tools specifically effective in the random-access configuration. It further includes new tools applicable to 360° video. The results indicate most substantial benefits for VVC in the test categories making use of these tools, demonstrating the versatility of the new video coding specification.

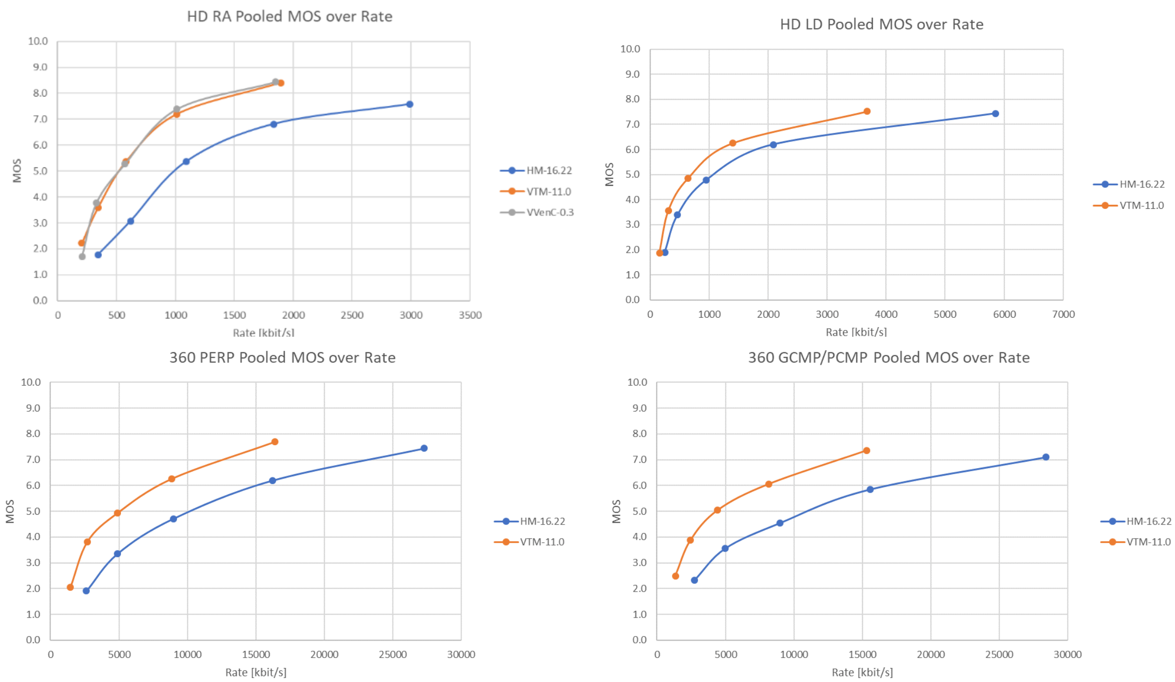

Figure 1: Pooled MOS over bit rate plots over the test sets for the four reported test categories[1]

1 Introduction

A major design goal for the development of the VVC standard was to achieve a substantial improvement in compression capability relative to its predecessor, the HEVC standard. This document is the second in a planned series of reports addressing a variety of test categories and embracing some of the available versatile tools provided by the VVC standard. It reports the results of a verification test to confirm that this goal was achieved and to estimate the magnitude of that achievement, following a test plan issued at the previous meeting [1].

A subjective evaluation was conducted at three test sites comparing the VVC Main 10 profile to the HEVC Main 10 profile for the HD SDR test category with random access and low delay configurations, and 360° video content with random access configuration.

2 Verification test logistics

The HD SDR subjective test was carried out at the following test sites:

- GBTech, Rome, IT
- Vabtech, London, UK
- RWTH Aachen University, Aachen, DE

The tests were conducted using the degradation category rating (DCR) test method [2] with an 11-grade impairment scale [3]. The verification test environment and testing methodology are described in Annex A. The arrangements for the test sites are shown in Table 1.

[1] Pooling was done by computing the geometric mean of the bitrates and the arithmetic mean of the MOS scores across the test sequences of each test category.

Table 1: Test site information and setup

Test Site	**GBTech**	**Vabtech**	**RWTH Aachen University**
Display, size, connection (resolution setting)	LG 65” CX6LA (3840×2160)	LG 65” CX6LA (3840×2160)	Sony 55” PVM X550, Quad-SDI (3840×2160), LG 65” CX6LA, HDMI (3840×2160)
Viewing distance	1 viewer at 1.5H	1 viewer at 1.5H	1 viewer at 1.5H
Viewing angle	90° (at screen center)	90° (at screen center)	90° (at screen center)
Total number of viewers	30 (16 females, 14 males; age 18-24), all screened for visual acuity and normal colour vision.	29 (20 females, 9 males, age 18-24), all screened for visual acuity and normal colour vision.	23 (4 female, 19 male; ages 16-26) all screened for visual acuity and normal colour vision

GBTech and Vabtech performed formal visual assessments for all categories. RWTH Aachen University performed formal visual assessments in the HD RA category.

3 Verification test setup

3.1 Verification test content generation

All test sequences were provided in the Rec. ITU-R BT.709 colour space [14][15].

In the test, the HEVC bitstreams were encoded using the HEVC reference software HM16.22 [4]. For VVC, the VTM-11.0 reference software [5] was used in all test categories. In the HD RA category, the open source VVenC-0.3 [6][7][8] was used as a second VVC encoder implementation. The RA configuration used for both, HD RA and 360° video enables random access to the bitstream every 1.07 seconds, i.e. every 64 pictures for the 60 Hz test sequences and every 32 pictures for the 30Hz test sequences.

For HEVC, the random-access configuration provided with the configuration file cfg/encoder_randomaccess_main10.cfg of HM-16.22 was used with `TemporalFilter=1` for the HD RA test category. In the 360° video category, temporal filtering was switched off. For the VTM, the random-access configuration provided with the configuration file cfg/encoder_randomaccess_vtm_gop32.cfg of VTM-11.0 was employed, with temporal filtering switched off in the 360° video category. These selected HM and VTM configurations result in the application of very similar configurations and very similar searching and rate-distortion optimization techniques in the HEVC and VVC contexts (using fixed QP settings and greedy optimization techniques with Lagrange multiplier $D + \lambda \cdot R$ decision making), thus maximizing the ability to test the capability of the differing syntax and decoding process features of the tested HEVC and VVC profiles in a controlled manner.

For the 360° video test category, VTM 11.0 with 360Lib 11.0 [10] was used for generating the VVC bitstreams, in two projection formats, padded equirectangular (PERP) and generalized cubemap projection (GCMP). The HEVC streams were generated with HM 16.22 and a patched version of 360Lib 5.0 in two projection formats, PERP and padded cubemap projection (PCMP).

For the VVenc-0.3 software [6][7][8], the Medium configuration has been used. VVenC was operated with perceptual QP adaptation and without rate control. Temporal filtering was switched on. In this configuration, the VVenC-0.3 encoding speed is reportedly more than 100 times faster than the VTM [9]. For the purpose of the verification tests, the QP values were selected such that the VVenC bit rate was approximately the same or lower than the VTM bit rate.

3.2 Test method and test design

The test sequences were evaluated using the 11-grade scale as specified in Rec. ITU-R BT.500-14, shown in Figure 2 below.

Score	Impairment item	
10	Imperceptible	
9	Slightly perceptible	somewhere
8		everywhere
7	Perceptible	somewhere
6		everywhere
5	Clearly perceptible	somewhere
4		everywhere
3	Annoying	somewhere
2		everywhere
1	Severely annoying	somewhere
0		everywhere

Figure 2: Meaning of the 11 grades numerical scale as specified in Rec. ITU-R BT.500-14 Table 2-4 [2]

A total of 9 test sessions were designed: three for HD LD, three for HD RA and three for 360; all the test sessions were designed inserting a "stabilization phase" ass suggested in ITU-R BT.500-14, and were not longer than 12 minutes, to avoid fatigue impact.

4 Test sequences and rate point selection

Five bit-rate points for each test sequence were selected for the quality assessment of the test sequences in all test categories. The bit-rate points were chosen such that the VTM/HM pair for a bit-rate point would represent approximately the same quality while at the same time allowing for approximate bit-rate matching of each HM bit-rate point with the next VTM bit-rate point. Thereby both an assessment of bit-rate savings at similar quality and an assessment of quality improvement at similar bit rates are enabled.

