TV 3.0 Project - Phase 3 - Real-Time Video Coding Subjective Quality Assessment

29 February 2024

Brazilian Digital Terrestrial Television System Forum

Brazilian Ministry of Communications

Brazilian National Education and Research Network (RNP, *Rede Nacional de Ensino e Pesquisa*)

University of Brasília (UnB, *Universidade de Brasília*)

Table of Contents

1 Introduction

The SBTVD Forum was created by the Brazilian Presidential Decree # 5 820 / 2006, to advise the Brazilian Government regarding policies and technical issues related to the approval of technical innovations, specifications, development, and implementation of the Brazilian Digital Terrestrial Television System (SBTVD). The SBTVD Forum is composed of representatives of the broadcasting, academia, transmission, reception, and software industry sectors, and has the participation of Brazilian Government representatives as non-voting members.

Free-to-air terrestrial television is the main audiovisual distribution platform in Brazil, covering almost all Brazilian households and used in more than 70% of them. It secures to most of the Brazilian population a free-of-charge, universal and democratic access to information and entertainment, made by Brazilians for Brazilians. It is, therefore, an important social cohesion, and national/cultural identity factor.

For its first generation Digital Terrestrial Television system, after thorough testing and careful studies, the Brazilian Government adopted in June 2006 the ISDB-T standard, incorporating technological innovations that were deemed relevant, such as MPEG-4 AVC (H.264) video coding, MPEG-4 AAC audio coding, an appropriate closed caption character set for the Brazilian Portuguese, and a new middleware for interactive applications (Ginga).

The SBTVD Forum developed the first SBTVD standards, which were published in 2007, allowing the official opening of transmissions in that same year. Since then, the standards have been continuously revised and updated by the Forum. The technological innovations proposed by Brazil were incorporated into the International ISDB-T standard, which is currently adopted by 20 countries.

In 2016, Brazil started a safe and gradual analog TV switch-off process that was designed to ensure that no one would be deprived of the terrestrial free-to-air TV. The process was divided into three stages.

In the first stage (2016 to 2019), the analog television switch-off was performed in all the state capitals, metropolitan areas, and other areas where it was required to release the 700 MHz band. During the first stage, 1 366 cities in 47 different clusters were impacted, accounting for nearly 128 million people (62% of the population). More than 12 million Digital TV reception kits were distributed to low-income families. The analog switch-off had no significant impact on the free-to-air terrestrial TV audience.

In the second stage (2019 to 2023), the analog television switch-off was performed in 2 939 municipalities that, on 01 September 2020, had no digital terrestrial television stations in operation. In this stage, 1 565 municipalities were selected for a government-sponsored DTTB expansion program called "Digitaliza Brasil", which provided shared DTTB transmission infrastructure in each municipality and distributed additional 1.5 million Digital TV reception kits for low-income families.

In the third stage (2024 to 2025), the analog television switch-off will be performed in the remaining 1 265 municipalities up to June 2025. After the implementation of Digital Terrestrial Television Brazil adopted an industrial policy that determined that all flat-panel TVs manufactured from 2012 must have an integrated Digital TV receiver and from 2013 no more CRT TVs were manufactured. Therefore, it is anticipated, based on the expected product lifetimes, that after 2023 Brazil will have nearly all its TV sets already equipped with an integrated Digital TV receiver, thus facilitating the analog television switch-off without additional Digital TV reception kits distribution.

As the Brazilian digital television switch-over began, the SBTVD Forum started considering the next steps for the evolution of Brazilian Television. The analog TV (that we conventionally call "TV 1.0"), which started in Brazil in 1950, was black and white with monophonic sound. Then, some backward-compatible improvements (that we conventionally call "TV 1.5"), such as color (in the 1970s), stereo sound and closed caption (in the 1980s) were added to it. In 2007, the first generation of Digital Terrestrial Television (that we conventionally call "TV 2.0") was introduced in Brazil, bringing high-definition video, surround sound, mobile reception, and interactivity. Since then, the technological landscape has changed a lot. The rhythm of development and introduction of innovations is increasingly accelerated. These innovations create new consumption habits and increase the expectations of technological service users regarding the quality and convenience of these services. Since the introduction of SBTVD, new immersive audio and video formats have emerged, and are already present in the new TV sets available in the market. The TV sets currently available have resolution and contrast greater than those supported in the first-generation SBTVD standard. That is the opposite of the market situation when Digital TV was launched in Brazil, as the HDTV sets were not as common at the time. The availability and the speed of Internet access in Brazil, especially in metropolitan areas, increased significantly, enabling the consumption of on-demand audiovisual content. This connectivity is already in use by TV sets (Smart TVs) and by broadcasters' Over-The-Top (OTT) offers. However, in the first generation SBTVD standard, there was not an integration between the broadcasting service and the Internet content offer. Furthermore, new techniques for signal coding, transport, and modulation were also developed, allowing greater efficiency in audiovisual transmission. Many Digital Terrestrial Television systems have also been evolving, including in this evolution not only enhancements in quality and efficiency but also new convergent services between broadcast and broadband. Based on this technological landscape, the SBTVD Forum recognized the necessity to evolve the SBTVD. It also acknowledged that changing the physical layer, the transport layer, and/or audiovisual coding would not be backward-compatible. Nevertheless, the transition to a new generation of Digital Terrestrial Television is a long process, based on the investments required for both broadcasters and consumers and the expected life span of TV transmitters and receivers. It was, therefore, deemed necessary to increase the lifespan of the existing Digital Terrestrial Television system as much as possible through a backward-compatible evolution (a project we called "TV 2.5") and to start the development of the next generation Digital Terrestrial Television system (the project we called "TV 3.0").

The "TV 2.5" project comprised two aspects: broadcast-broadband integration and audiovisual quality. The first aspect involved the development of a new receiver profile for the middleware Ginga (receiver profile D, a.k.a. "DTV Play"), addressing use cases such as on-demand video, synchronized companion device, audiovisual enhancement over the Internet, and targeted content. The second aspect was addressed through the introduction of three new optional immersive audio codecs (MPEG-H Audio, E-AC-3 JOC, and AC-4) while retaining MPEG-4 AAC main audio for backward compatibility, and through the introduction of two new optional HDR video formats (SL-HDR1 dynamic metadata and HLG "preferred transfer characteristics" signaling) while keeping MPEG-4 AVC (H.264) / 8-bit / BT.709 / 1 080i for backward-compatibility. The revision of the SBTVD standards containing both "TV 2.5" aspects has already been published (available at https://forumsbtvd.org.br/legislacao-e-normas-tecnicas/normas-tecnicas-da-tv-digital/english/).

For the "TV 3.0" project, the SBTVD Forum, after agreeing on its requirements (use cases and corresponding technical specifications), decided to release a Call for Proposals (available at https://forumsbtvd.org.br/wp-content/uploads/2020/07/SBTVDTV-3-0-CfP.pdf) for any interested organization to submit its proposed candidate technologies for any of the system components or sub-components. The new system is expected to start operating within the next few years, but based on the Brazilian experience with the transition from analog to digital television, the complete transition from the current SBTVD to TV 3.0 is expected to last at least 15+ years.

As described in the aforementioned Call for Proposals document, the response to the Call for Proposals was divided into two phases.

Phase 1 responses comprised the identification of each proposed candidate technology and appropriate contact persons and filling the compliance form of the components or sub-components corresponding to the proposed candidate technology.

Phase 2 responses included providing the full specification of the proposed candidate technology, adhering to the SBTVD Forum Intellectual Property Rights Policy, and the additional requirements considering general information and resources needed for evaluating and comparing the proposed candidate technologies. The "TV 3.0 CfP Phase 2 / Testing and Evaluation" document (available at https://forumsbtvd.org.br/wp-content/uploads/2021/03/SBTVD-TV_3_0-P2_TE_2021-03-15.pdf) provides further information on Phase 2, along with the test procedures for evaluating and comparing the proposals of candidate technologies.

The Call for Proposals was open from 17 July 2020 to 30 November 2020. It received in total, considering its 6 system components (Over-the-air Physical Layer, Transport Layer, Video Coding, Audio Coding, Captions, and Application Coding), 36 responses from 21 different organizations worldwide. Some similar proposals were merged for the sake of Phase 2 testing and evaluation, resulting in 30 candidate technologies.

Phase 2 tests were funded by the Brazilian Ministry of Communications through the Brazilian National Council for Scientific and Technological Development (CNPq, *Conselho Nacional de Desenvolvimento Científico e Tecnológico*). The Video Coding candidate technologies were tested by the University of Brasilia (UnB) from 05 July 2021 to 03 December 2021. Phase 2 Video Coding testing and evaluation report is available at https://forumsbtvd.org.br/wp-content/uploads/2021/12/SBTVD-TV_3_0-VC-Report.pdf.

At the end of Phase 2, the SBTVD Forum concluded that it was necessary to carry out a real-time video coding subjective quality assessment using the video coding technologies selected for the TV 3.0 Project (VVC and LCEVC) to determine the required bit rate for the Over-the-air Physical Layer to enable delivering audiovisual quality superior to that of the first-generation Brazilian DTTB system.

In April 2023, the Brazilian Presidential Decree # 11 484, which provides the guidelines for the evolution of the Brazilian Digital Terrestrial Television System and for ensuring the availability of radio spectrum for its deployment, was published (available at http://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/decreto/D11484.htm, in Portuguese only).

It establishes that the next-generation Digital Terrestrial Television Broadcasting (DTTB) system in Brazil, called TV 3.0, shall have the following characteristics:

I. audiovisual quality superior to that of the first-generation Brazilian DTTB system;
II. fixed reception, with external and internal antenna, and mobile reception;
III. integration between contents transmitted by the broadcasting service and over the internet;
IV. app-based user interface;
V. content segmentation according to viewers' geographic location;
VI. customization of content according to viewers' preferences;
VII. optimized use of the radio frequency spectrum destined for terrestrial television broadcasting; and
VIII. new forms of access to cultural, educational, artistic, and informative content.

The characteristic "I" has a direct relationship with the Video Coding and is in line with the requirements defined in the Call for Proposals. To ensure the fulfillment of this requirement, a subjective assessment was set-up to compare the maximum quality achievable with real-time video coding for the first-generation Brazilian DTTB system (H.264 High profile @ L4, 1 920 × 1 080/59.94/I, 4:2:0, 8-bit, BT.709, SL-HDR1, at 14 Mbps) and the quality resulting from real-time video coding using the technologies selected for TV 3.0 at different configurations and bitrates. The results of this test were meant to be considered for the determination of the minimum bitrate of the main over-the-air physical layer configuration to be tested in the field.

In Phase 2, 11 use cases for video coding were evaluated, ranging from 720p resolution to 8K resolution, and varying characteristics such as dynamic range, frame rate, and scalability, among others. Due to the scale and the large number of use cases and test candidates, in Phase 2, all tests regarding video content quality were performed using objective analysis. In the video coding aspect, Phase 2 tests enabled the selection of a few technologies and use cases so that a subjective analysis would be feasible. This subjective analysis is the topic of the Phase 3 tests, which only involve subjective experiments of a few codecs and use cases.

The Phase 3 Real-Time Video Coding Subjective Quality Assessment was carried out by the University of Brasília from June 2023 to January 2024, under the coordination of the SBTVD Forum and funded by the Brazilian Ministry of Communications through the Brazilian National Education and Research Network (RNP, *Rede Nacional de Ensino e Pesquisa*).

Further information regarding the TV 3.0 Project can be obtained at: https://forumsbtvd.org.br/tv3_0/.

This document contains the results of the Phase 3 Real-Time Video Coding Subjective Quality Assessment that were considered for the determination of the minimum bitrate of the main over-the-air physical layer configuration to be tested in the field.

It is important to highlight that, in this process, the SBTVD Forum has a propositional role, with the Brazilian Government alone being responsible for making any decisions about the standards applicable to the broadcasting service in Brazil.

2 Research Team

This report was developed by the following team at the University of Brasília:

Eduardo Peixoto Fernandes da Silva, Universidade de Brasília - http://lattes.cnpq.br/7720922538480924

Pedro Garcia Freitas, Universidade de Brasília - http://lattes.cnpq.br/0669244312752630

Mylene Christine Queiroz Farias, Universidade de Brasília - http://lattes.cnpq.br/4465619366143200

José Edil Guimarães de Medeiros, Universidade de Brasília - http://lattes.cnpq.br/2307490672260330

Fernando Cardoso Guimarães, Universidade de Brasília - https://lattes.cnpq.br/8587888778545035

Felipe Sidrone da Silva Dionísio, Universidade de Brasília - http://lattes.cnpq.br/4451878253778940

Gabriel Correia Lima da Cunha e Menezes, Universidade de Brasília - http://lattes.cnpq.br/0543326076269968

Victor Orfeu Melo, Universidade de Brasília - http://lattes.cnpq.br/2846252961723018

We also had the collaboration of SBTVD Forum members, in particular:

Luiz Fausto – Chair, Technical Module, SBTVD Forum

Carlos Cosme – Chair, Audio & Video Coding Workgroup, SBTVD Forum

2.1 Partners

We want to thank the companies that provided laboratory equipment for the tests:

- Globo
- MainConcept
- V-Nova
- Philips
- Ateme

3 Glossary

1 080i	Interlaced video with a resolution of 1 920 x 1 080 pixels
1 080p	Progressive video with a resolution of 1 920 x 1 080 pixels
2 160p	Progressive video with a resolution of 3 840 x 2 160 pixels
4K	Progressive video with a resolution of 3 840 x 2 160 pixels
720p	Picture resolution of 1 280 x 720 pixels
AAC	Advanced Audio Coding

AC-4 Audio Compression 4

AVC Advanced Video Coding

BT.2020 Recommendation ITU-R BT.2020: Parameter values for ultra-high-definition television systems for production and international programme exchange

BT.709 Recommendation ITU-R BT.709: Parameter values for the HDTV standards for production and international programme exchange

CfP Call for Proposals

CNPq *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (Brazilian National Council for Scientific and Technological Development)

CRF Constant Rate Factor

CRT Cathode Ray Tube

DTV Digital Television

E-AC-3 JOC Enhanced Audio Compression 3 Joint Object Coding

fps frames per second

GOP Group of Pictures

H.264 Recommendation ITU-T H.264: Advanced video coding for generic audiovisual services

H.266 Recommendation ITU-T H.266: Versatile video coding

HDMI High-Definition Multimedia Interface

HDR High Dynamic Range

HDR10 High Dynamic Range 10-bit

HDTV High-Definition Television

ISDB-T Integrated Services Digital Broadcasting-Terrestrial

ITU-R	International Telecommunications Union - Radiocommunication sector
ITU-T	International Telecommunications Union - Telecommunication standardization sector
kbps	kilobits per second
LCEVC	Low Complexity Enhancement Video Coding
MaxCLL	Maximum Content Light Level
MaxFALL	Maximum Frame Average Light Level
Mbps	Megabits per second
MOS	Mean Opinion Score
MPEG	Motion Picture Experts Group
OLED	Organic Light-Emitting Diode
OTT	Over-The-Top
RNP	*Rede Nacional de Ensino e Pesquisa* (Brazilian National Education and Research Network)
SBTVD	*Sistema Brasileiro de Televisão Digital* (Brazilian Digital Television System)
SDI	Serial Digital Interface
SDR	Standard Dynamic Range
SI	Spatial Index
SL-HDR	Single Layer HDR
SMPTE	Society of Motion Picture and Television Engineers
TI	Temporal Index
TV	TeleVision

VQEG Video Quality Experts Group

VUT Video Under Test

VVC Versatile Video Coding

4 TV 3.0 Architecture

The TV 3.0 system components described in this document reflect the reference TV 3.0 architecture, as depicted in Figure 1.

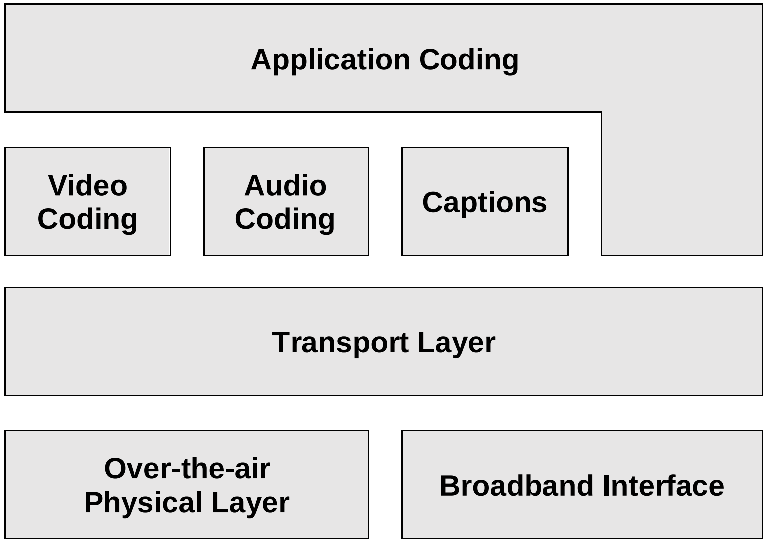


Figure 1 - TV 3.0 Architecture

For further information about TV 3.0 architecture, please refer to the TV 3.0 Call for Proposals (CfP) document (available at https://forumsbtvd.org.br/wp-content/uploads/2020/07/SBTVDTV-3-0-CfP.pdf).

5 TV 3.0 Phase 2 Video Coding Testing and Evaluation

Please refer to Section 4.4 of the "TV 3.0 CfP Phase 2 / Testing and Evaluation" document (available at https://forumsbtvd.org.br/wp-content/uploads/2021/03/SBTVD-TV_3_0-P2_TE_2021-03-15.pdf) for further information on the video coding testing and evaluation procedures performed in Phase 2.

Phase 2 Video Coding testing and evaluation report is available at https://forumsbtvd.org.br/wp-content/uploads/2021/12/SBTVD-TV_3_0-VC-Report.pdf.

6 TV 3.0 Phase 3 Real-Time Video Coding Subjective Quality Assessment Results

The scope of the current tests is to evaluate the overall subjective quality of each codec when working in a real-time encoding scenario. All encoders used, either for the VUTs or the reference videos, were used in a real-time configuration, and the input video was fed to the encoder using an SDI connection. Only the overall quality of the video streams is being evaluated. Participants are not asked to detail why they have rated a video in a certain way.

Section 6.1 introduces the experimental setup arranged for the subjective quality evaluation experiments and details the experimental protocol each subject is submitted during the experiments, Section 6.2 describes the analysis protocols executed to understand the data generated during the experiments, and Section 6.3 details the definition of the specific tests the team performed, their results and analysis.

6.1 Experimental Setup

This section details the contents used, the laboratory conditions, and the interface of the experiments.

6.1.1 Experimental Stimuli

The video dataset used for the experiments was prepared by the SBTVD Forum specifically for the tests. It comprises five different video streams, mostly camera-generated content with some computer-generated graphical overlays and high contrast colorful scenes, each one with a few scene cuts. Some content contains smooth gradients between similar colors to assist in detecting possible banding artifacts. Some of the content have scenes commonly transmitted through Brazilian television, such as football matches and carnival scenes. The scenes were chosen due to their diverse range of spatial and temporal information. The details of the video streams are shown in Table 1.

The spatial and temporal information for the scenes is shown in Figure 2, while Figure 3 shows the file size for each content when encoding with libx265 at CRF = 28. This is used as an indicator of the complexity of encoding each content. It can be seen that content globo01 needs almost four times the bitrate compared to content philips03 when using the same Quantization Parameter. A short description and a sample frame for each content are shown in Table 2.

Table 1 - Details of the contents used for the tests

Resolution (horizontal x vertical)	3 840 x 2 160
Aspect ratio	16:9 (square pixels)
Sampling ratio	YC_BC_R 4:2:2
Bit depth	10 bits
Frame rate	59.94 fps
Scan	Progressive
Transfer characteristics	Perceptual Quantization (max 1 000 cd/m²)
HDR Metadata	HDR10 (SMPTE ST 2086, MaxFALL and MaxCLL)
Color Gamut	Rec. ITU-R BT.2020
Duration	30 s
File format	Uncompressed MOV file

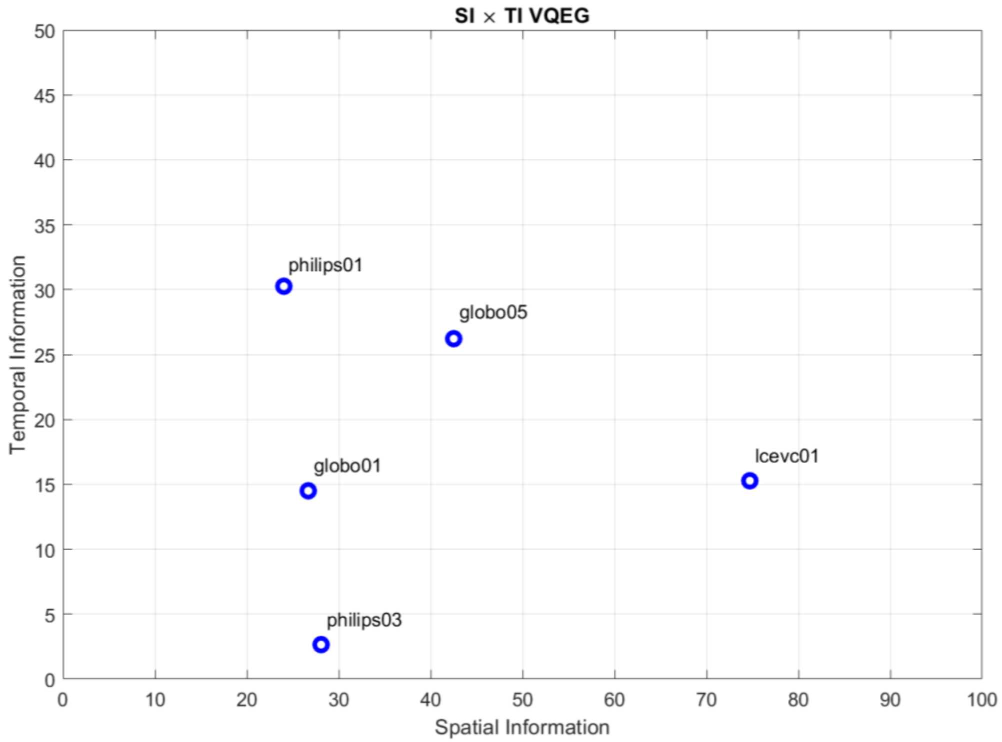


Figure 2 - Spatial and Temporal Information for the test contents used

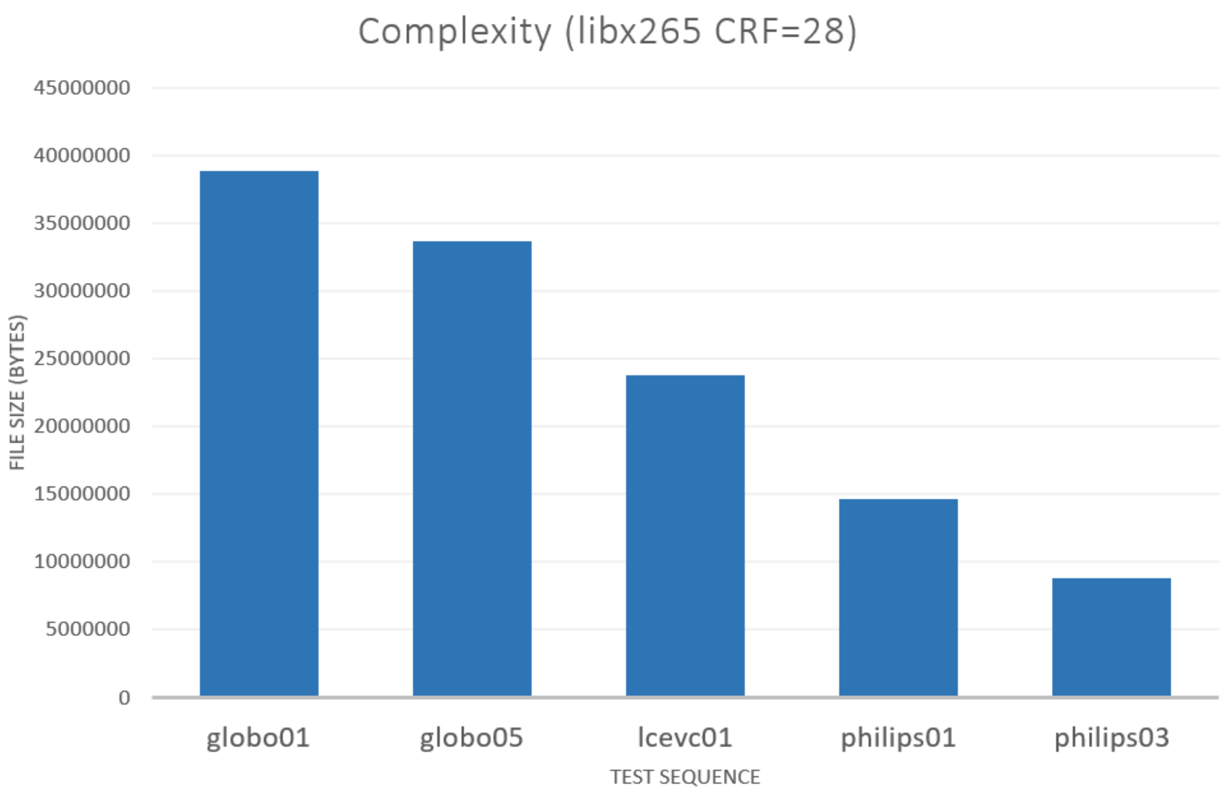
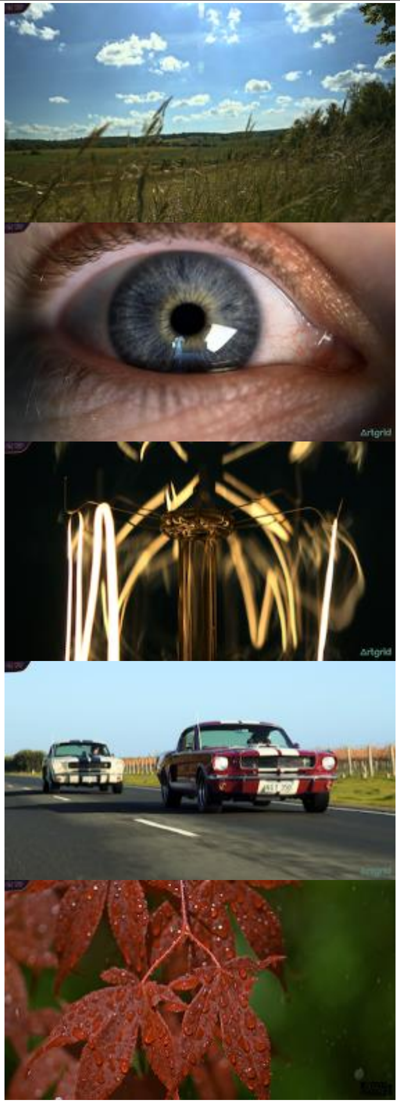


Figure 3 - Content file size when encoded with libx265 at CRF=28

Table 2 - Short description of the contents used for the tests

globo01		Excerpts from the 2022 Rio de Janeiro samba school parade. This is a high-movement colorful scene. There is a lot of content with high spatial information, including many human faces close to the camera.
globo05		Excerpts from the 2019 Conmebol Copa America (South America Football Championship) final match. There is a lot of high-speed movement and a large number of people moving around in the background.
lcevc01		A collection of short clips of different landscapes with direct sunlight and graphics overlay, including a few sunrise and sunset scenes, and a sunflower field with bees. There are also some Formula 1 results graphics overlay in the left part of the screen which contributes to a lot of high-frequency content.

philips01		A collection of short clips, including a wheat field, a close-up of a human eye, a close-up of an incandescent lamp, fast-moving vintage cars on a highway, and some bright red leaves with raindrops.

philips03		A collection of short clips, including images of a canal in the Netherlands, a close-up of a teenage girl's face, and a pier with moving water.

6.1.2 Experimental Conditions

The experiments were performed on a 465cm x 490cm laboratory located on the first floor of the SG-11 building at the University of Brasília Darcy Ribeiro Campus. The lab was available exclusively for the TV 3.0 Phase 3 experiments.

Figure 4 shows the relative position of the subject in relation to the screen. In this figure, H is the vertical size of the screen, which is 81 cm. Therefore, the viewing distance from the screen was around 131 cm.

The equipment setup included two sets of the following equipment:

1. Mac Studio (2022 model) desktop equipped with an Apple M1 Ultra and 128GB of RAM running MacOS Ventura 13.2.1 operating system and a Blackmagic Design Media Express 3.8.1 software;
2. Blackmagic Design UltraStudio 4k Mini; and
3. LG 65" OLED TV set model OLED65C2PSA.CWZQLJZ (firmware version 03.33.85) with a screen area of approximately 81cm x 142cm.

The LG TVs are consumer-grade products; thus, a limited set of configurations was available. Table 3 shows the configurations used for both TV sets. Table 4 shows a summary of the tests that were performed.

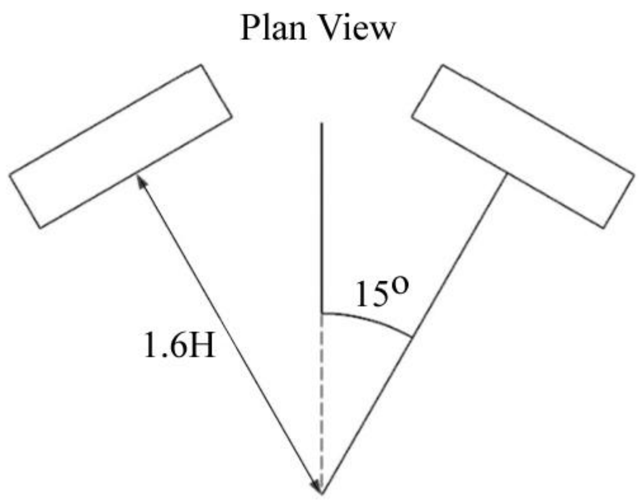


Figure 4 - Laboratory set-up for the subjective tests

Table 3 - TV sets configuration

Parameter		Setting
Brightness	OLED Pixel Brightness	80
	Adjust Contrast	95
	Black Level	50
	Auto Dynamic Contrast	Medium
	Dynamic Tone Mapping	On
	Peak Brightness	High
	Gamma	2.2
	Video Range	Auto
	Motion Eye Care	Off
Color	Color Depth	55
	Tint	0
	Color Gamut	Auto Detect
Color/Fine Tune	Color Adjustment	Off
Balance	Color Temperature	0
	Method	2 Points
	Point	High
	Red	0
	Green	0
	Blue	0
Clarity	Adjust Sharpness	25
	Super Resolution	Off
	Noise Reduction	Off
	MPEG Noise Reduction	Off
	Smooth Gradation	Off
	Real Cinema	Off
	TruMotion	Off

Table 4 - Configurations used for the tests

Test	Reference Video	Video Under Test (VUT)	Bitrates
VUT 1.1	1 080i H.264/AVC + SL-HDR1	720p H.266/VVC HDR10	[1.50 2.24 3.35 5.02 7.50] Mbps
VUT 1.2	1 080i H.264/AVC + SL-HDR1	1 080p H.266/VVC HDR10	[2.00 2.99 4.47 6.69 10.00] Mbps
VUT 1.3	1 080p H.266/VVC HDR10	2 160p H.266/VVC HDR10	[4.00 5.98 8.94 13.37 20.00] Mbps
VUT 2.3[1]	1 080i H.264/AVC + SL-HDR1	1 080p H.266/VVC HDR10 (Base Layer)	[2.40 3.60 5.40 8.00 12.00] Mbps
VUT 2.4[1]	1 080p H.266/VVC HDR10	2 160p [LCEVC (Enhancement Layer) + 1 080p H.266/VVC HDR10 (Base Layer)]	[2.67 3.99 5.96 8.92 13.33] Mbps

It should be noted that the TV sets used in the test have a progressive 3 840 x 2 160 display. The TV set upscales any input signal with a lower resolution to its display native resolution. Likewise, the TV set deinterlaces any non-progressive input signal to display it.

