**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION  
ORGANISATION INTERNATIONALE DE NORMALISATION  
ISO/IEC JTC 1/SC 29/AG 5  
MPEG VISUAL QUALITY ASSESSMENT**

**ISO/IEC JTC 1/SC 29/AG 5 N178**

**October 2025, Geneva, CH**

|  |  |
| --- | --- |
| **Title** | **Final report on subjective assessment of MPEG multilayer video coding technology** |
| **Editors** | **P. de Lagrange, InterDigital, O. Chubach, MediaTek, S. Iwamura, NHK** |
| **Status** | **Approved** |
| **S/N** | **25451** |

# Introduction

In this document, the results for the subjective testing according to the multilayer video coding test plan document[[1]](#footnote-1) AG5N160 [1] are reported. The test exercised three types of MPEG multilayer video coding technologies which have different characteristics in terms of features, performance, and complexity, and are capable of dual-layer coding with an HD base layer and a UHD enhancement layer (2x spatial scalability). These comprise SHVC (the scalable extension of HEVC), the VVC multilayer profile, and LCEVC.

It is noted that the three coding technologies used in this test were designed for different target usage scenarios. SHVC was designed as a relatively straightforward extension of the HEVC standard, using the HEVC core coding tools for coding its enhancement layers, with a base layer that could be coded with any single-layer video coding technology. VVC represents the next generation of advancement in compression capability beyond HEVC. Because of reference picture resampling included in VVC, its enhancement layer coding uses the same core coding tools that are used for VVC single-layer coding. LCEVC was designed as a low-complexity enhancement-layer coding technology for use with a base layer that could be coded with any single-layer video coding technology, using relatively simple coding tools in the enhancement layer. As reflected by the “low complexity” in its name, LCEVC has relatively low complexity for its enhancement layer decoding, and it is also designed to have relatively low complexity for encoding as well, as it defines a limited number of relatively simple coding tools.

The focus was to assess the enhancement layer coding performance at good to very good quality, using encoder configurations of the technologies under test that were aligned as much as possible. The configurations were chosen to be independent of visual optimization tools that are available and can be different in the different standards under test.

To isolate the impact of the different technologies of enhancement layer representation, the base layer was as common as possible, coded using VVC but each having sample alignment best fitting the different standards. The base layer quantization step was selected to provide acceptable visual quality when upscaled to the UHD target resolution, while still leaving headroom to evaluate the improvement brought by enhancement layers.

A single enhancement layer was used. To avoid perceptual biases, upsampling used for inter-layer prediction was aligned in terms of perceived sharpness and visual optimization was disabled for all standards (such as quantization matrices for all standards, adaptive quantization step size for VVC and SHVC, and some upsampling processing for LCEVC). No denoising pre-filtering was used for the enhancement layers as it was not available in all test models. The precise settings used in the tests are reported in the Annex to this document.

Two rate points were tested using two configurations with approximately 10% and 50% of the overall bitrate allocated to the enhancement layer, respectively. The amount of enhancement layer data was understood to be a key parameter of dual-layer coding, and the intention was to represent different use cases, where the behaviour of the technologies under test may differ, and where the balance of expectations between coding performance, complexity and functionality may also differ.

The evaluation was performed by formal subjective visual testing according to the DCR / DSIS protocols [3][4].

# Test description

The test configuration applied 2x spatial scalability with:

* two layers for VVC Multi-Layer profile
* LCEVC combined with a VVC base layer
* SHVC combined with a VVC base layer

The resolution of the tested content was HD (base layer) + UHD (enhancement layer). The base layer was intended to be common to all enhancement layers (EL). An upsampled version of the base layer (upsampled from HD to UHD) was included as a test point.

The quality of the base layer was chosen to be in the range of MOS 6 to 7 on the 11-grade scale when compared to an uncompressed original (high quality yet allowing sufficient headroom to see the enhancement). The purpose of this test was to evaluate how various enhancement layers (obtained with the tested standards) allow to visually improve the base layer coded at this quality and understand the behavior of the tested standards in different conditions.

Two test cases were considered:

* 50% of EL corresponds to a simplified use case of quality scalability, where significant improvement is expected as in e.g. adaptive streaming, with a focus on coding efficiency. The goal of such layered coding is to save space in cache storage. The 50% ratio was similarly used in the SHVC Verification Tests. The bitrate ratio also resembles a scenario indicated by ATSC in their liaison letter sent to MPEG [3].
* 10% of EL corresponds to a use case of lightweight enhancement of an existing bitstream, providing improved visual quality of the content at moderate cost. Such ratio was, for example, tested by SBTVD for LCEVC when combined with VVC [4].

The test scenario included SDR, HDR PQ and HDR HLG test sequences, and thereby constitutes a scenario different from previous verification tests. The test focused on settings where the enhancement layer makes a significant difference. Settings relevant for adaptive resolution (when the upsampled base layer would be better than full-resolution coding at the same bitrate) were not part of this test.

The test sequences are reported in Table 1.

**Table 1 Test sequences**

|  |  |  |  |
| --- | --- | --- | --- |
| **Test sequence & format** | **Thumbnail** | **Test sequence & format** | **Thumbnail** |
| BodeMuseum  3840×2160, 60fps, 600 frames  SDR, BT.2020 color  f2b01ec32bd014ecc17cc0db069cb734 | A building with a dome on the side of it  AI-generated content may be incorrect. | Metro  3840×2160, 60fps, 600 frames  SDR, BT.709 color  ed4ac35cd1d659a3bf8cd846f5be0efc | People walking in a building  AI-generated content may be incorrect. |
| OberbaumSpree  3840×2160, 60fps, 600 frames  SDR, BT.2020 color  a5200ad0aee110ed6bd3c560edd96198 | A city next to a body of water  AI-generated content may be incorrect. | SubwayTree  3840×2160, 60fps, 600 frames  SDR, BT.2020 color  fc6a517ab5ce9ec0e28dd1c770c51fc8 | A bridge with a railing  AI-generated content may be incorrect. |
| WaterFront  3840×2160, 24fps, 240 frames  SDR, BT.709 color  1297ba70dea887e8138bbb2119581ebc | A small town near a body of water  AI-generated content may be incorrect. | FootballLargeAdvert  3840×2160, 60fps, 600 frames  HLG, BT.2020 color  cbb3532a286f568c0b0a32f6986c969e | A crowd of people in a stadium  AI-generated content may be incorrect. |
| H3-AMS01  3840×2160, 60fps, 600 frames  HLG, BT.2020 color  02be2d01e46a4d2316fc0e951f94f742 | A group of buses in a city  AI-generated content may be incorrect. | H3-AMS02  3840×2160, 60fps, 600 frames  HLG, BT.2020 color  d701f761b1f44a687fc51eb8c0ab9fbe | A boat on a river  AI-generated content may be incorrect. |
| H3-AMS05  3840×2160, 60fps, 600 frames  HLG, BT.2020 color  bd26ce99ea7da819d56a1a562e8e48df | A row of buildings next to a canal  AI-generated content may be incorrect. | WomenFootball  3840×2160, 50fps, 500 frames  HLG, BT.2020 color  71cc16a0f97a124b8f64de9aafed0c04 | A group of people playing football  AI-generated content may be incorrect. |
| GreenMountains1  3840×2160, 25fps, 250 frames  PQ, BT.2020 color  8e315d6d6a8d6891330008cc84807b0c | A high angle view of a mountain  AI-generated content may be incorrect. | KitchenDressin  3840×2160, 60fps, 600 frames  PQ, BT.2020 color  eefaabe95d4bff35c222973d4407990e | A person sitting in a kitchen  AI-generated content may be incorrect. |
| RiverPlate1  3840×2160, 60fps, 600 frames  PQ, BT.2020 color  0d95e0c1d4c559abfb9c56b44f71a5d3 | A group of logos on a wall  AI-generated content may be incorrect. | TiergartenParkway  3840×2160, 60fps, 600 frames  PQ, BT.2020 color  410e5bb4900840ba3642ba7678c53938 | A row of trees on a road  AI-generated content may be incorrect. |
| WalkInPark  3840×2160, 60fps, 600 frames  PQ, BT.2020 color  2b9c43c6e0f4b4b251756573ec2792a9 | A group of people walking on a path  AI-generated content may be incorrect. |  |  |

For VTM and SHM, the QP of enhancement layer was selected to get the closest match to the 10% and 50% EL bitrate targets. For LTM, SW2 was adjusted to get the closest match to the 10% and 50% EL bitrate targets[[2]](#footnote-2). SW1 was set to 32767 following the rule used in LCEVC verification tests and considering the base QPs used in this test. This effectively disables LCEVC layer 1, in line with the rationale to use a single enhancement layer.

Table 2 – Quantization parameters

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Test sequence** | **Base layer QP** | **SHVC** | | **VVC** | | **LCEVC** | |
| **EL QP** | **EL rate** | **EL QP** | **EL rate** | **SW2** | **EL rate** |
| BodeMuseum | 29 | 32  39 | 53%  10% | 34  41 | 50%  11% | 2625  3125 | 49%  12% |
| Metro | 29 | 31  38 | 54%  9% | 33  39 | 47%  9% | 2250  2938 | 52%  9% |
| OberbaumSpree | 29 | 31  35 | 51%  10% | 33  37 | 49%  10% | 2313  2813 | 51%  10% |
| SubwayTree | 31 | 34  39 | 52%  10% | 35  41 | 53%  9% | 2625  3250 | 52%  9% |
| WaterFront | 27 | 34  42 | 49%  10% | 35  43 | 50%  11% | 3125  3938 | 49%  11% |
| FootballLargeAdvert | 31 | 32  40 | 54%  10% | 34  40 | 47%  10% | 2438  3188 | 49%  10% |
| H3-AMS01 | 29 | 34  41 | 50%  9% | 35  42 | 51%  12% | 2688  3250 | 50%  12% |
| H3-AMS02 | 29 | 34  39 | 46%  10% | 35  41 | 47%  10% | 2500  3063 | 48%  10% |
| H3-AMS05 | 29 | 33  39 | 53%  9% | 34  41 | 49%  9% | 2250  2875 | 50%  9% |
| WomenFootball | 31 | 33  37 | 46%  10% | 34  38 | 47%  8% | 2063  2750 | 48%  8% |
| GreenMountains1 | 27 | 30  34 | 51%  11% | 31  35 | 54%  11% | 1625  2375 | 51%  10% |
| KitchenDressin | 27 | 29  39 | 47%  10% | 30  39 | 52%  10% | 2125  3125 | 53%  10% |
| RiverPlate1 | 27 | 30  36 | 54%  10% | 31  38 | 50%  11% | 2375  3313 | 49%  11% |
| TiergartenParkway | 29 | 33  36 | 50%  14% | 33  38 | 48%  10% | 2125  2750 | 47%  10% |
| WalkInPark | 27 | 31  34 | 47%  12% | 30  34 | 48%  10% | 1563  2313 | 49%  10% |

For the evaluation, the following software versions were used:

* SHM version 12.4
* VTM version 23.8
* LTM version 7.0

In addition, reconstructed and upscaled videos of the base layer streams were evaluated as an anchor point. The LTM base layer streams were used for this purpose.