4.1 HD random access

Table 2: HD RA test sequences

Test sequence	Resolution	fps	Frames	md5
BarScene	1920×1080	60	600	89b3195543e8e9a4014d38d425d4c9a3
DrivingPOV	1920×1080	60	600	71aa6ab712e2152c1d4142db8350ca34
Meridian2	1920×1080	60	600	514bdb1932c11352f024edc63e524d1d
Metro	1920×1080	60	600	1030673bd52a31705ccbba36a2959f4f

Table 3: QP settings for HM, VTM, VVenC for the HD RA test sequences

Sequence	HM QPs	VTM QPs	VVenC QPs
BarScene	42, 38, 33, 28, 23	43, 39, 35, 31, 26	47, 43, 39, 35, 30
DrivingPOV	39, 34, 30, 27, 24	42, 38, 34, 30, 26	43, 39, 35, 31, 27
Meridian2	39, 34, 30, 27, 25	42, 38, 34, 30, 26	47, 44, 39, 35, 31
Metro	39, 35, 31, 27, 24	42, 38, 34, 30, 26	44, 39, 35, 31, 27

4.2 HD low delay

Table 4: HD LD test sequences

Test sequence	Resolution	fps	Frames	md5
Beatriz	1920×1080	50	500	fe74cd5046fa033b4f743f42b29e69cd
OfficeWalkAtWall	1920×1080	30	300	529c15491ea8e1eb0320244a6ff902bb
OfficeWalkCeiling	1920×1080	30	300	265906bc8f65441fb53d167483b9f726
DOTA2	1920×1080	60	600	be1c5d02a8fb298e26e5c5b890451413
EuroTruckSimulator2	1920×1080	60	600	f6850dbfff967945c0a273a374f28abd
STARCRAFT	1920×1080	60	600	e3d024bd65f4483fd5895f0e0ca44433

The test sequences Beatriz, OfficeWalkAtWall, and OfficeWalkCeiling represent conversational video content. The test sequences DOTA2, EuroTruckSimulator2, and Starcraft represent gaming-type content.

Table 5: QP settings for HM and VTM for the HD LD test sequences

Sequence	HM QPs	VTM QPs
Beatriz	41, 37, 33, 28, 23	42, 38, 34, 29, 24
OfficeWalkAtWall_	46, 42, 38, 32, 26	47, 44, 40, 34, 27
OfficeWalkCeiling	45, 40, 34, 30, 24	46, 42, 36, 32, 26
DOTA2	38, 34, 29, 24, 20	40, 36, 31, 26, 22
EuroTruckSimulator2	45, 40, 36, 32, 26	46, 41, 37, 33, 27
STARCRAFT	38, 34, 30, 26, 21	40, 36, 32, 28, 23

4.3 360° video

Table 6: 360° video test sequences

Test sequence	Resolution	Fps	Frames	md5
GT_Sheriff	4320×2160	30	300	0f8ecc5f4976d7cf056cdf78c4dc3f2c
HarborBiking2	8192×4096	30	300	746b0aba98bca07dae646c75348eed87
KiteFliteWalking2	8192×4096	30	300	8a102140ff5c9aa3b3f515c512511a59
SkateBoardAtBridge	6144×3072	30	300	3dd5bef4f2ab2c887d4730332ee4d5de

Dynamic viewports with 78.1×49.1 degrees of field of view were used to generate viewport video at the resolution of 1920×1080 and viewed on conventional HD displays. The viewpaths proposed in [11] have been applied.

Table 7: Coding resolution for different input resolutions and coding projection formats

	PCMP/GCMP (before padding)	PERP (before padding)
4K (4320x2160)	1184×1184 per face --CodingFaceWidth=1184 --CodingFaceHeight=1184	4096×2048 --CodingFaceWidth=4096 --CodingFaceHeight=2048 --WrapAroundOffset=4096 (only for VTM)
6K & 8K	1280×1280 per face --CodingFaceWidth=1280 --CodingFaceHeight=1280	4432×2216 --CodingFaceWidth=4432 --CodingFaceHeight=2216 --WrapAroundOffset=4432 (only for VTM)

Table 8: QP settings for HM and VTM for the 360° video test sequences in PERP format

Test sequence	**HM QPs**	**VTM QPs**
GT_Sheriff	36, 32, 28, 24, 20	38, 34, 30, 26, 22
HarborBiking2	35, 31, 27, 23, 20	38, 34, 30, 26, 22
KiteFliteWalking2	40, 36, 32, 28, 24	42, 38, 34, 30, 26
SkateBoardAtBridge	39, 35, 31, 27, 24	42, 38, 34, 30, 26

Table 9: QP settings for HM and VTM for the 360° video test sequences in GCMP/PCMP formats

Test sequence	**HM QPs (PCMP)**	**VTM QPs (GCMP)**
GT_Sheriff	39, 35, 31, 27, 22	42, 38, 34, 30, 26
HarborBiking2	35, 31,27, 23, 20	38, 34, 30, 26, 22
KiteFliteWalking2	40, 36, 32, 28, 24	42, 38, 34, 30, 26
SkateBoardAtBridge	38, 34, 30, 27, 23	42, 38, 34, 30, 26

5 Results and analysis

The measured MOS values of the reconstructed video on the 11-grade scale are plotted over the bit rate of the corresponding bitstream. The ±95% confidence intervals for the MOS values are indicated.

For all test categories, the bit rate and MOS differences for all bit-rate points are collected and the Bjøntegaard delta rate relative to HM-16.22 based on bit rate and MOS is reported. The Bjøntegaard delta rate has been computed with the RDPlot tool [19] using the attached csv-files. The bit-rate savings is computed as the difference between the VVC bit-rate point and the corresponding HEVC bit-rate point relative to the HEVC bit rate. The MOS difference is reported as a number if the value is larger than the maximum of the VVC and the HEVC confidence intervals. Otherwise, "< CI" is indicated. The results are reported relative to the HM for both the VTM and VVenC encoders, where applicable.

5.1 HD random access

At the 21st JVET meeting, a misconfiguration of the temporal filtering in the HM-16.22 bitstreams for the BarScene, DrivingPOV, and Metro sequences was reported [20][21]. A set of tests for this category were performed after the meeting such that a comparison can be provided relative to HM-16.22 with temporal filtering on and off, respectively. It is noted that a direct comparison of the MOS results for the visual tests with TF on and TF off is not considered appropriate since they have not been evaluated in the same test sessions.

5.1.1 Results with temporal filtering on

5.1.1.1 MOS plots

Table 10: Bitrate, MOS and confidence intervals (HM-16.22 with temporal filtering on)

TF on		HM-16.22			VTM-11.0			VVenC-0.3		
Sequence	**RP**	**Rate [kbit/s]**	**MOS**	**CI**	**Rate [kbit/s]**	**MOS**	**CI**	**Rate [kbit/s]**	**MOS**	**CI**
BarScene	R1	1632.12	7.6	0.4	885.32	8.7	0.3	887.23	8.3	0.2
BarScene	R2	805.03	7.0	0.4	445.30	7.5	0.4	454.87	7.5	0.4
BarScene	R3	432.30	4.9	0.5	272.71	5.2	0.5	267.96	5.8	0.4
BarScene	R4	240.62	2.0	0.3	171.91	3.6	0.5	174.91	3.7	0.6
BarScene	R5	154.20	1.2	0.4	109.90	1.5	0.3	110.65	1.1	0.3
DrivingPOV	R1	8868.10	8.2	0.4	5316.45	8.5	0.3	5048.55	8.8	0.3
DrivingPOV	R2	5453.87	7.5	0.5	2823.59	7.8	0.3	2761.01	8.1	0.4
DrivingPOV	R3	3420.69	6.0	0.7	1548.06	5.9	0.6	1515.37	5.2	0.6
DrivingPOV	R4	1936.71	3.0	0.5	872.60	4.3	0.6	796.95	4.1	0.5
DrivingPOV	R5	954.20	2.3	0.4	495.55	3.9	0.5	469.65	2.2	0.4
Meridian2	R1	991.30	6.5	0.5	710.48	7.7	0.4	662.28	8.0	0.4
Meridian2	R2	696.10	5.4	0.4	370.97	6.1	0.4	354.00	6.2	0.3
Meridian2	R3	435.35	4.6	0.5	207.18	4.5	0.3	195.51	4.3	0.3
Meridian2	R4	246.59	3.6	0.3	121.74	3.2	0.4	117.76	3.0	0.3
Meridian2	R5	128.51	1.7	0.4	75.01	1.9	0.4	81.54	1.8	0.3
Metro	R1	5576.69	8.0	0.4	3856.09	8.8	0.4	3968.93	8.6	0.3
Metro	R2	3719.49	7.3	0.4	2234.19	7.4	0.4	2351.65	7.8	0.3
Metro	R3	2207.46	6.0	0.5	1309.61	6.0	0.6	1351.25	5.8	0.6
Metro	R4	1281.47	3.7	0.4	771.21	3.3	0.5	731.19	4.2	0.5
Metro	R5	745.79	1.9	0.4	442.74	1.6	0.4	439.83	1.6	0.2