6.1.3 Experimental Protocol

We have performed 5 different independent tests, each one with 30 subjects, properly screened for normal visual acuity and normal color vision. For each test, we have used 5 different contents, provided by the SBTVD Forum and detailed in Section 6.1.1. For each test, 5 different bitrates were used for the video under test (VUT), while the reference video was kept constant for all bitrates. The goal of each test was to find the bitrate at which a particular quality was achieved for all contents.

For all tests, the reference was always displayed on the left screen, and this was known by the participant. For each VUT, we have used three different pseudo-random sequences to display the 25 videos of each session.

The experimental session comprises two distinct segments: training and subjective evaluation. In the initial training phase, the conductor outlines the experiment's objectives and procedures, gathers demographic information from participants, and conducts a training session to get them acquainted with the experimental tasks. Before beginning the experiment, subjects are asked to read a 4-page document and fill in some information. The first two pages of the document contain an explanation of the experimental protocol and the risks associated with it. Subsequently, participants are requested to sign a consent form and provide details about their age, gender, technical expertise, and familiarity with visual impairments.

In the training session, each subject is shown 3 video samples with varying quality levels that correspond to different encoding parameters. The goal of the training session is to provide the subjects an opportunity to familiarize themselves with the experimental task and to understand the quality rating

[1] Note that the same encodings were used for VUTs 2.3 and 2.4, and only the decoding was carried out differently (decoding only the VVC base layer in VUT 2.3 and decoding the VVC base layer with the LCEVC enhancement layer in VUT 2.4). In the first case, the bitrates shown in Table 4 reflect the average rates assigned only to the VVC base layer, while in the second case, the bitrates reflect the total average rates used by the combination of the VVC base layer (used in VUT 2.3) and the LCEVC enhancement layer.

scale. After the training, the experimenter may answer questions from the subject. Then, the experimenter takes note of the VUT information as well as the pseudo-random sequence number (used to analyze the data) and starts the subjective evaluation phase.

During the subjective evaluation phase, participants assess a series of 25 test video sequences, each lasting around 30 seconds. To record their quality scores, participants utilize paper forms and pens. The evaluation form, based on the stimulus-comparison method with adjectival categorical judgment as described in Recommendation ITU-R BT.500-15, Section A4-4.1 from Annex 4 to Part 2, contains a 7-level quality scale, as depicted in Table 5, for each test sequence to be evaluated. Prior to the presentation of each video, a 3-second preparation countdown is displayed on a gray screen. Following each video, a gray screen prompts participants to complete the evaluation form. The pace of the video exhibition is manually controlled by the experimenter, who adjusts the experiment's speed based on the capacities of the participants.

Table 5 - Subjective evaluation scale

Textual Scale (in Portuguese)	Textual Scale	Numeric Scale
Muito Pior	Much Worse	-3
Pior	Worse	-2
Pouco Pior	Slightly Worse	-1
Igual	Equal	0
Pouco Melhor	Slightly Better	1
Melhor	Better	2
Muito Melhor	Much Better	3

6.2 Data Analysis Protocol

This section details how the experimental data gathered during the experiments was processed. In the first part, the subjective scores for each VUT were processed to generate an output target bitrate. This target bitrate is the minimum bitrate for which a target quality score is achieved for all contents. Then, the videos are encoded at this target bitrate and the output bitstream is analyzed in order to assess the minimum and maximum bitrates to be used by the codec configuration. To achieve these target bitrates, some processing steps are performed. These steps include removing outliers and computing the mean opinion scores (MOS), performing an experimental model regression, and analyzing the video stream data.

6.2.1 Outlier Removal and Mean Opinion Score Determination

In conducting the statistical analysis of the experimental data for this study, we used the following protocol to ensure the reliability and validity of the obtained results. The protocol involves several key steps, starting with the identification and removal of outliers. Preliminary data inspection used statistical measures, such as the computation of Z-scores, to flag potential outliers. In this work, outliers are defined as data points lying outside predetermined thresholds and are an indicator of measurement errors. The outlier detection step was based on Recommendation ITU-R BT.500-15. Following the recommendation, the threshold used was 2 standard deviations. The subsequent removal of these outliers was performed systematically to prevent their undue influence on the subsequent statistical analysis phases. Following outlier removal, a critical step of the statistical analysis is the transformation of the collected experimental scores. Then, the results were further analyzed by finding the mean opinion scores (MOS) and confidence intervals (CI), as described in Recommendation ITU-T P.1401.

6.2.2 Experimental Model Regression

After removing the outliers and determining the MOS for each video, a regression to a logistic function is applied to enhance the interpretability and reliability of the scores. This transformation adjusts raw scores, improving the model fit to the underlying experimental data. The ultimate goal of this transformation is to provide an accurate representation of the relationship between objective and subjective variables, capturing non-linear patterns. In this context, we fitted a logistic function to the MOS versus bitrate relationship. The logistic function is a monotonic function given by the following equation:

$$MOS_p = \frac{b_1}{1 + exp(-b_2 \times (bitrate - b_3))},$$

where MOS_p is the predicted MOS as a function of the *bitrate* and b_1, b_2, and b_3 are parameters obtained via least squares fit[2]. In summary, the above predictive model was used to evaluate various metrics on the sets of subjectively measured MOS and the corresponding bitrate. Using this model, we predicted the bitrates where a given encoded video is better than a reference video, comparing the studied codecs.

6.2.3 Video Stream Analysis Protocol

Once a target bitrate is found for each VUT, each content was encoded at the defined target bitrate using the same codec and parameters as that VUT. We analyze the bitstream to find the minimum and maximum bitrates of each GOP. This is not the ideal methodology to analyze the video stream bitrate variation over time and should be considered only as an indication of this variation.

6.3 VUTs

The results and analysis for each VUT are detailed in this section.

[2] https://www.itu.int/md/T01-SG09-C-0060

6.3.1 VUT 1.1 - VVC 720p

6.3.1.1 VUT 1.1 Definition

The goal of this VUT was to test the VVC encoder working at 720p resolution. The reference video was encoded with H.264 + SL-HDR, which corresponds to the best quality possible under the current TV 2.5 specifications. The complete details are found in Table 6. Since the VUT was encoded with a lower resolution than the reference video, the quality target considered was “slightly worse” (-1).

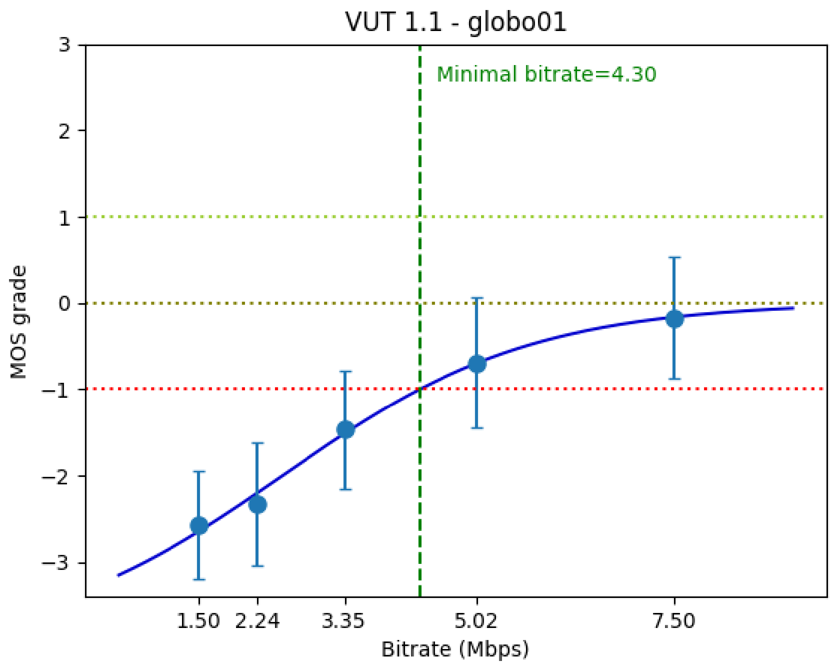
Table 6 - VUT 1.1 encoding details

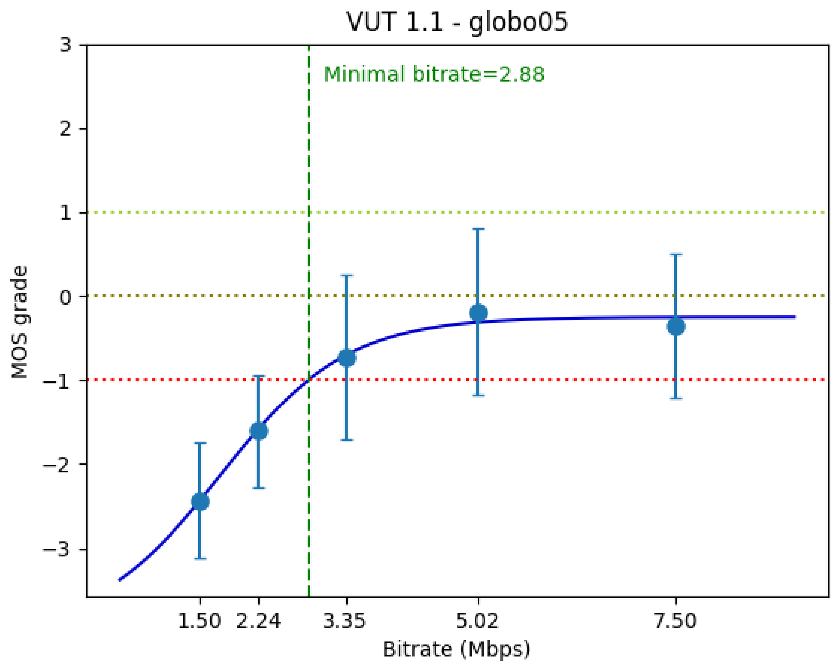
	Reference Video	VUT 1.1
Content label	1 080i	720p
Resolution	1 920 x 1 080	1 280 x 720
Frame rate	29.97 fps	59.94 fps
Scan	Interlaced	Progressive
Bit depth	8 bits	10 bits
Color gamut	BT.709	BT.2020
HDR Mode	SL-HDR1	HDR10
Codec	H.264/AVC	H.266/VVC
GOP size	60 frames (2 seconds)	120 frames (2 seconds)
Encoder	MainConcept Live Encoder v0.0.0.961	Ateme TitanLive Innovation v 4.1.31.911
Encoder type	Real-time	Real-time
Bitrate	14 Mbps	[1.50 2.24 3.35 5.02 7.50] Mbps

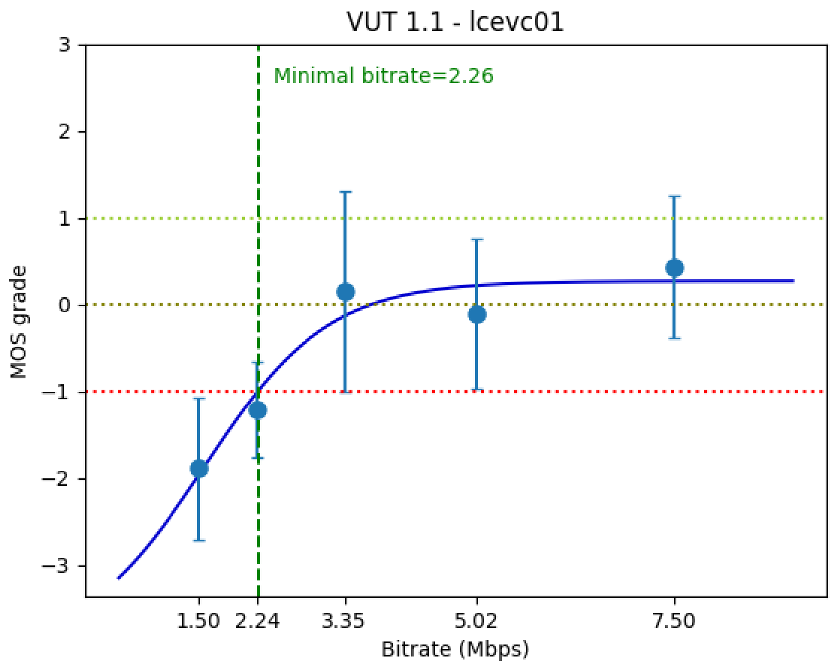
6.3.1.2 VUT 1.1 Experimental Findings

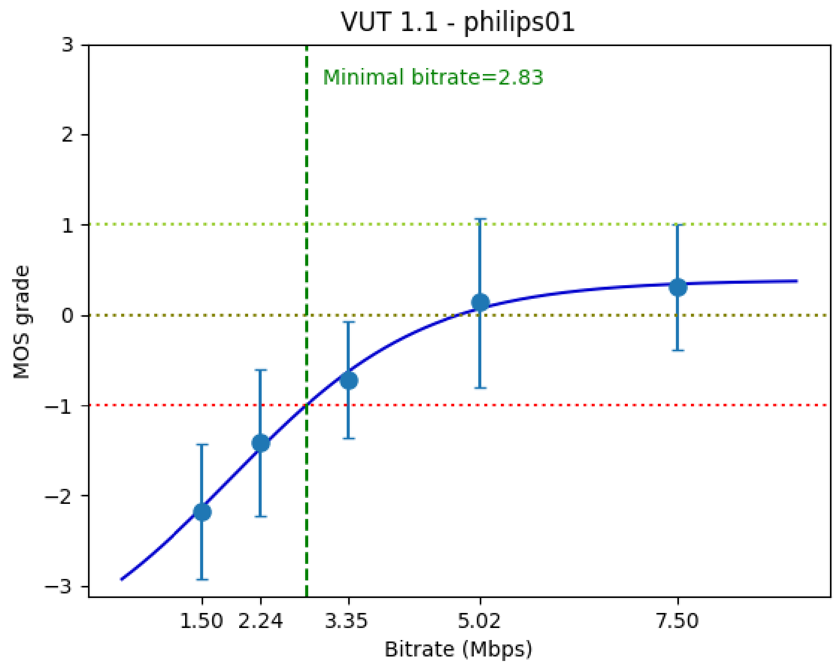
Threshold σ=-1

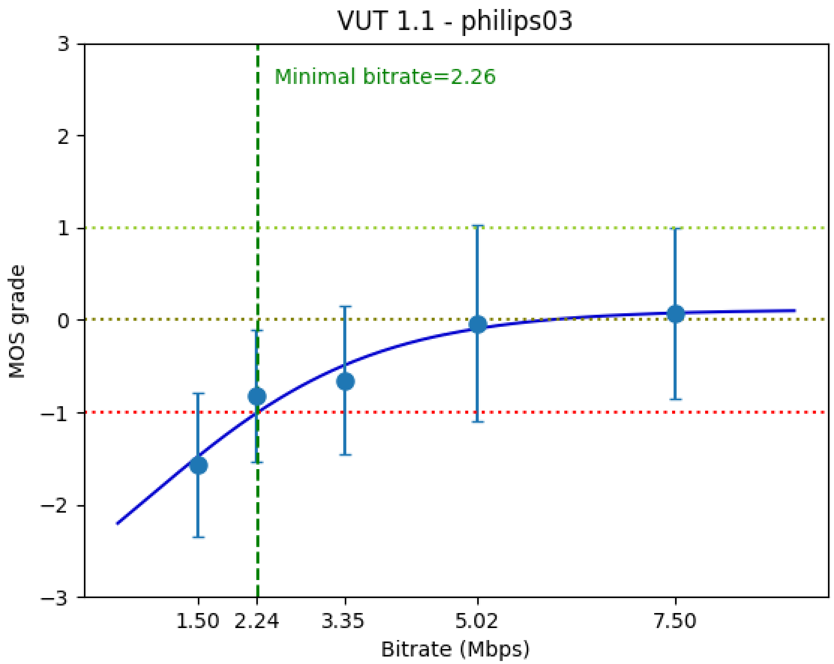
Table 7 - VUT 1.1 results targeting a MOS grade of -1 (slightly worse)





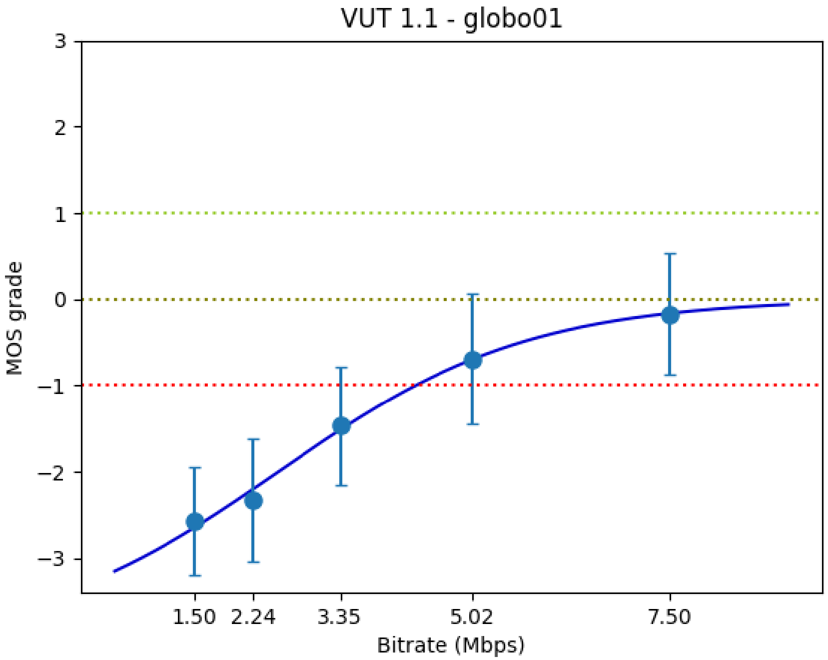
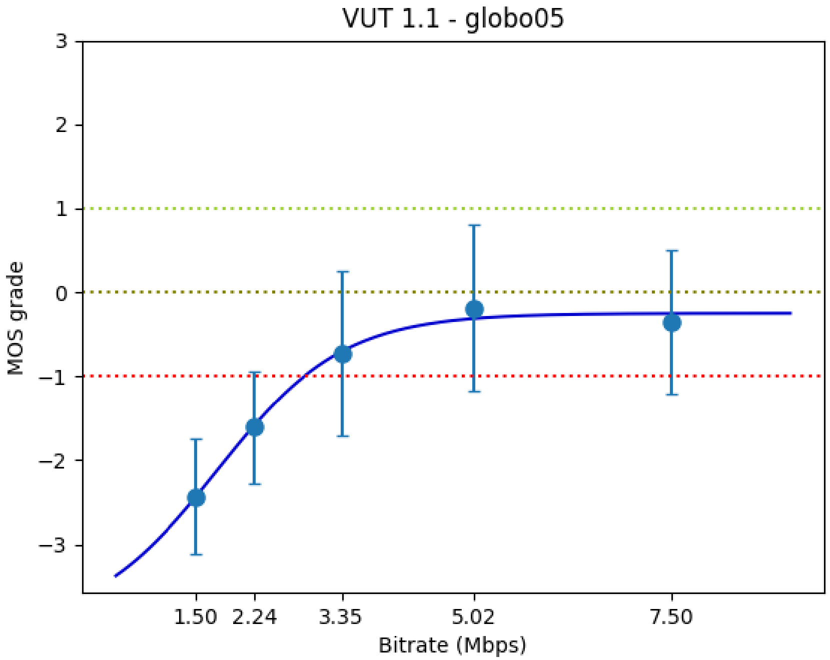

VUT 1.1 - lcevc01
Minimal bitrate=2.26
MOS grade
Bitrate (Mbps)
1.50 2.24 3.35 5.02 7.50
3 2 1 0 −1 −2 −3

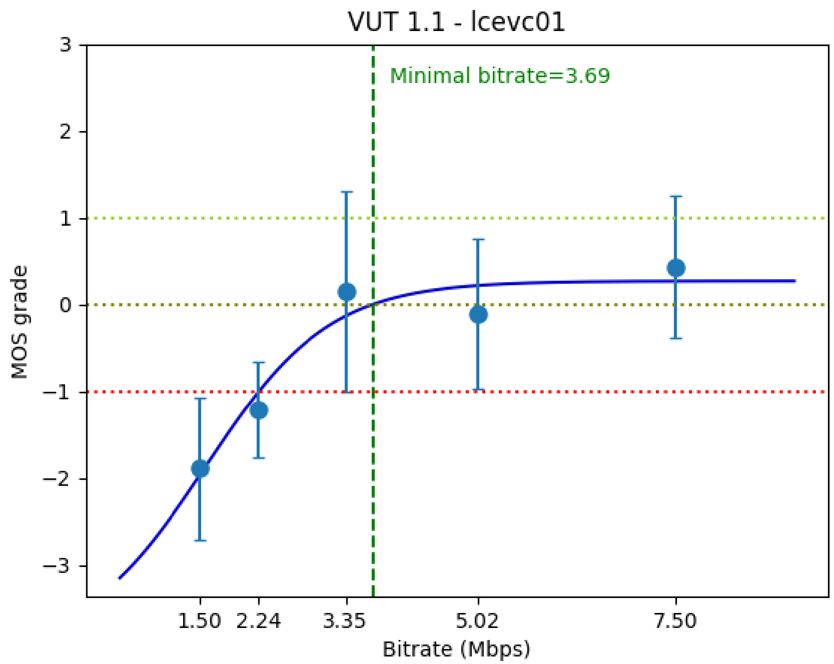

VUT 1.1 - philips01
Minimal bitrate=2.83
MOS grade
Bitrate (Mbps)
1.50 2.24 3.35 5.02 7.50
3 2 1 0 −1 −2 −3

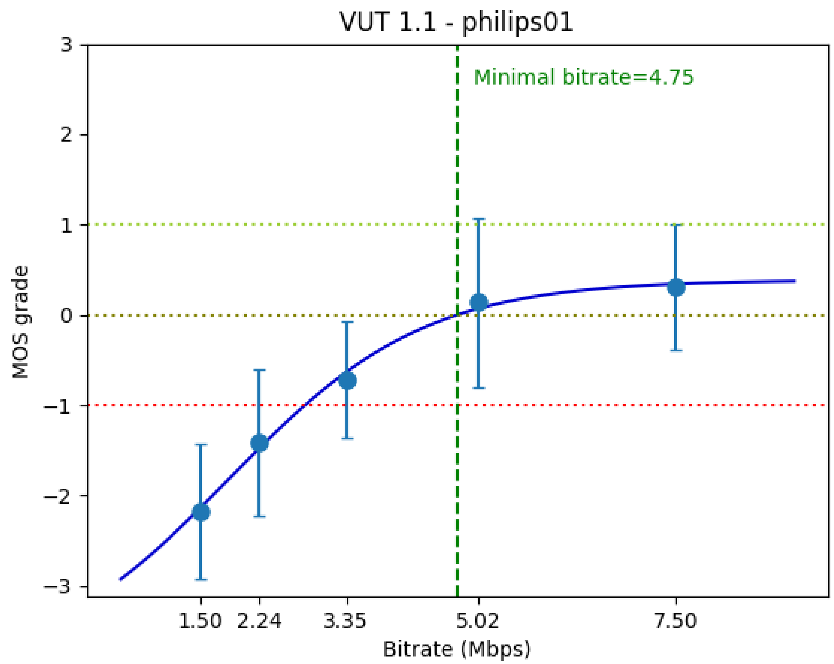

VUT 1.1 - philips03
Minimal bitrate=2.26
MOS grade
3
2
1
0
−1
−2
−3
1.50
2.24
3.35
5.02
7.50
Bitrate (Mbps)

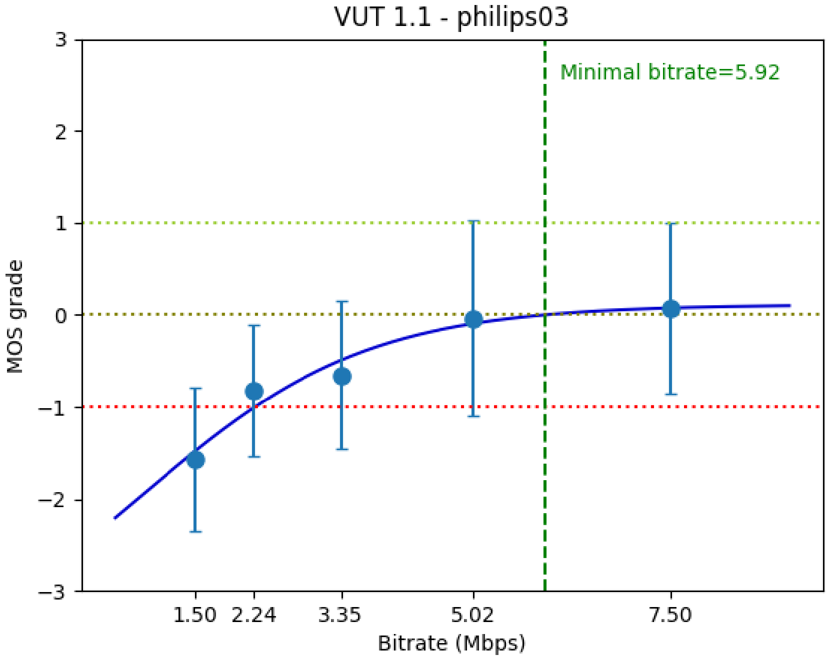
<u>Threshold σ=0</u>

Table 8 - VUT 1.1 results targeting a MOS grade of 0 (same quality)

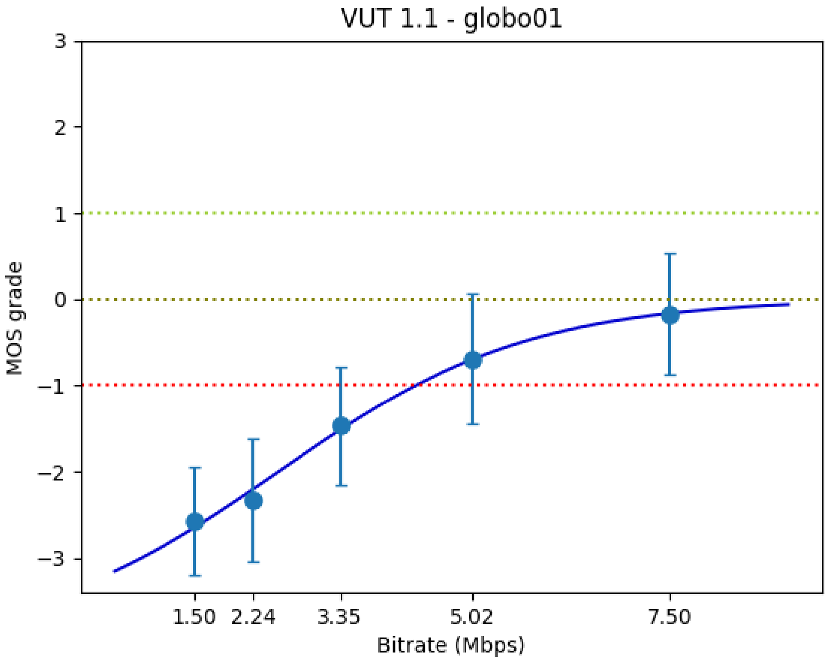

VUT 1.1 - lcevc01
Minimal bitrate=3.69
MOS grade
3
2
1
0
−1
−2
−3
1.50
2.24
3.35
5.02
7.50
Bitrate (Mbps)

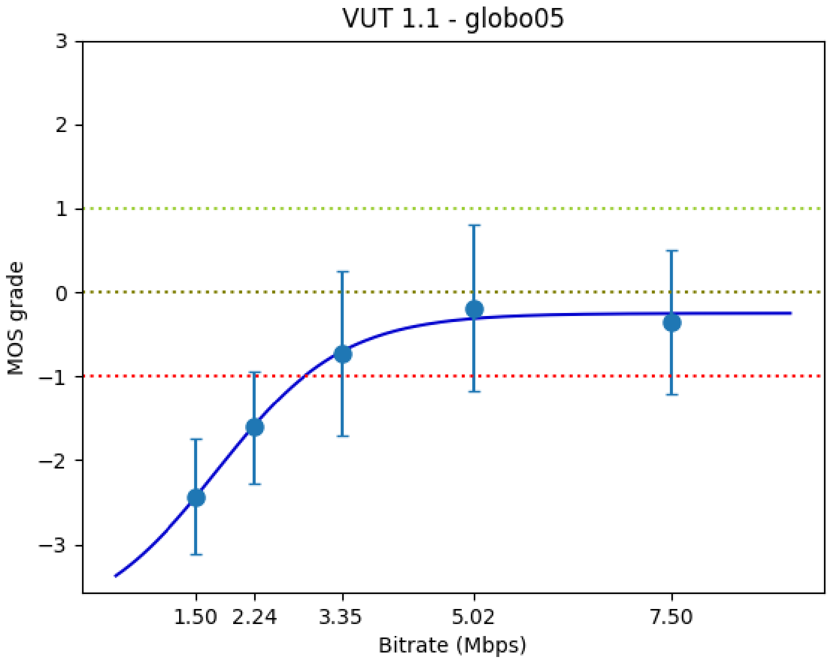

VUT 1.1 - philips01
Minimal bitrate=4.75
MOS grade
3
2
1
0
−1
−2
−3
1.50
2.24
3.35
5.02
7.50
Bitrate (Mbps)

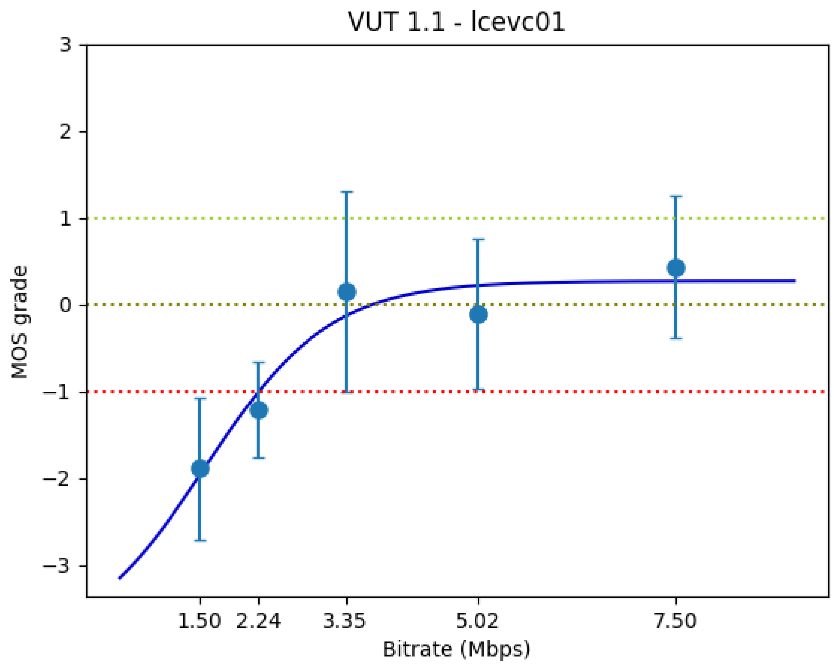

VUT 1.1 - philips03
Minimal bitrate=5.92
MOS grade
3
2
1
0
−1
−2
−3
1.50
2.24
3.35
5.02
7.50
Bitrate (Mbps)

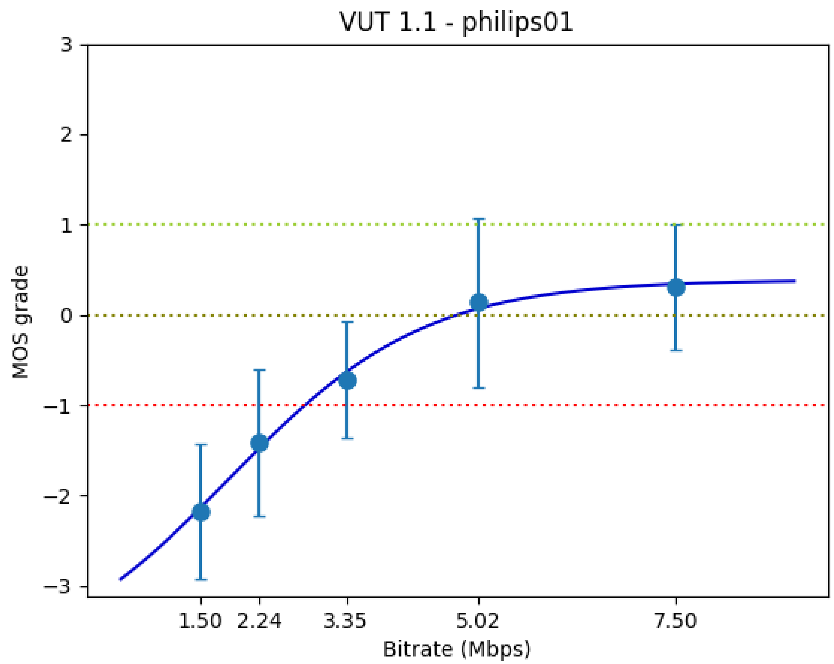
Threshold σ=1

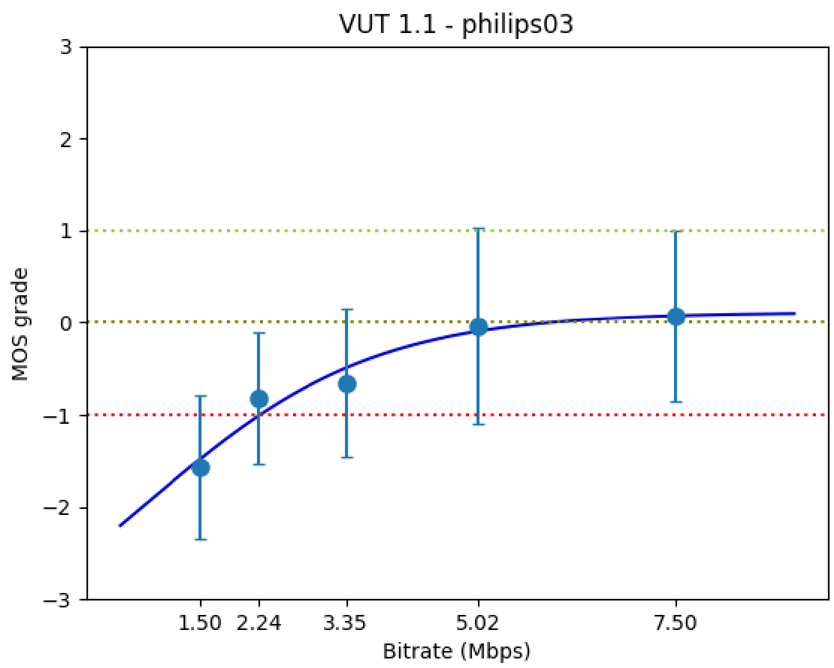
Table 9 - VUT 1.1 results targeting a MOS grade of 1 (slightly better)






VUT 1.1 - lcevc01
MOS grade
3
2
1
0
−1
−2
−3
1.50
2.24
3.35
5.02
7.50
Bitrate (Mbps)


VUT 1.1 - philips01
MOS grade
3
2
1
0
−1
−2
−3
1.50
2.24
3.35
5.02
7.50
Bitrate (Mbps)



6.3.1.3 VUT 1.1 Output bitrate analysis

The bitrates for which each target quality is achieved are shown in Table 10.