Bitstreams were generated by InterDigital and MediaTek. Crosschecks were performed by Ericsson, InterDigital, and MediaTek [5][6][7].

# Logistics

As recommended by AG 5 at its 19th meeting, RWTH Aachen University and VABTech performed the subjective visual tests.

RWTH coordinated the test effort including all administrative matters. The test was independently performed in the two labs in order to achieve two independent measurements for all test points.

## RWTH test setup

| **Test Site** | **RWTH Aachen University** |
| --- | --- |
| **Display, size (resolution setting), connection** | * 1× LG OLED77 (3840×2160), HDMI * 1× LG OLED65CX (3840×2160), HDMI * 2× LG OLED55G19LA (3840×2160), HDMI   The source signal was sent from the video board of the PC in parallel to the displays via an HDMI splitter. |
| **Viewing distance** | 3 viewers at 1.5H for the 77” and 65” displays, 2 viewers at 1.5H for the 55” displays. The HD video signal was displayed centered to the UHD screen with a mid-grey padding for the unused area. |
| **Viewing angle** | 90°, 70° |
| **Total number of viewers** | 36 viewers (8 females, 28 males; age 20‑27, 9 different nationalities), all screened for visual acuity and normal colour vision |

The test subjects were paid for their participation in the visual tests.

## VABTech test setup

| **Test Site** | **GBTech** |
| --- | --- |
| **Display, size (resolution setting), connection** | * 2× LG OLED65CX (3840×2160), HDMI. * Each display was connected by an HDMI cable to a different server; the two servers were identical and run the same video player. * All scores were collected by means of paper scoring sheets; then the scores were inserted in an Excel spreadsheet by the test administrators. |
| **Viewing distance** | 2 viewers at 1.5H per display, the HD video signal was displayed centered to the UHD screen with a mid-grey padding for the unused area. |
| **Viewing angle** | 70° |
| **Total number of viewers** | 19 viewers (7 females, 12 males; age 18‑23, all Italian), all screened for visual acuity and normal colour vision. |

All test subjects were paid for their participation to the testing experiment

# Test design

## Test sessions

The tests were conducted according to the DCR of ITU-T P.910 using the 11-grade impairment scale shown below.

Ein Bild, das Text, Screenshot, Schrift, Zahl enthält.

Automatisch generierte Beschreibung

***Figure 1: 11 grades impairment scale***

### RWTH Aachen University

The viewers were called for separate viewing sessions for each category in which the respective test sessions were conducted. The instructions, training, and viewing sessions were conducted by an experienced student worker.

#### Design of test sessions

For each category, SDR, HDR-HLG, HDR-PQ, two test sessions were designed with 27 basic test cells (BTCs) each, resulting in a total of 6 test sessions. The test sessions included a stabilization phase of three BTCs and one BTC evaluating the uncompressed sequence to enable checking of the viewers’ attention.

#### Conduction of tests

The tests were performed with up to 20 viewers per viewing session, who were split into two groups of max 10 viewers for the training and the test sessions. The training as well as the tests were conducted in an interleaved fashion where one group was performing the session while the other group was resting, respectively. In each session, the viewers were provided with voting sheets, a pen and a clipboard to note down their scores. The viewers of the resting group were asked to enter their scores of the previous test sessions into a database for further processing.

#### Training

At the start of the viewing session, the participants were instructed on the test. The concept of the evaluated MPEG multilayer video coding technology with a HD base and UHD enhancement layer, as well as the size of the sessions, was described. The DCR test protocol was explained by providing the participants with the viewers’ instructions available in AG5N150 [8]. After answering potential questions from the viewers, they were shown the training session. The training sessions consisted of five BTCs. For each category the training session comprised the coding technologies under test using the 10% and 50% bitrate points in a balanced way.

### VABTech

The viewing subjects were trained by means of a short practice (training) session. The video material used for the training session were 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.

Considering the total number of test points, the test coordinator decided to divide the video clips representing all the test points (Anchor + Proposals) in two separates test sessions. Each test session was created following what is recommended in Rec. ITU-R BT.500, i.e. adding at the beginning of each test session a “stabilization phase” made of three BTCs; the stabilization BTCs were selected representing best, mid and worst quality. In this way, the viewing subjects had an indication of the range of quality they were expected to evaluate during that session. The scores of the stabilization phase were discarded. Consistency of the behaviour of the subjects were checked by inserting in the session a BTC in which the uncompressed video (SRC) is compared to itself.

# Test results and graphs

Overall, more than 10 000 opinion scores () were recorded in each of the laboratories during this test effort. The acquired results were first independently and carefully processed by each of the labs with respect to outlier detection and consistency analysis.

## Data processing

### RWTH Aachen University

At RWTH, post-screening according to ITU-R BT.500 A1-2.3 was applied to the collected scores. The processing included outlier rejection based on the Pearson correlation (A1-2.3.3) and kurtosis-based post-screening applicable for test following the DSIS/DCR protocol (A1-2.3.1). The processing steps were applied on a per-session basis leading to 16.7% of the raw data being excluded from further evaluation.

### VABTech

At VABtech, post-test screening of the raw data produced by the viewers was done to exclude the outliers. This screening was according to what is specified in ITU-R BT.500.