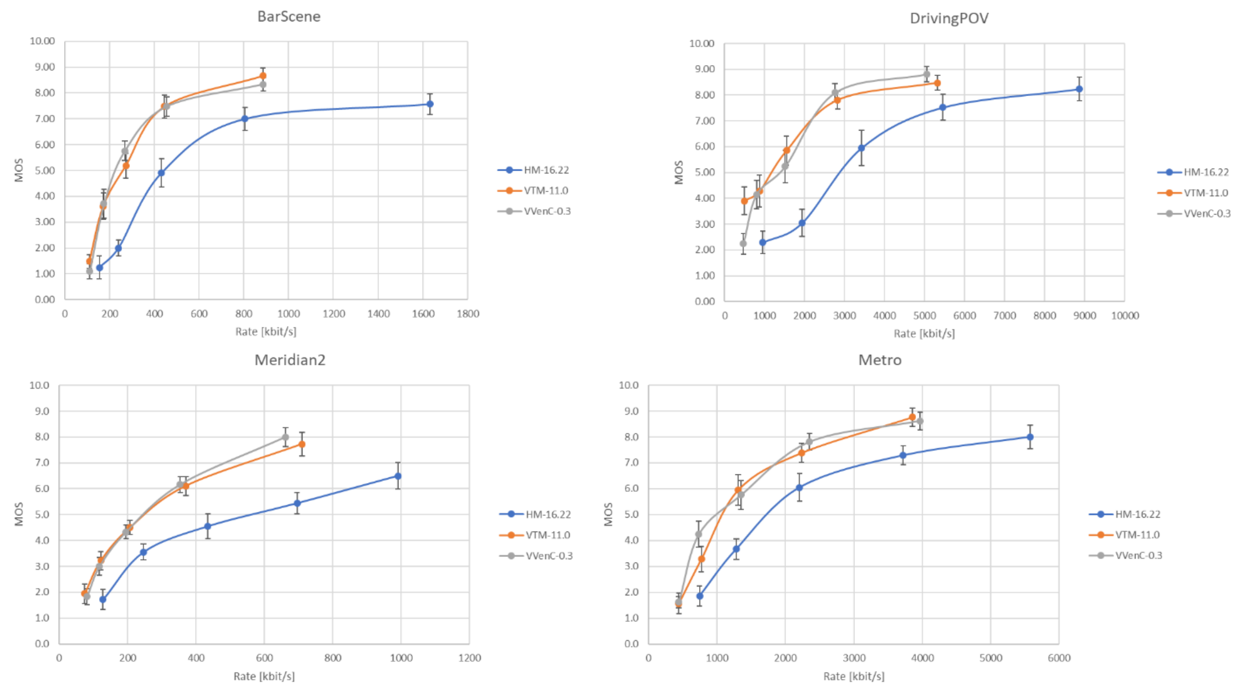

Figure 3: Collection of the MOS-over-rate plots for HM with TF on, VTM, and VVenC for the HD RA test sequences

5.1.1.2 Analysis

Table 11: Bjøntegaard delta rate relative to HM-16.22 with TF on based on bit rate and MOS

BD-Rate	**VTM-11.0**	**VVenC-0.3**
BarScene	-48%	-50%
DrivingPOV	-58%	-60%
Meridian2	-50%	-49%
Metro	-38%	-45%
Overall	**-49%**	**-51%**

Table 12: Bit-rate savings and MOS deltas for the bit-rate points

VTM / HM (TF on)		Rate Diff.	ΔMOS
BarScene	R1	-45.8%	1.1
BarScene	R2	-44.7%	0.5
BarScene	R3	-36.9%	< CI
BarScene	R4	-28.6%	1.6
BarScene	R5	-28.7%	< CI
DrivingPOV	R1	-40.0%	< CI
DrivingPOV	R2	-48.2%	< CI
DrivingPOV	R3	**-54.7%**	< CI
DrivingPOV	R4	**-54.9%**	1.2
DrivingPOV	R5	-48.1%	1.6
Meridian2	R1	-28.3%	1.2
Meridian2	R2	-46.7%	0.7
Meridian2	R3	**-52.4%**	< CI
Meridian2	R4	**-50.6%**	< CI
Meridian2	R5	-41.6%	< CI
Metro	R1	-30.9%	0.8
Metro	R2	-39.9%	< CI
Metro	R3	-40.7%	< CI
Metro	R4	-39.8%	< CI
Metro	R5	-40.6%	< CI

VVenC / HM(TF on)		Rate Diff.	ΔMOS
BarScene	R1	-45.6%	0.8
BarScene	R2	-43.5%	0.5
BarScene	R3	-38.0%	0.9
BarScene	R4	-27.3%	1.7
BarScene	R5	-28.2%	< CI
DrivingPOV	R1	-43.1%	0.6
DrivingPOV	R2	-49.4%	0.6
DrivingPOV	R3	**-55.7%**	-0.7
DrivingPOV	R4	**-58.9%**	1.1
DrivingPOV	R5	**-50.8%**	< CI
Meridian2	R1	-33.2%	1.5
Meridian2	R2	-49.1%	0.7
Meridian2	R3	**-55.1%**	< CI
Meridian2	R4	**-52.2%**	-0.6
Meridian2	R5	-36.5%	< CI
Metro	R1	-28.8%	0.6
Metro	R2	-36.8%	0.5
Metro	R3	-38.8%	< CI
Metro	R4	-42.9%	0.6
Metro	R5	-41.0%	< CI

The figures reported in Table 11 and Table 12 indicate significant compression performance improvements for VVC compared to its predecessor HEVC. It is observed that the VVC encoder implementation provided by VVenC provides more gain over the HM than the VTM for the vast majority of bit-rate points.

5.1.2 Results with temporal filtering off

5.1.2.1 MOS plots

Table 13: Bitrate, MOS and confidence intervals (HM-16.22 with temporal filtering off)

TF off		HM-16.22			VTM-11.0		
Sequence	**RP**	**Rate [kbit/s]**	**MOS**	**CI**	**Rate [kbit/s]**	**MOS**	**CI**
BarScene	R1	2001.43	7.5	0.3	885.32	7.0	0.5
BarScene	R2	835.54	6.0	0.4	445.30	5.9	0.4
BarScene	R3	439.74	4.2	0.4	272.71	4.8	0.3
BarScene	R4	242.67	2.4	0.3	171.91	3.3	0.3
BarScene	R5	155.35	1.5	0.3	109.90	1.7	0.3
DrivingPOV	R1	9754.28	6.9	0.5	5316.45	7.1	0.4
DrivingPOV	R2	5820.71	6.1	0.6	2823.59	6.3	0.6
DrivingPOV	R3	3583.84	4.9	0.6	1548.06	4.9	0.4
DrivingPOV	R4	1992.94	3.2	0.5	872.60	3.8	0.5
DrivingPOV	R5	968.88	2.7	0.6	495.55	2.4	0.4
Meridian2	R1	1105.73	8.7	0.4	710.48	8.0	0.5
Meridian2	R2	750.33	7.7	0.6	370.97	7.5	0.4
Meridian2	R3	456.86	5.5	0.5	207.18	5.5	0.6
Meridian2	R4	253.60	3.6	0.4	121.74	3.2	0.4
Meridian2	R5	130.31	1.7	0.3	75.01	2.2	0.3
Metro	R1	5730.31	7.7	0.3	3856.09	7.1	0.5
Metro	R2	3797.83	6.3	0.4	2234.19	6.1	0.5
Metro	R3	2245.85	5.2	0.4	1309.61	4.7	0.3
Metro	R4	1300.87	3.9	0.4	771.21	3.7	0.5
Metro	R5	755.42	2.2	0.2	442.74	1.7	0.3