Table 10 - Achieved Target Bitrates per content (the output target bitrate is highlighted)

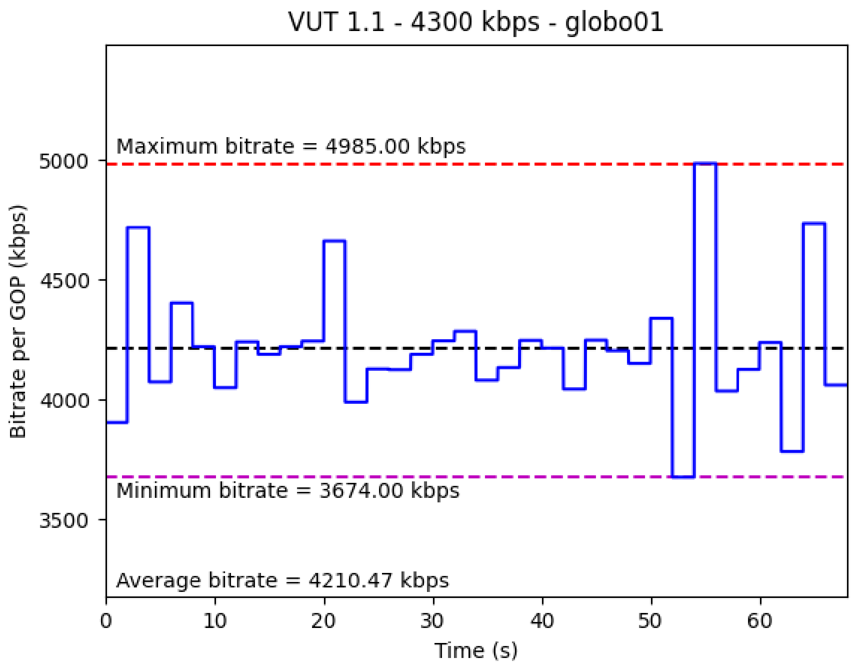
Target Quality	globo01	globo05	philips01	philips03	lcevc01
-1	**4.30 Mbps**	2.88 Mbps	2.83 Mbps	2.26 Mbps	2.26 Mbps
0	Not achieved	Not achieved	4.75 Mbps	5.92 Mbps	3.69 Mbps
1	Not achieved	Not achieved	Not achieved	Not achieved	Not achieved

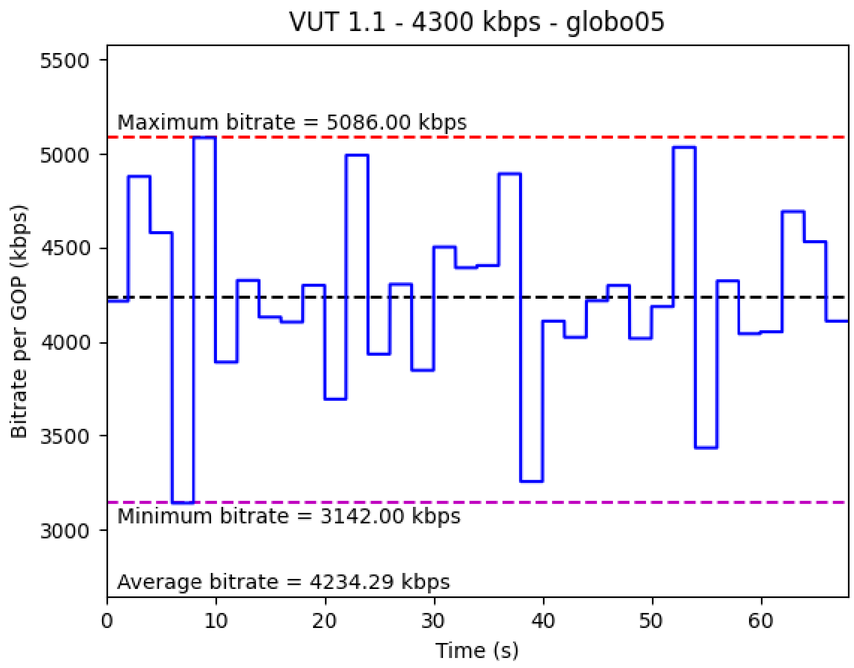
Following the results from the previous sections, the output bitrate for VUT 1.1 was found to be 4.3 Mbps. We have then encoded the stream with the same configuration as Section 6.3.1.1 at 4.3 Mbps and we have carried out the analysis of the bitrate of each GOP.

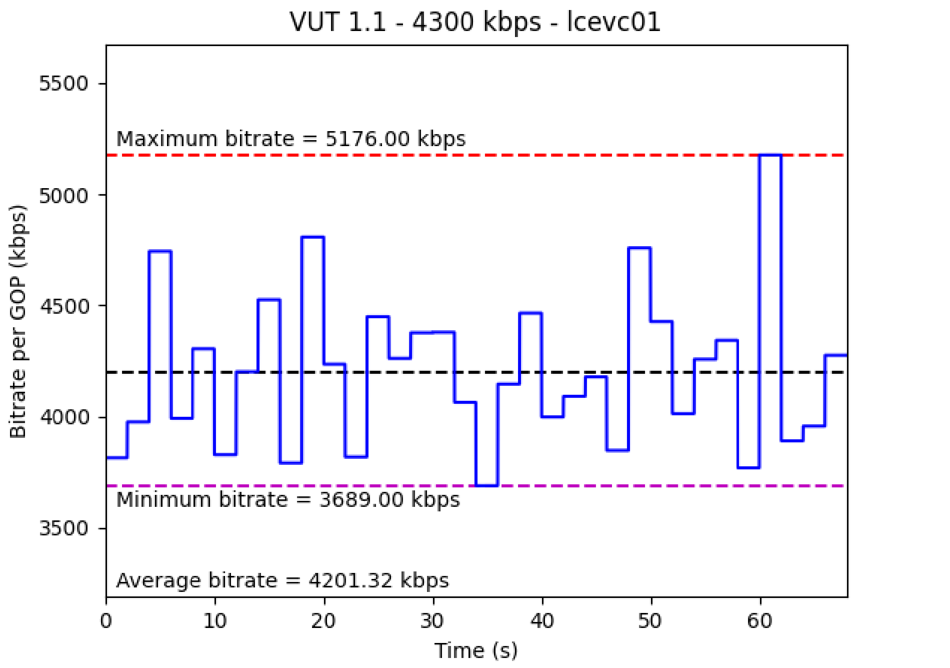
6.3.1.4 VUT 1.1 Bitrate per GOP analysis

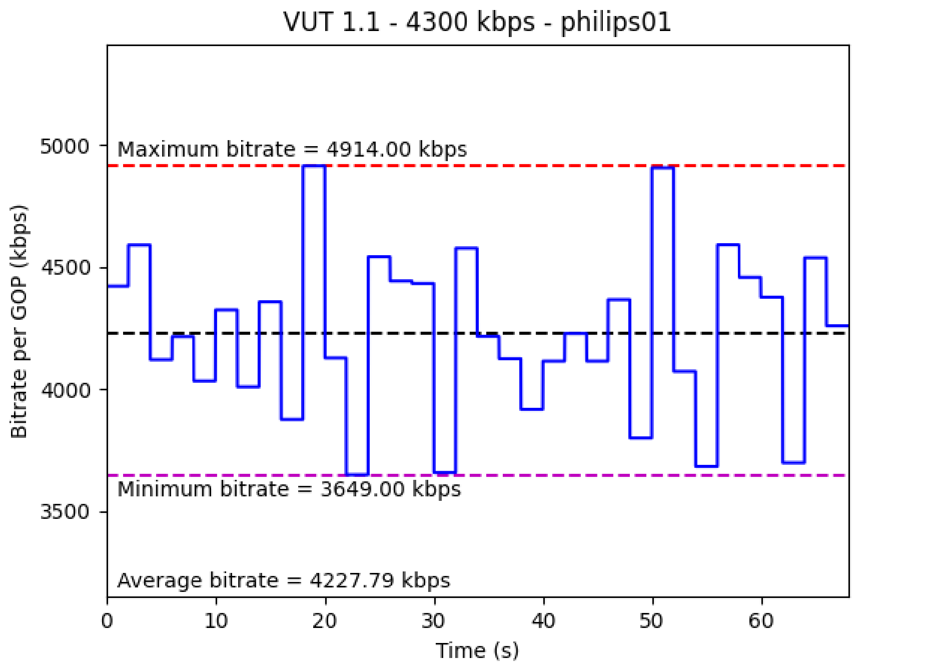
We have also carried an analysis of the bitrate per GOP for the stream at 4.3 Mbps described above. The main results are shown in Table 11.

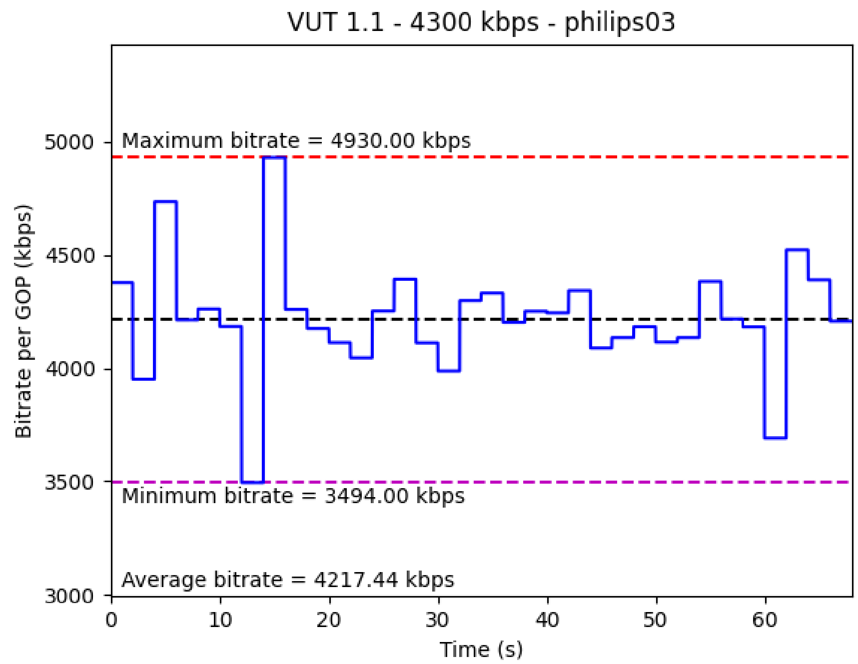
Table 11 - VUT 1.1 Bitrate per GOP when encoded with the output bitrate of 4.3 Mbps






VUT 1.1 - 4300 kbps - lcevc01
Maximum bitrate = 5176.00 kbps
Minimum bitrate = 3689.00 kbps
Average bitrate = 4201.32 kbps
Bitrate per GOP (kbps)
Time (s)
5500
5000
4500
4000
3500
0
10
20
30
40
50
60


VUT 1.1 - 4300 kbps - philips01
Maximum bitrate = 4914.00 kbps
Minimum bitrate = 3649.00 kbps
Average bitrate = 4227.79 kbps
Bitrate per GOP (kbps)
Time (s)
5000
4500
4000
3500
0
10
20
30
40
50
60



The average, minimum and maximum bitrates per GOP are shown in Table 12. All rates are given in kbps.

Table 12 - VUT 1.1 Average, Minimum and Maximum bitrates per GOP when encoded with the output bitrate of 4.3 Mbps

Content	TARGET Rate	Average Rate	Minimum Rate	Maximum Rate
globo01	4300	4210.47	3674	4985
globo05	4300	4234.29	3142	5086
lcevc01	4300	4201.32	3689	5176
philips01	4300	4227.79	3649	4914
philips03	4300	4217.44	3494	4930

As can be seen from the table, the average bitrate per GOP is very close to the target bitrate.

6.3.1.5 VUT 1.1 Analysis and Conclusions

Following the results shown in the previous sections, the output target bitrate for VUT 1.1 for a target quality of -1 (“slightly worse”) was found to be 4.30 Mbps. An output target bitrate for a target quality of 0 (“same quality”) could not be achieved for some contents, namely globo01 and globo05. This is coherent with the complexity shown in Figure 3, where these two contents showed the highest complexity. Moreover, these contents have a larger number of human faces in the scene at the same time. It is known that human faces draw a lot of attention from the viewers, and when reducing the resolution to 720p the quality impairment in the faces is noticeable, and this shows in the overall quality scores given. An output target bitrate for a target quality of 1 (“slightly higher”) could not be achieved for any content - this is likely because the VUT was encoded at 720p while the reference was a 1 080i stream.

When encoding with the target bitrate of 4.30 Mbps, the average bitrate output by the encoder is very similar to the desired bitrate as shown in Table 12.

6.3.2 VUT 1.2 VVC 1 080p

6.3.2.1 VUT 1.2 Definition

The goal of this VUT was to test the VVC encoder working at 1 080p resolution. The reference video was encoded with H.264 + SL-HDR, which corresponds to the best quality possible under the current TV 2.5 specifications. The complete details are found in Table 13. Since the VUT was encoded with the same resolution as the reference video, the quality target considered was “the same” (0).

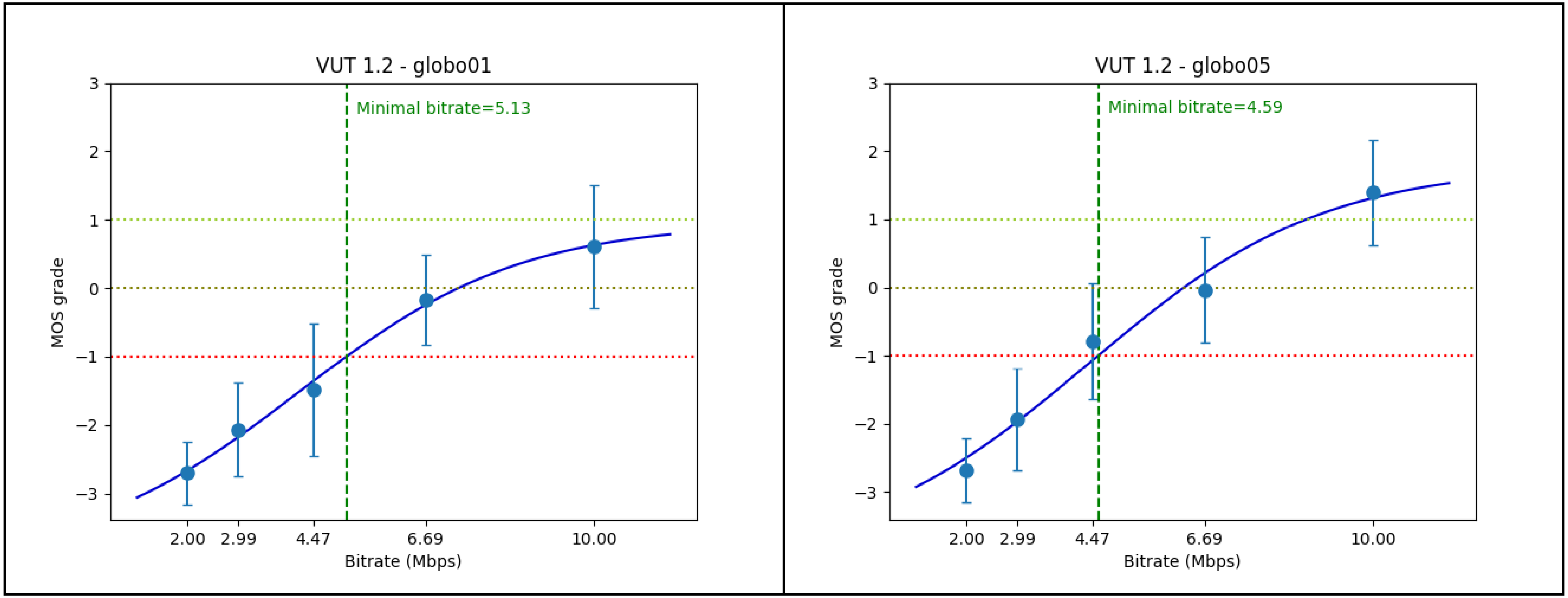
Table 13 - VUT 1.2 encoding details

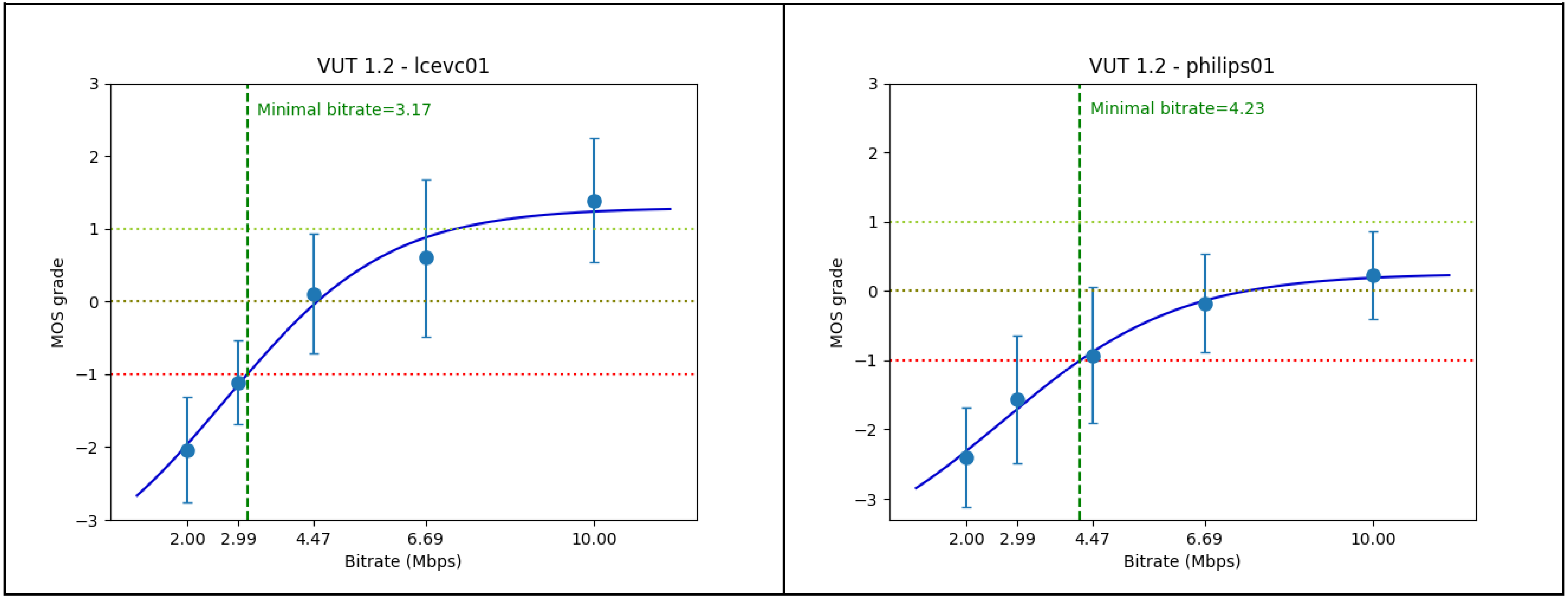
	Reference Video	VUT 1.2
Content label	1 080i	1 080p
Resolution	1 920 x 1 080	1 920 x 1 080
Frame rate	29.97 fps	59.94 fps
Scan	Interlaced	Progressive
Bit depth	8 bits	10 bits
Color gamut	BT.709	BT.2020
HDR Mode	SL-HDR1	HDR10
Codec	H.264/AVC	H.266/VVC
GOP size	60 frames (2 seconds)	120 frames (2 seconds)
Encoder	MainConcept Live Encoder v0.0.0.961	Ateme TitanLive Innovation v 4.1.31.911
Encoder type	Real-time	Real-time
Bitrate	14 Mbps	[2.00 2.99 4.47 6.69 10.00] Mbps

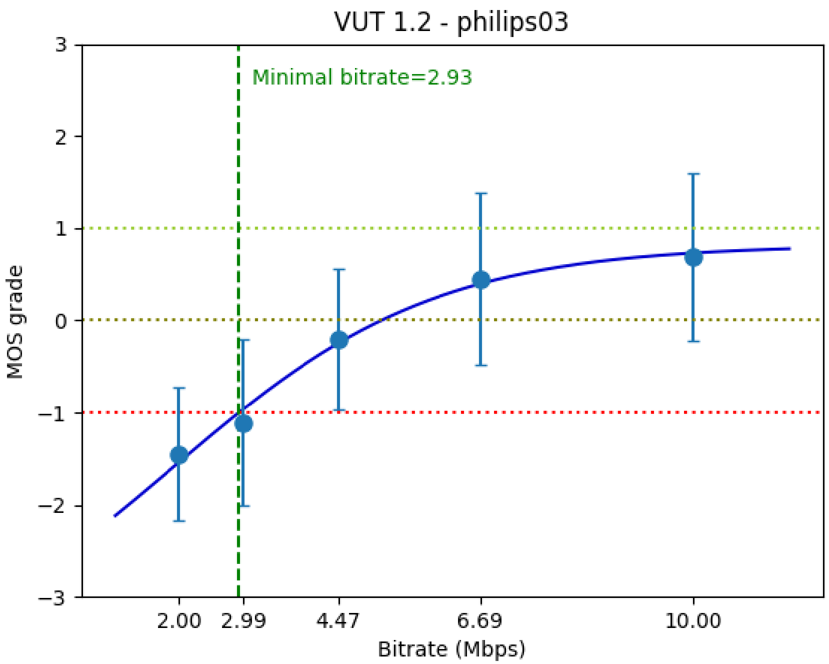
6.3.2.2 VUT 1.2 Experimental Findings

Threshold σ=-1

Table 14 - VUT 1.2 results targeting a MOS grade of -1 (slightly worse)

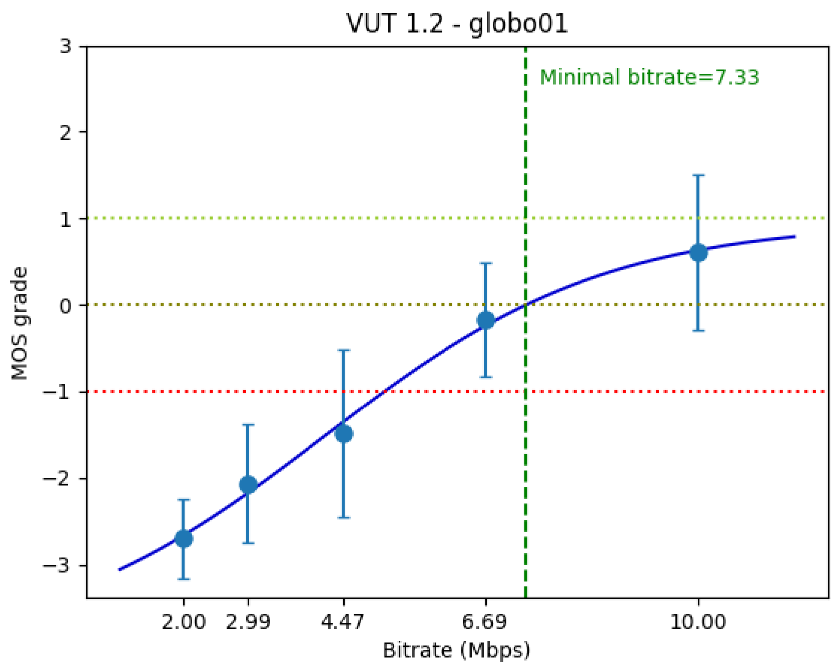


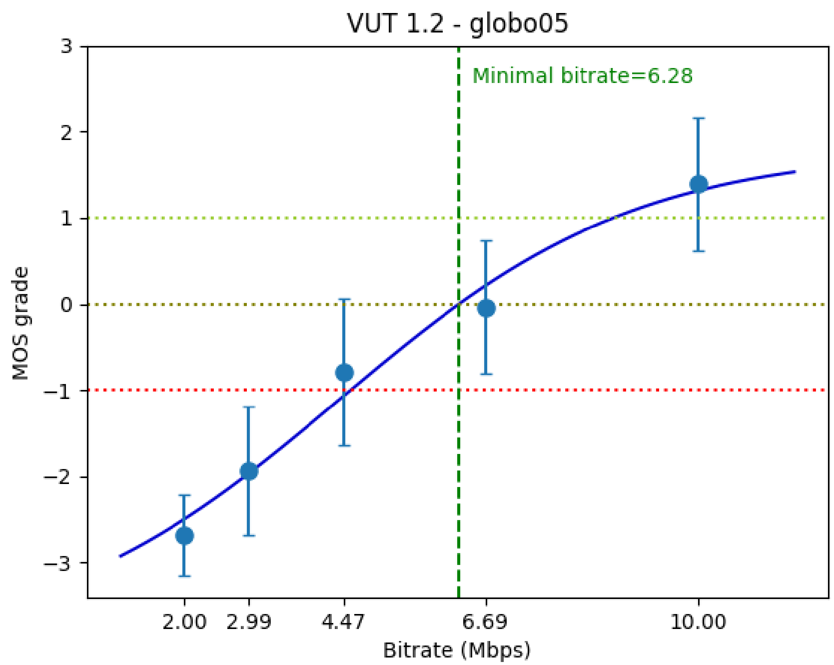

VUT 1.2 - lcevc01
Minimal bitrate=3.17
MOS grade
Bitrate (Mbps)
2.00 2.99 4.47 6.69 10.00
VUT 1.2 - philips01
Minimal bitrate=4.23
MOS grade
Bitrate (Mbps)
2.00 2.99 4.47 6.69 10.00