## MOS results and MOS-over-rate plots

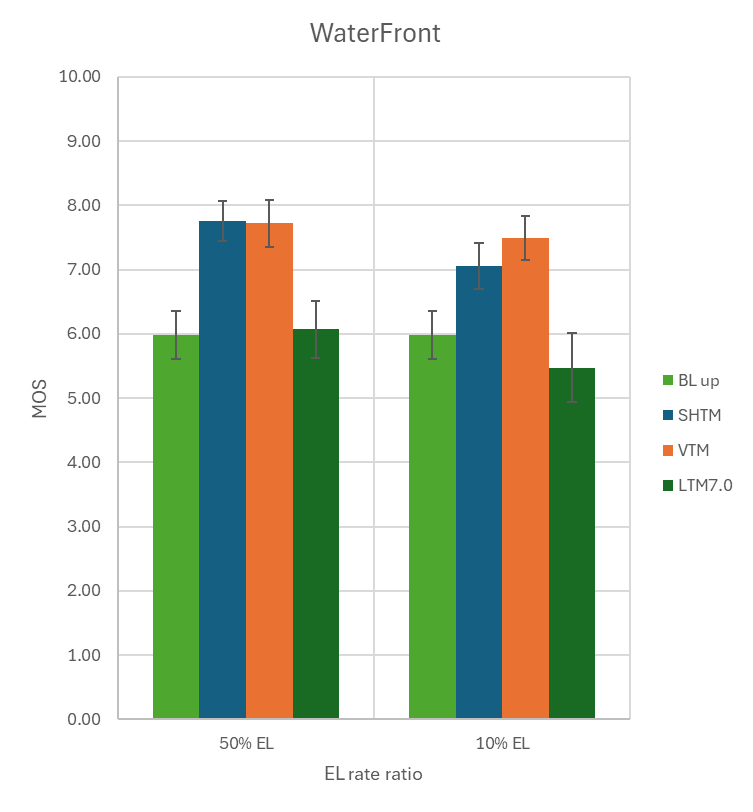
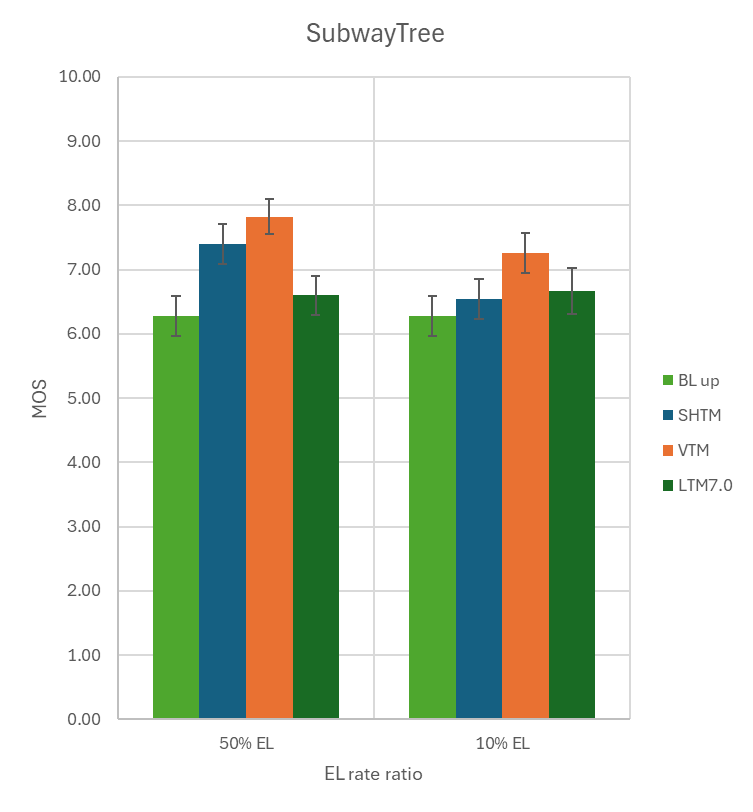
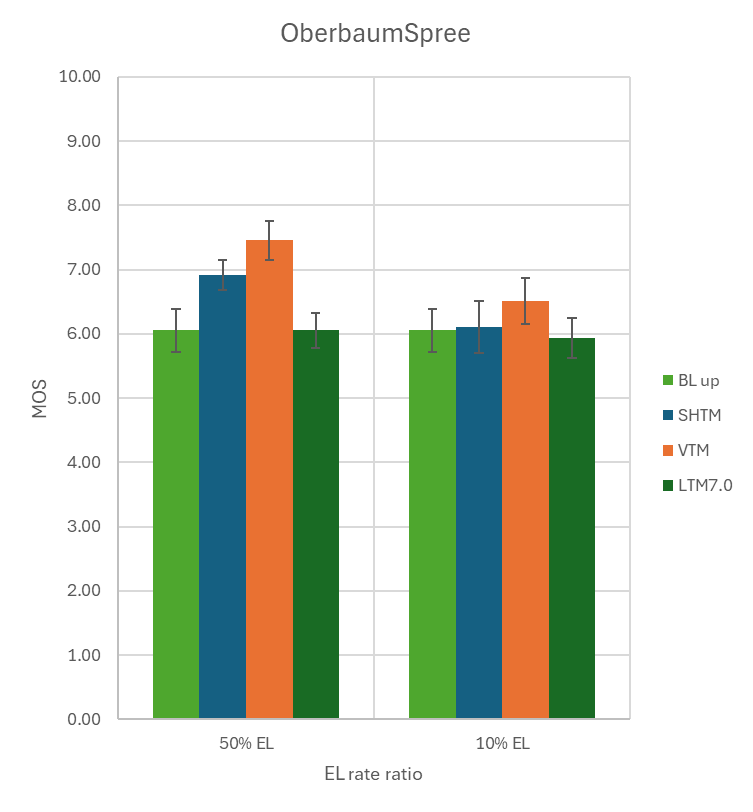
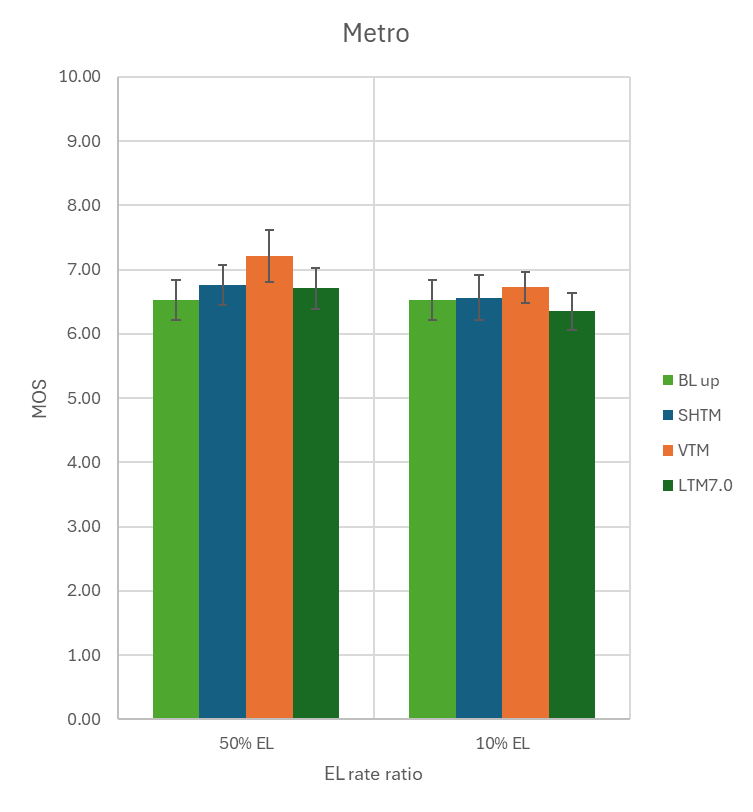
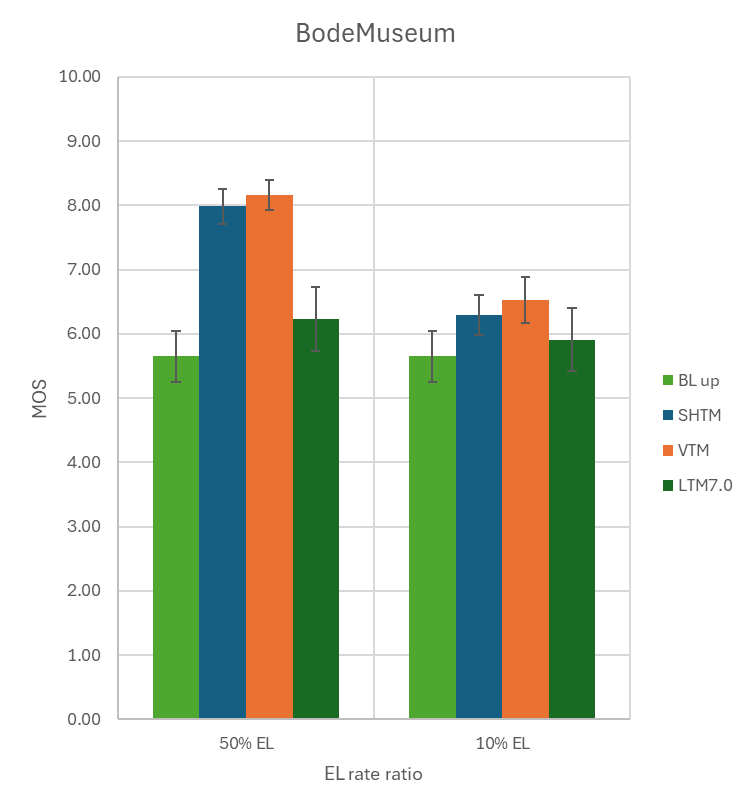
After completion of the data acquisition, the results were processed independently in both laboratories and then analyzed. Both achieved mostly consistent results, with many test points showing the same ranking of the coding technologies under test. For some cases, the order of the coding technologies was different. Due to differences in the training process, the number range of both laboratories were slightly different. The overall average across all recorded opinion scores was 6.1 at RWTH and 6.9 at VABtech. In order to accommodate all expressed viewer opinions, it was decided to merge the consolidated data of both laboratories for reporting and to compute the MOS and confidence interval over the pool of all viewers.

The plots presented below show the resulting overall MOS scores with the 95% confidence interval (CI) marked for both, the positive and negative direction. For each category, the resulting MOS and CI values are reported in the tables and plots below. The results are shown in a grouped manner for the 10% and 50% enhancement layer (EL) cases (E10/E50).

In the table, BL up indicates the result for the upscaled video of the base layer bitstream.