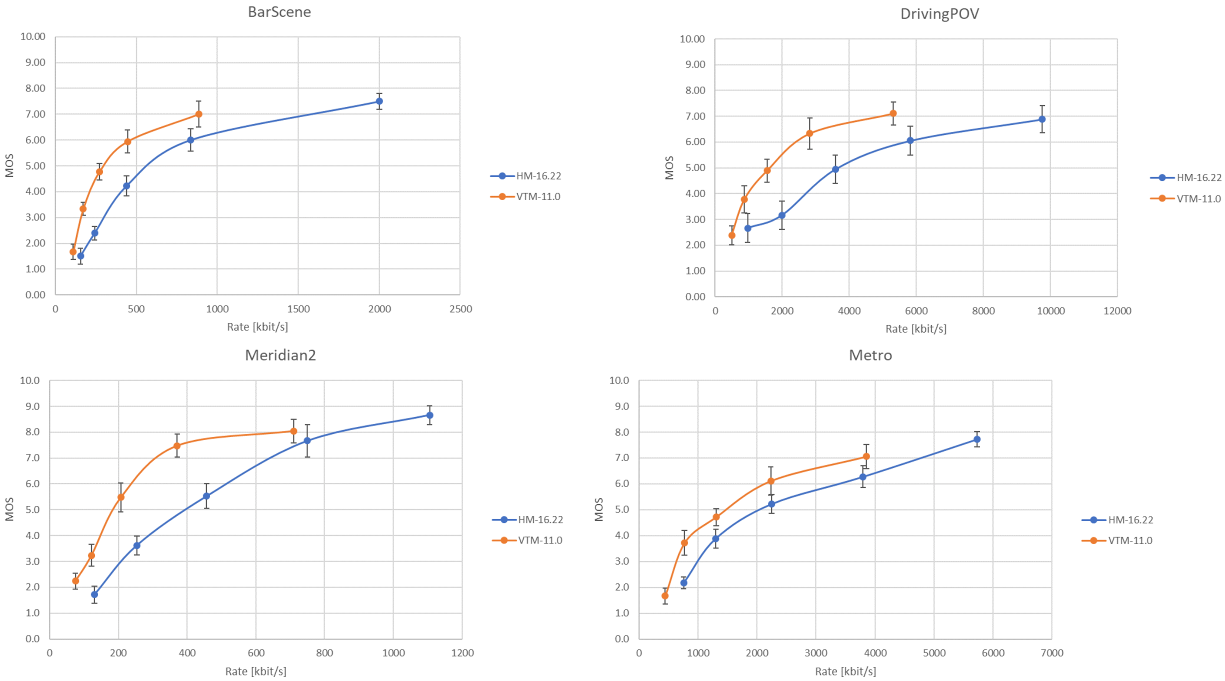

Figure 4: Collection of the MOS-over-rate plots for HM with TF off, VTM, and VVenC for the HD RA test sequences

5.1.2.2 Analysis

Table 14: Bit-rate savings and MOS deltas for the bit-rate points and Bjøntegaard delta rate relative to HM-16.22 with TF off based on bit rate and MOS

VTM / HM		**Rate Diff.**	**ΔMOS**
BarScene	R1	**-55.8%**	-0.5
BarScene	R2	-46.7%	< CI
BarScene	R3	-38.0%	0.6
BarScene	R4	-29.2%	0.9
BarScene	R5	-29.3%	< CI
DrivingPOV	R1	-45.5%	< CI
DrivingPOV	R2	**-51.5%**	< CI
DrivingPOV	R3	**-56.8%**	< CI
DrivingPOV	R4	**-56.2%**	0.6
DrivingPOV	R5	-48.9%	< CI
Meridian2	R1	-35.7%	-0.6
Meridian2	R2	**-50.6%**	< CI
Meridian2	R3	**-54.7%**	< CI
Meridian2	R4	**-52.0%**	< CI
Meridian2	R5	-42.4%	0.5
Metro	R1	-32.7%	-0.7
Metro	R2	-41.2%	< CI
Metro	R3	-41.7%	-0.5
Metro	R4	-40.7%	< CI
Metro	R5	-41.4%	-0.5

BD-Rate	**VTM-11.0**
BarScene	-46%
DrivingPOV	-60%
Meridian2	-50%
Metro	-33%
Overall	**-47%**

5.2 HD low delay

5.2.1 MOS plots

Table 15: Bitrate, MOS and confidence intervals

		HM-16.22			VTM-11.0		
Sequence	**RP**	**Rate [kbit/s]**	**MOS**	**CI**	**Rate [kbit/s]**	**MOS**	**CI**
Beatriz	R1	3464.52	7.9	0.2	2238.70	8.5	0.2
Beatriz	R2	1039.33	6.2	0.3	728.57	6.4	0.3
Beatriz	R3	412.42	5.0	0.4	292.22	5.3	0.3
Beatriz	R4	230.26	3.3	0.3	164.37	3.8	0.3
Beatriz	R5	134.78	2.5	0.3	97.45	2.6	0.3
DOTA2	R1	9561.39	7.6	0.4	6465.95	7.6	0.2
DOTA2	R2	4149.42	6.5	0.3	2701.41	6.7	0.4
DOTA2	R3	1695.16	5.0	0.4	1045.92	4.8	0.4
DOTA2	R4	724.95	3.8	0.3	470.77	3.6	0.3
DOTA2	R5	396.25	1.4	0.2	250.30	1.4	0.3
EuroTruckSimulator2	R1	28789.65	7.5	0.3	18490.75	7.6	0.2
EuroTruckSimulator2	R2	6899.45	6.3	0.3	4635.84	6.2	0.3
EuroTruckSimulator2	R3	3021.16	4.8	0.2	2009.09	5.1	0.2
EuroTruckSimulator2	R4	1308.97	3.8	0.3	870.75	4.2	0.4
EuroTruckSimulator2	R5	505.90	2.1	0.3	309.93	2.1	0.3
OfficeWalkAtWall	R1	1599.54	6.5	0.2	906.67	6.4	0.3
OfficeWalkAtWall	R2	628.82	5.2	0.3	493.83	5.5	0.3
OfficeWalkAtWall	R3	298.70	3.7	0.3	265.23	4.0	0.2
OfficeWalkAtWall	R4	186.89	2.3	0.2	144.97	2.4	0.3
OfficeWalkAtWall	R5	132.88	1.3	0.3	90.57	1.2	0.3
OfficeWalkCeiling	R1	3441.61	6.9	0.2	1989.26	7.0	0.4
OfficeWalkCeiling	R2	1330.18	5.8	0.2	783.94	5.7	0.4
OfficeWalkCeiling	R3	757.67	4.9	0.2	470.44	4.6	0.4
OfficeWalkCeiling	R4	351.93	3.0	0.4	217.02	3.1	0.3
OfficeWalkCeiling	R5	204.73	1.7	0.3	128.45	1.4	0.3
STARCRAFT	R1	7661.82	8.4	0.2	5142.52	8.1	0.2
STARCRAFT	R2	3382.69	7.3	0.3	2111.71	7.2	0.2
STARCRAFT	R3	1607.93	5.5	0.3	957.00	5.5	0.3
STARCRAFT	R4	726.86	4.3	0.4	450.49	4.4	0.3
STARCRAFT	R5	344.48	2.5	0.3	217.05	2.7	0.4