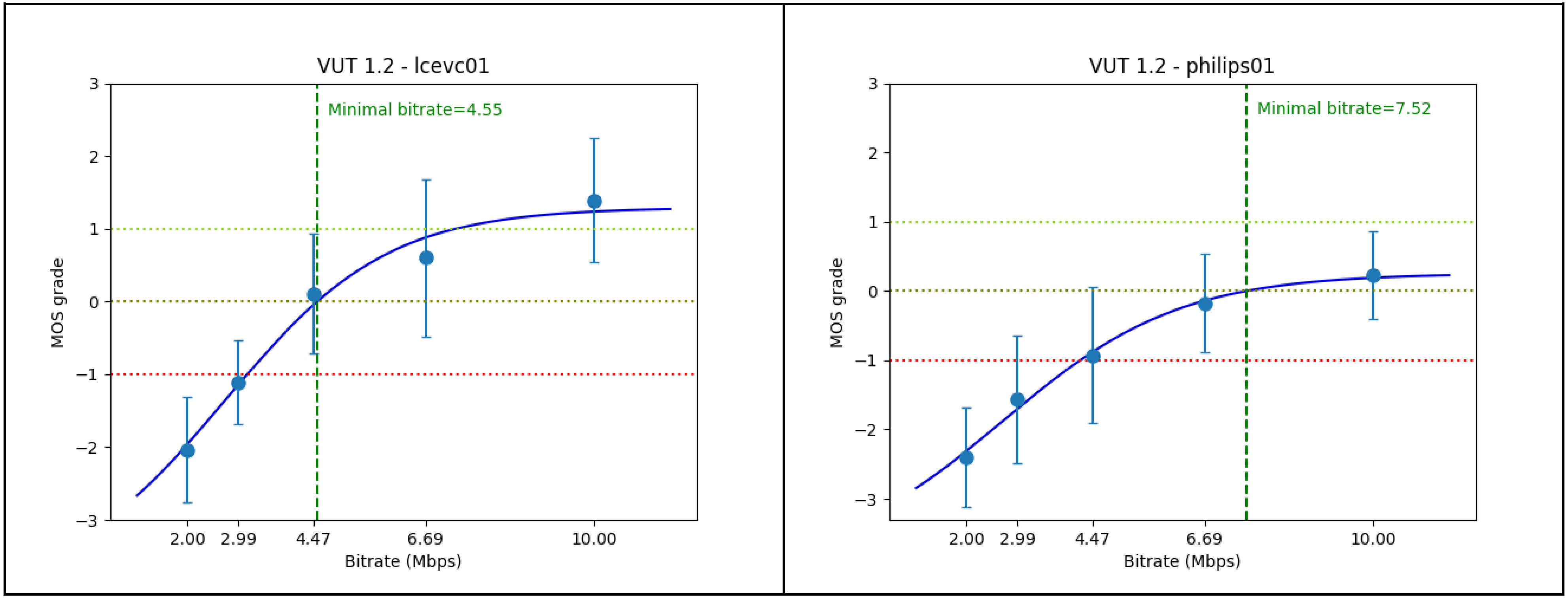

VUT 1.2 - philips03
Minimal bitrate=2.93
3
2
1
0
−1
−2
−3
MOS grade
2.00
2.99
4.47
6.69
10.00
Bitrate (Mbps)

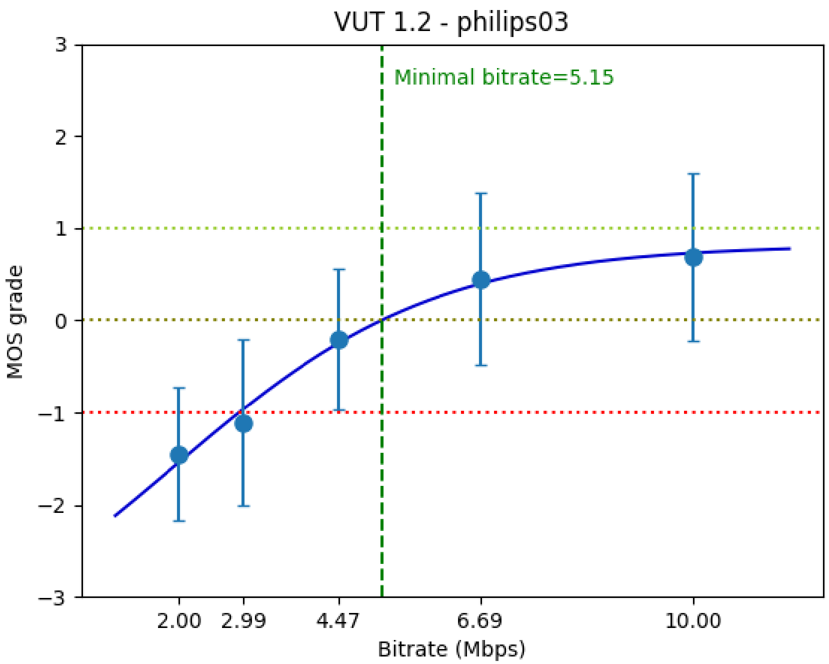
Threshold σ=0

Table 15 - VUT 1.2 results targeting a MOS grade of 0 (same quality)



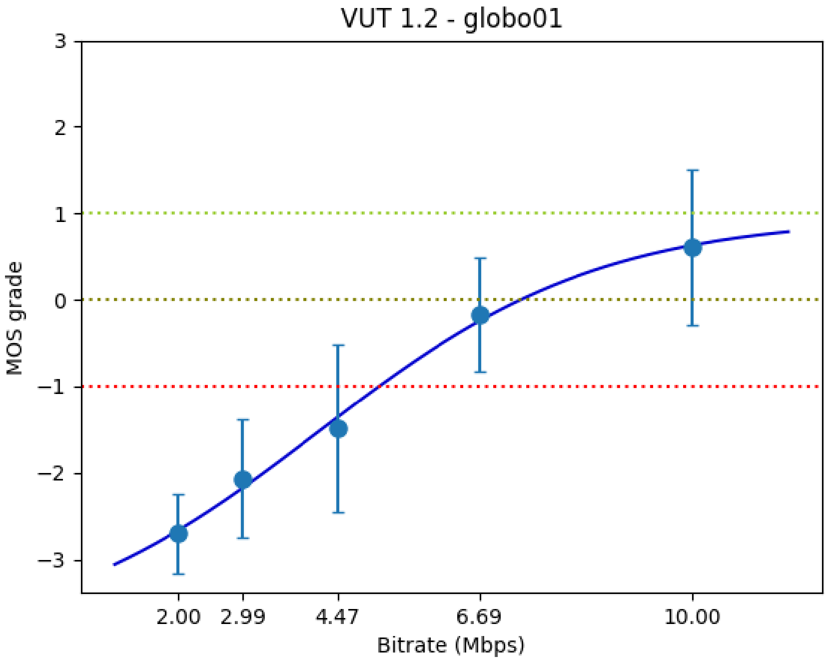


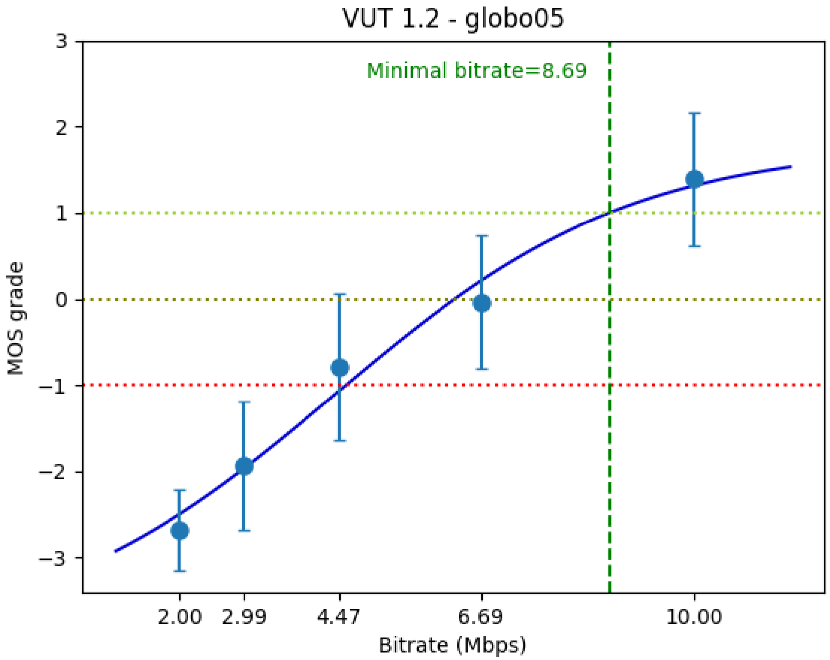

VUT 1.2 - lcevc01
Minimal bitrate=4.55
MOS grade
Bitrate (Mbps)
2.00
2.99
4.47
6.69
10.00
3
2
1
0
−1
−2
−3
VUT 1.2 - philips01
Minimal bitrate=7.52
MOS grade
Bitrate (Mbps)
2.00
2.99
4.47
6.69
10.00
3
2
1
0
−1
−2
−3

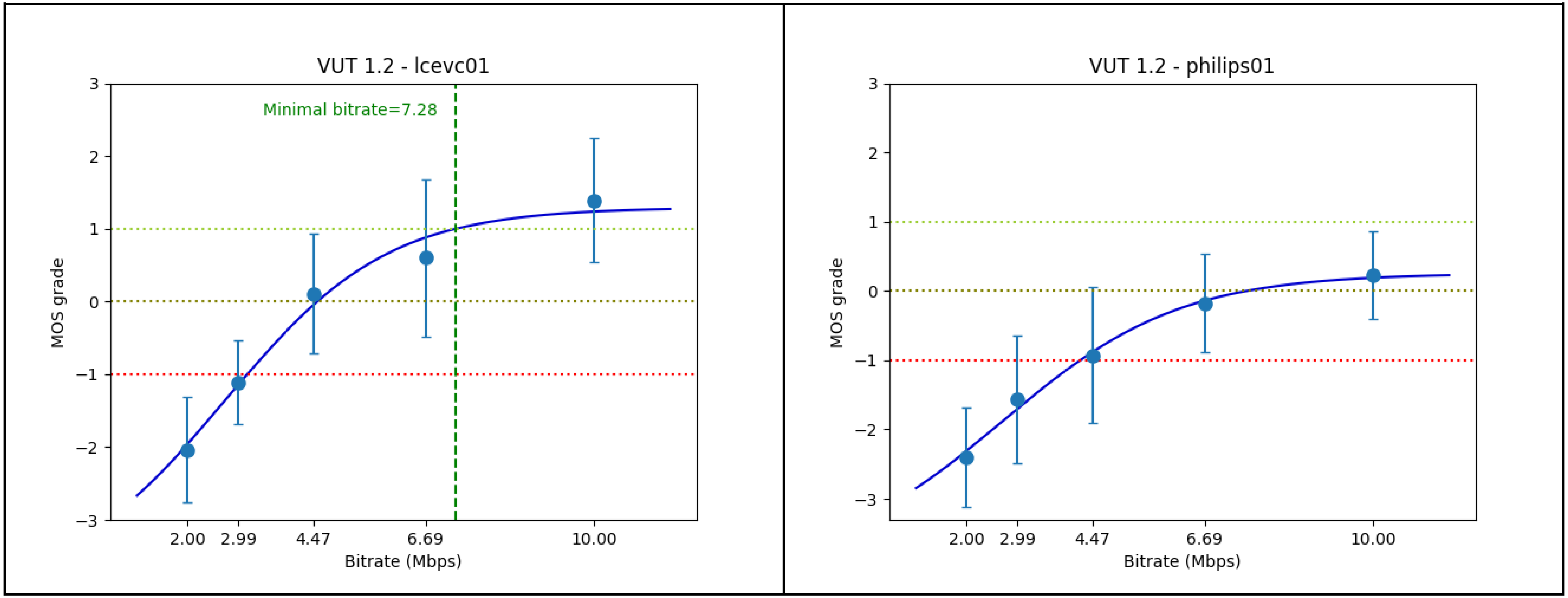

VUT 1.2 - philips03
Minimal bitrate=5.15
MOS grade
Bitrate (Mbps)
3
2
1
0
−1
−2
−3
2.00
2.99
4.47
6.69
10.00

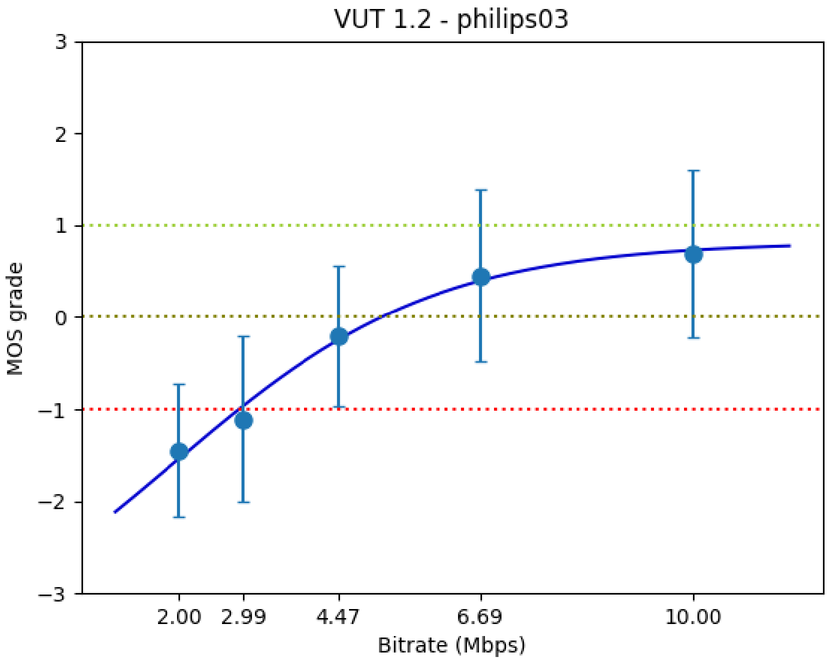
Threshold σ=1

Table 16 - VUT 1.2 results targeting a MOS grade of 1 (slightly better)






VUT 1.2 - lcevc01
Minimal bitrate=7.28
MOS grade
Bitrate (Mbps)
2.00
2.99
4.47
6.69
10.00
VUT 1.2 - philips01
MOS grade
Bitrate (Mbps)
2.00
2.99
4.47
6.69
10.00



6.3.2.3 VUT 1.2 Output bitrate analysis

The bitrates for which each target quality is achieved are shown in Table 17.

Table 17 - Achieved Target Bitrates per content (the output target bitrate is highlighted)

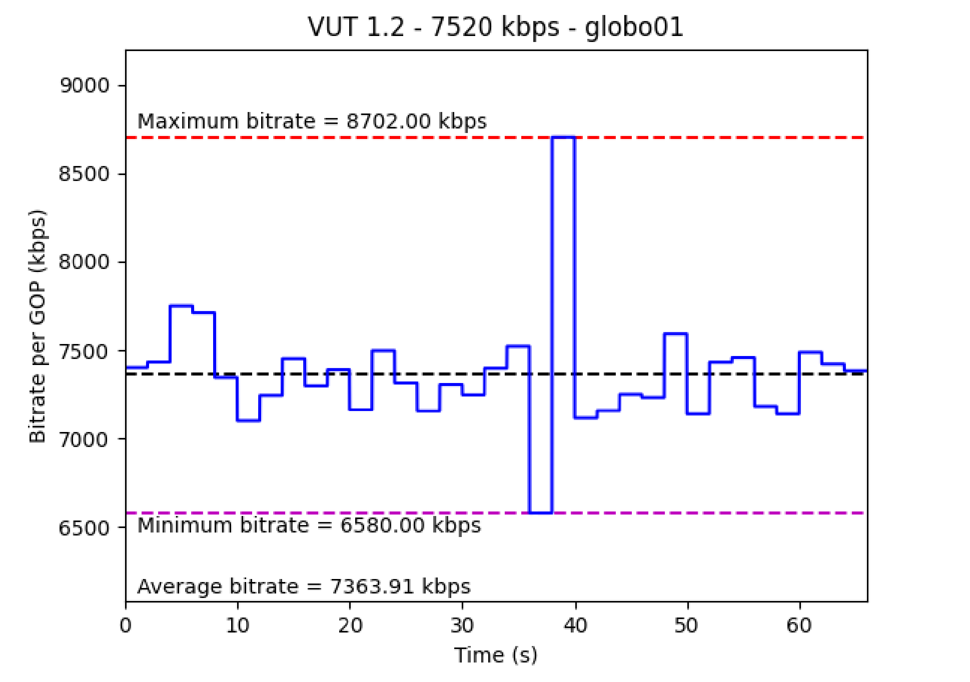
Target Quality	globo01	globo05	philips01	philips03	lcevc01
-1	5.13 Mbps	4.59 Mbps	4.23 Mbps	2.93 Mbps	3.17 Mbps
0	7.33 Mbps	6.28 Mbps	**7.52 Mbps**	5.15 Mbps	4.55 Mbps
1	Not achieved	8.69 Mbps	Not achieved	Not achieved	7.28 Mbps

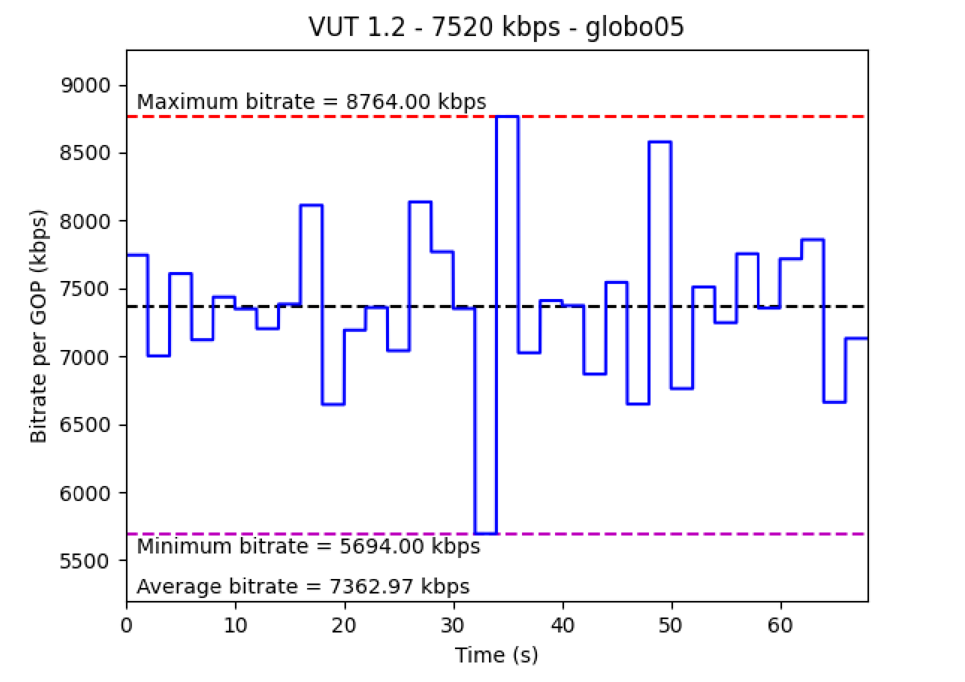
Following the results from the previous sections, the output bitrate for VUT 1.2 was found to be 7.52 Mbps. We have then encoded the stream with the same configuration as Section 6.3.2.1 at 7.52 Mbps and we have carried out the analysis of the bitrate of each GOP.

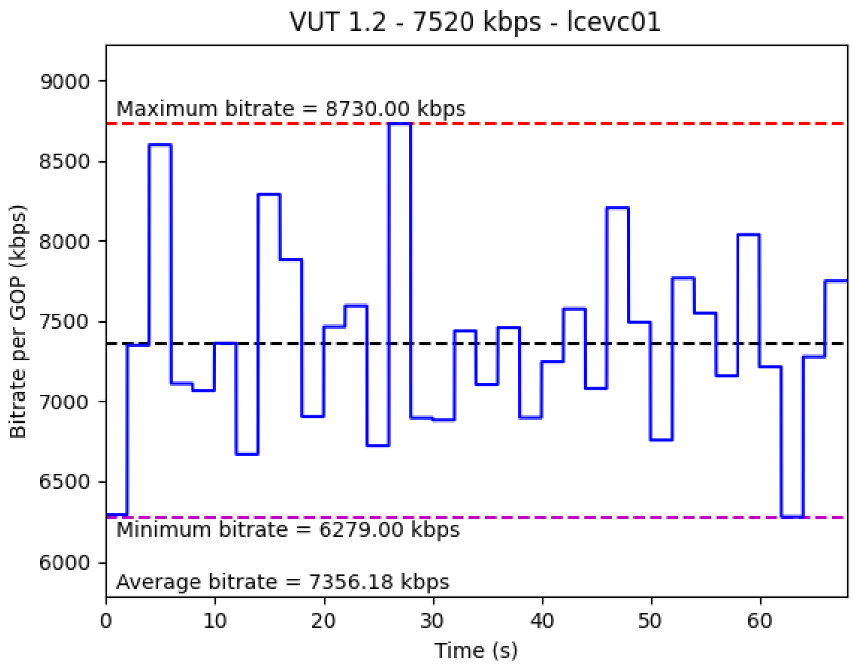
6.3.2.4 VUT 1.2 Bitrate per GOP analysis

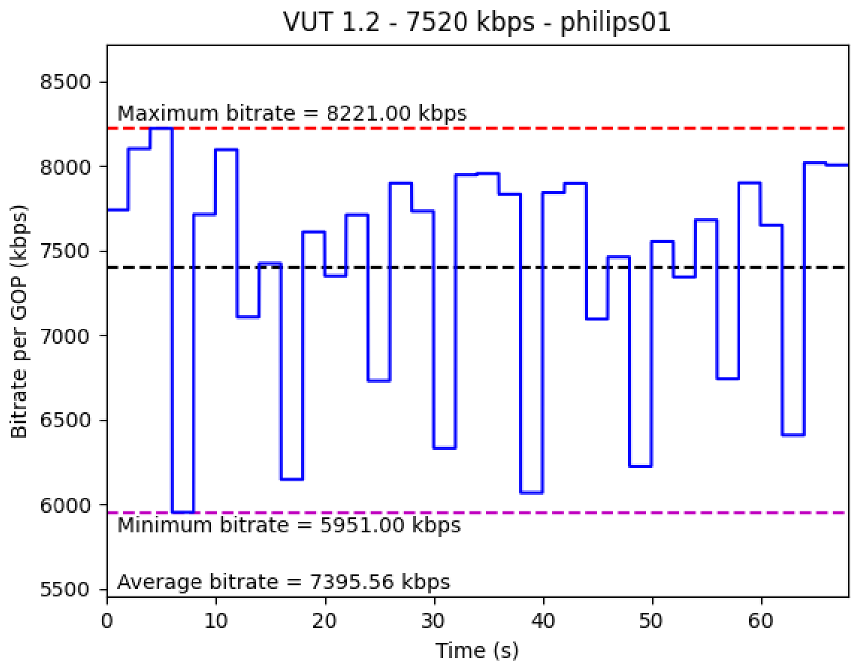
We have carried an analysis of the bitrate per GOP for the stream at 7.52 Mbps described above. The main results are shown in Table 18.

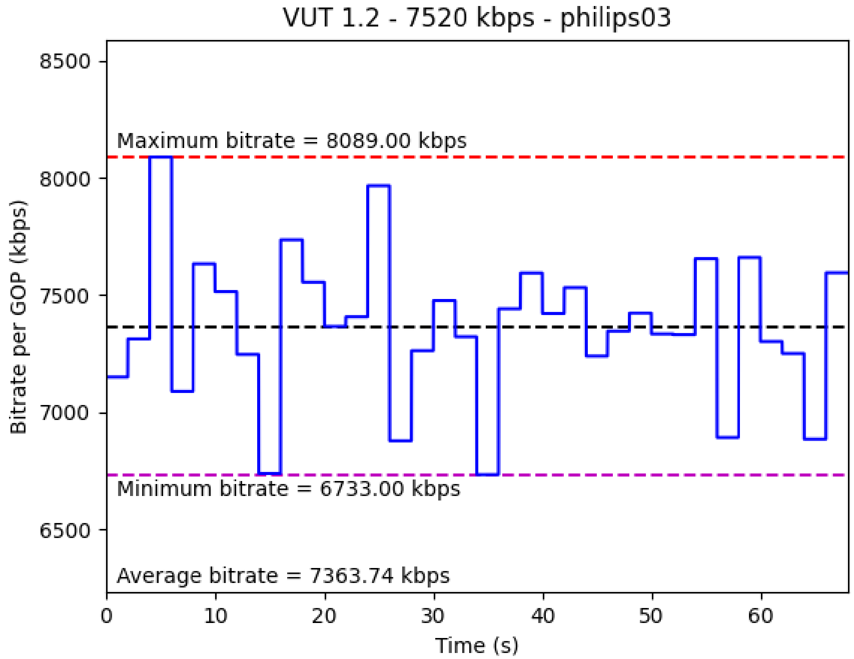
Table 18 - VUT 1.2 Bitrate per GOP when encoded with the output bitrate of 7.52 Mbps






VUT 1.2 - 7520 kbps - lcevc01
Maximum bitrate = 8730.00 kbps
Minimum bitrate = 6279.00 kbps
Average bitrate = 7356.18 kbps
Bitrate per GOP (kbps)
Time (s)
9000
8500
8000
7500
7000
6500
6000
0
10
20
30
40
50
60


VUT 1.2 - 7520 kbps - philips01
Maximum bitrate = 8221.00 kbps
Minimum bitrate = 5951.00 kbps
Average bitrate = 7395.56 kbps
Bitrate per GOP (kbps)
Time (s)
8500
8000
7500
7000
6500
6000
5500
0
10
20
30
40
50
60



The average, minimum and maximum bitrates per GOP are shown in Table 19. All rates are given in kbps.

Table 19 - VUT 1.2 Average, Minimum and Maximum bitrates per GOP when encoded with the output bitrate of 7.52 Mbps

Content	TARGET Rate	Average Rate	Minimum Rate	Maximum Rate
globo01	7520	7363.91	6580	8702
globo05	7520	7362.97	5694	8764
lcevc01	7520	7356.18	6279	8730
philips01	7520	7395.56	5951	8221
philips03	7520	7363.74	6733	8089

As can be seen from the table, the average bitrate per GOP is very close to the target bitrate.

6.3.2.5 VUT 1.2 Analysis and Conclusions

As depicted in Table 17, the output target bitrate for VUT 1.2 for a target quality of -1 (“slightly worse”) was found to be 5.13. This result is based on the content "globo01" that demanded the highest bitrate. The output target bitrate for a target quality of 0 (“same quality”) was achieved at 7.52 Mbps for the "philips01" video sequence. An output target bitrate for a target quality of 1 (“slightly higher”) could not be achieved for "globo01", "philips01" and "philips03" contents. Similarly to VUT 1.1, as described in Section 6.3.1, we encoded the VUT 1.2 videos with the found target bitrate. However, differently from VUT 1.1, in this experiment, we get the quality target as 0 ("the same") since VUT 1.2 was encoded with the same resolution of the reference. Using this configuration, the selected bitrate of 7.52 Mbps achieves an average rate value close to 7.3 Mbps as shown in Table 19.

6.3.3 VUT 1.3 VVC 2 160p

6.3.3.1 VUT 1.3 Definition

The goal of this VUT was to test the VVC encoder working at 2 160p resolution. Following the results of the VUT 1.2 on Section 6.3.2, the reference video for this VUT was encoded with the VVC at 1 080p resolution, at 7.52 Mbps. The complete details are found in Table 20. Since the VUT was encoded with a higher resolution than the reference video, the quality target considered was “the same” (0). Note that using the highest bitrate corresponding to the target score 0 ("the same") among the five clips in the test material means that this VUT would provide a similar subjective quality for the clip with the highest required bitrate, and somewhat higher score in the other clips while not necessarily reaching the score 1 ("slightly better") in all clips.