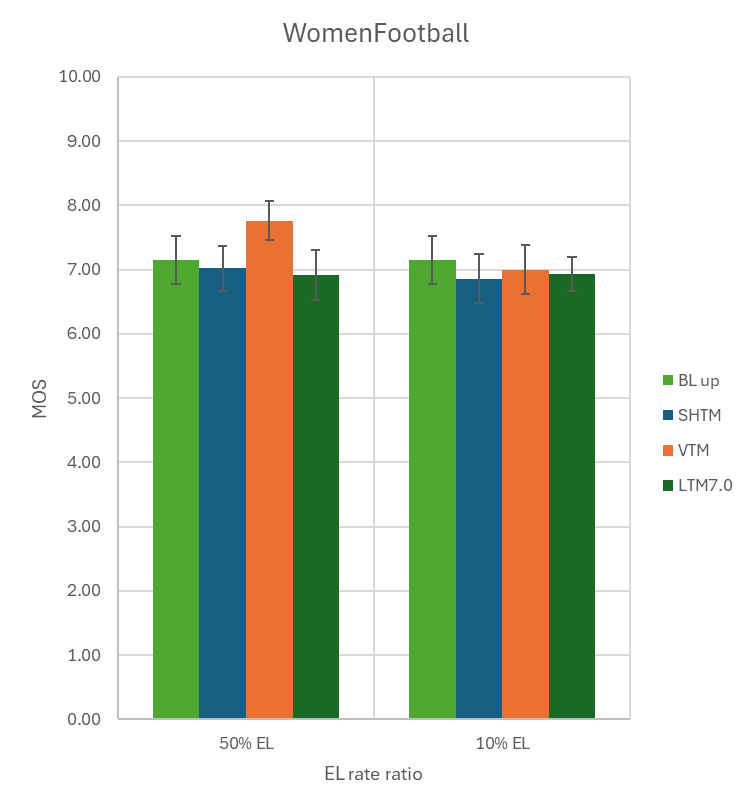
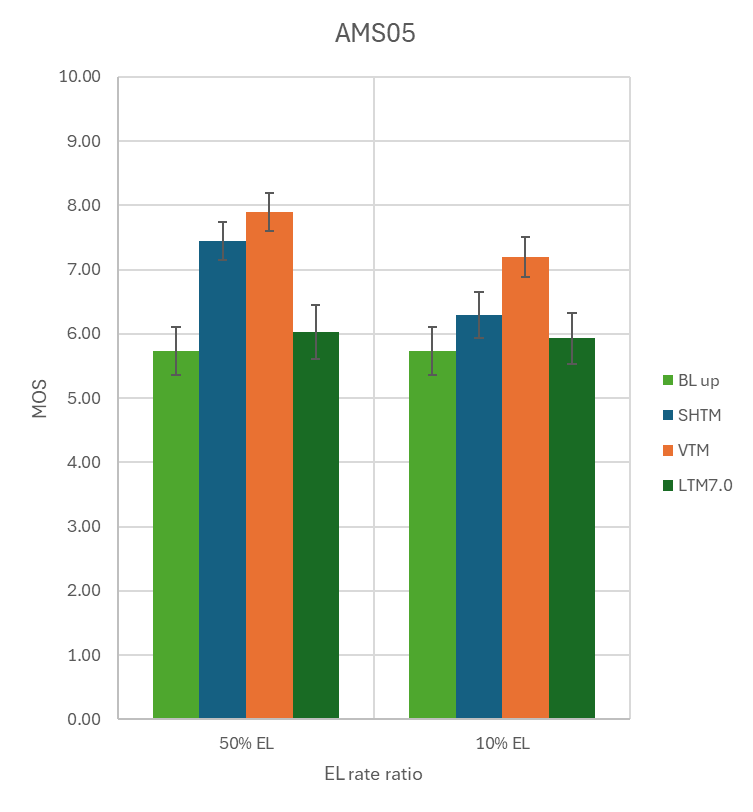
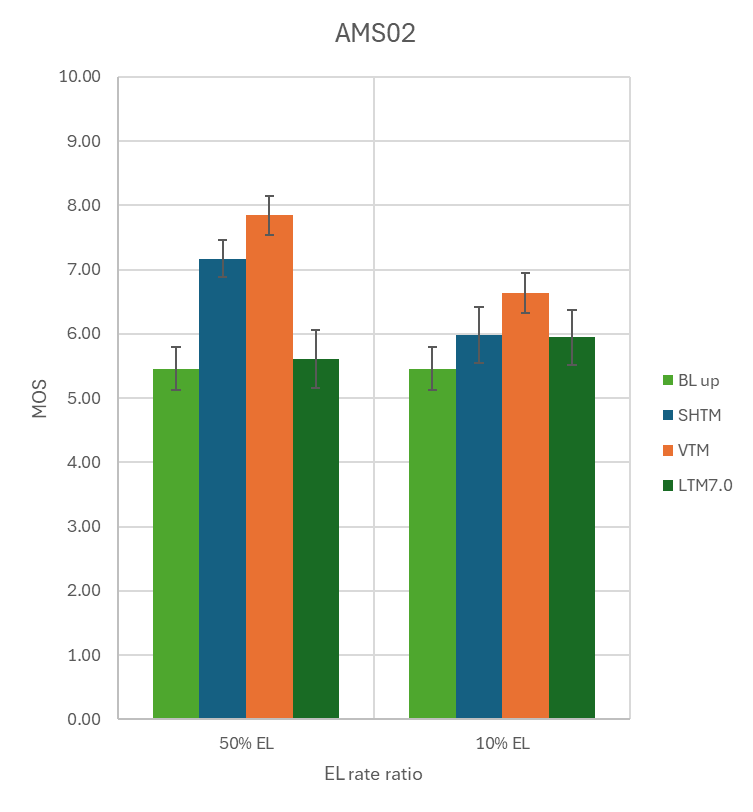
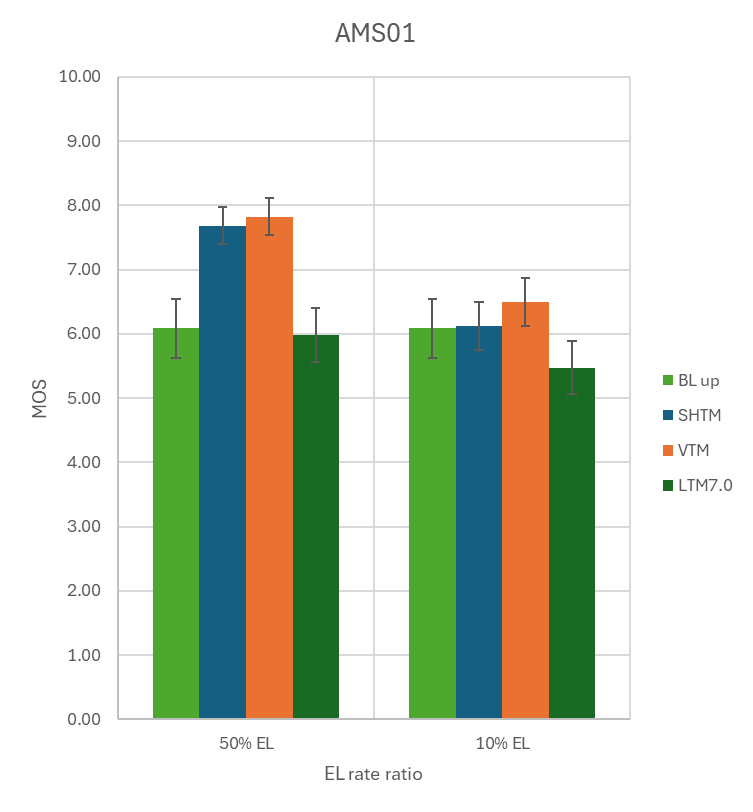
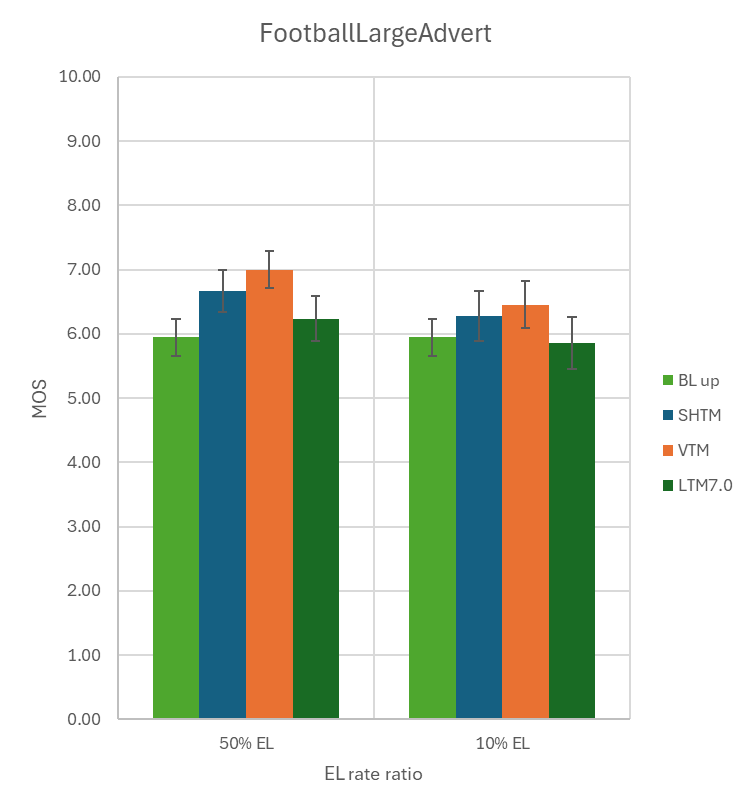
### SDR test sequences

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | BL up |  |  | SHM |  | VTM |  | LTM7.0 |  |
| **SeqName** | **MOS** | **CI** | **Config** | **MOS** | **CI** | **MOS** | **CI** | **MOS** | **CI** |
| BodeMuseum | 5.65 | 0.40 | E50 | 7.98 | 0.28 | 8.16 | 0.23 | 6.24 | 0.50 |
|  |  |  | E10 | 6.30 | 0.31 | 6.53 | 0.36 | 5.91 | 0.49 |
| Metro | 6.53 | 0.32 | E50 | 6.75 | 0.31 | 7.21 | 0.41 | 6.71 | 0.32 |
|  |  |  | E10 | 6.56 | 0.36 | 6.73 | 0.24 | 6.35 | 0.29 |
| OberbaumSpree | 6.05 | 0.33 | E50 | 6.91 | 0.23 | 7.45 | 0.30 | 6.05 | 0.27 |
|  |  |  | E10 | 6.11 | 0.41 | 6.51 | 0.35 | 5.93 | 0.31 |
| SubwayTree | 6.28 | 0.31 | E50 | 7.40 | 0.31 | 7.82 | 0.28 | 6.60 | 0.30 |
|  |  |  | E10 | 6.54 | 0.31 | 7.25 | 0.31 | 6.67 | 0.36 |
| WaterFront | 5.98 | 0.37 | E50 | 7.75 | 0.31 | 7.72 | 0.37 | 6.07 | 0.44 |
|  |  |  | E10 | 7.05 | 0.36 | 7.49 | 0.34 | 5.47 | 0.54 |