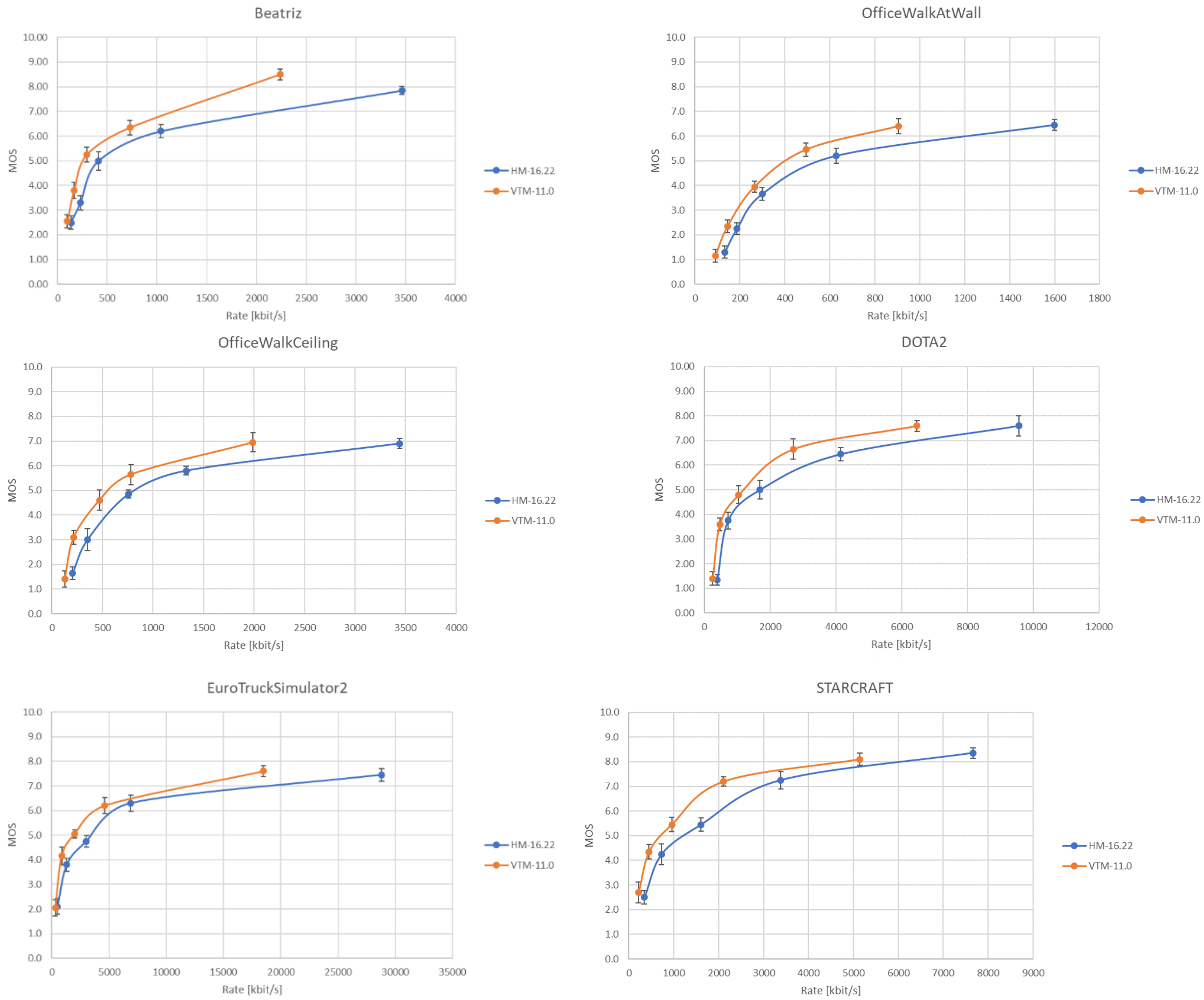

Figure 5: Collection of the MOS-over-rate plots for HM and VTM for the HD LD test sequences

5.2.2 Analysis

Table 16: Bit-rate savings and MOS deltas for the bit-rate points and Bjøntegaard delta rate relative to HM-16.22 based on bit rate and MOS

VTM / HM		Rate Diff.	ΔMOS
Beatriz	R1	-35.4%	0.7
Beatriz	R2	-29.9%	< CI
Beatriz	R3	-29.1%	< CI
Beatriz	R4	-28.6%	0.5
Beatriz	R5	-27.7%	< CI
DOTA2	R1	-32.4%	< CI
DOTA2	R2	-34.9%	< CI
DOTA2	R3	-38.3%	< CI
DOTA2	R4	-35.1%	< CI
DOTA2	R5	-36.8%	< CI
EuroTruckSimulator2	R1	-35.8%	< CI
EuroTruckSimulator2	R2	-32.8%	< CI
EuroTruckSimulator2	R3	-33.5%	0.3
EuroTruckSimulator2	R4	-33.5%	< CI
EuroTruckSimulator2	R5	-38.7%	< CI
OfficeWalkAtWall	R1	-43.3%	< CI
OfficeWalkAtWall	R2	-21.5%	< CI
OfficeWalkAtWall	R3	-11.2%	0.3
OfficeWalkAtWall	R4	-22.4%	< CI
OfficeWalkAtWall	R5	-31.8%	< CI
OfficeWalkCeiling	R1	-42.2%	< CI
OfficeWalkCeiling	R2	-41.1%	< CI
OfficeWalkCeiling	R3	-37.9%	< CI
OfficeWalkCeiling	R4	-38.3%	< CI
OfficeWalkCeiling	R5	-37.3%	< CI
STARCRAFT	R1	-32.9%	-0.3
STARCRAFT	R2	-37.6%	< CI
STARCRAFT	R3	-40.5%	< CI
STARCRAFT	R4	-38.0%	< CI
STARCRAFT	R5	-37.0%	< CI

BD-Rate	VTM-11.0
Beatriz	-41%
DOTA2	-34%
EuroTruckSimulator2	-42%
OfficeWalkAtWall	-28%
OfficeWalkCeiling	-37%
STARCRAFT	-38%
Overall	**-37%**

BD-Rate	VTM-11.0
Conversational	-35%
Gaming	-38%

The figures reported in Table 16 indicate significant compression performance improvements for VVC compared to its predecessor HEVC. The gains are found to be a bit lower than in the RA configuration since some the tools which were introduced for motion compensation in VVC cannot be applied due to the fact that in the low delay configuration, motion compensated prediction can only be applied from past reference frames. Still, the compression benefits typically exceed the gains reported from objective measurements using PSNR [22]. As noted previously, the test sequences Beatriz, OfficeWalkAtWall, and OfficeWalkCeiling represent conversational video content while the test sequences DOTA2, EuroTruckSimulator2, and Starcraft represent gaming-type content. The average bit-rate savings over the three test sequences in each test category are estimated to be 35% for test sequences with conversational content and 38% for test sequences with gaming-type content.

5.3 360° video PERP

5.3.1 MOS plots

Table 17: Bitrate, MOS and confidence intervals

		HM-16.22			**VTM-11.0**		
Sequence	**RP**	**Rate [kbit/s]**	**MOS**	**CI**	**Rate [kbit/s]**	**MOS**	**CI**
GTSheriff PERP	R1	14329.25	7.2	0.3	9944.65	7.6	0.2
GTSheriff PERP	R2	9116.19	5.7	0.3	6270.49	5.9	0.3
GTSheriff PERP	R3	5450.60	4.6	0.3	3693.01	4.5	0.3
GTSheriff PERP	R4	3017.93	2.8	0.4	1971.90	3.4	0.4
GTSheriff PERP	R5	1552.64	1.2	0.3	982.52	1.2	0.4
HarborBiking2 PERP	R1	37020.82	7.3	0.3	21192.80	7.4	0.4
HarborBiking2 PERP	R2	22369.36	6.6	0.3	11001.24	6.1	0.3
HarborBiking2 PERP	R3	11418.08	5.1	0.4	5639.45	5.3	0.3
HarborBiking2 PERP	R4	6158.57	3.9	0.3	3058.94	4.0	0.2
HarborBiking2 PERP	R5	3299.98	2.4	0.3	1720.04	3.1	0.4
KiteFliteWalking2 PERP	R1	44195.85	7.3	0.4	26852.30	7.6	0.3
KiteFliteWalking2 PERP	R2	24447.12	5.6	0.3	14680.56	6.0	0.3
KiteFliteWalking2 PERP	R3	13767.46	4.1	0.5	8115.46	4.6	0.3
KiteFliteWalking2 PERP	R4	7440.57	2.6	0.4	4431.64	3.6	0.3
KiteFliteWalking2 PERP	R5	3912.59	1.6	0.3	2299.22	1.3	0.2
SkateBoardAtBridge PERP	R1	23712.84	7.9	0.3	12720.90	8.3	0.3
SkateBoardAtBridge PERP	R2	13851.80	6.9	0.3	6113.80	7.1	0.4
SkateBoardAtBridge PERP	R3	7544.34	5.1	0.4	3365.60	5.3	0.3
SkateBoardAtBridge PERP	R4	4139.93	4.1	0.3	1949.24	4.3	0.2
SkateBoardAtBridge PERP	R5	2380.18	2.4	0.4	1135.77	2.6	0.4