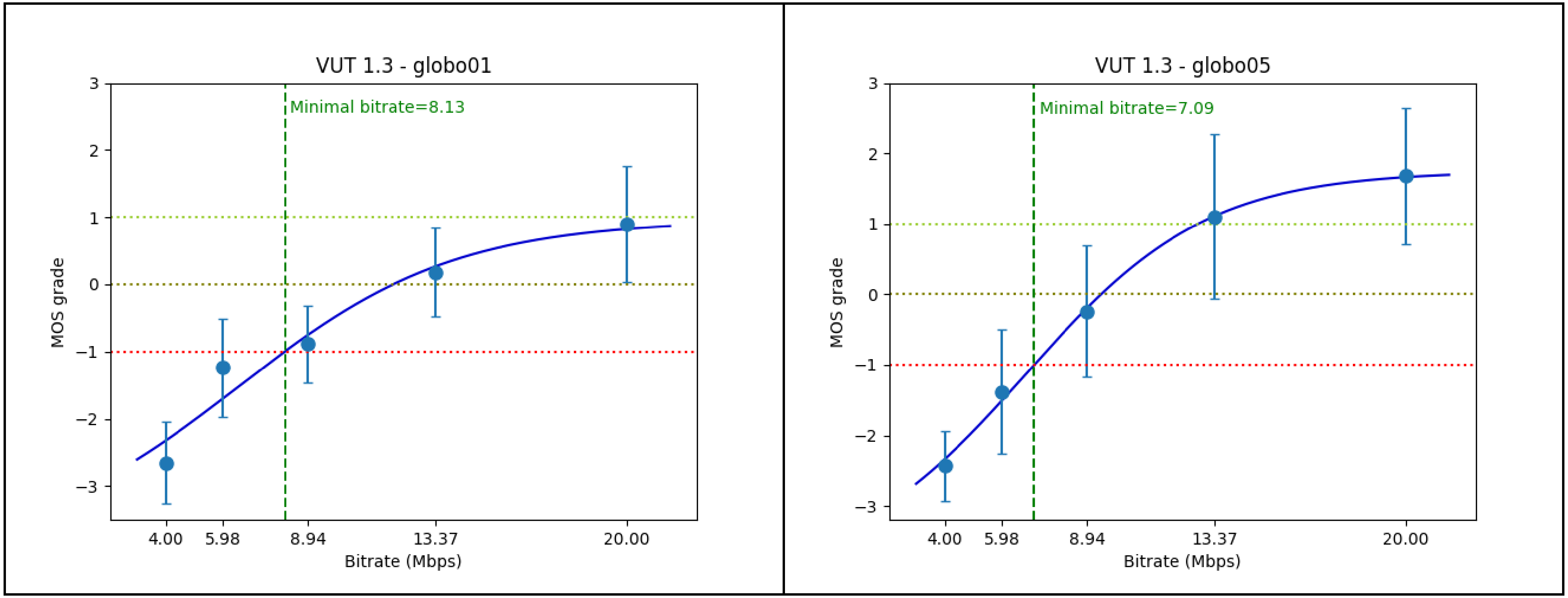
Table 20 - VUT 1.3 encoding details

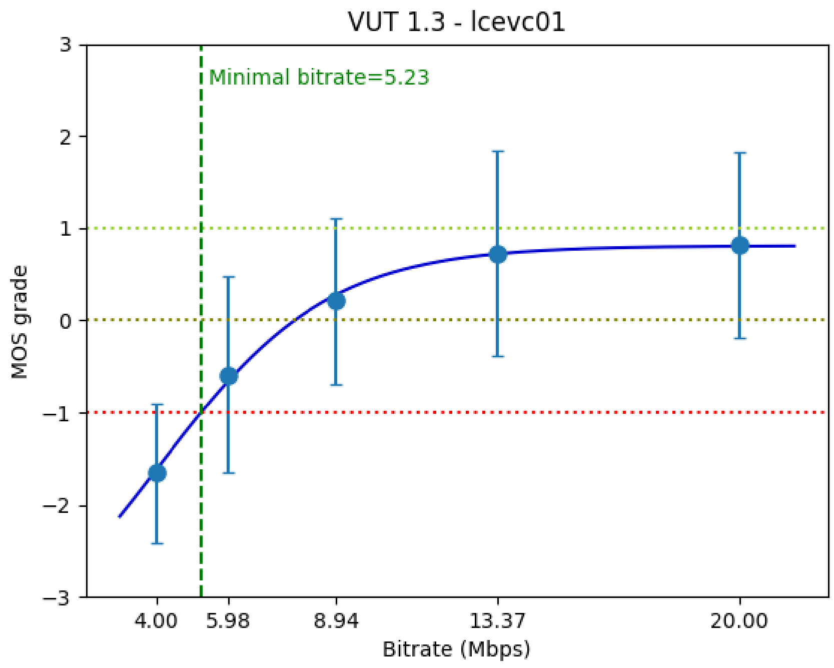
	Reference Video	VUT 1.3
Content label	1 080p	2 160p
Resolution	1 920 x 1 080	3 840 x 2 160
Frame rate	59.94 fps	59.94 fps
Scan	Progressive	Progressive
Bit depth	10 bits	10 bits
Color gamut	BT.2020	BT.2020
HDR Mode	HDR10	HDR10
Codec	H.266/VVC	H.266/VVC
GOP size	120 frames (2 seconds)	120 frames (2 seconds)
Encoder	Ateme TitanLive Innovation v 4.1.31.911	Ateme TitanLive Innovation v 4.1.31.911
Encoder type	Real-time	Real-time
Bitrate	7.52 Mbps	[4.00 5.98 8.94 13.37 20.00] Mbps

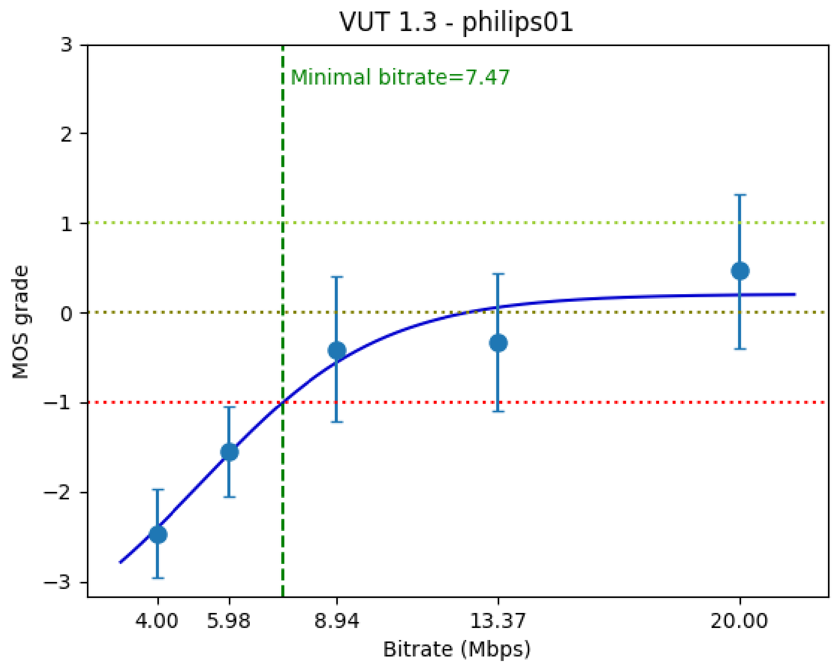
6.3.3.2 VUT 1.3 Experimental Findings

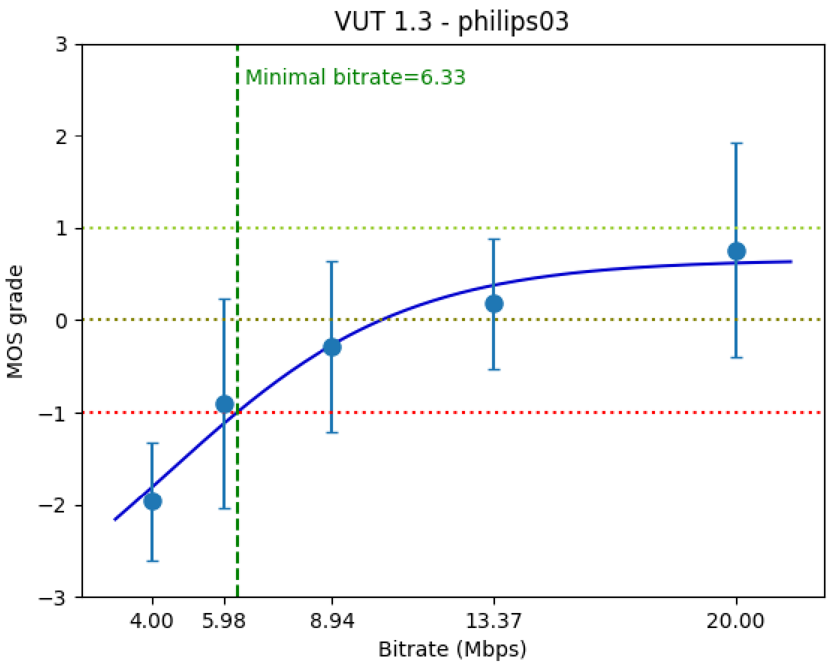
Threshold σ=-1

Table 21 - VUT 1.3 results targeting a MOS grade of -1 (slightly worse)



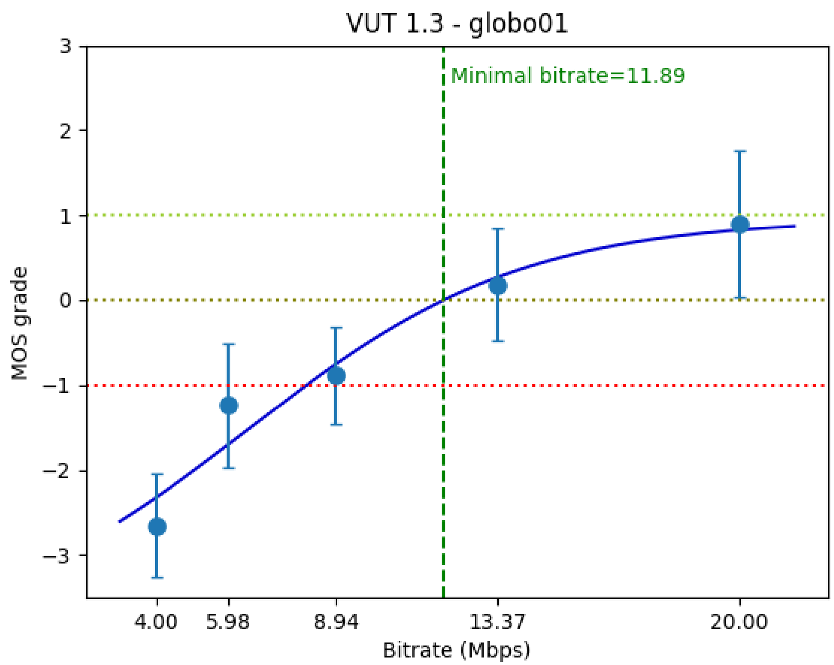

VUT 1.3 - lcevc01
Minimal bitrate=5.23
MOS grade
Bitrate (Mbps)
4.00
5.98
8.94
13.37
20.00
3
2
1
0
−1
−2
−3

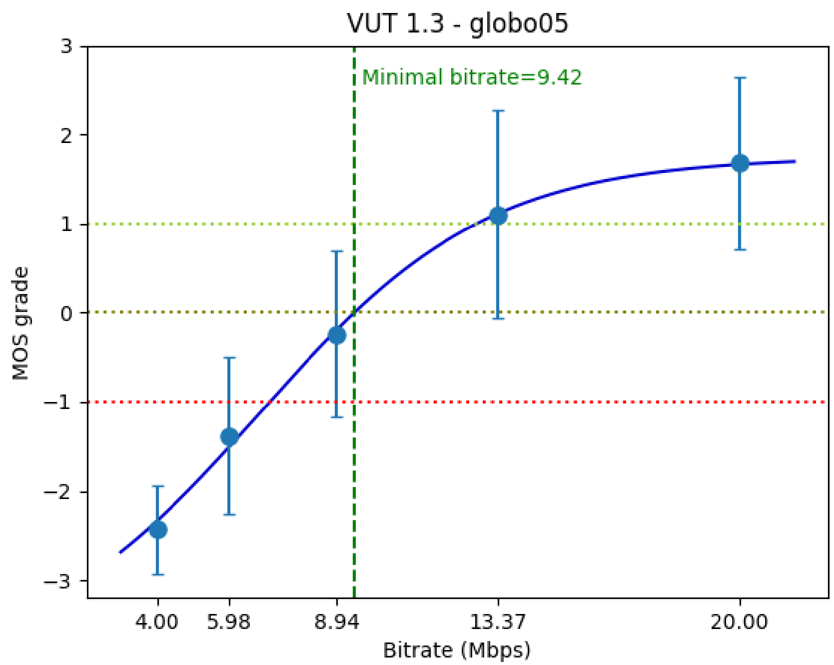

VUT 1.3 - philips01
Minimal bitrate=7.47
MOS grade
Bitrate (Mbps)
4.00
5.98
8.94
13.37
20.00
3
2
1
0
−1
−2
−3

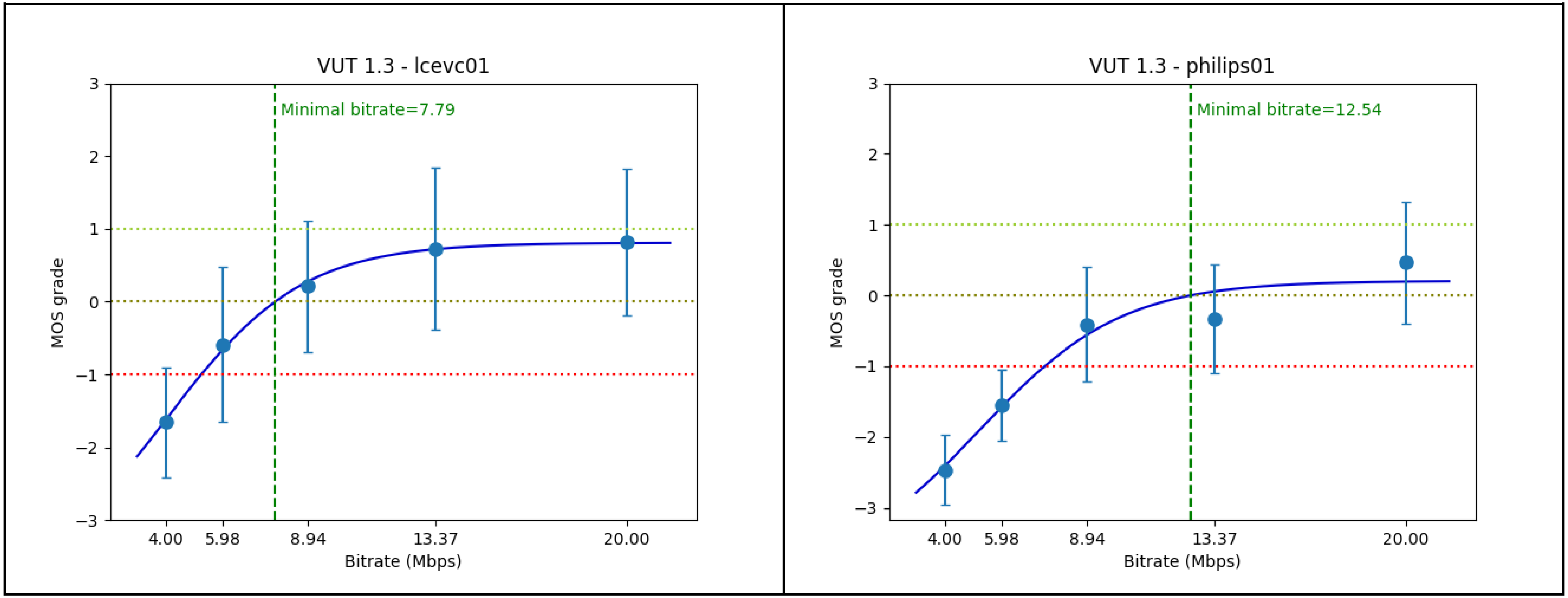

VUT 1.3 - philips03
Minimal bitrate=6.33
3
2
1
0
−1
−2
−3
MOS grade
4.00
5.98
8.94
13.37
20.00
Bitrate (Mbps)

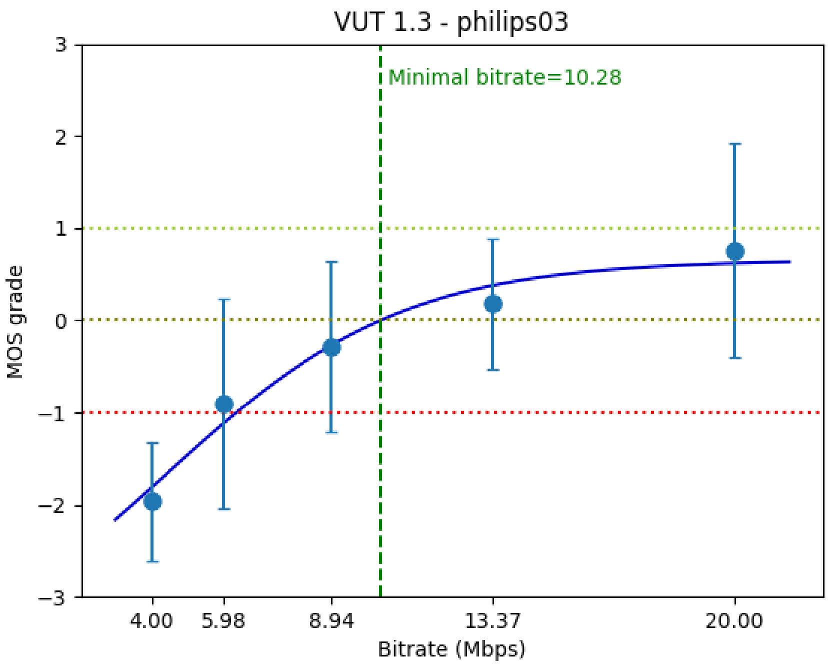
Threshold σ=0

Table 22 - VUT 1.3 results targeting a MOS grade of 0 (same quality)



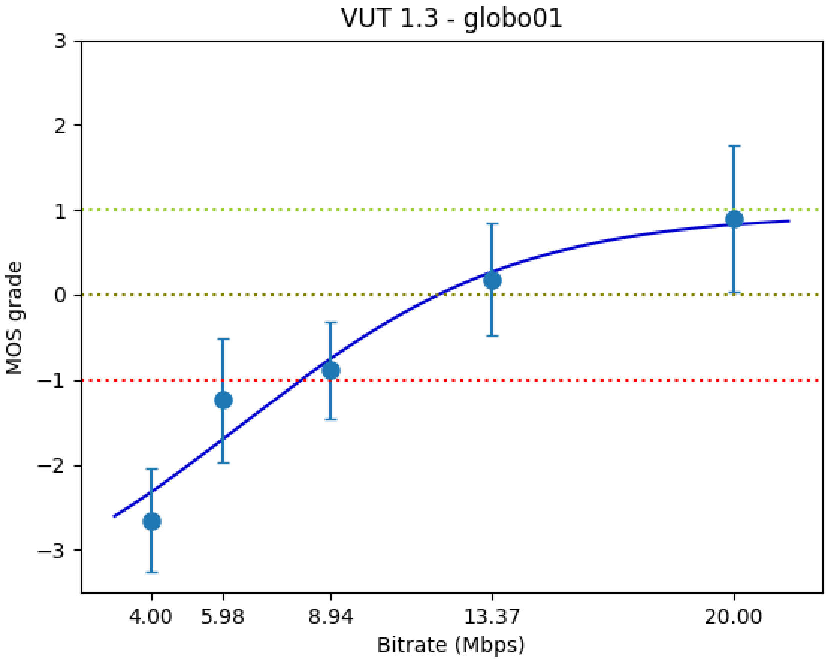


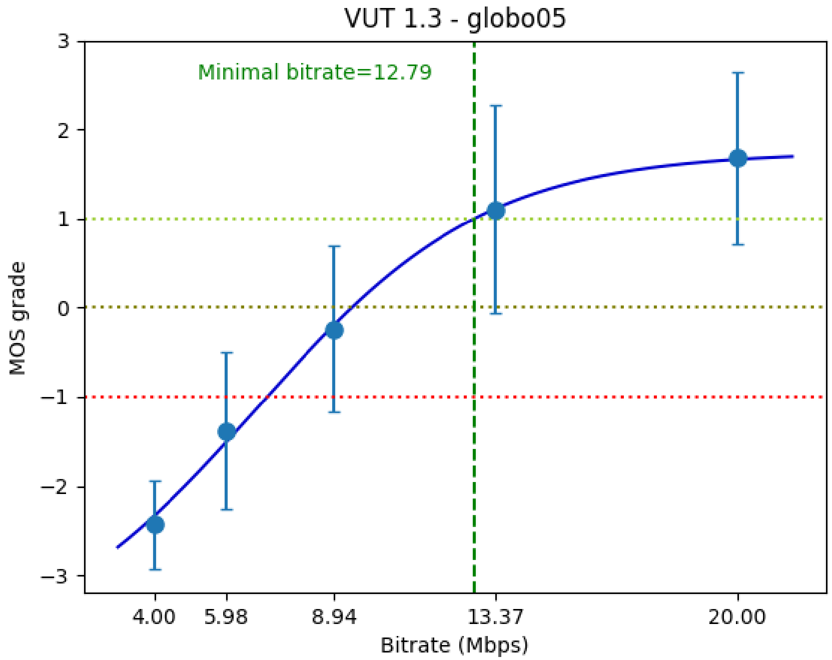

VUT 1.3 - lcevc01
Minimal bitrate=7.79
MOS grade
Bitrate (Mbps)
4.00
5.98
8.94
13.37
20.00
VUT 1.3 - philips01
Minimal bitrate=12.54
MOS grade
Bitrate (Mbps)

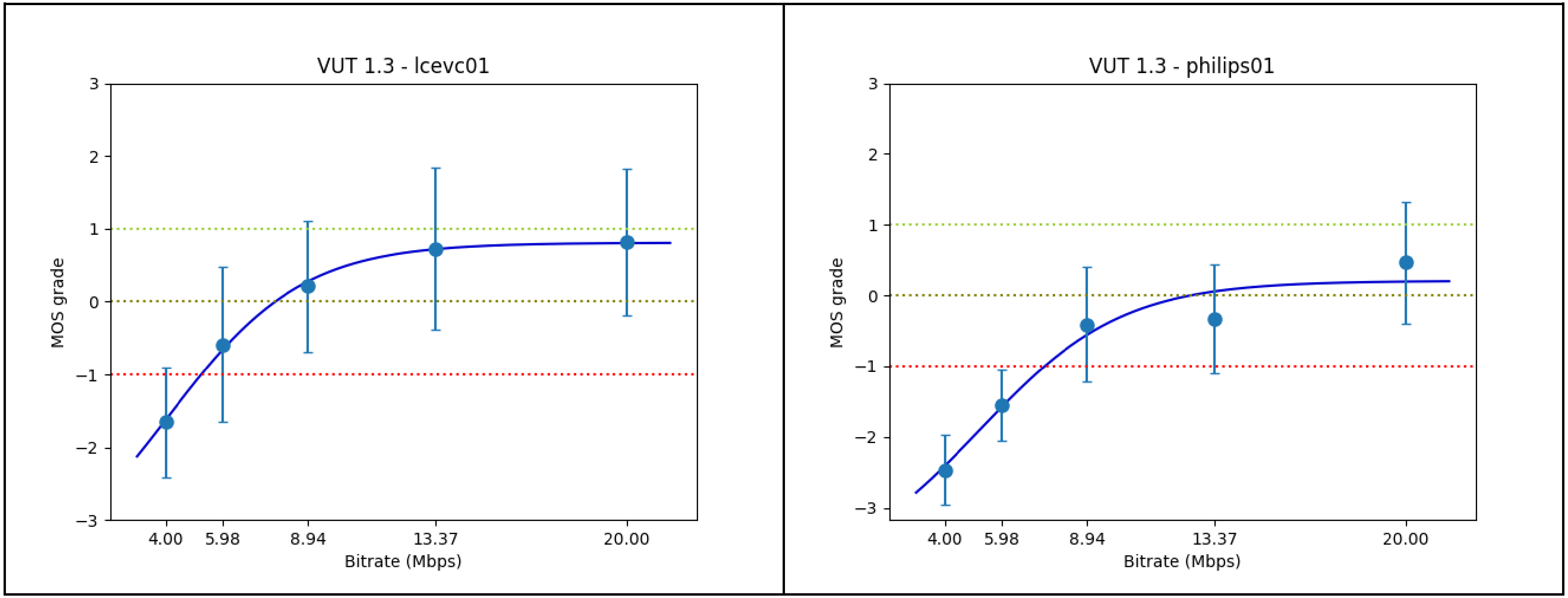

VUT 1.3 - philips03
Minimal bitrate=10.28
MOS grade
Bitrate (Mbps)
3
2
1
0
−1
−2
−3
4.00
5.98
8.94
13.37
20.00

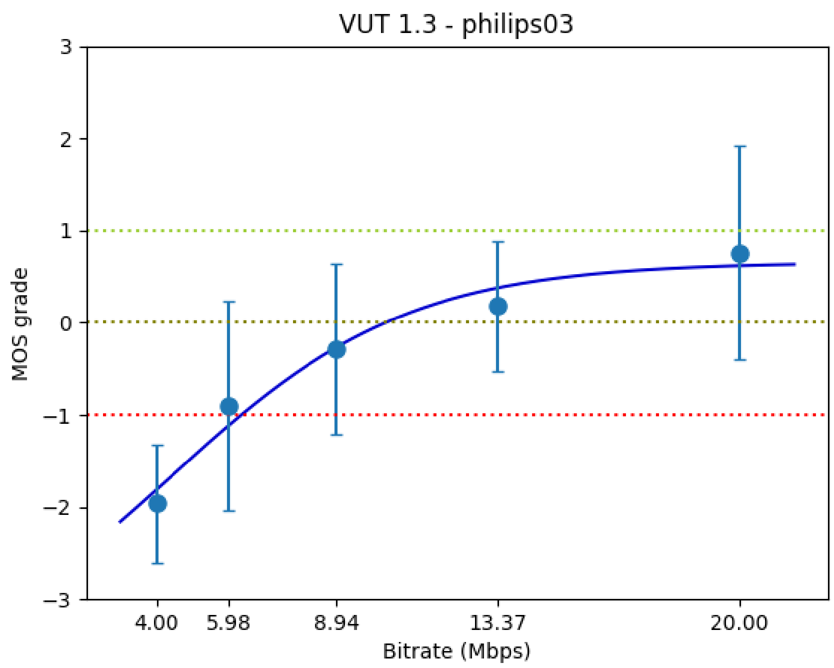
Threshold σ=1

Table 23 - VUT 1.3 results targeting a MOS grade of 1 (slightly better)






VUT 1.3 - lcevc01
MOS grade
Bitrate (Mbps)
4.00
5.98
8.94
13.37
20.00
3
2
1
0
−1
−2
−3
VUT 1.3 - philips01
MOS grade
Bitrate (Mbps)
4.00
5.98
8.94
13.37
20.00
3
2
1
0
−1
−2
−3



6.3.3.3 VUT 1.3 Output bitrate analysis

The bitrates for which each target quality is achieved are shown in Table 24.

Table 24 - Achieved Target Bitrates per content (the output target bitrate is highlighted)

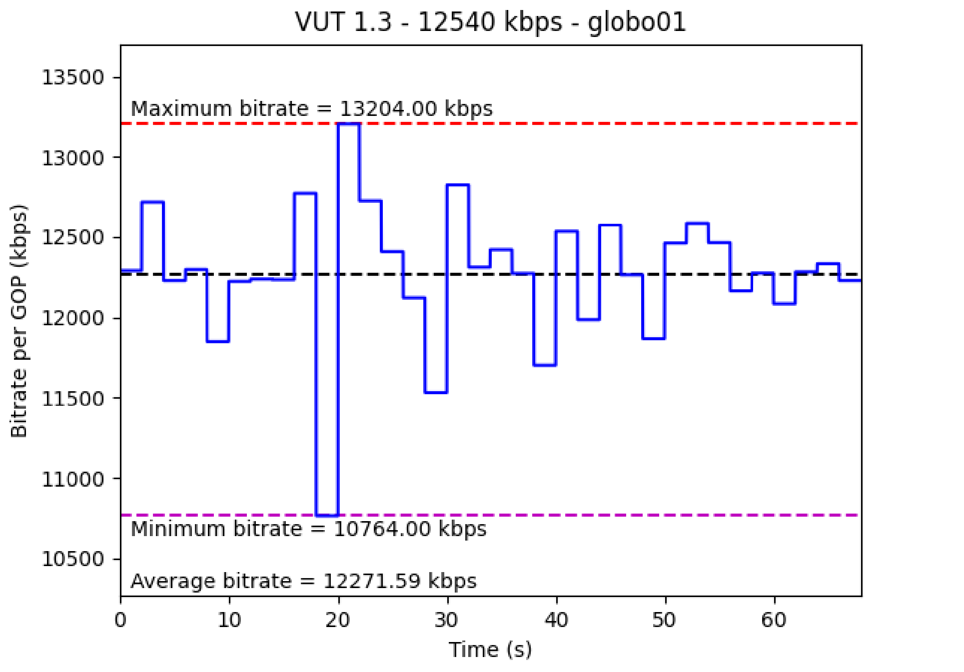
Target Quality	globo01	globo05	philips01	philips03	lcevc01
-1	8.13 Mbps	7.09 Mbps	7.47 Mbps	6.33 Mbps	5.23 Mbps
0	11.89 Mbps	9.42 Mbps	**12.54 Mbps**	10.28 Mbps	7.79 Mbps
1	Not achieved	12.79 Mbps	Not achieved	Not achieved	Not achieved

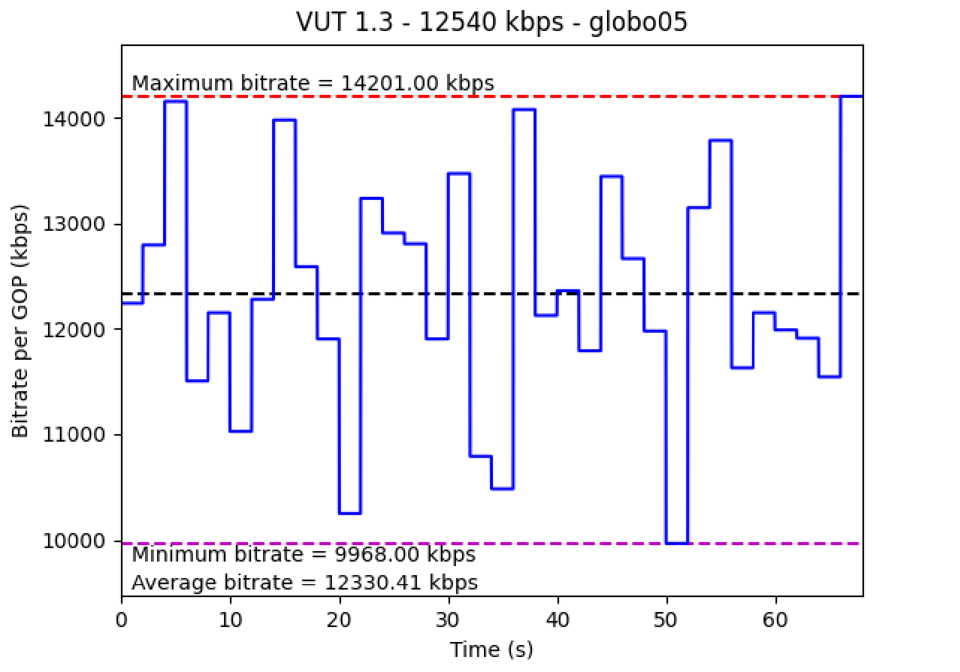
Following the results from the previous sections, the output bitrate for VUT 1.3 was found to be 12.54 Mbps. We have then encoded the stream with the same configuration as Section 6.3.3.1 at 12.54 Mbps and we have carried out the analysis of the bitrate of each GOP.

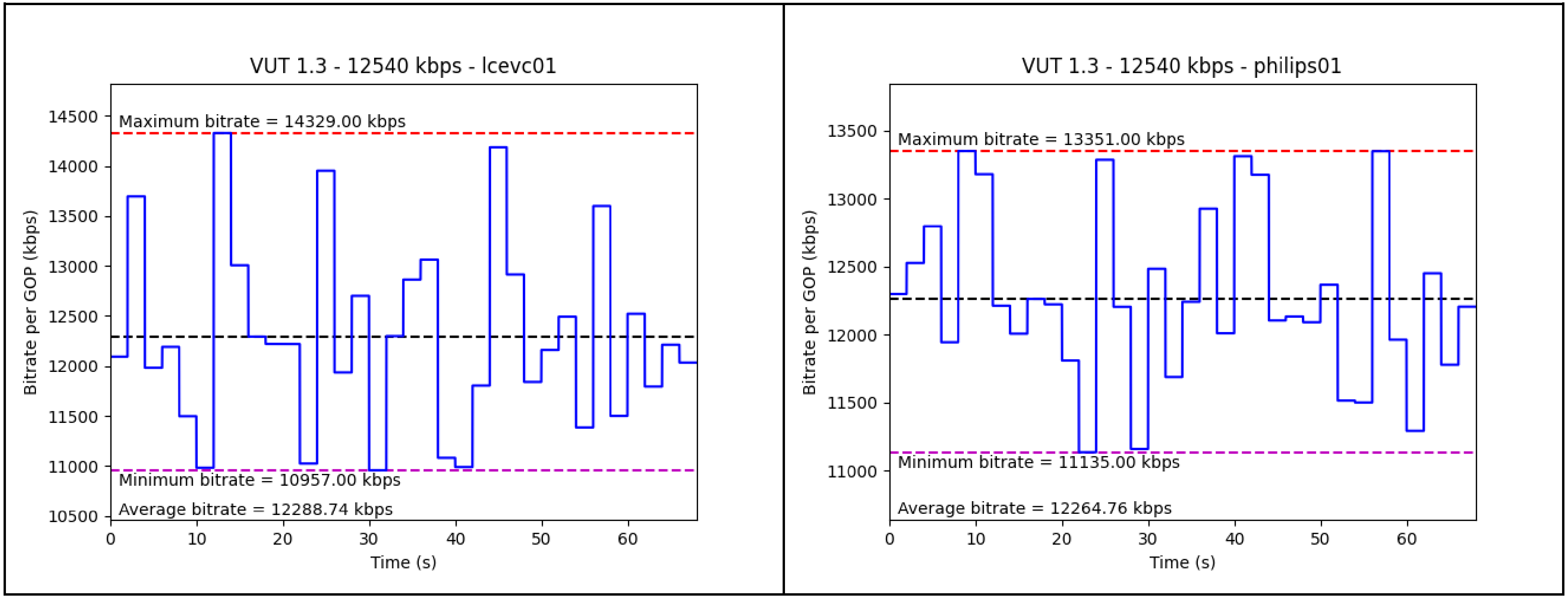
6.3.3.4 VUT 1.3 Bitrate per GOP analysis

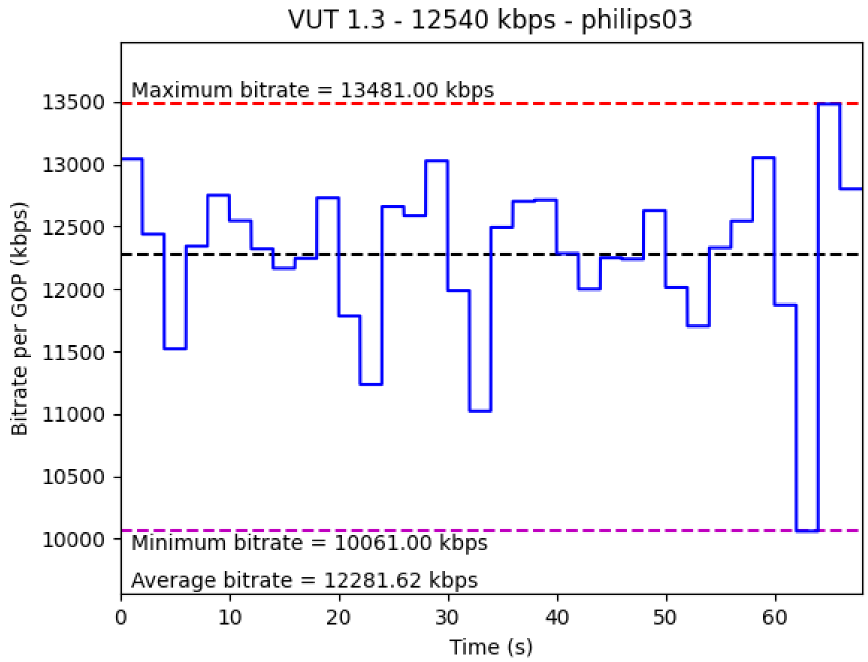
We have also carried an analysis of the bitrate per GOP for the stream at 12.54 Mbps described above. The main results are shown in Table 25.