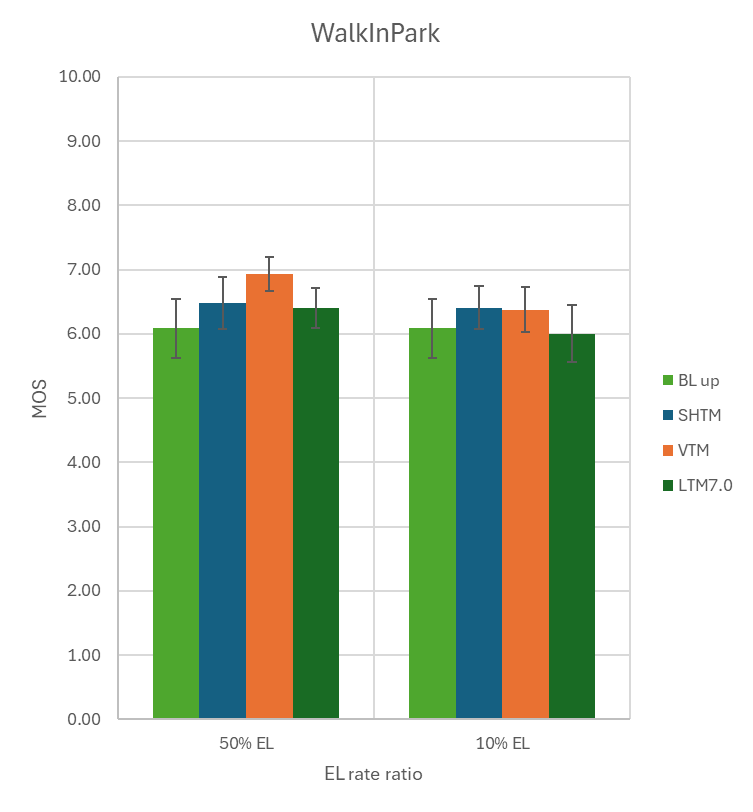
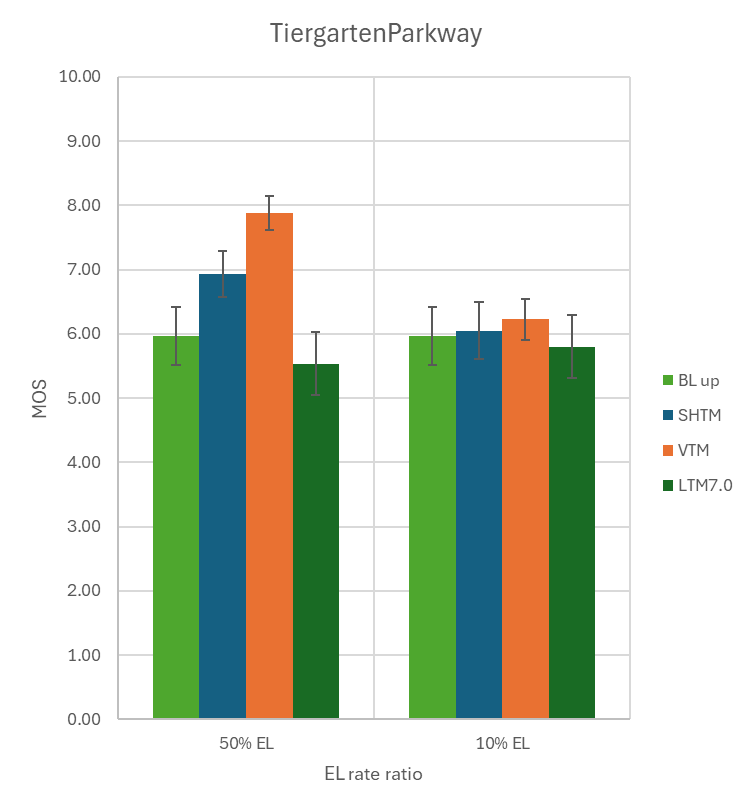
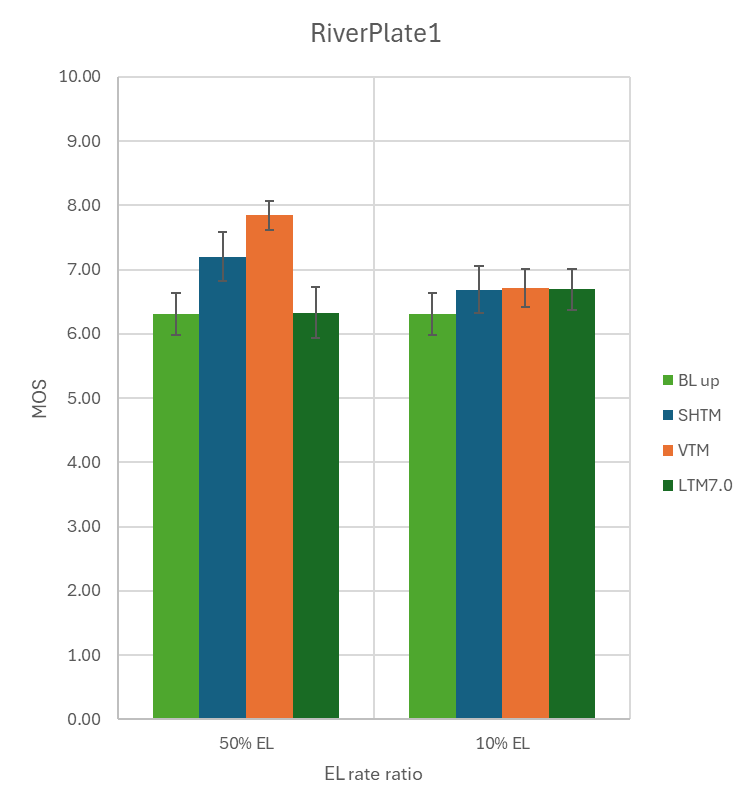
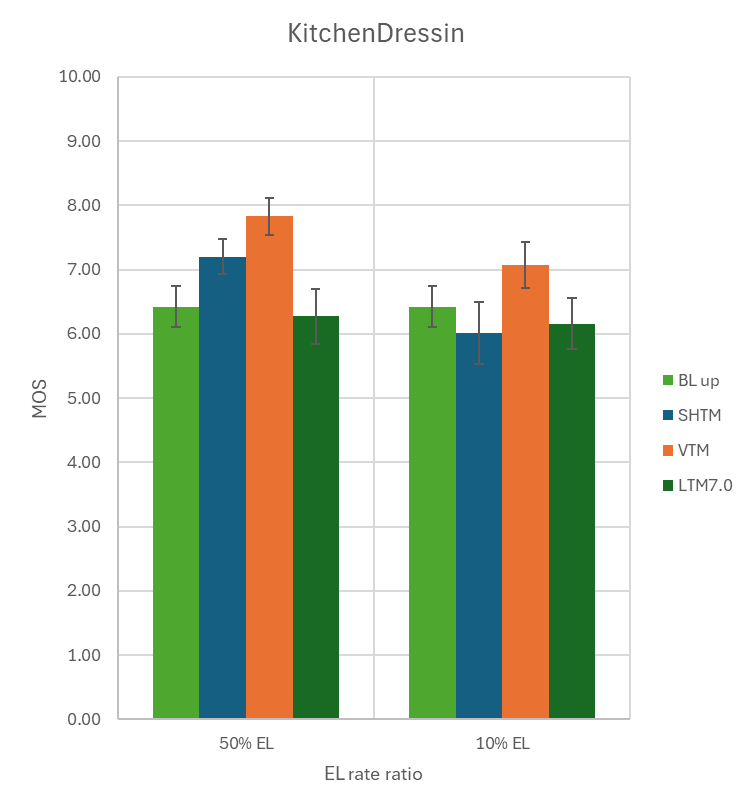
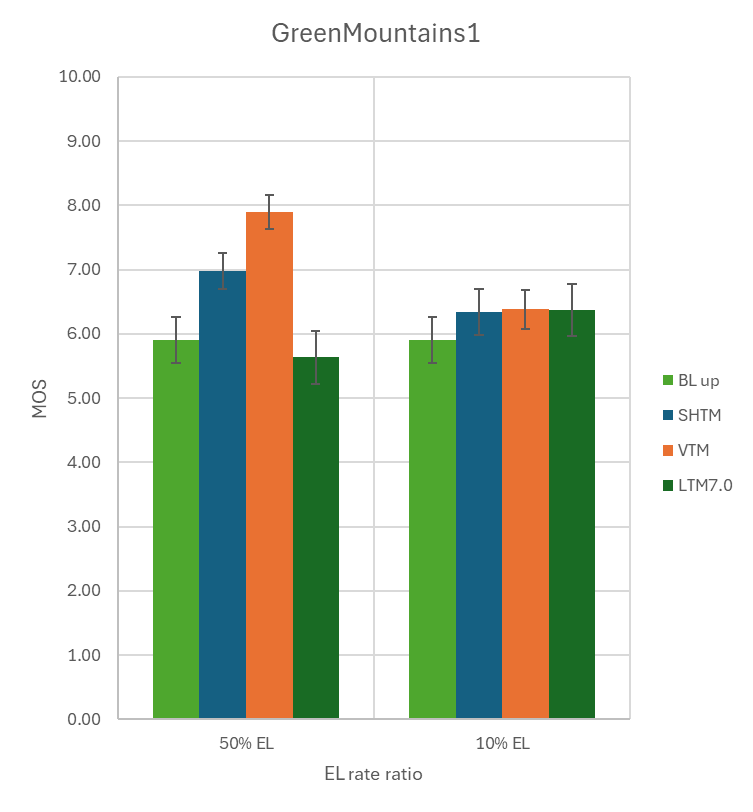
### HDR HLG test sequences

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | BL up |  |  | SHM |  | VTM |  | LTM7.0 |  |
| **SeqName** | **MOS** | **CI** | **Config** | **MOS** | **CI** | **MOS** | **CI** | **MOS** | **CI** |
| FootballLargeAdvert | 5.95 | 0.29 | E50 | 6.67 | 0.32 | 7.00 | 0.29 | 6.24 | 0.35 |
|  |  |  | E10 | 6.27 | 0.39 | 6.46 | 0.36 | 5.86 | 0.40 |
| AMS01 | 6.08 | 0.46 | E50 | 7.68 | 0.29 | 7.82 | 0.29 | 5.98 | 0.42 |
|  |  |  | E10 | 6.12 | 0.37 | 6.49 | 0.37 | 5.47 | 0.42 |
| AMS02 | 5.46 | 0.34 | E50 | 7.17 | 0.29 | 7.84 | 0.31 | 5.61 | 0.45 |
|  |  |  | E10 | 5.98 | 0.44 | 6.63 | 0.31 | 5.95 | 0.43 |
| AMS05 | 5.74 | 0.38 | E50 | 7.44 | 0.30 | 7.90 | 0.29 | 6.04 | 0.42 |
|  |  |  | E10 | 6.29 | 0.36 | 7.19 | 0.31 | 5.93 | 0.40 |
| WomenFootball | 7.15 | 0.37 | E50 | 7.02 | 0.34 | 7.76 | 0.30 | 6.91 | 0.39 |
|  |  |  | E10 | 6.86 | 0.38 | 7.00 | 0.38 | 6.93 | 0.27 |



### HDR PQ test sequences

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | BL up |  |  | SHM |  | VTM |  | LTM7.0 |  |
| **SeqName** | **MOS** | **CI** | **Config** | **MOS** | **CI** | **MOS** | **CI** | **MOS** | **CI** |
| GreenMountains1 | 5.90 | 0.36 | E50 | 6.98 | 0.28 | 7.90 | 0.27 | 5.64 | 0.41 |
|  |  |  | E10 | 6.34 | 0.36 | 6.38 | 0.30 | 6.37 | 0.40 |
| KitchenDressin | 6.42 | 0.32 | E50 | 7.20 | 0.28 | 7.83 | 0.29 | 6.27 | 0.42 |
|  |  |  | E10 | 6.02 | 0.48 | 7.07 | 0.36 | 6.16 | 0.40 |
| RiverPlate1 | 6.31 | 0.33 | E50 | 7.20 | 0.39 | 7.85 | 0.23 | 6.33 | 0.40 |
|  |  |  | E10 | 6.69 | 0.37 | 6.71 | 0.29 | 6.69 | 0.32 |
| TiergartenParkway | 5.97 | 0.45 | E50 | 6.93 | 0.36 | 7.88 | 0.26 | 5.53 | 0.49 |
|  |  |  | E10 | 6.05 | 0.44 | 6.22 | 0.32 | 5.80 | 0.49 |
| WalkInPark | 6.09 | 0.46 | E50 | 6.48 | 0.40 | 6.93 | 0.26 | 6.41 | 0.31 |
|  |  |  | E10 | 6.41 | 0.34 | 6.37 | 0.35 | 6.00 | 0.44 |



# Conclusions

This is a report of the results of the visual assessment according to the test plan. Additional information highlighting the individual properties of the multilayer video coding technologies under test is provided in the annexes. These include aspects such as complexity, flexibility / standard-agnosticism of the technology for the base layer, further applicable tools and aspects of the investigated multilayer video coding technologies.