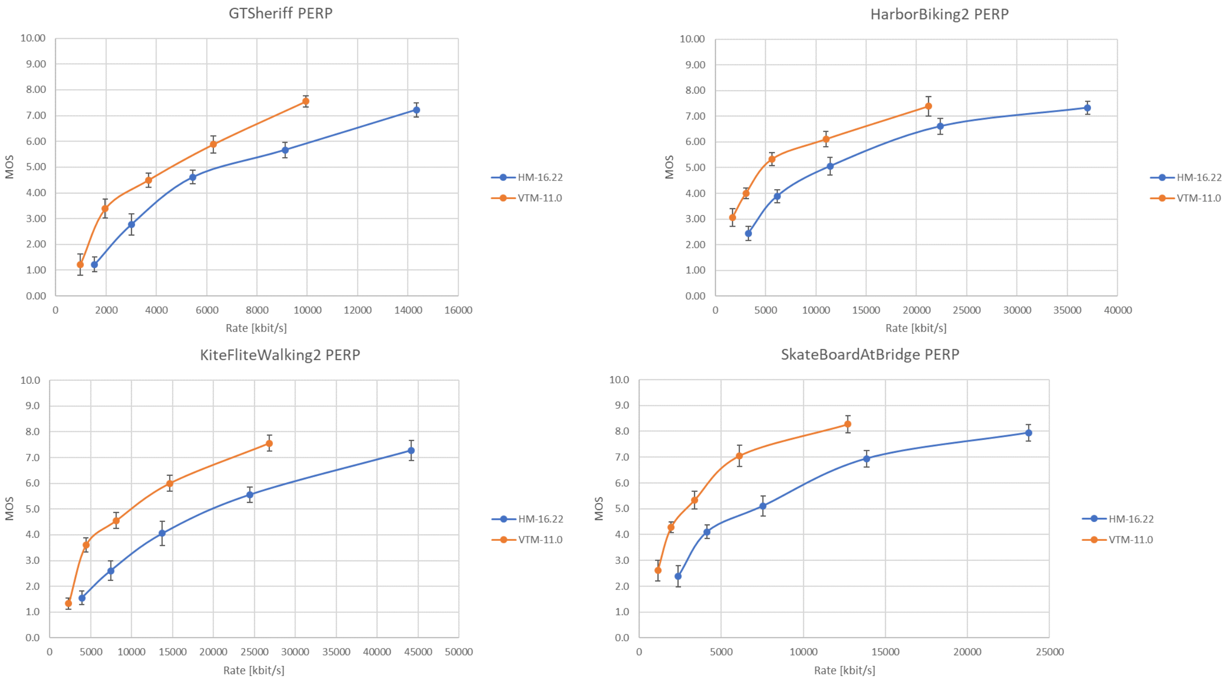

Figure 6: Collection of the MOS-over-rate plots for HM and VTM for the 360° video test sequences in PERP format

5.3.2 Analysis

Table 18: Bit-rate savings and MOS deltas for the bit-rate points and Bjøntegaard delta rate relative to HM-16.22 based on bit rate and MOS

VTM / HM		Rate Diff.	ΔMOS
GTSheriff PERP	R1	-30.6%	0.3
GTSheriff PERP	R2	-31.2%	< CI
GTSheriff PERP	R3	-32.2%	< CI
GTSheriff PERP	R4	-34.7%	0.6
GTSheriff PERP	R5	-36.7%	< CI
HarborBiking2 PERP	R1	-42.8%	< CI
HarborBiking2 PERP	R2	**-50.8%**	-0.5
HarborBiking2 PERP	R3	**-50.6%**	< CI
HarborBiking2 PERP	R4	**-50.3%**	< CI
HarborBiking2 PERP	R5	-47.9%	0.6
KiteFliteWalking2 PERP	R1	-39.2%	< CI
KiteFliteWalking2 PERP	R2	-39.9%	0.4
KiteFliteWalking2 PERP	R3	-41.1%	0.5
KiteFliteWalking2 PERP	R4	-40.4%	1.0
KiteFliteWalking2 PERP	R5	-41.2%	< CI
SkateBoardAtBridge PERP	R1	-46.4%	0.3
SkateBoardAtBridge PERP	R2	**-55.9%**	< CI
SkateBoardAtBridge PERP	R3	**-55.4%**	< CI
SkateBoardAtBridge PERP	R4	**-52.9%**	< CI
SkateBoardAtBridge PERP	R5	**-52.3%**	< CI

BD-Rate	VTM-11.0
GTSheriff PERP	-40%
HarborBiking2 PERP	-49%
KiteFliteWalking2 PERP	-53%
SkateBoardAtBridge PERP	-58%
Overall	-50%

The results reported in Table 18 indicate significant compression performance improvements for VVC compared to its predecessor HEVC with bitrate savings of more than 30% at all rate points. Parts of the gains can be attributed to specific coding tools introduced for 360° video compression in VVC. It is noted that for all but one rate point, the MOS values for HM-16.22 and VTM-11.0 are within the confidence interval or indicate additional subjective improvements for VVC over HEVC. Thereby, the overall measured bitrate savings estimates exceed the rate savings reported for the individual rate points.

5.4 360° video GCMP/PCMP

5.4.1 MOS plots

Table 19: Bitrate, MOS and confidence intervals

		HM-16.22			**VTM-11.0**		
Sequence	**RP**	**Rate [kbit/s]**	**MOS**	**CI**	**Rate [kbit/s]**	**MOS**	**CI**
GTSheriff GCMP/PCMP	R1	12410.55	6.8	0.4	6453.12	7.0	0.4
GTSheriff GCMP/PCMP	R2	6681.26	5.2	0.4	3765.94	4.9	0.5
GTSheriff GCMP/PCMP	R3	3883.18	3.3	0.3	2007.93	3.5	0.2
GTSheriff GCMP/PCMP	R4	2082.09	2.1	0.2	1030.80	2.1	0.3
GTSheriff GCMP/PCMP	R5	1084.82	0.9	0.3	528.13	0.7	0.2
HarborBiking2 GCMP/PCMP	R1	33892.76	7.2	0.2	20300.06	7.5	0.2
HarborBiking2 GCMP/PCMP	R2	20408.64	6.2	0.2	10444.99	6.3	0.4
HarborBiking2 GCMP/PCMP	R3	10661.14	5.4	0.4	5472.74	5.4	0.3
HarborBiking2 GCMP/PCMP	R4	6003.60	4.2	0.2	3056.05	4.2	0.3
HarborBiking2 GCMP/PCMP	R5	3368.34	2.8	0.3	1770.47	2.9	0.3
KiteFliteWalking2 GCMP/PCMP	R1	48497.65	7.2	0.3	31115.35	7.5	0.3
KiteFliteWalking2 GCMP/PCMP	R2	27042.35	5.4	0.5	17155.24	6.4	0.4
KiteFliteWalking2 GCMP/PCMP	R3	15438.58	4.3	0.3	9572.71	5.7	0.3
KiteFliteWalking2 GCMP/PCMP	R4	8472.12	3.5	0.2	5273.53	4.4	0.3
KiteFliteWalking2 GCMP/PCMP	R5	4507.54	1.9	0.3	2743.25	2.7	0.3
SkateBoardAtBridge GCMP/PCMP	R1	31931.28	7.2	0.2	13536.16	7.4	0.3
SkateBoardAtBridge GCMP/PCMP	R2	15921.67	6.6	0.3	6607.03	6.6	0.3
SkateBoardAtBridge GCMP/PCMP	R3	10138.68	5.2	0.2	3691.58	5.6	0.3
SkateBoardAtBridge GCMP/PCMP	R4	5793.22	4.4	0.3	2167.74	4.8	0.2
SkateBoardAtBridge GCMP/PCMP	R5	3322.45	3.7	0.3	1274.39	3.7	0.3