Table 25 - VUT 1.3 Bitrate per GOP when encoded with the output bitrate of 12.54 Mbps






VUT 1.3 - 12540 kbps - lcevc01
Maximum bitrate = 14329.00 kbps
Minimum bitrate = 10957.00 kbps
Average bitrate = 12288.74 kbps
Bitrate per GOP (kbps)
Time (s)
VUT 1.3 - 12540 kbps - philips01
Maximum bitrate = 13351.00 kbps
Minimum bitrate = 11135.00 kbps
Average bitrate = 12264.76 kbps
Bitrate per GOP (kbps)
Time (s)



The average, minimum and maximum bitrates per GOP are shown in Table 26. All rates are given in kbps.

Table 26 - VUT 1.3 Average, Minimum and Maximum bitrates per GOP when encoded with the output bitrate of 12.54 Mbps

Content	TARGET Rate	Average Rate	Minimum Rate	Maximum Rate
globo01	12540	12271.59	10764	13204
globo05	12540	12330.41	9968	14201
lcevc01	12540	12288.74	10957	14329
philips01	12540	12264.76	11135	13351
philips03	12540	12281.62	10061	13481

As can be seen from the table, the average bitrate per GOP is very close to the target bitrate.

6.3.3.5 VUT 1.3 Analysis and Conclusions

As depicted in Table 24, the output target bitrate for VUT 1.3 for a target quality of -1 (“slightly worse”) was found to be 8.13 Mbps. This result is based on the content "globo01" that demanded the highest bitrate. The output target bitrate for a target quality of 0 (“same quality”) was achieved at 12.54 Mbps for the "philips01" video sequence. The output target bitrate for a target quality of 1 (“slightly higher”) could not be achieved for "globo01", "philips01", "philips03", and "lcevc01" contents. Even though VUT 1.3 was encoded with a higher resolution than the reference video, we considered as output target bitrate the highest rate achieved with “same quality” (0). Note that using the highest bitrate corresponding to the target score 0 ("the same") among the five clips in the test material means that this VUT would provide a similar subjective quality for the clip with the highest required bitrate, and somewhat higher score in the other clips while not necessarily reaching the score 1 ("slightly better") in all clips.

6.3.4 VUT 2.3 VVC Base Layer 1 080p

6.3.4.1 VUT 2.3 Definition

The goal of this VUT was to test the base layer of the VVC + LCEVC encoder. The base layer is actually encoded with the VVC codec, working at 1 080p resolution. However, since the encoders used in VUTs 1.2 and 2.3 were from different manufacturers and used different configurations, and since the goal was to evaluate the complete VVC + LCEVC codec, this VUT was included in the tests although it cannot be directly compared with VUT 1.2. The reference video was encoded with H.264 + SL-HDR, which corresponds to the best quality possible under the current TV 2.5 specifications. The complete details are found in Table 27. Since the VUT was encoded with the same resolution as the reference video, the quality target considered was “the same” (0).

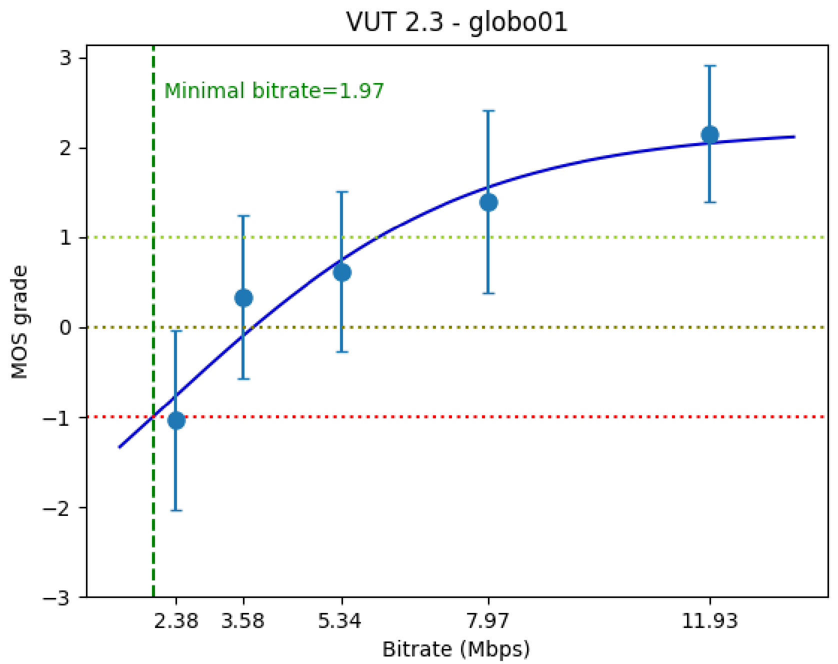
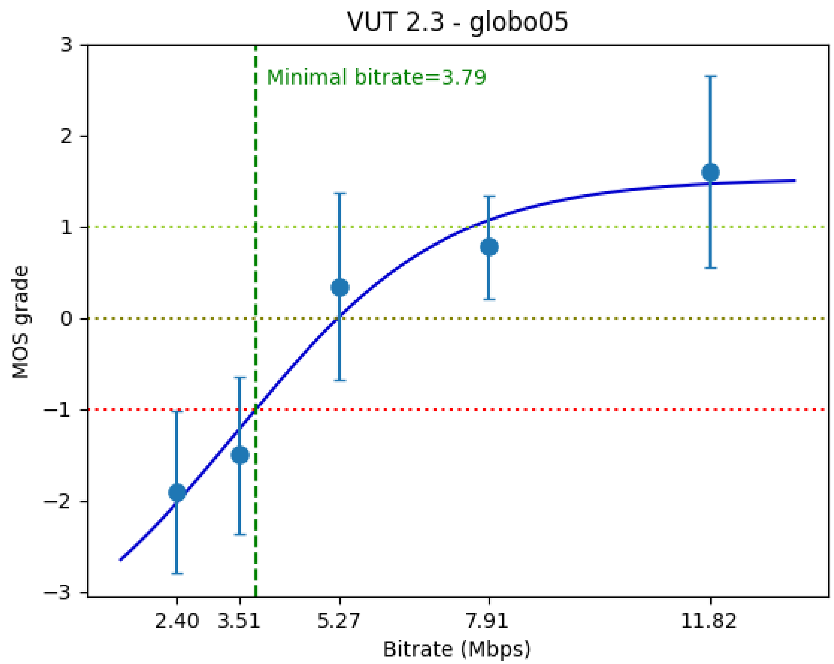
Table 27 - VUT 2.3 encoding details

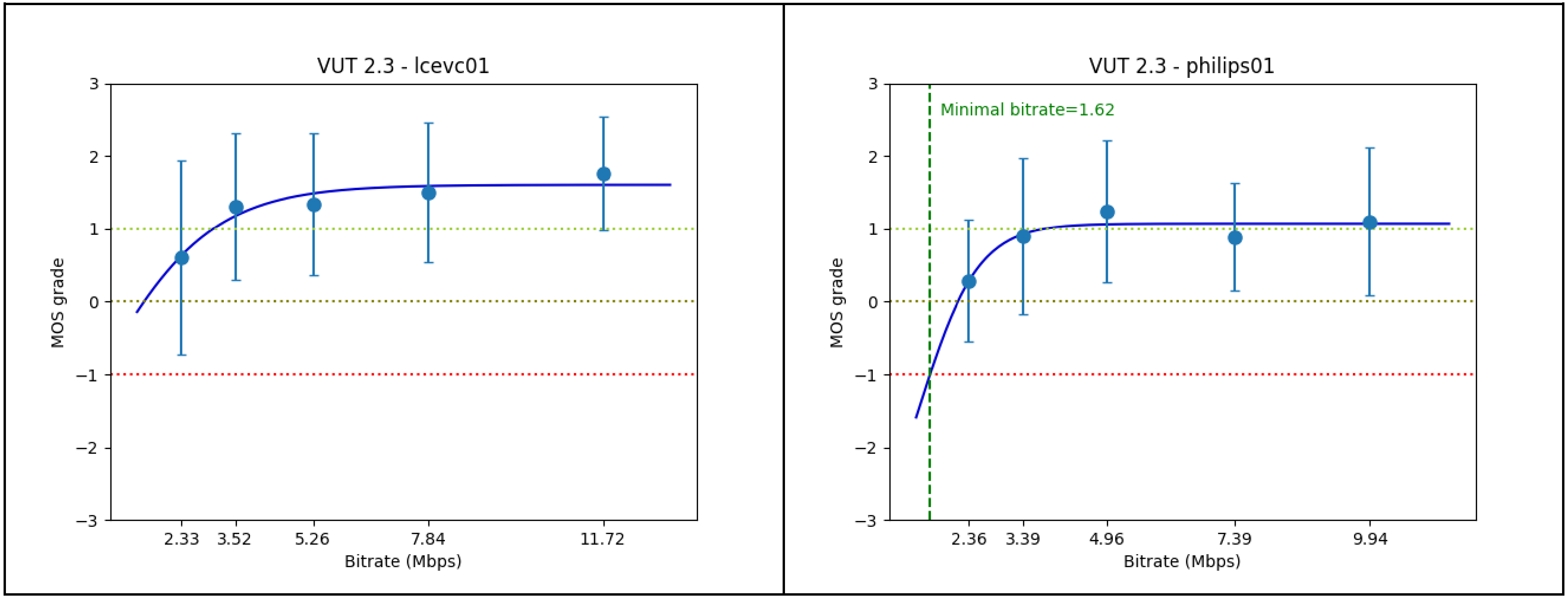
	Reference Video	VUT 2.3
Content label	1 080i	1 080p
Resolution	1 920 x 1 080	1 920 x 1 080
Frame rate	29.97 fps	59.94 fps
Scan	Interlaced	Progressive
Bit depth	8 bits	10 bits
Color gamut	BT.709	BT.2020
HDR Mode	SL-HDR1	HDR10
Codec	H.264/AVC	H.266/VVC
GOP size	60 frames (2 seconds)	360 frames (6 seconds)
Encoder	MainConcept Live Encoder v0.0.0.961	MainConcept Live Encoder v 0.0.0.2340
Encoder type	Real-time	Real-time
Bitrate	14 Mbps	[2.40 3.60 5.40 8.00 12.00] Mbps

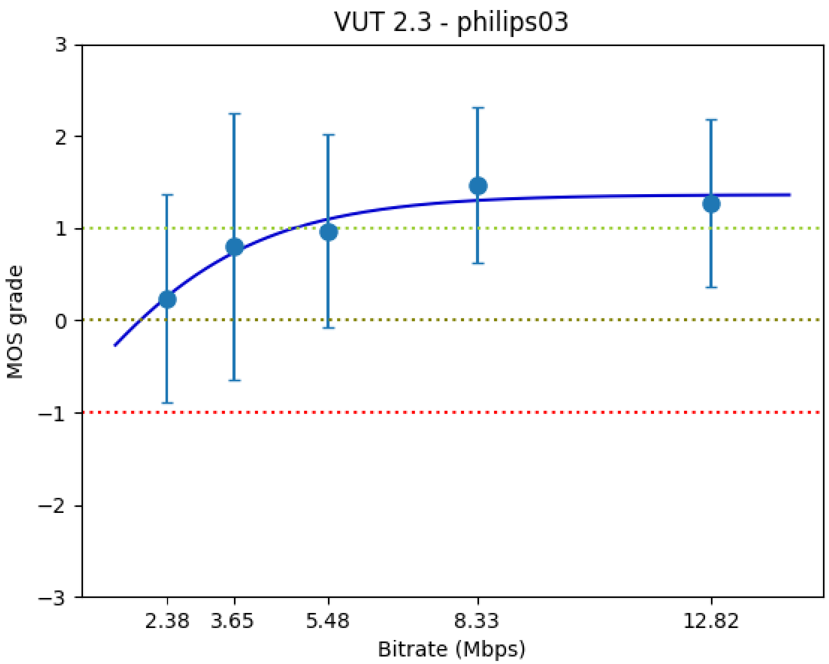
6.3.4.2 VUT 2.3 Experimental Findings

Threshold σ=-1

Table 28 - VUT 2.3 results targeting a MOS grade of -1 (slightly worse)

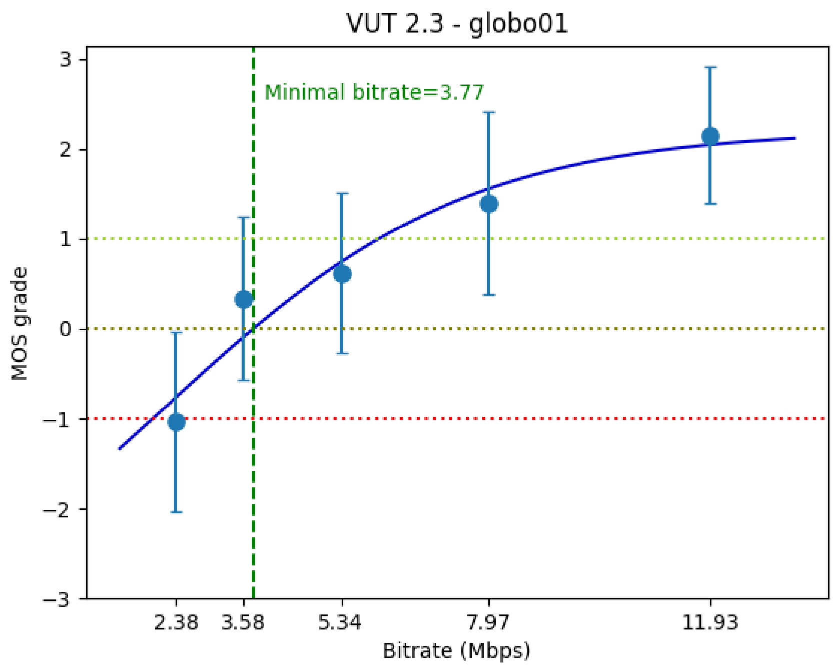
	

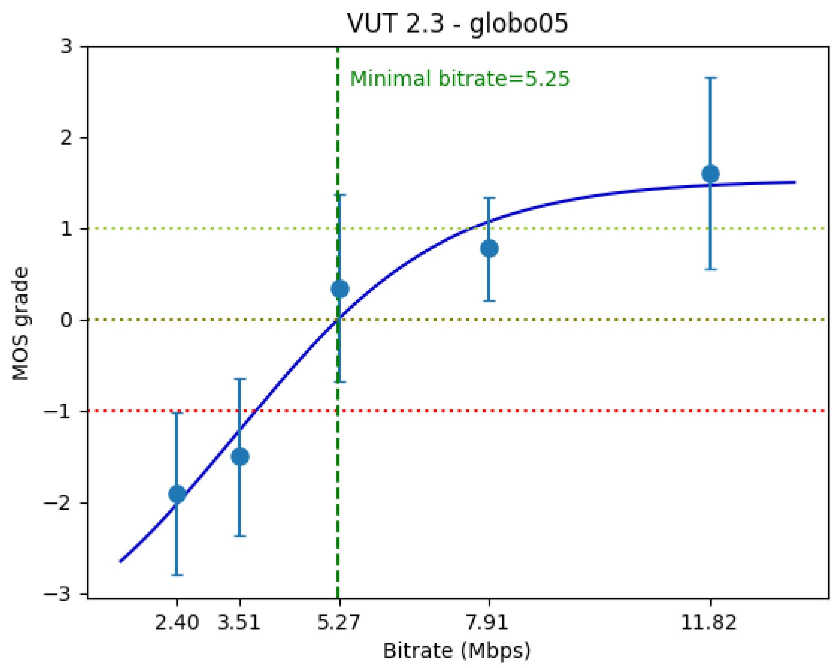

VUT 2.3 - lcevc01
MOS grade
Bitrate (Mbps)
2.33
3.52
5.26
7.84
11.72
3
2
1
0
−1
−2
−3
VUT 2.3 - philips01
Minimal bitrate=1.62
MOS grade
Bitrate (Mbps)
2.36
3.39
4.96
7.39
9.94

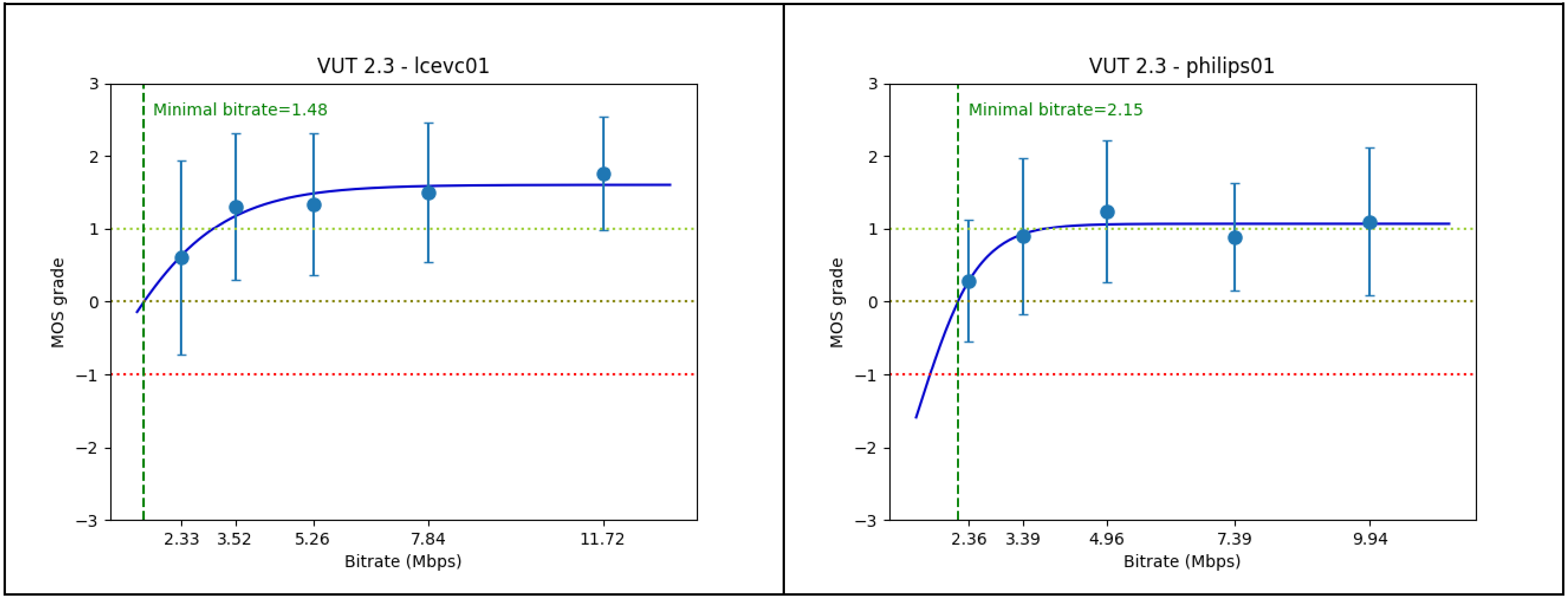

VUT 2.3 - philips03
MOS grade
Bitrate (Mbps)
3
2
1
0
−1
−2
−3
2.38
3.65
5.48
8.33
12.82

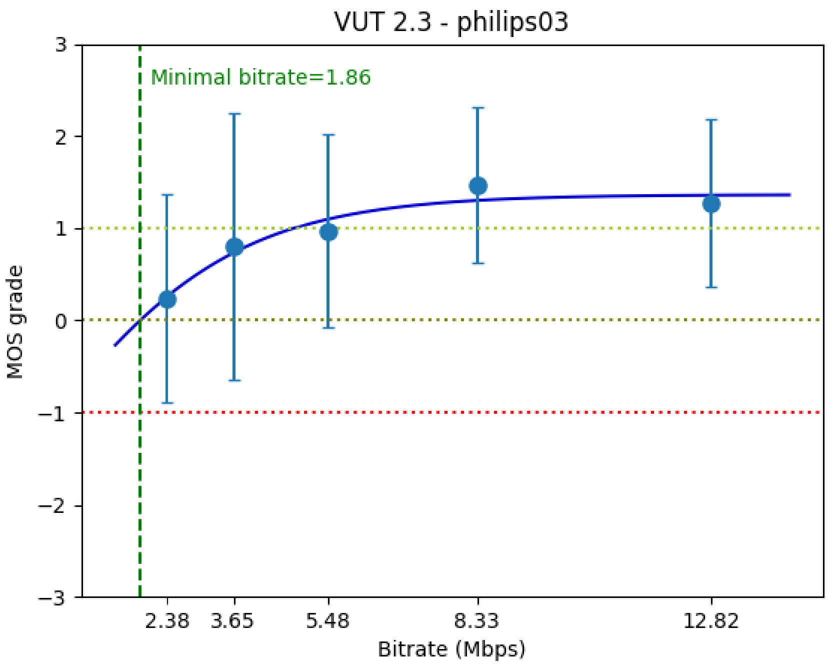
Threshold σ=0

Table 29 - VUT 2.3 results targeting a MOS grade of 0 (same quality)





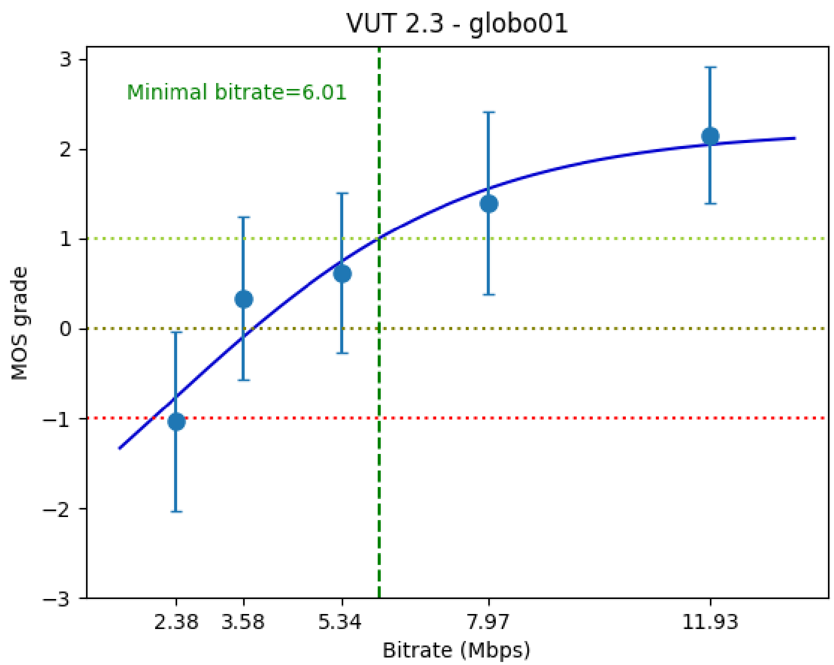

VUT 2.3 - lcevc01
Minimal bitrate=1.48
MOS grade
Bitrate (Mbps)
2.33
3.52
5.26
7.84
11.72
VUT 2.3 - philips01
Minimal bitrate=2.15
MOS grade
Bitrate (Mbps)
2.36
3.39
4.96
7.39
9.94

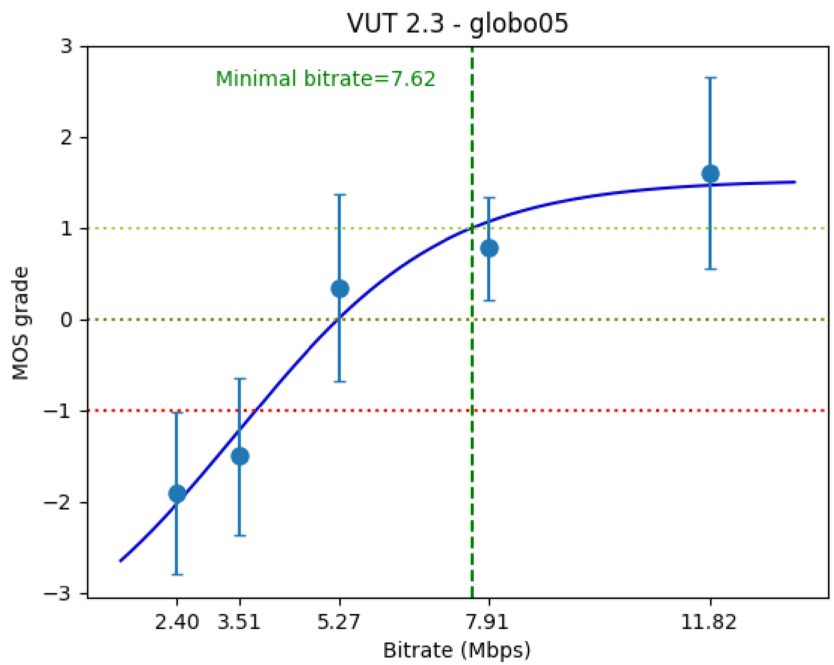


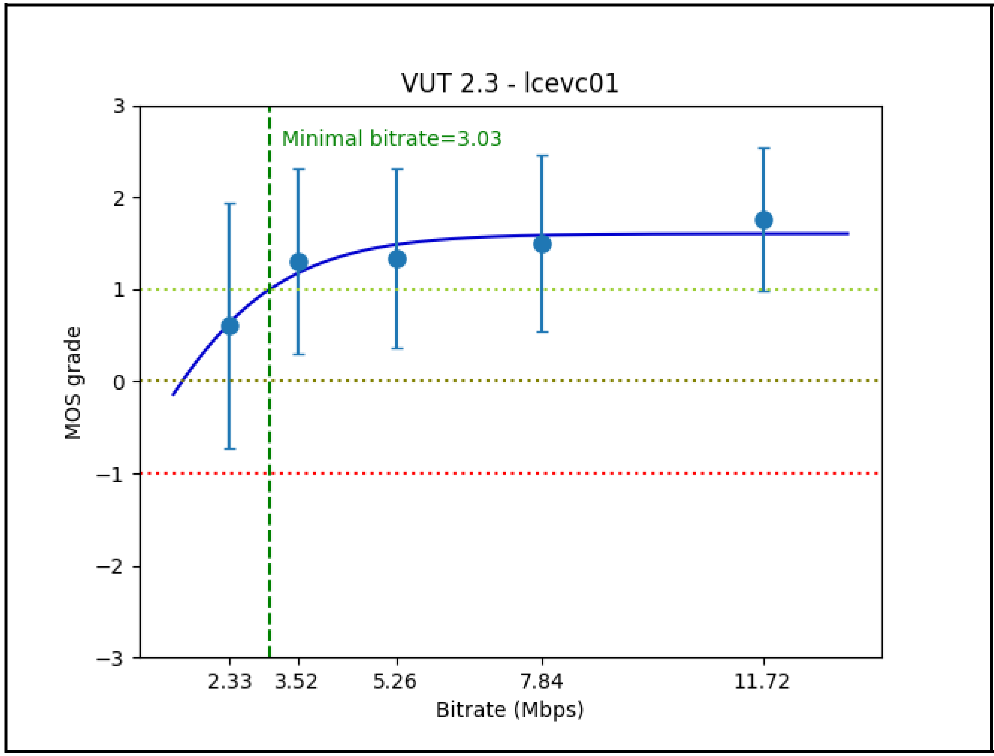
VUT 2.3 - philips03
Minimal bitrate=1.86
MOS grade
Bitrate (Mbps)
2.38
3.65
5.48
8.33
12.82
3
2
1
0
−1
−2
−3

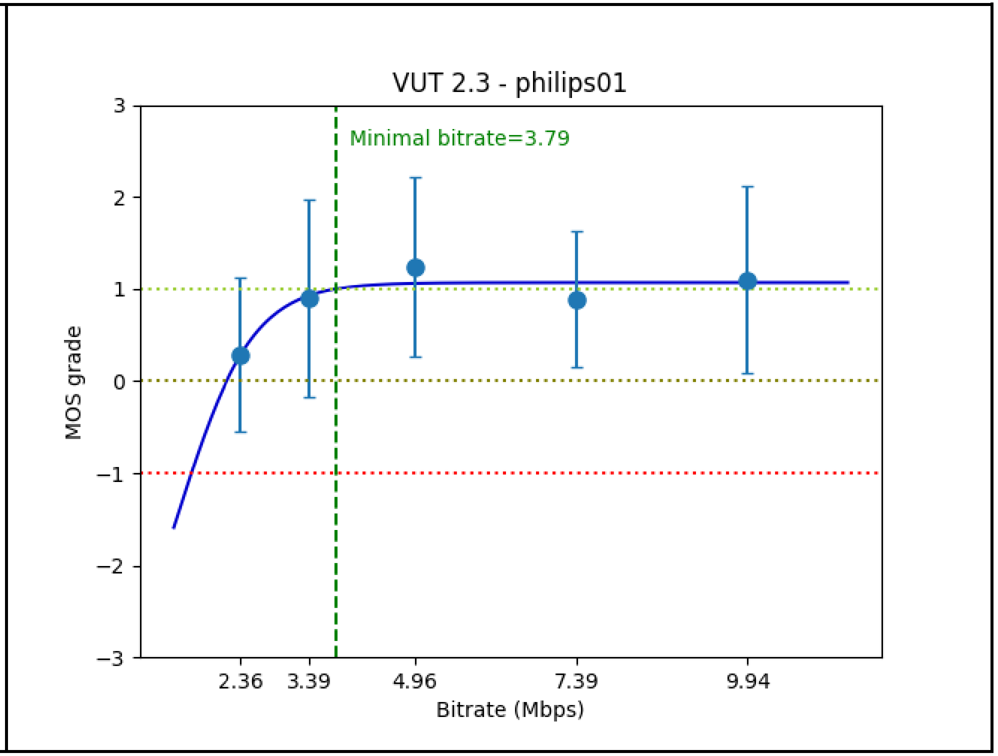
<u>Threshold σ=1</u>

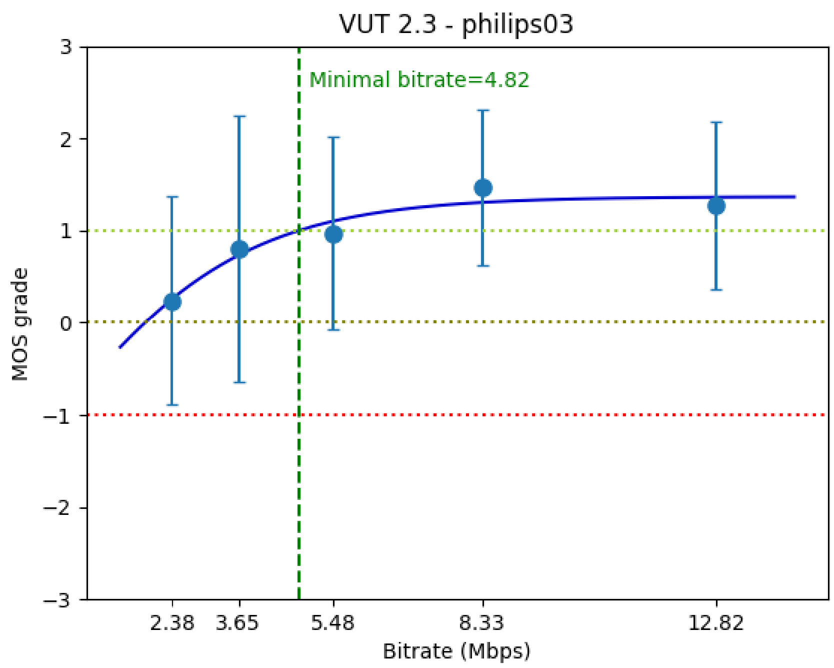
Table 30 - VUT 2.3 results targeting a MOS grade of 1 (slightly better)






VUT 2.3 - lcevc01
Minimal bitrate=3.03
MOS grade
Bitrate (Mbps)
3
2
1
0
−1
−2
−3
2.33
3.52
5.26
7.84
11.72


VUT 2.3 - philips01
Minimal bitrate=3.79
MOS grade
Bitrate (Mbps)
3
2
1
0
−1
−2
−3
2.36
3.39
4.96
7.39
9.94



6.3.4.3 VUT 2.3 Output bitrate analysis

The bitrates for which each target quality is achieved are shown in Table 31.