# References

1. ISO/IEC JTC 1/SC 29/AG 5, “Test plan for assessing MPEG multilayer video coding technology,” Doc. AG5N160, 19th meeting, online, Mar 2025.
2. ISO/IEC JTC 1/SC 29/AG 5, “Corrigendum to test plan for assessing MPEG multilayer video coding technology,” Doc. AG5N167, 20th meeting, Daejeon, KR, Jul 2025.
3. Recommendation ITU-T P.910 (2021), *Subjective video quality assessment methods for multimedia applications*.
4. Recommendation ITU-R BT.500-14 (2019), *Methodologies for the subjective assessment of the quality of television images.*
5. P. de Lagrange, “[AHG ML] Bitstreams and cross-checks for ML tests,” doc. m72913, AG 5, Daejeon, KR, Jun 2025.
6. O. Chubach, “[AHG ML] VTM and LTM bitstream generation for verification testing of MPEG multilayer video coding technologies,” doc. m72852, AG 5, Daejeon, KR, Jun 2025.
7. K. Andersson, “Cross-check of AG5N160,” Doc. m72850, AG 5, Daejeon, KR, Jun 2025.
8. ISO/IEC JTC 1/SC 29/AG 5, “CVQM - Dataset of compressed video for study of quality metrics (version 3),” Doc. AG5N150, 18th meeting, Geneva, CH, Jan 2025.
9. ITU-T Technical Paper HSTP-ASC (2020), *Working practices using objective metrics for evaluation of video coding efficiency experiments*.

# Annex 1 – Test software and configuration

## Base layer

An example VTM command line is given below:

EncoderApp -c cfg/encoder\_randomaccess\_vtm.cfg -q <QP> -c resHD\_60.cfg \  
-i <hd\_input.yuv> -ip <IP> -b <output.bin>

<QP> is set according to Table 2.

<IP> is set to the multiple of 32 closest to frame rate.

encoder\_randomaccess\_vtm.cfg is found in the VTM code tree.

resHD\_60.cfg contains the following (example for 10-bit 60fps test content):

InputBitDepth : 10 # Input bitdepth

InputChromaFormat : 420 # Ratio of luminance to chrominance samples

FrameRate : 60 # Frame Rate per second

FrameSkip : 0 # Number of frames to be skipped in input

SourceWidth : 1920 # Input frame width

SourceHeight : 1080 # Input frame height

FramesToBeEncoded : 600 # Number of frames to be coded

Level : 4.1

The base layer for VVC multi-layer coding, when encoded separately, is required to additionally use --MaxNumALFAPS=4, to prevent errors with the APS overriding in the EL. In this test, the same base layer is used for VVC and SHVC. The base layer for LCEVC does not use the -–MaxNumALFAPS option.

For BT.2020 SDR content, the “--VerCollocatedChroma=1” option is added.

For HDR PQ content, the “-c cfg/per-class/classH1.cfg” option is added. For HDR HLG content, the “-c cfg/per-class/classH2.cfg” option is added. Those configuration files are found in the VTM code tree.

## VVC enhancement layer

An example VTM command line is given below, following the software manual:

EncoderApp -c cfg/encoder\_randomaccess\_vtm.cfg \  
-c cfg/multi-layer/two\_layers\_scalable.cfg --DMVREncMvSelect=1 \  
-l0 -q <QP\_BL> -l0 -c resHD\_60.cfg -l0 -i <hd\_input.yuv> \  
-l1 -q <QP\_EL> -l1 -c resUHD\_60.cfg -l1 -i <uhd\_input.yuv> \  
-l1 -c cfg/multi-layer/layer1\_gop\_randomaccess.cfg -l1 --Level=6.1 \  
-l1 --TemporalFilter=0 -ip <IP> -b <output.bin> -o /dev/null

To reuse a base layer encoding, “–l0 -–debugBitstream <baselayer.bin>” can be used.

<IP> and <QP\_BL> shall match the base layer settings described in the previous section. <QP\_EL> is adjusted to match the target bitrate, as described in Table 2.

cfg/multi-layer/layer1\_gop\_randomaccess.cfg is found in the VTM code tree.

resUHD\_60.cfg is similar to the resHD\_60.cfg shown in the previous section, but with UHD resolution (width 3840, height 2160, and level 5.1).

For BT.2020 SDR content, the “--VerCollocatedChroma=1” option is added.

For HDR PQ content, the “-c cfg/per-class/classH1.cfg” option is added. For HDR HLG content, the “-c cfg/per-class/classH2.cfg” option is added. Those configuration files are found in the VTM code tree.

## SHVC enhancement layer

An example SHM command line is given below (for a 60 fps test sequences), following the software manual:

TAppEncoder -c cfg/encoder\_randomaccess\_scalable10.cfg -c <testpoint.cfg> \  
-c cfg/layers\_avcbase.cfg -ibl <decoded\_baselayer.yuv> -q1 <QP\_EL> \  
-b <output.bin>

encoder\_randomaccess\_scalable10.cfg and cfg/layers\_avcbase.cfg are found in the SHM code tree.

For BT.2020 color content (either SDR or HDR), the “--PhaseVerChroma1=0” option is added.

testpoint.cfg contains the following:

FrameSkip: 0

FramesToBeEncoded: 600

Level0: 5.1

Level1: 4.1

Level2: 5.1

InputFile0: <hd\_input.yuv> # Note: unused (base layer is external)

FrameRate0: 60

SourceWidth0: 1920

SourceHeight0: 1080

IntraPeriod0: <IP>

ConformanceMode0: 1

QP0: <QP\_BL> # Note: unused (base layer is external)

InputBitDepth0: 10

InternalBitDepth0: 10

RepFormatIdx0: 0

LayerPTLIndex0: 1

OutputBitDepth0: 10

InputFile1: <uhd\_input.yuv>

FrameRate1: 60

SourceWidth1: 3840

SourceHeight1: 2160

IntraPeriod1: <IP>

ConformanceMode1: 1

QP1: <QP\_EL>

InputBitDepth1: 10

InternalBitDepth1: 10

RepFormatIdx1: 1

LayerPTLIndex1: 2

OutputBitDepth1: 10

<IP> is set according to the same rule as VVC. <QP\_EL> is adjusted to match the target bitrate, as described in Table 2.

## LCEVC enhancement layer

An example LTM command line is given below (for a 60 fps test sequences), following the software manual:

ModelEncoder -i <uhd\_input.yuv> -o <output.bin> --output\_recon=<rec.yuv> \  
-w 3840 -h 2160 -r 60 -l 600 -f yuv420p10 --encapsulation=nal \  
-b vvc --base <baselayer.bin> --base\_recon <decoded\_baselayer.yuv> \  
--qp=<QP\_BL> --intra\_period=<IP> --base\_depth=10 --parameter\_config=default \  
--dithering\_control=false --dithering\_type=none \  
--level\_1\_filtering\_enabled=false --quant\_matrix\_mode=custom \  
--qm\_coefficient\_1="0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0" \  
--qm\_coefficient\_2=”0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0” \  
--cq\_step\_width\_loq\_0=<SW2> --cq\_step\_width\_loq\_1=<SW1> \  
--predicted\_residual=false --encoding\_upsample=adaptivecubic \  
--upsampling\_coefficients="1945 14997 3825 493"

<QP\_BL> and <IP> match the base layer encoding.

<SW1> and <SW2> are adjusted to match the target bitrate, as described in Table 2.

It is noted that this command line does not include the encoder parameter --priority\_mode=mode\_0\_0.

Note: the following change (bug fix) is first made to the LTM. In file encoder/src/FileEncoder.cpp, the line:

parameters\_["encoding\_upsample"].get\_vector<unsigned>(upsampling\_coefficients\_, 4);

is replaced with:

parameters\_["upsampling\_coefficients"].get\_vector<unsigned>(upsampling\_coefficients\_, 4);

# Annex 2 – Selected features

This section reports selected features for each of the tested standards. More detailed description and additional information can be obtained from the corresponding specification documents.