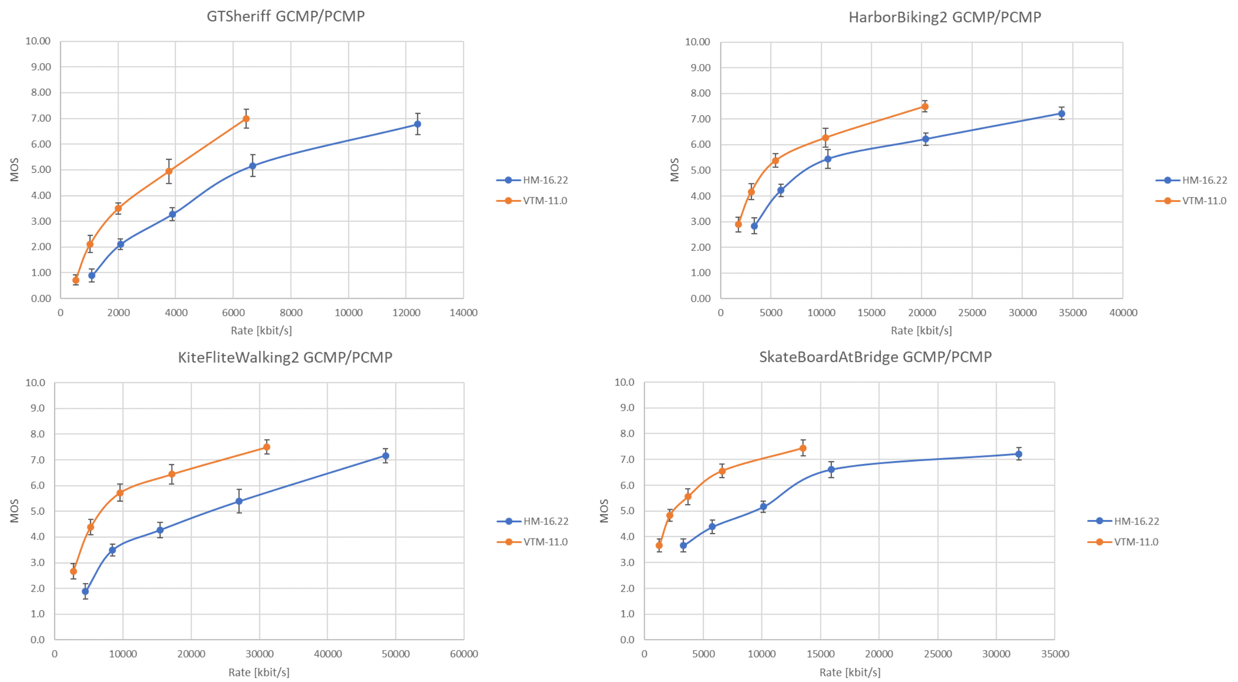

Figure 7: Collection of the MOS-over-rate plots for HM and VTM for the 360° video test sequences in PCMP format for HM and GCMP format for VTM

5.4.2 Analysis

Table 20: Bit-rate savings and MOS deltas for the bit-rate points and Bjøntegaard delta rate relative to HM-16.22 based on bit rate and MOS

VTM / HM		Rate Diff.	ΔMOS
GTSheriff GCMP/PCMP	R1	-48.0%	< CI
GTSheriff GCMP/PCMP	R2	-43.6%	< CI
GTSheriff GCMP/PCMP	R3	-48.3%	< CI
GTSheriff GCMP/PCMP	R4	**-50.5%**	< CI
GTSheriff GCMP/PCMP	R5	**-51.3%**	< CI
HarborBiking2 GCMP/PCMP	R1	-40.1%	0.3
HarborBiking2 GCMP/PCMP	R2	-48.8%	< CI
HarborBiking2 GCMP/PCMP	R3	-48.7%	< CI
HarborBiking2 GCMP/PCMP	R4	-49.1%	< CI
HarborBiking2 GCMP/PCMP	R5	-47.4%	< CI
KiteFliteWalking2 GCMP/PCMP	R1	-35.8%	0.3
KiteFliteWalking2 GCMP/PCMP	R2	-36.6%	1.1
KiteFliteWalking2 GCMP/PCMP	R3	-38.0%	1.4
KiteFliteWalking2 GCMP/PCMP	R4	-37.8%	0.9
KiteFliteWalking2 GCMP/PCMP	R5	-39.1%	0.8
SkateBoardAtBridge GCMP/PCMP	R1	**-57.6%**	< CI
SkateBoardAtBridge GCMP/PCMP	R2	**-58.5%**	< CI
SkateBoardAtBridge GCMP/PCMP	R3	**-63.6%**	0.4
SkateBoardAtBridge GCMP/PCMP	R4	**-62.6%**	0.4
SkateBoardAtBridge GCMP/PCMP	R5	**-61.6%**	< CI

BD-Rate GCMP/PCMP	VTM-11.0
GTSheriff	-47%
HarborBiking2	-50%
KiteFliteWalking2	-62%
SkateBoardAtBridge	-67%
Overall	-56%

The results reported in Table 20 indicate significant compression performance improvements for VVC compared to its predecessor HEVC which are even higher than in the PERP format for the same test sequences. Again. parts of the gains can be attributed to specific coding tools introduced for 360° video compression in VVC. Bitrate savings of more than 35% are observed at all rate points. It is noted that for all rate points, the MOS values for HM-16.22 and VTM-11.0 are within the confidence interval or indicate (partially significant) additional subjective improvements for VVC over HEVC. Thereby, the overall measured bitrate savings estimates exceed the rate savings reported for the individual rate points.

6 Conclusions

The results of a visual assessment of VVC encoders compared to an HEVC Main 10 profile reference software encoder by naïve test subjects are reported. The assessment included four test sequences in the HD random access category, a total of six test sequences in the HD low delay category (including three test sequences of conversational and three test sequences of gaming-type content), and four test sequences in the 360° video category. The 360° video test sequences were assessed using both, the PERP and the GCMP/PCMP projection formats which are representing relevant use cases for 360° video coding. The results demonstrate the substantial compression benefit provided by VVC. VVC provides several new compression tools which specifically become effective for random access configurations and for the challenging signal types of 360° video. The reported mean opinion score (MOS) figures indicate a significant improvement of VVC relative to HEVC in all categories resulting in overall average bit-rate savings estimates of 49% and 51% for VTM-11.0 and VVenC-0.3 in the HD RA category as well as 37% for VTM-11.0 in the HD LD category. For 360° video with PERP and GCMP/PCMP, overall average bit-rate savings estimates of 50% and 56% are reported for VTM-11.0. The reported results document the substantial compression benefit provided by the new scheme.

7 References

[1] M. Wien, V. Baroncini, A. Segall, and Y. Ye, "VVC verification test plan (Draft 5)," Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 5, output document JVET-U2021, 21st JVET meeting, by teleconference, Jan. 2021.

[2] Recommendation ITU-T P.910 (2008), *Subjective video quality assessment methods for multimedia applications*.

[3] Recommendation ITU-R BT.500-14 (2019), *Methodologies for the subjective assessment of the quality of television images*.

[4] HEVC Test Model, https://vcgit.hhi.fraunhofer.de/jct-vc/HM/-/tree/HM-16.22, online, accessed Oct. 2020.

[5] VVC Test Model, https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tree/VTM-11.0, online, accessed Jan. 2021.

[6] J. Brandenburg, A. Wieckowski, T. Hinz, A. Henkel, V. George, I. Zupancic, C. Stoffers, B. Bross, H. Schwarz, and D. Marpe, "Towards Fast and Efficient VVC Encoding," IEEE 22nd Workshop on Multimedia Signal Processing (MMSP 2020), Tampere, Finland, 2020.

[7] A. Wieckowski, J. Brandenburg, C. Bartnik, V. George, J. Güther, G. Hege, C. Helmrich, A. Henkel, T. Hinz, C. Lehmann, C. Stoffers, I. Zupancic, B. Bross, H. Schwarz, and D. Marpe, "Open optimized VVC encoder (VVenC) and decoder (VVdeC) implementations," Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29 input document JVET-T0099, 20th JVET meeting, by teleconference, Oct. 2020.