Table 31 - Achieved Target Bitrates per content (the output target bitrate is highlighted)

Target Quality	globo01	globo05	philips01	philips03	lcevc01
-1	1.97 Mbps	3.79 Mbps	1.62 Mbps	Achieved for all bitrates	Achieved for all bitrates
0	3.77 Mbps	**5.25 Mbps**	2.15 Mbps	1.86 Mbps	1.48 Mbps
1	6.01 Mbps	7.62 Mbps	3.79 Mbps	4.82 Mbps	3.03 Mbps

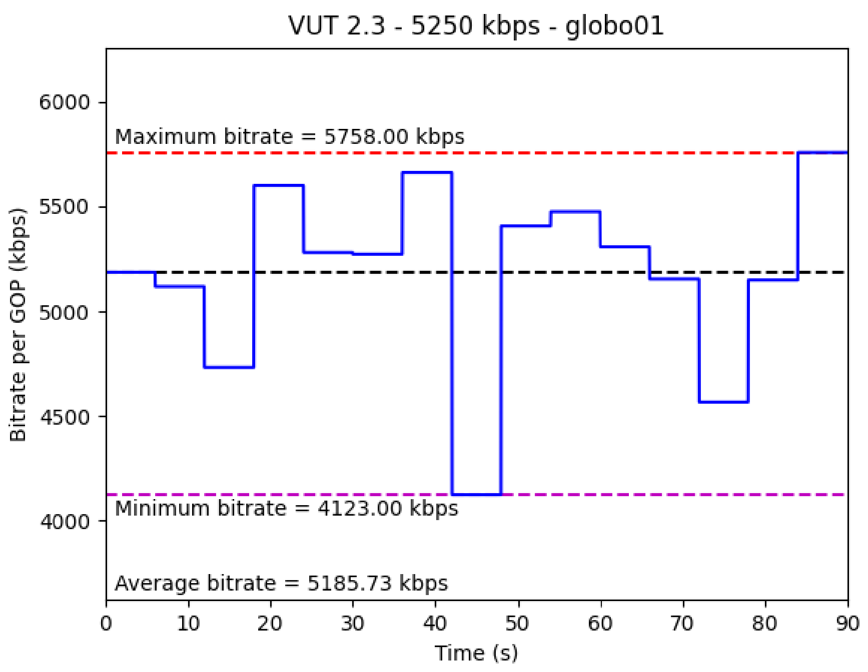
Following the results from the previous sections, the output bitrate for VUT 2.3 was found to be 5.25 Mbps. We have then encoded the stream with the same configuration as Section 6.3.4.1 at 5.25 Mbps and we have carried out the analysis of the bitrate of each GOP.

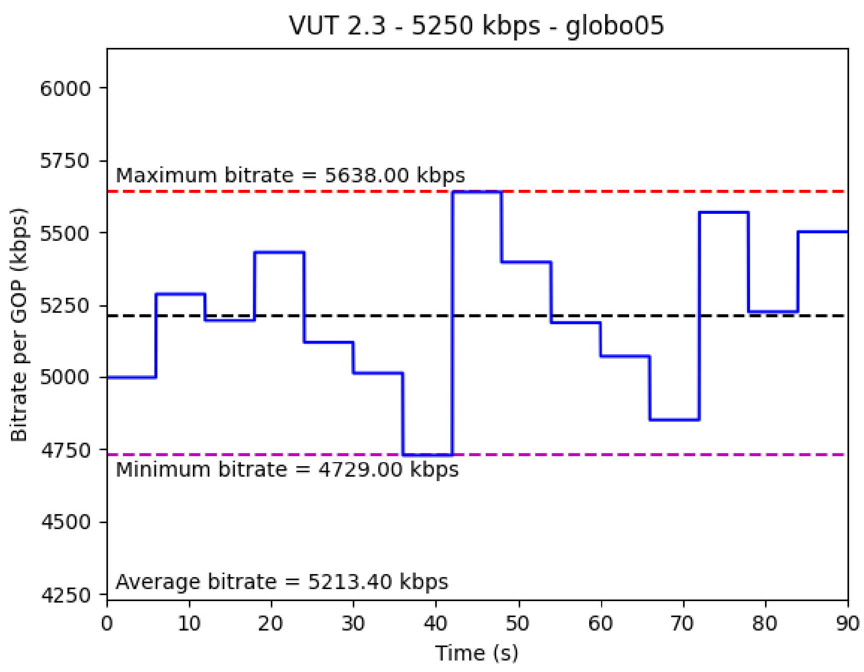
It should be noted that a direct comparison of the VVC encoders used in VUT 1.2 and VUT 2.3 must account for the fact that the two encoders use different configurations. For instance, the encoder used in VUT 2.3 uses a GOP size of 360 frames, while the encoder used in VUT 1.2 uses a GOP size of 120 frames. The coding structure used by the two encoders is also different, as well as the optimization algorithms used.

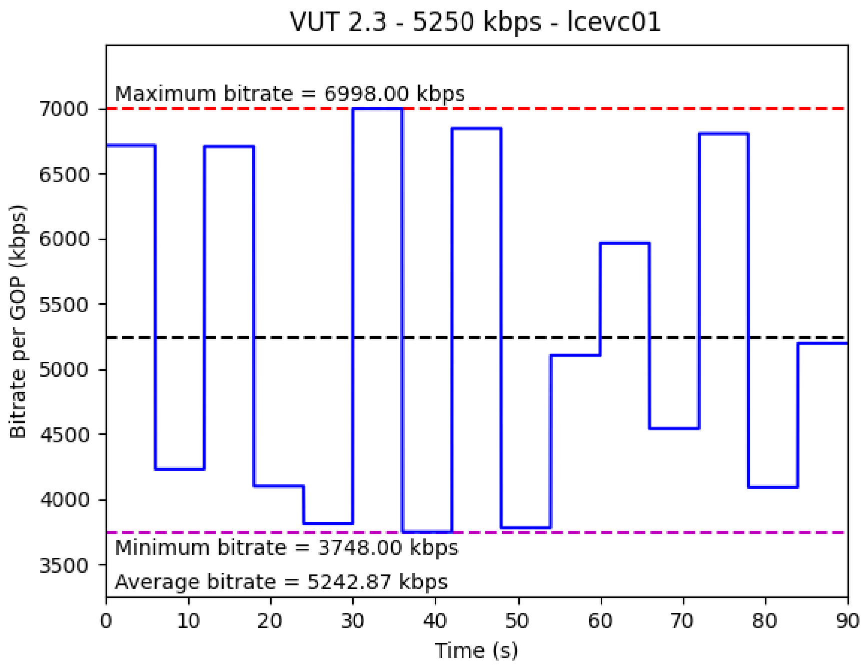
6.3.4.4 VUT 2.3 Bitrate per GOP analysis

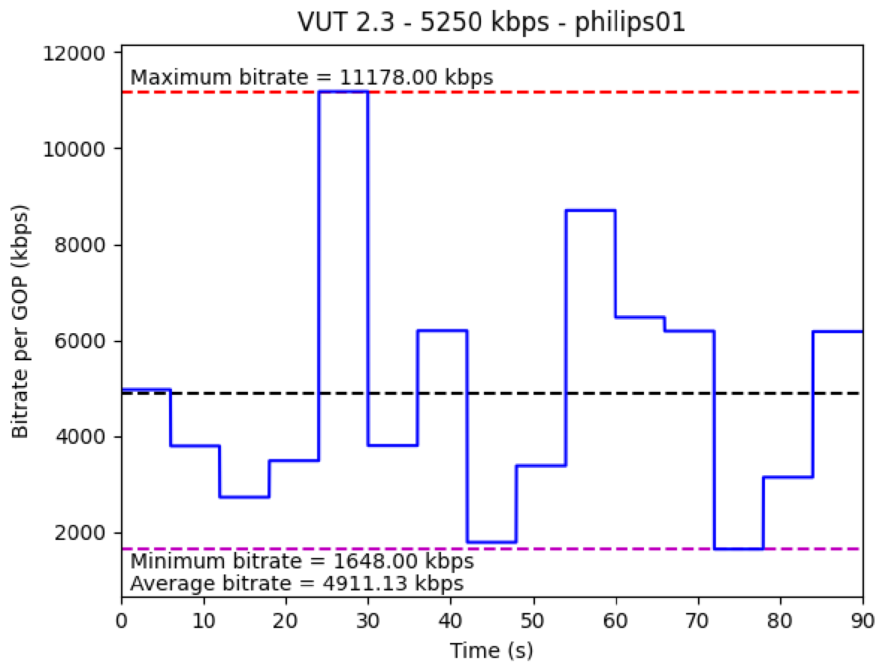
We have carried an analysis of the bitrate per GOP for the stream at 5.25 Mbps described above. The main results are shown in Table 32.

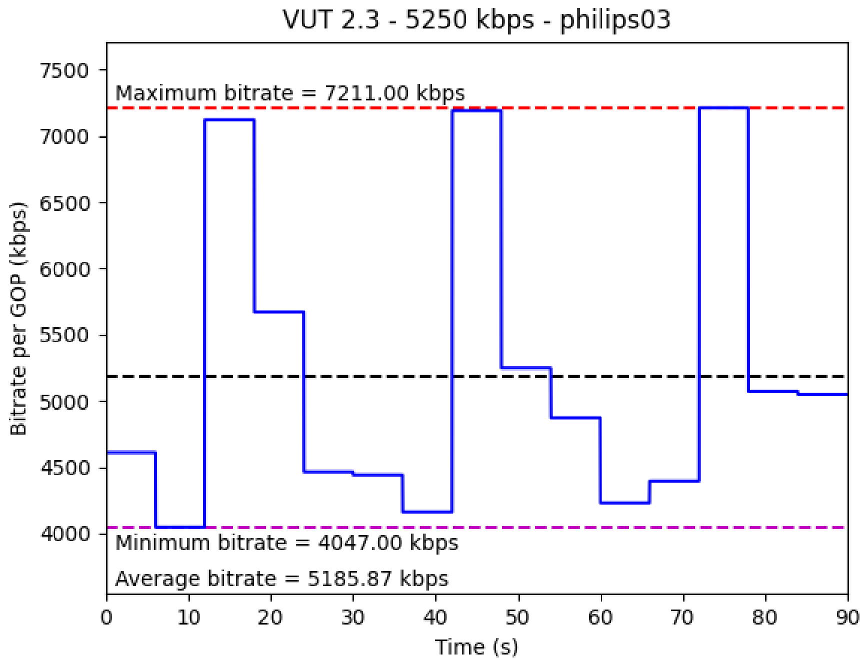
Table 32 - VUT 2.3 Bitrate per GOP when encoded with the output bitrate of 5.25 Mbps






VUT 2.3 - 5250 kbps - lcevc01
Maximum bitrate = 6998.00 kbps
Minimum bitrate = 3748.00 kbps
Average bitrate = 5242.87 kbps
Bitrate per GOP (kbps)
Time (s)


VUT 2.3 - 5250 kbps - philips01
Maximum bitrate = 11178.00 kbps
Minimum bitrate = 1648.00 kbps
Average bitrate = 4911.13 kbps
Bitrate per GOP (kbps)
Time (s)



The average, minimum and maximum bitrates per GOP are shown in Table 33. All rates are given in kbps.

Table 33 - VUT 2.3 Average, Minimum and Maximum bitrates per GOP when encoded with the output bitrate of 5.25 Mbps

Content	TARGET Rate	Average Rate	Minimum Rate	Maximum Rate
globo01	5250	5185.73	4123	5758
globo05	5250	5213.40	4729	5638
lcevc01	5250	5242.87	3748	6998
philips01	5250	4911.13	1648	11178
philips03	5250	5185.87	4047	7211

As can be seen from the table, the average bitrate per GOP is very close to the target bitrate.

6.3.4.5 VUT 2.3 Analysis and Conclusions

As depicted in Table 31, the output target bitrate for VUT 2.3 for a target quality of -1 (“slightly worse”) was found to be 3.79 Mbps. This result is based on the content "globo05", which demanded the highest bitrate. Noticeably, no values less than -1 were achieved for the "philips03" and "lcevc" sequences, i.e., all target qualities were achieved for all bitrates used in the VUT 2.3 experiments. In a similar manner, the output target bitrate for a target quality of 0 (“same quality”) and 1 (“slightly higher”) were both achieved with the "globo05" video sequence, corresponding to 5.25 Mbps and 7.62 Mbps, respectively. Since VUT 2.3 was encoded with the same resolution as the reference video, we considered the output target bitrate as the highest rate achieved with “the same quality” (0). After coding the contents using this bitrate (i.e., 5.25 Mbps), we observed an average bitrate around 5.2 Mbps. One particular pattern in VUT 2.3 experimental data is the higher variance in the average rate when compared to VUT 1.1, 1.2, and 1.3.

VUT 2.3 achieves a target quality of 0 ("same quality") at a lower bitrate (5.25 Mbps) than VUT 1.2 (7.52 Mbps). It should be noted that a direct comparison of the VVC encoders used in VUT 1.2 and VUT 2.3 must account for the fact that the two encoders use different configurations. For instance, the encoder used in VUT 2.3 uses a GOP size of 360 frames, while the encoder used in VUT 1.2 uses a GOP size of 120 frames. The coding structure used by the two encoders is also different, as well as the optimization algorithms used. In addition, VUT 2.3 coding allowed for much greater bitrate variability. However, real-time video coding applications for over-the-air broadcasting usually require smaller GOP sizes and less bitrate variability.

6.3.5 VUT 2.4 VVC+LCEVC 2 160p

6.3.5.1 VUT 2.4 Definition

The goal of this VUT was to test the VVC+LCEVC encoder working at 2 160p resolution. Following the results of the VUT 1.2 on Section 6.3.2, the reference video for this VUT was encoded with the VVC at 1 080p resolution, at 7.52 Mbps. The complete details are found in Table 34. Since the VUT was encoded with a higher resolution than the reference video, the quality target considered was “the same” (0). Note that using the highest bitrate corresponding to the target score 0 ("the same") among the five clips in the test material means that this VUT would provide a similar subjective quality for the clip with the highest required bitrate, and somewhat higher score in the other clips while not necessarily reaching the score 1 ("slightly better") in all clips.

Table 34 - VUT 2.4 encoding details

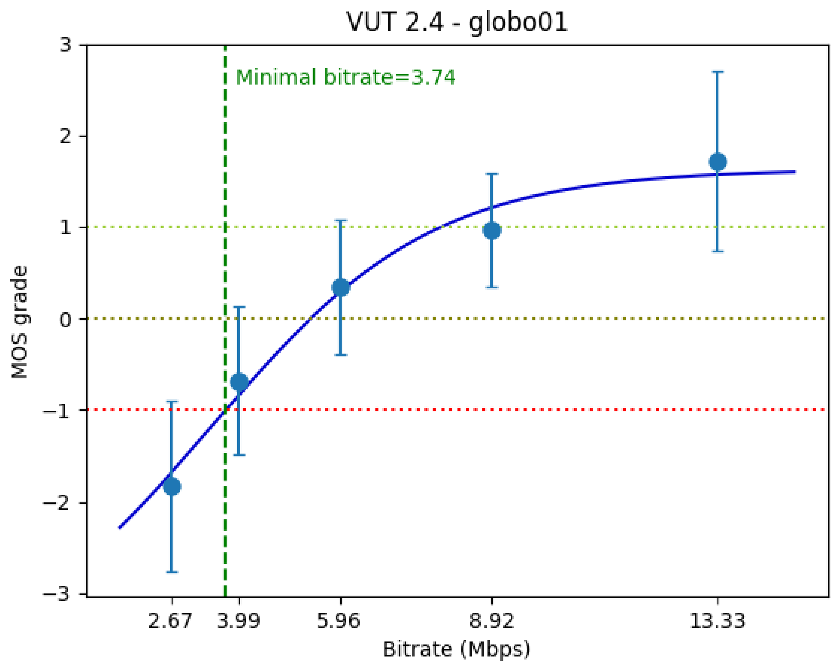
	Reference Video	VUT 2.4
Content label	1 080p	2 160p
Resolution	1 920 x 1 080	3 840 x 2 160
Frame rate	59.94 fps	59.94 fps
Scan	Progressive	Progressive
Bit depth	10 bits	10 bits
Color gamut	BT.2020	BT.2020
HDR Mode	HDR10	HDR10
Codec	H.266/VVC	H.266/VVC + LCEVC
GOP size	120 frames (2 seconds)	360 frames (6 seconds)
Encoder	Ateme TitanLive Innovation v 4.1.31.911	MainConcept Live Encoder v 0.0.0.2340
Encoder type	Real-time	Real-time
Bitrate	7.52 Mbps	[2.67 3.99 5.96 8.92 13.33] Mbps[3]

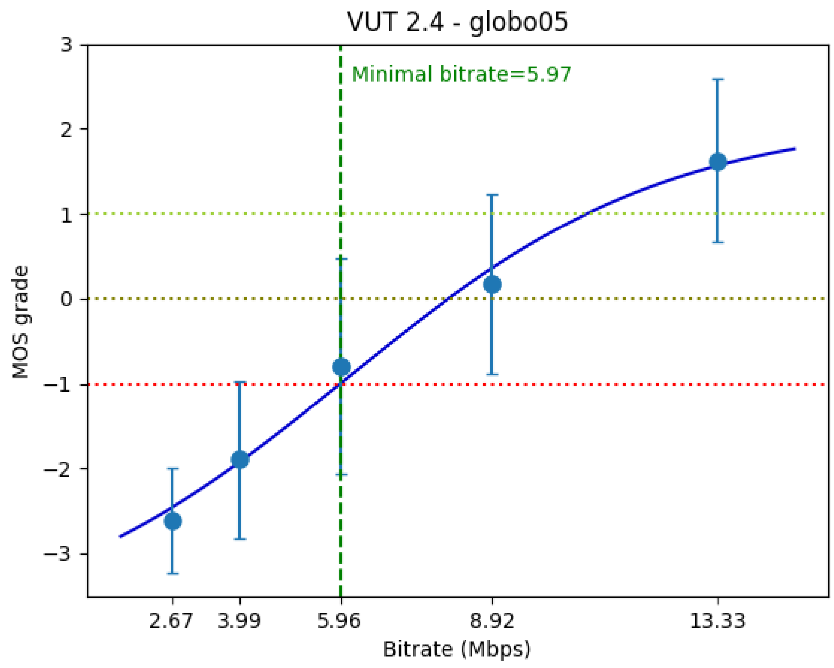
[3] Note that the same encodings were used for VUTs 2.3 and 2.4, and only the decoding was carried out differently (decoding only the VVC base layer in VUT 2.3 and decoding the VVC base layer with the LCEVC enhancement layer in VUT 2.4). In the first case, the bitrates shown in Table 27 reflect the average rates assigned only to the VVC base layer. In the second case, the bitrates shown in Table 34 reflect the total average rates used by the combination of the VVC base layer and the LCEVC enhancement layer.

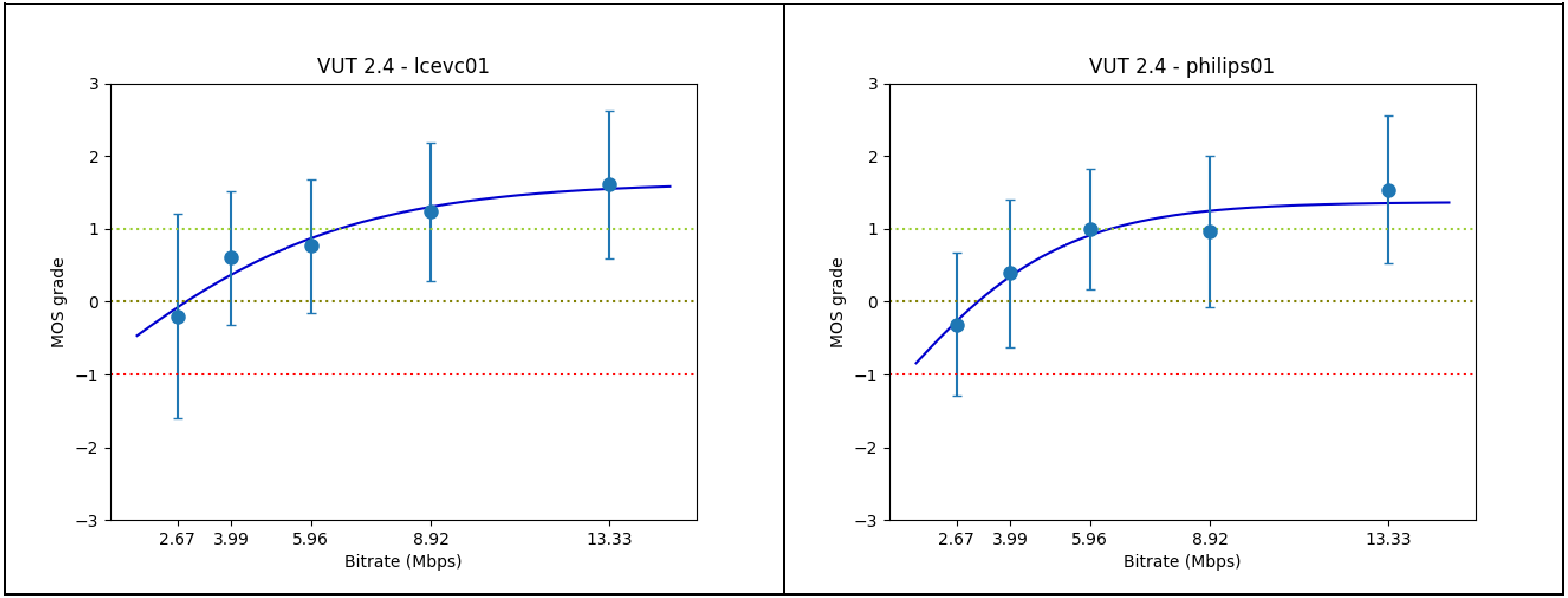
6.3.5.2 VUT 2.4 Experimental Findings

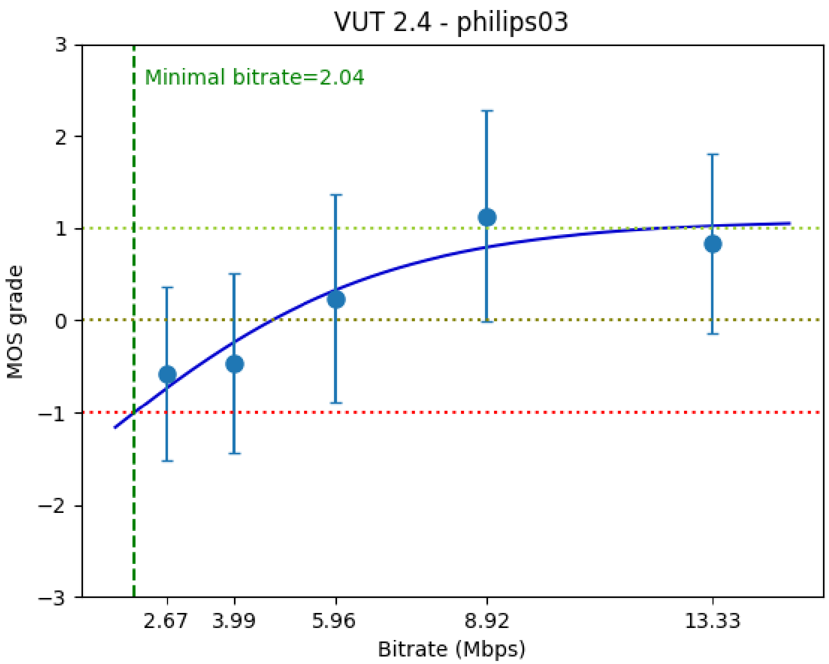
Threshold σ=-1

Table 35 - VUT 2.4 results targeting a MOS grade of -1 (slightly worse)



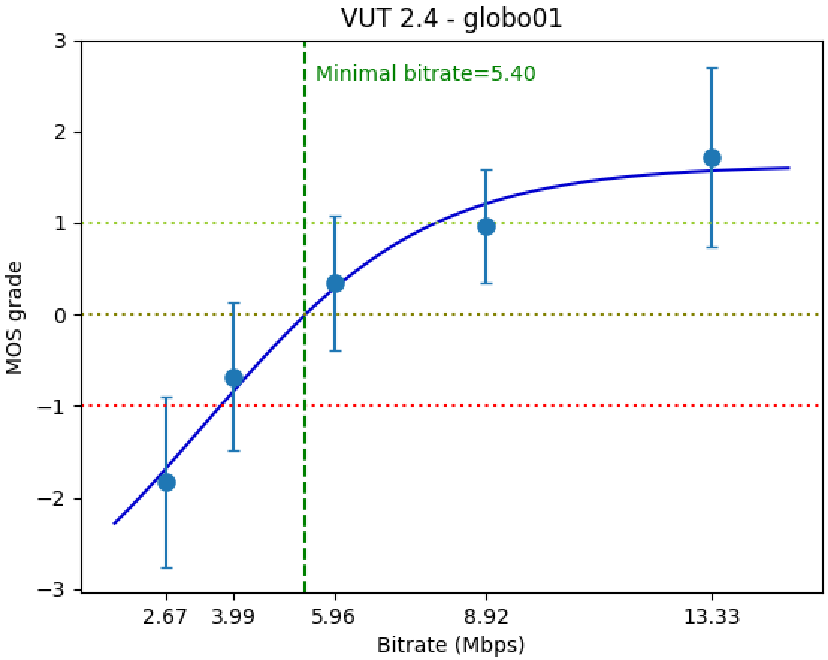


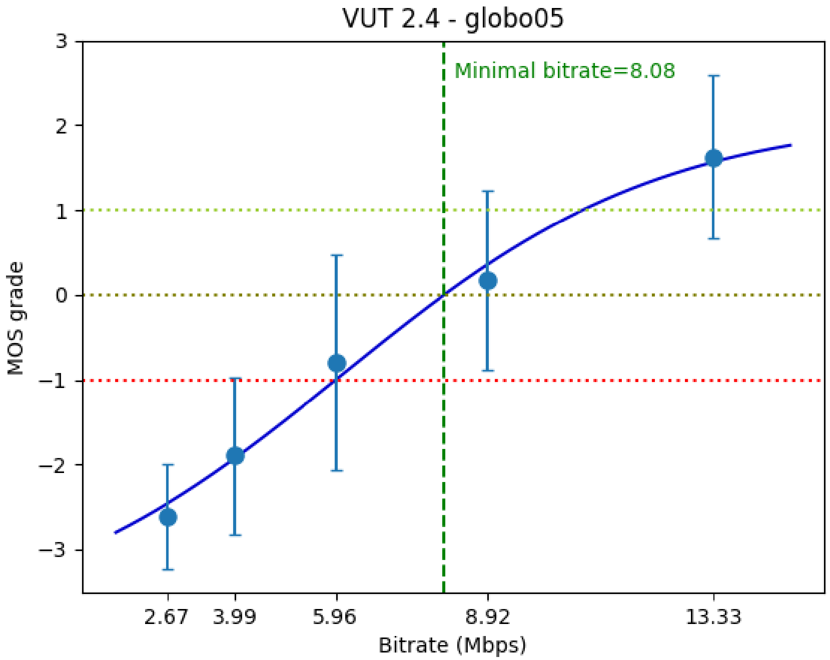

VUT 2.4 - lcevc01
MOS grade
Bitrate (Mbps)
2.67
3.99
5.96
8.92
13.33
3
2
1
0
−1
−2
−3
VUT 2.4 - philips01
MOS grade
Bitrate (Mbps)

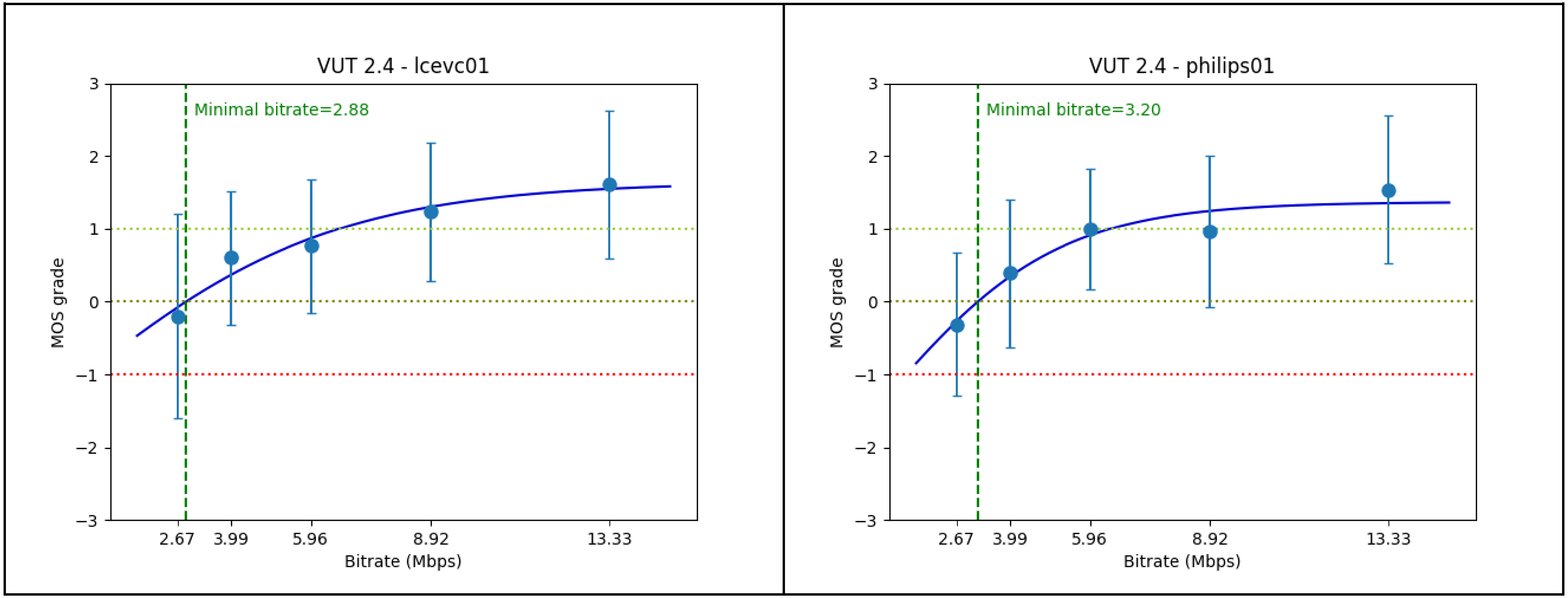

VUT 2.4 - philips03
Minimal bitrate=2.04
MOS grade
3
2
1
0
−1
−2
−3
2.67
3.99
5.96
8.92
13.33
Bitrate (Mbps)