Table A1. Selected features for the tested standards

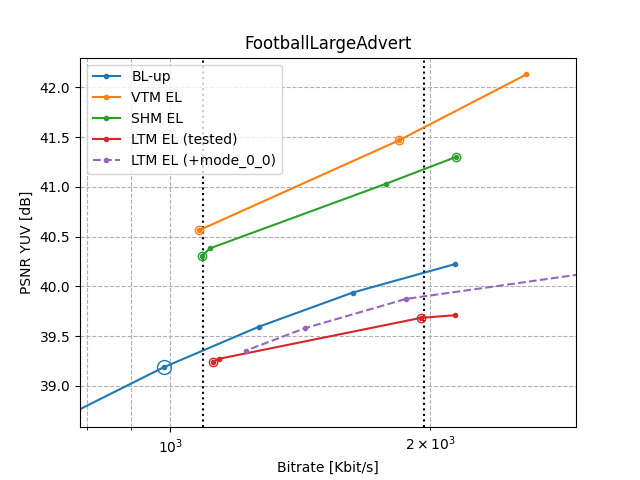
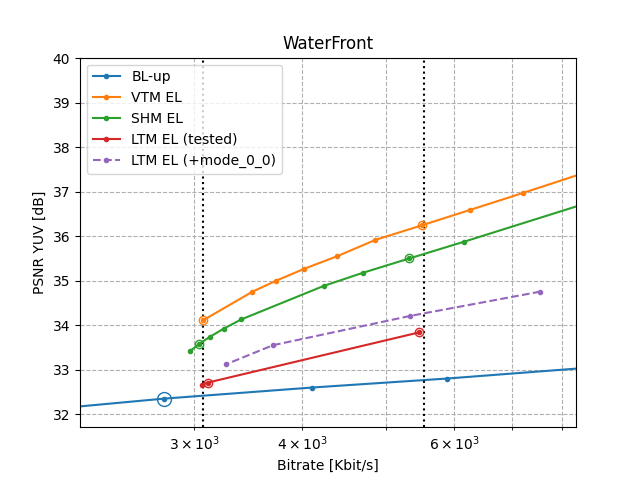
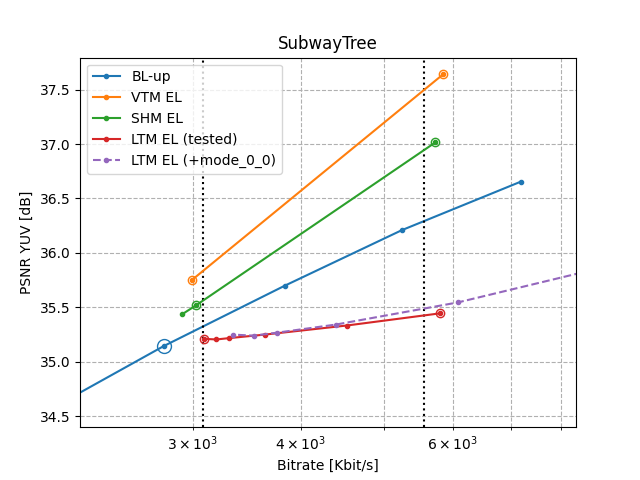
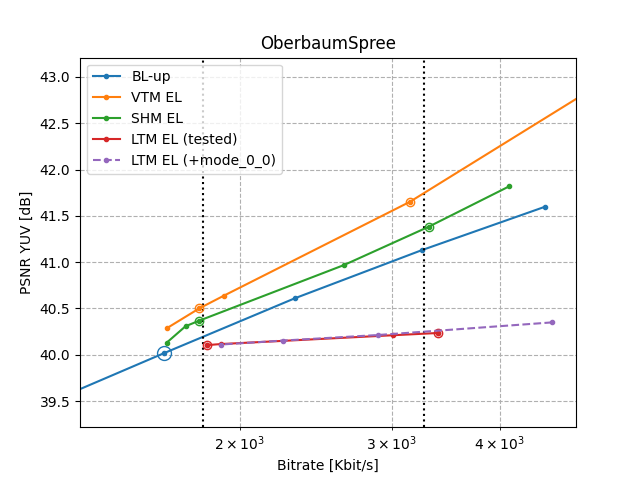
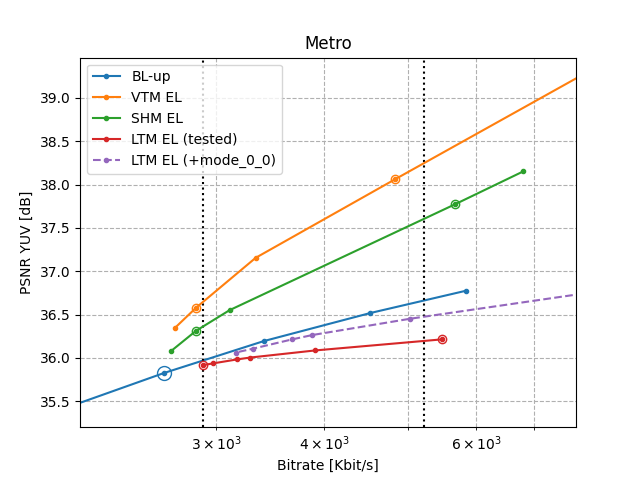
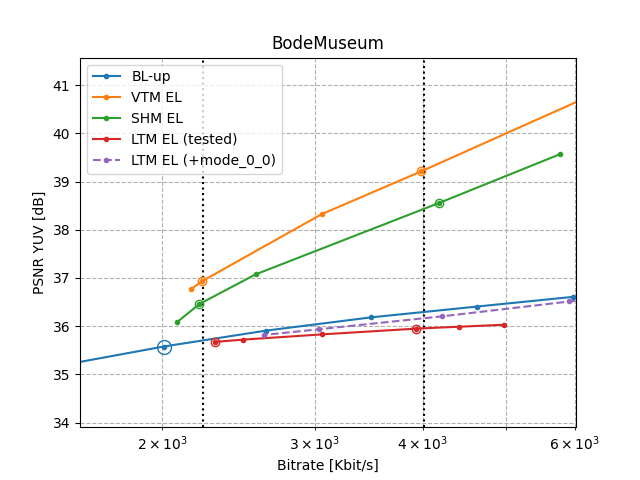
|  |  |  |  |
| --- | --- | --- | --- |
|  | VVC | SHVC | LCEVC |
| External base layer | Not supported | Supported | Always |
| Number of layers | 1 to 64 | 1 to 63 | 1 layer consisting of 1 or 2 sublayers (L1 and/or L2), on top of base |
| Layers video output | Output layer sets: combinations of layers, several outputs possible | Output layer sets | Single output |
| Type of scalability | Spatial, temporal, quality, content, multiview | Spatial, temporal, quality, content, multiview, color | Spatial, quality |
| Scaling ratio | Fractional, from 0.5x to 8x | Fractional, >= 1.0x | Per sublayer: 1x, 2x horizontal, 2x horizontal+vertical |
| Scaling window | Supported | Supported | Not supported |
| Inter-layer luma sample alignment | Full sample | Flexible | Centered-samples |
| Inter-layer chroma sample alignment | Common vertical and horizontal collocated flags | Flexible | Fixed |
| Upsampling filters (inter-layer) | Fixed (lanczos), 8 taps | Fixed (lanczos), 8 taps | Flexible (4 predefined + user-defined, 4 taps) |
| Inter-layer texture prediction | Yes | Yes | Yes |
| Inter-layer motion prediction | Yes (when base is same size) | Yes (when base is HEVC) | No |
| Picture buffers | Multiple reference picture buffers (coded size), shared across layers | Multiple reference picture buffers (coded size), per layer | 1 residual temporal buffer (L2 coded size) |
| Partitioning | CTU 128 + Hierarchical QT/BT/TT + SBT, ISP | CTU 64 + Hierarchical QT | Uniform 2x2 or 4x4 (transform size) |
| Intra prediction | DC, planar + 65 directional modes + IBC, MIP | DC, planar + 33 directional modes | No |
| Inter prediction | Motion compensation with resampling, merge, affine, GPM, CIIP, SBTMVP | Motion compensation, merge, AMVP | Residual copy from L2 temporal buffer |
| Transforms | 3 DCT variants, secondary transform | 2 DCT variants | Hadamard-based |
| Entropy coding | CABAC | CABAC | Huffman-based, RLE |
| Quantization matrices | Flat or user-defined | Flat, default (non-flat), or user-defined | Default (non-flat) or user-defined. |
| Local adaptation of quantization | Yes | Yes | No |
| Post-filtering | Non-normative, supported by optional data (SEI messages) | Non-normative, supported by optional data (SEI messages) | Non-normative, optional dithering parameter |