[8] Fraunhofer HHI, "Fraunhofer Versatile Video Encoder (VVenC)," version 0.3, https://github.com/fraunhoferhhi/vvenc/tags, online, accessed Mar. 2021.

[9] C. Helmrich, B. Bross, J. Pfaff, H. Schwarz, D. Marpe, and T. Wiegand, "Information on and analysis of the VVC encoders in the UHD SDR verification test," Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29 input document JVET-T0103, 20th JVET meeting, by teleconference, Oct. 2020.

[10] Y. Ye, J. Boyce, "Algorithm descriptions of projection format conversion and video quality metrics in 360Lib Version 8", Joint Video Exploration Team of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-L1004, October 2018, Macau, CN.

[11] M. Wien, V. Baroncini, Y. Ye, "Status Report on 360º video Verification Test Preparation," Joint Video Experts Team, Doc. JVET-S0146, teleconference, June 2020.

[12] Recommendation ITU-R BT.2100-2 (2018), *Image parameter values for high dynamic range television for use in production and international programme exchange*.

[13] SMPTE ST 2084, *High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays*, 2014.

[14] SMPTE ST 2036-1, *Ultra High Definition Television – Image Parameter Values for Program Production*, 2014.

[15] ETSI TS 101 154, *Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcast and Broadband Applications*, 2019.

[16] ITU-T Tech. Paper HSTP-VID-WPOM and ISO/IEC TR 23008-8 (*Eds.*: K. Andersson, F. Bossen, J.-R. Ohm, A. Segall, R. Sjöberg, J. Ström, G. J. Sullivan, A. Tourapis), Working practices using objective metrics for evaluation of video coding efficiency experiments, July 2020.

[17] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," ITU-T SG 16 Q.6 input document VCEG-M33, 13th VCEG meeting, Austin, Texas, USA, Apr. 2001.

[18] G. Bjøntegaard, "Improvements of the BD-PSNR model," ITU-T SG16 Q.6 document VCEG-AI11, 35th VCEG meeting, Berlin, Germany, July 2008.

[19] RDPlot, https://github.com/IENT/RDPlot, accessed 2021-06-18.

[20] J.-R. Ohm, G. J. Sullivan, "Meeting Report of the 22nd Meeting of the Joint Video Experts Team (JVET), by teleconference, 20–28 April 2021", Joint Video Exploration Team of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-V1000, teleconference, April 2021.

[21] V. Baroncini and M. Wien, "Report on VVC compression performance verification testing in the SDR HD and 360 Video categories", Joint Video Exploration Team of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-V0174, teleconference, April 2021.

[22] F. Bossen, X. Li, K. Sühring, K. Sharman, and V. Seregin, "JVET AHG report: Test model software development (AHG3) ", Joint Video Exploration Team of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-V0003, teleconference, April 2021.

Acknowledgements

The verification test coordinators wish to thank the organizations who contributed to and supported the VVC verification tests, particularly including Alibaba Inc., Bytedance Inc., Ericsson, Fraunhofer HHI, GBTech, Huawei, InterDigital, MediaTek, RWTH Aachen University, Tencent, Vabtech.

Annex A

The same evaluation method as for the HEVC verification tests is adopted for the VVC verification tests. The following description is based on JCTVC-Q1011 [A4] with minor adaptations.

A.1 Test method

The test method adopted for this evaluation is degradation category rating (DCR) [A1].

A.1.1 Degradation Category Rating (DCR)

This test method is commonly adopted when the material to be evaluated shows a range of visual quality that well distributes across all quality scales. All the video material used for these tests consist of video clips of 10 seconds duration.

This method has been used under the schema of evaluation of the quality; for this reason, a quality rating scale made of 11 levels was adopted, ranging from "0" (lowest quality) to "10" (highest quality), see also Figure 2.

The structure of the basic test cell (BTC) of the DCR method was made by two consecutive presentations of the video clip under test; at first the original version of the video clip is displayed, immediately afterwards the coded version of the video clip is presented; then a message displays for 5 seconds asking the viewers to vote. The presentation of the video clips is preceded by a mid-grey screen displaying “Source” for the original and “Test” for the coded version of the sequence under test for one second.

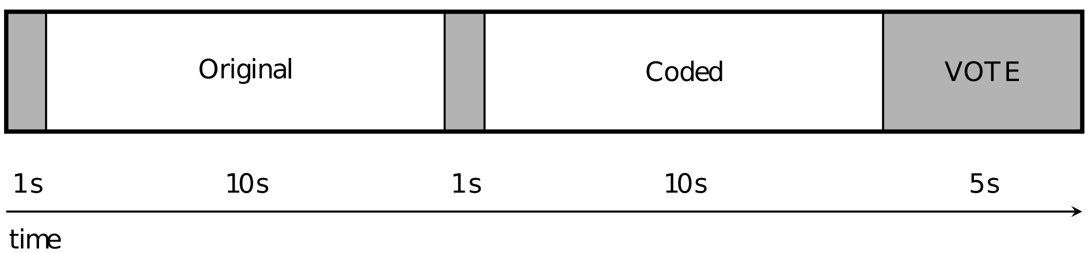

Figure 8 – DCR BTC

A.2 How to express the visual quality opinion with DCR

The viewers were asked to express their vote putting a mark on a scoring sheet.

The scoring sheet for a DCR test is made of a section for each BTC; each section has a box wherein which the viewer shall write the score ranging from 0 to 10. By writing a score of “10”, the subject will express an opinion of “best” quality, while by writing a score of “0” the subject will express an opinion of “worst” quality, as shown in Figure 2.

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 box of the scoring sheet.

A.4 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 demonstrating the range of qualities to be expected in 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 qualities 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.

A.5 The laboratory setup

The laboratories for subjective assessments were arranged according to [A1], except for the selection of the display and the video play-out server. Play-out of the HD video clips was done at the native resolution without upscaling, with the HD video being centered on the UHD screen with a mid-gray surrounding.

The PCs used to play the video sequence supported the display of 10 bit UHD at 30 and 60 frames per second, without any limitation, or without introducing any additional temporal or visual degradation. At GBTech and Vabtech, the connection between the PC and the display was provided by a 10 bit-capable HDMI connection. At RWTH Aachen University, the display was connected via quad-link SDI.

A.6 Viewing environment

The viewing distance was 1.5H, where H is equal to the height of the active part of the screen, depending on the size of the active part of the screen and its native resolution.

The test laboratories were protected from external visual or audio pollution. Internal general light was low (just enough to allow the viewing subjects to fill out the scoring sheets) and a uniform light was placed behind the monitor, in a way no direct light hits the viewing subjects seated in front of the screen; the light behind the monitor must be dimmed to an intensity as specified in Table 4 of Recommendation ITU-T P.911 ("Typical viewing and listening conditions as used in audio-visual quality assessment"). No other light source was admitted, and in particular any light source directed to the screen or creating reflections.

A.7 Overall test effort and subjects' involvement

Each viewing session did not run for more than 20 minutes and the same viewing subject did not participate to the test run for more than six hours in total. Young people were hired as test subjects, selecting them for an age from 16 to 30, mostly students of scientific faculties. Viewing subjects were compensated for their participation to the testing activities.

A.8 Statistical analysis and presentation of the results

The data collected from the score sheets, filled out by the viewing subjects, were stored in an Excel spread sheet. For each coding condition the Mean Opinion Score (MOS) and associated Confidence Interval (CI) values were computed in the spread-sheets.

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

From the "raw" data subject reliability should be calculated and the method used to assess subject reliability should be reported. Some criteria for subjective reliability are given in [A2] and [A3].

A.9 References

[A1] Recommendation ITU-T P.910 (2008), *Subjective video quality assessment methods for multimedia applications*.

[A2] *Pseudo Isochromatic Plates*, engraved and printed by *The Beck Engraving Co., Inc.*, Philadelphia and New York, United States.

[A3] KIRK (R.E.): *Experimental Design – Procedures for the Behavioural Sciences*, 2nd Edition, *Brooks/Cole Publishing Co*., California, 1982.

[A4] Tan, T. K.; Mrak, M.; Baroncini, V.; and Ramzan, N., "Report on HEVC compression performance verification testing," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 output document JCTVC-Q1011, Apr. 2014.