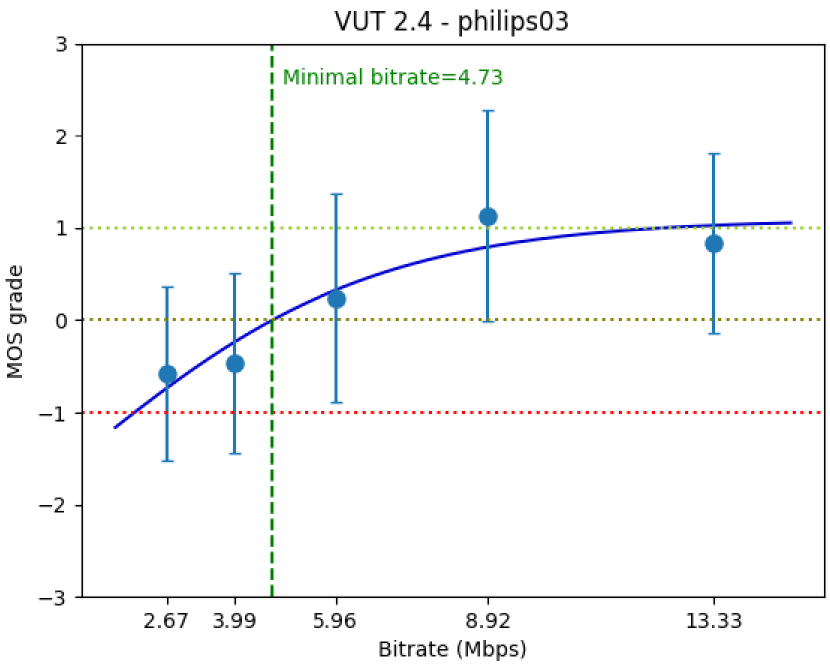
Threshold σ=0

Table 36 - VUT 2.4 results targeting a MOS grade of 0 (same quality)



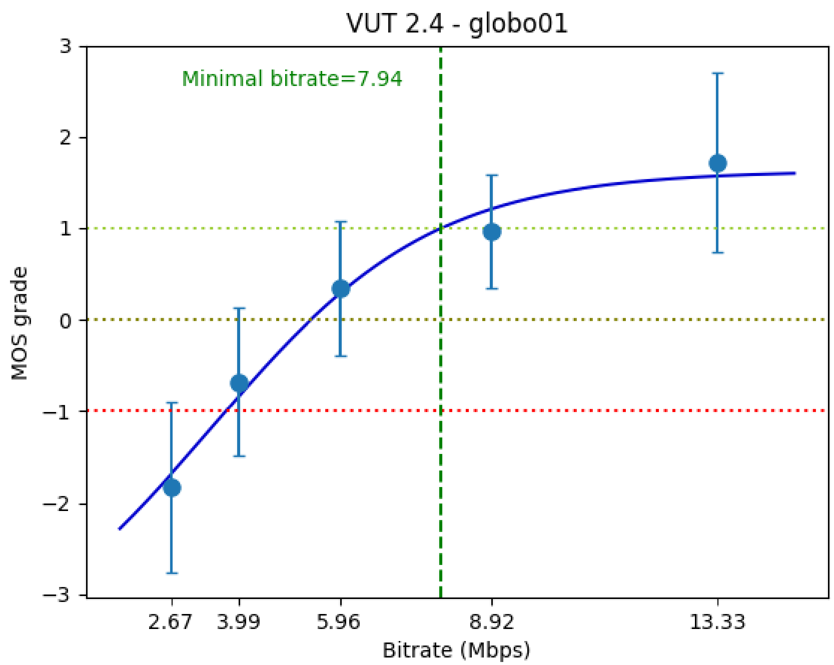


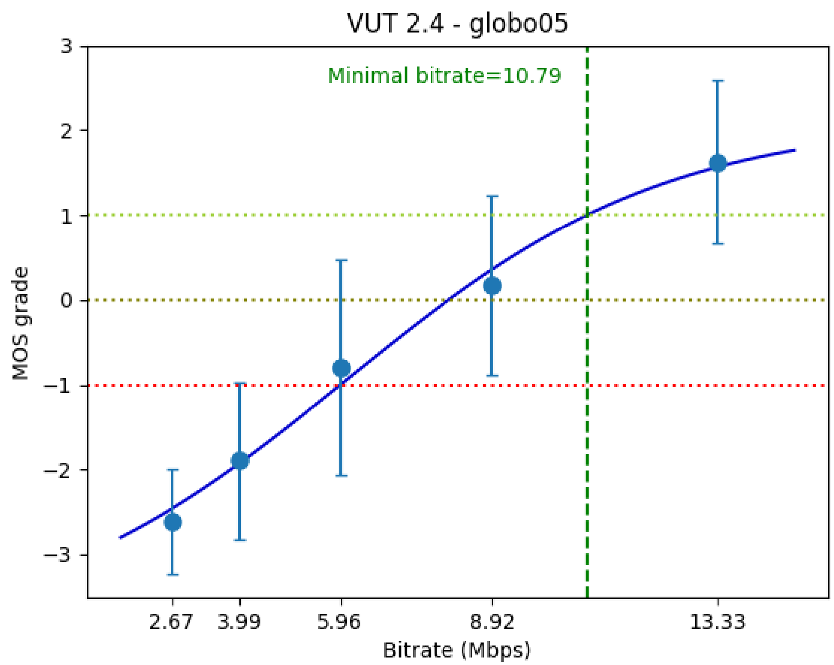

VUT 2.4 - lcevc01
Minimal bitrate=2.88
MOS grade
Bitrate (Mbps)
2.67
3.99
5.96
8.92
13.33
VUT 2.4 - philips01
Minimal bitrate=3.20
MOS grade
Bitrate (Mbps)
2.67
3.99
5.96
8.92
13.33

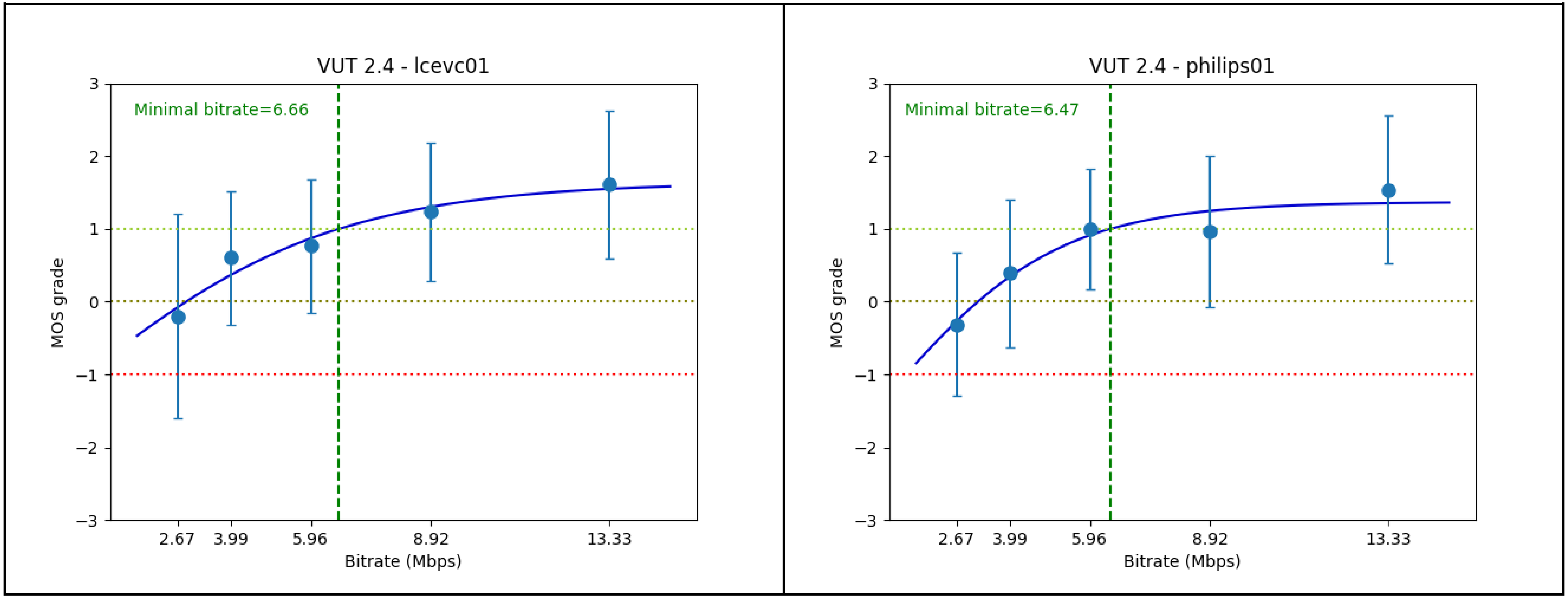

VUT 2.4 - philips03
Minimal bitrate=4.73
MOS grade
Bitrate (Mbps)
3
2
1
0
−1
−2
−3
2.67
3.99
5.96
8.92
13.33

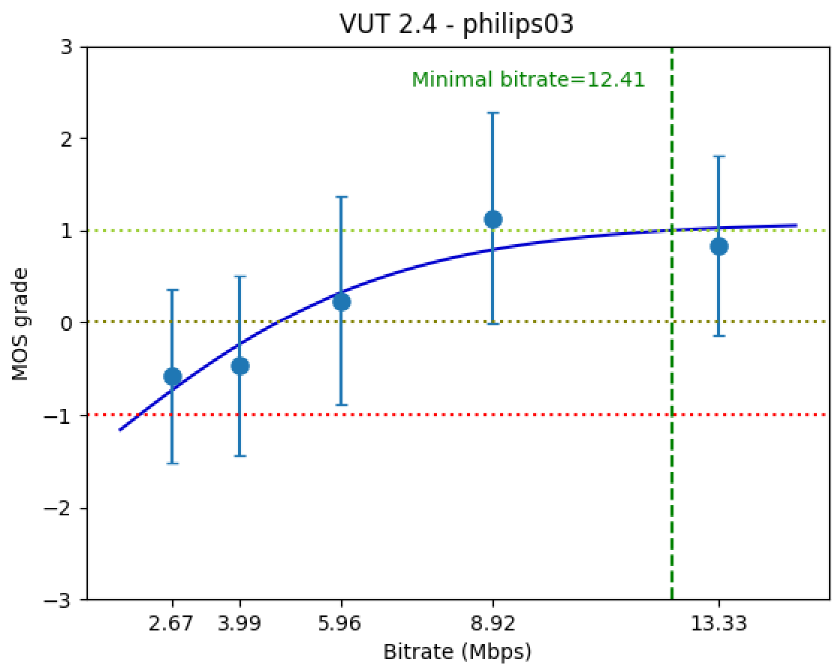
Threshold σ=1

Table 37 - VUT 2.4 results targeting a MOS grade of 1 (slightly better)






VUT 2.4 - lcevc01
Minimal bitrate=6.66
MOS grade
Bitrate (Mbps)
2.67
3.99
5.96
8.92
13.33
3
2
1
0
−1
−2
−3
VUT 2.4 - philips01
Minimal bitrate=6.47
MOS grade
Bitrate (Mbps)
2.67
3.99
5.96
8.92
13.33
3
2
1
0
−1
−2
−3



6.3.5.3 VUT 2.4 Output bitrate analysis

The bitrates for which each target quality is achieved are shown in Table 38.

Table 38 - Achieved Target Bitrates per content (the output target bitrate is highlighted)

Target Quality	globo01	globo05	philips01	philips03	lcevc01
-1	3.74 Mbps	5.97 Mbps	Achieved for all bitrates	2.04 Mbps	Achieved for all bitrates
0	5.40 Mbps	**8.08 Mbps**	3.20 Mbps	4.73 Mbps	2.88 Mbps
1	7.94 Mbps	10.79 Mbps	6.47 Mbps	12.41 Mbps	6.66 Mbps

Following the results from the previous sections, the output bitrate for VUT 2.4 was found to be 8.08 Mbps. We have then encoded the stream with the same configuration as Section 6.3.5.1 at 8.08 Mbps and we have carried out the analysis of the bitrate of each GOP.

It should be noted that a direct comparison of the VVC encoders used in VUT 1.3 and VUT 2.4 must account for the fact that the two encoders use different configurations. For instance, the encoder used in VUT 2.4 uses a GOP size of 360 frames, while the encoder used in VUT 1.3 uses a GOP size of 120 frames. The coding structure used by the two encoders is also different, as well as the optimization algorithms used.

For a target output of 8.08 Mbps, we have carried out an analysis on the bitrates used by the base-layer and the enhancement layer for the VVC+LCEVC codec. These results are shown in Table 39. It can be seen that, in this configuration, the average bitrate used for the base layer was between 85.83% (7.02 Mbps, for lcevc01 sequence) and 91.12% (7.22 Mbps, for globo05 sequence). As the base layer uses the same encoder, in the same configuration, as VUT 2.3, but with an average bitrate significantly higher than the output bitrate for VUT 2.3, it is expected that the average quality of the base layer of VUT 2.4 encoded at the output bitrate of 8.08 Mbps to be higher than the average quality of VUT 2.3 encoded at the output bitrate of 5.25 Mbps. All quality results for VUT 2.4 reported in this document refer to the full VVC+LCEVC decoding.

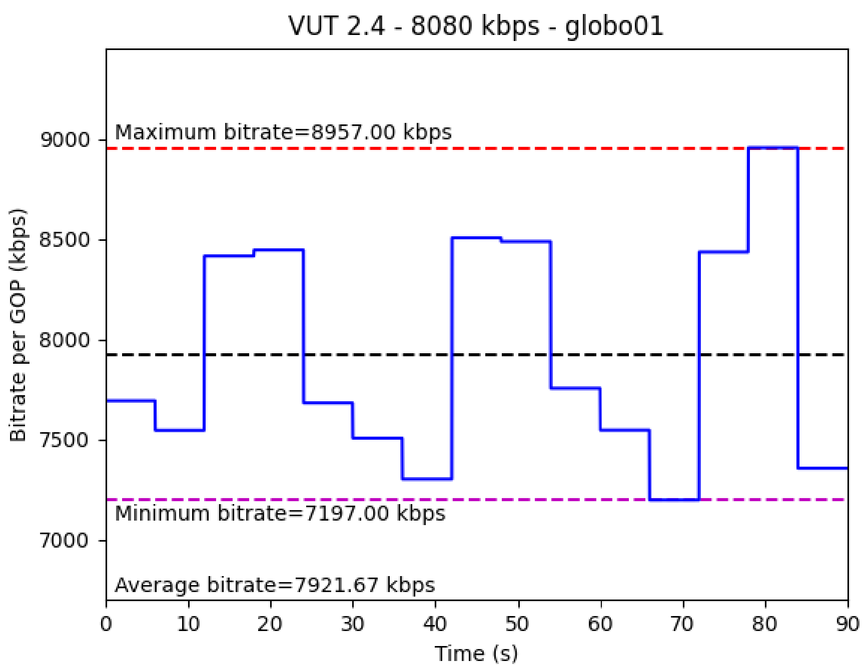
Table 39 - Bitrates (in kbps) used in each layer for VUT 2.4 encoded at 8.08 Mbps

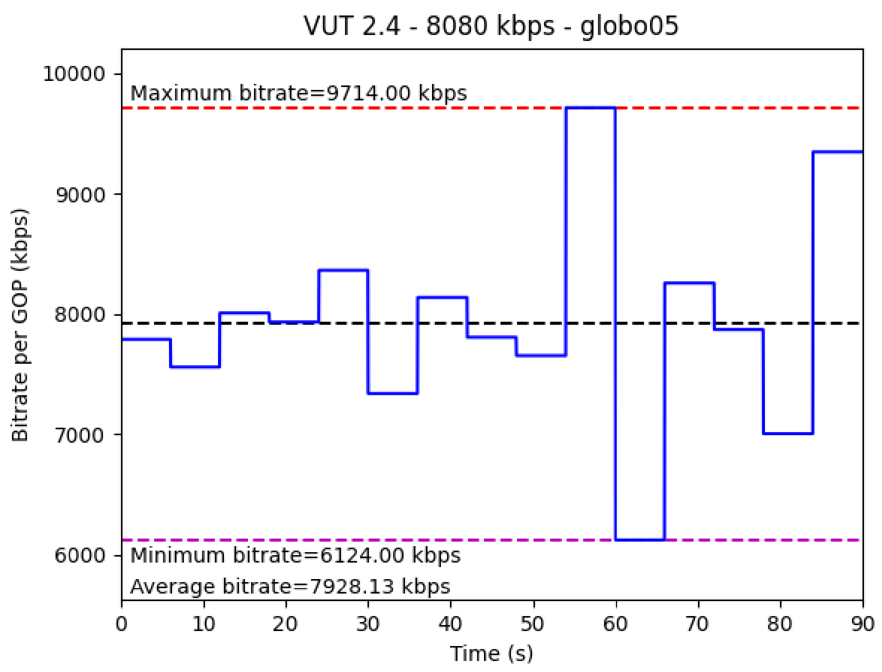
	VVC Base Layer	LCEVC Enhancement Layer	Total Bitrate (VVC+LCEVC)
globo01	7190.67 (90.77%)	731.00 (9.23%)	7921.67 (100%)
globo05	7223.93 (91.12%)	704.20 (8.88%)	7928.13 (100%)
lcevc01	7023.53 (85.83%)	1154.94 (14.17%)	8182.67 (100%)
philips01	7001.53 (90.26%)	755.87 (9.74%)	7757.40 (100%)
philips03	7166.13 (90.64%)	739.87 (9.36%)	7906.00 (100%)

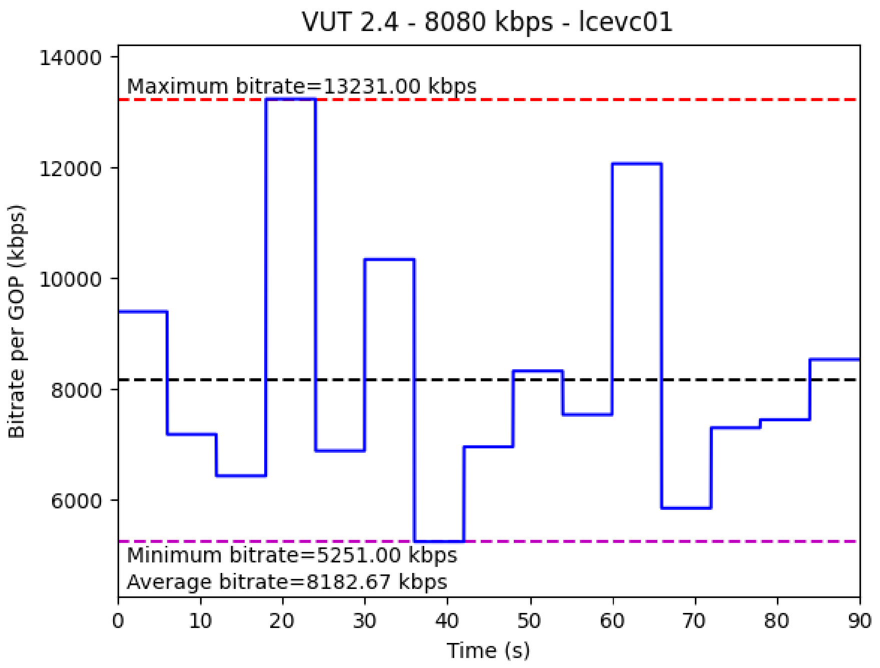
6.3.5.4 VUT 2.4 Bitrate per GOP analysis

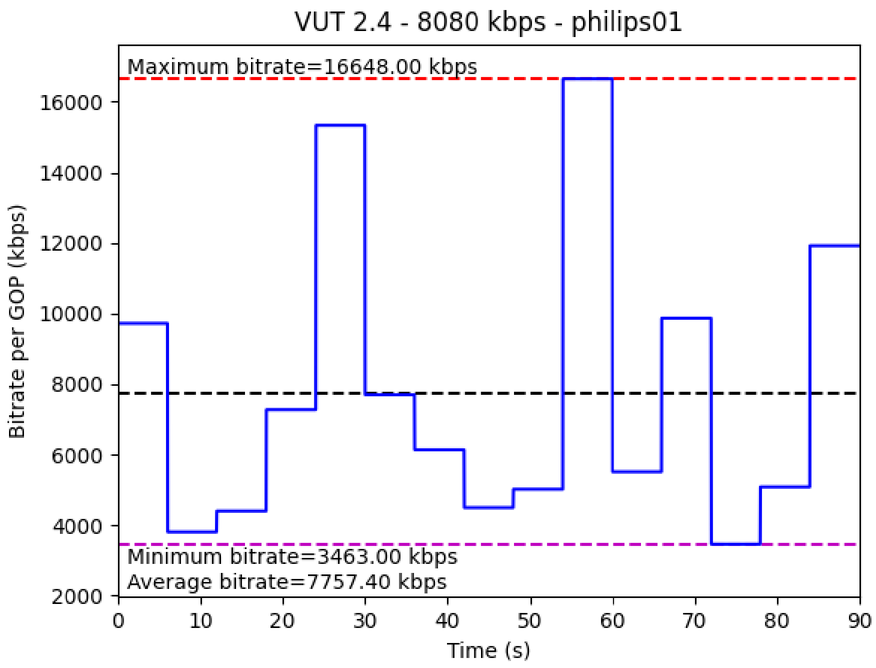
We have carried an analysis of the bitrate per GOP for the stream at 8.08 Mbps described above. The main results are shown in Table 40.

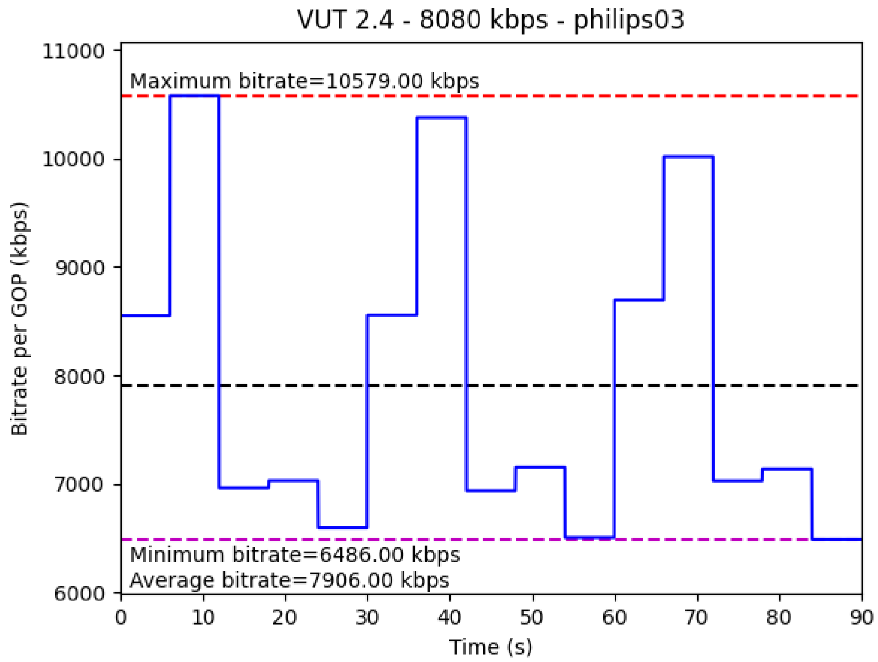
Table 40 - VUT 2.4 Bitrate per GOP when encoded with the output bitrate of 8.08 Mbps






VUT 2.4 - 8080 kbps - lcevc01
Maximum bitrate=13231.00 kbps
Minimum bitrate=5251.00 kbps
Average bitrate=8182.67 kbps
Bitrate per GOP (kbps)
Time (s)


VUT 2.4 - 8080 kbps - philips01
Maximum bitrate=16648.00 kbps
Minimum bitrate=3463.00 kbps
Average bitrate=7757.40 kbps
Bitrate per GOP (kbps)
Time (s)



The average, minimum and maximum bitrates per GOP are shown in Table 41. All rates are given in kbps.

Table 41 - VUT 2.4 Average, Minimum and Maximum bitrates per GOP when encoded with the output bitrate of 8.08 Mbps

Content	TARGET Rate	Average Rate	Minimum Rate	Maximum Rate
globo01	8080	7921.67	7197	8957
globo05	8080	7928.13	6124	9714
lcevc01	8080	8182.67	5251	13231
philips01	8080	7757.40	3463	16648
philips03	8080	7906.00	6486	10579

As can be seen from the table, the average bitrate per GOP is very close to the target bitrate.

6.3.5.5 VUT 2.4 Analysis and Conclusions

As depicted in Table 38, the output target bitrate for VUT 2.4 for a target quality of -1 ("slightly worse") and 0 ("same") were found to be 5.97 Mbps and 8.08 Mbps, respectively. These results correspond to the "globo05" video sequence. The target quality of 1 (“slightly higher”) was achieved at 12.41 Mbps with the "philips03" video sequence. Even though VUT 2.4 was encoded with a higher resolution than the reference video, we considered as output target bitrate the highest rate achieved with “same quality” (0). Note that using the highest bitrate corresponding to the target score 0 ("the same") among the five clips in the test material means that this VUT would provide a similar subjective quality for the clip with the highest required bitrate, and somewhat higher score in the other clips while not necessarily reaching the score 1 ("slightly better") in all clips.

VUT 2.4 achieves a target quality of 0 ("same quality") at a lower bitrate (8.08 Mbps) than VUT 1.3 (12.54 Mbps). It should be noted that a direct comparison of the VVC encoders used in VUT 1.3 and VUT 2.4 must account for the fact that the two encoders use different configurations. For instance, the encoder used in VUT 2.4 uses a GOP size of 360 frames, while the encoder used in VUT 1.3 uses a GOP size of 120 frames. The coding structure used by the two encoders is also different, as well as the optimization algorithms used. In addition, VUT 2.4 coding allowed for much greater bitrate variability. However, real-time video coding applications for over-the-air broadcasting usually require smaller GOP sizes and less bitrate variability.

6.4 Summary of the Results

The results of the previous sections are summarized in Table 42. The main output for each VUT is the minimum bitrate for which that encoding configuration achieves that particular quality for all 5 contents tested. Not all VUTs are able to achieve some of the target grades due to the resolution of the sequences tested and the capabilities of the encoders used. The target grade intended is highlighted in the table. Also, it is important to notice that VUTs 1.1, 1.2, and 2.3 use 1 080i H.264/AVC encoded references, whereas VUTs 1.3 and 2.4 use a 1 080p VVC encoded reference, and thus these tests are not directly comparable.

It should be noted that a direct comparison of the encoders used in VUTs 1.2/1.3 and VUTs 2.3/2.4 must account for the fact that they use different configurations. For instance, the encoder used in VUTs 2.3/2.4 uses a GOP size of 360 frames, while the encoder used in VUTs 1.2/1.3 uses a GOP size of 120 frames. The coding structure used by the encoders is also different, as well as the optimization algorithms used. In addition, VUTs 2.3/2.4 coding allowed for much greater bitrate variability. However, real-time video coding applications for over-the-air broadcasting usually require smaller GOP sizes and less bitrate variability.

Table 42 - Summary of the results

VUT	Encoder	Reference Video	Target Grade -1	Target Grade 0	Target Grade 1
VUT 1.1	VVC 720p	1 080i H.264 + SL-HDR1 at 14 Mbps	**4.30 Mbps**	Not achieved	Not achieved
VUT 1.2	VVC 1 080p	1 080i H.264 + SL-HDR1 at 14 Mbps	5.13 Mbps	**7.52 Mbps**	Not achieved
VUT 1.3	VVC 2 160p	1 080p VVC + HDR10 at 7.52 Mbps	8.13 Mbps	**12.54 Mbps**	Not achieved
VUT 2.3	VVC Base Layer 1 080p	1 080i H.264 + SL-HDR1 at 14 Mbps	3.79 Mbps	**5.25 Mbps**	7.62 Mbps
VUT 2.4	VVC + LCEVC 2 160p	1 080p VVC + HDR10 at 7.52 Mbps	5.97 Mbps	**8.08 Mbps**	12.41 Mbps

These bitrate values were determined considering worst-case situations, to ensure that the TV 3.0 physical layer has sufficient capacity to transport HD HDR video at a quality higher than the maximum possible in the first-generation Brazilian DTTB system and/or UHD HDR video at a quality higher than its HD HDR video. Note that using the highest bitrate corresponding to the target score 0 ("the same") among the five clips in the test material means that a given VUT would provide a similar subjective quality for the clip with the highest required bitrate, and somewhat higher score in the other clips while not necessarily reaching the score 1 ("slightly better") in all clips.

The starting point for the quality comparison to establish the required bitrate for HD HDR video was the highest quality currently possible for real-time video coding in the first-generation Brazilian DTTB system, at 14 Mbps, which already provides a very high quality. This represents the case of commercial broadcasters operating with a single HD program service on the ISDB-T 12-segment layer, with 64QAM modulation and 3/4 FEC, on a 6 MHz channel. Public broadcasters in Brazil operate with up to 4 program services per channel and some commercial broadcasters operate with 16QAM modulation and/or with more than one program service in the case of agreements with the government to broadcast educational program services. In these cases, broadcasters use significantly lower video bitrates. Furthermore, different program genres have different video coding complexities and video quality requirements. In this test, an upper limit was sought for the bitrate requirement in high-quality live content. The VVC HD HDR video under test was encoded with a significantly more complex configuration than the reference H.264 HD HDR video (progressive rather than interlaced, thus with twice the frame rate, 10-bit rather than 8-bit, wide color gamut rather than BT.709).

Finally, it is worth noting that VVC technology is very recent and some of the first real-time VVC encoder implementations were used in this test. It is expected that the performance of real-time VVC encoders will improve over time, as it did with previous generations of video codecs. Multi-pass real-time encoding can also improve performance but with the penalty of increased latency. Offline VVC encoding (file-based encoding that can take much longer than the content's duration) performs much better than real-time VVC encoding and can be used for content that is not produced and broadcast live.

TV 3.0 standardization will not prescribe the bitrate to be used for video coding.