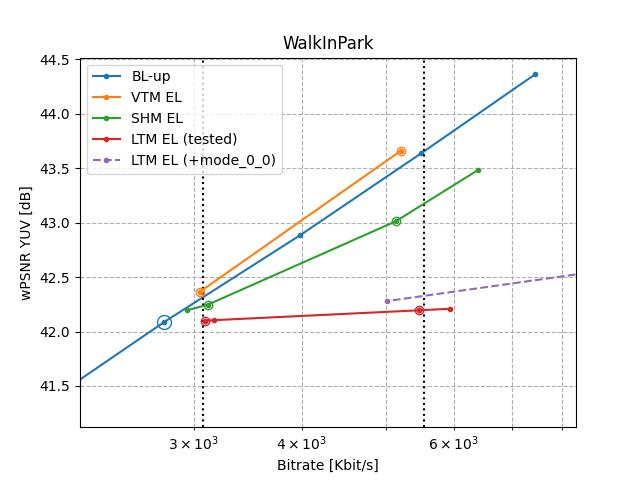
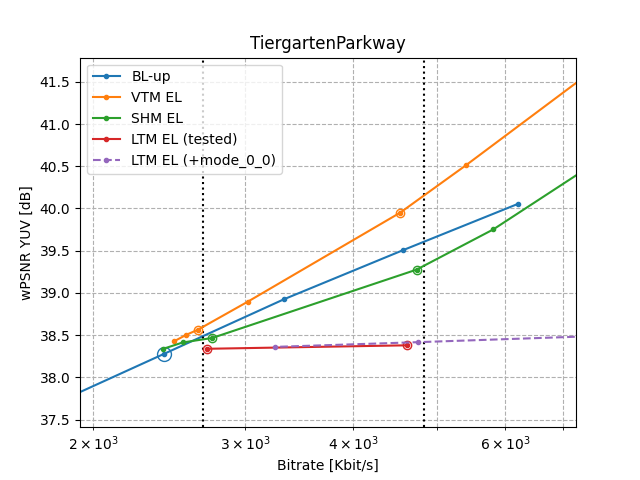
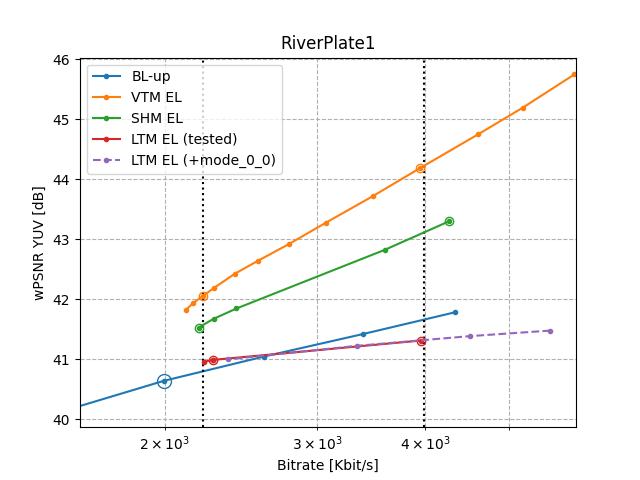
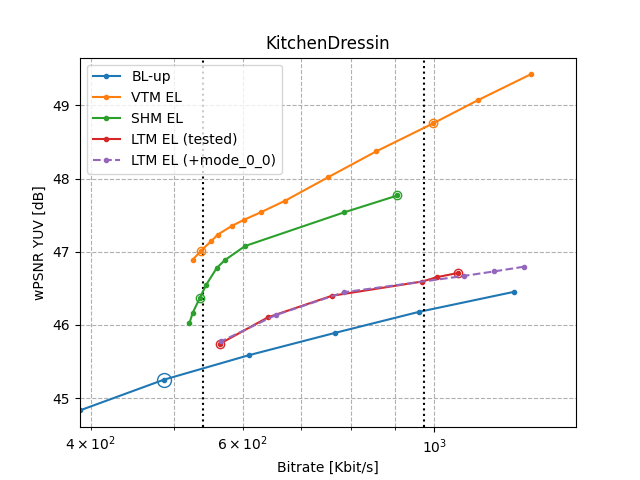
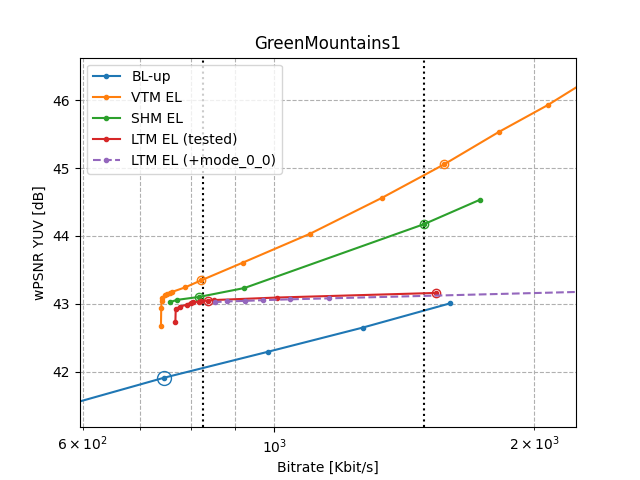
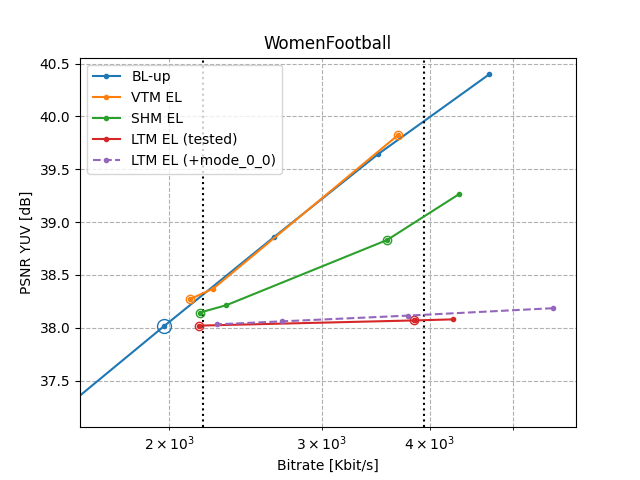
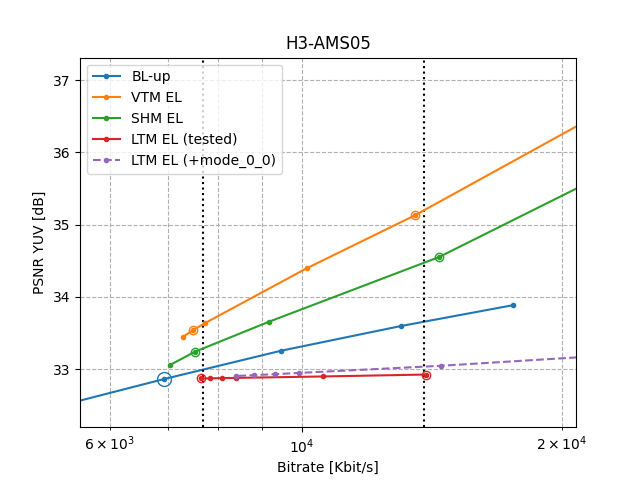
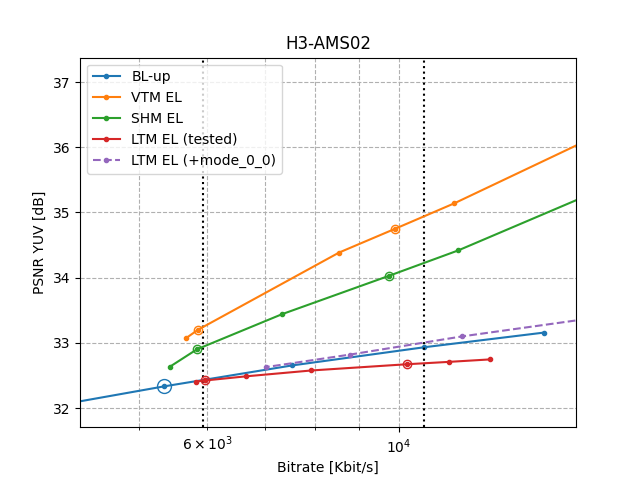
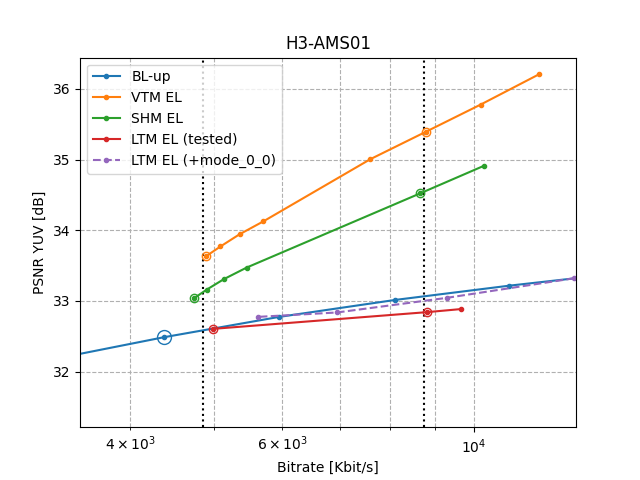
# Annex 3 – Objective metrics

This section reports PSNR for SDR and HDR HLG and wPSNR for HDR PQ. The value is computed as it suggested in HSTP-VID-WPOM [9].

In addition to settings which were tested, complementary objective results are provided with the same LTM and VTM configurations + mode\_0\_0 encoder setting used for LTM instead of default.

In the following plots, the circled points correspond to those that were used for visual assessment. The vertical dashed lines correspond to the 10% and 50% EL targets. BL-up corresponds to the upsampled LTM base layer.





# Annex 4 – Decoder runtime

This section reports information about the decoding time for enhancement layer of VTM 23.8 ML, SHM 12.4 and LTM 7.0.

The timing data is described below.

* **User Time (in seconds)**: Time spent by the CPU to execute the process

The base and enhancement layer processing times were measured using the time utility on a Linux system. Both processes were executed sequentially, ensuring that only one process was active at any given time.

The machine specifications are as follows:

* **CPU**: AMD Ryzen 9 7950X3D 16-Core Processor,
* **Thread(s) per core**: 2
* **Total RAM**: 64 GiB
* **OS**: Ubuntu 22.04.5 LTS
* **Kernel**: (GNU/Linux 5.15.0-134-generic x86\_64)

**Table A2. Enhancement layer decoding time VTM 23.8, SHM12.4 and LTM 7.0**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sequence name** | **Class** | **Enh %** | **EnhT[s] VTM 23.8 ML** | **EnhT[s] SHM 12.4 ML** | **EnhT[s] LTM7.0** |
| RiverPlate\_1 | PQ | E50 | 67.87 | 77.79 | 41.35 |
| E10 | 74.62 | 70.50 | 42.10 |
| GreenMountain2 | PQ | E50 | 28.32 | 32.35 | 12.14 |
| E10 | 37.39 | 29.05 | 12.07 |
| WalkInThePark | PQ | E50 | 72.08 | 74.52 | 30.52 |
| E10 | 66.95 | 69.82 | 30.97 |
| KitchenDressing | PQ | E50 | 48.94 | 67.71 | 28.16 |
| E10 | 78.00 | 70.19 | 29.17 |
| TiergartenParkway | PQ | E50 | 82.07 | 83.48 | 30.42 |
| E10 | 69.89 | 70.06 | 30.58 |
| FootballLargeAdvert | HLG | E50 | 60.38 | 69.70 | 29.77 |
| E10 | 60.86 | 69.24 | 30.11 |
| H3\_AMS01 | HLG | E50 | 76.74 | 74.27 | 31.74 |
| E10 | 70.39 | 92.15 | 31.64 |
| H3\_AMS02 | HLG | E50 | 71.74 | 72.25 | 31.29 |
| E10 | 69.69 | 73.64 | 31.27 |
| H3\_AMS05 | HLG | E50 | 95.86 | 72.88 | 33.12 |
| E10 | 81.62 | 77.28 | 32.84 |
| WomenFootball | HLG | E50 | 59.02 | 63.92 | 25.69 |
| E10 | 59.32 | 55.04 | 25.96 |
| WaterFront | SDR | E50 | 30.08 | 28.98 | 17.39 |
| E10 | 27.86 | 28.80 | 17.19 |
| BodeMuseum | SDR | E50 | 59.03 | 72.28 | 31.47 |
| E10 | 58.36 | 74.07 | 29.24 |
| Metro | SDR | E50 | 66.58 | 70.54 | 30.38 |
| E10 | 82.77 | 69.83 | 30.75 |
| OberbaumSpree | SDR | E50 | 64.20 | 75.79 | 29.95 |
| E10 | 70.96 | 73.99 | 29.62 |
| SubWayTree | SDR | E50 | 72.77 | 71.41 | 30.35 |
| E10 | 77.78 | 70.33 | 30.91 |
| **Average** |  | **E50** | **63.71** | **67.19** | **28.92** |
| **E10** | **65.76** | **66.27** | **28.96** |

1. An editorial corrigendum has been published as AG5N167. [↑](#footnote-ref-1)
2. A step size of 62 was used for this purpose instead of the coarser value of 250 mentioned in AG5N160. [↑](#footnote-ref-2)