Original Manuscript ID: **electronics-2835939**

Original Article Title: "**Performance Comparison of VVC, AV1, HEVC and AVC for High Resolutions**"

To: **Electronics Editor**

Re: **Responses to reviewers**

Dear Editor,

thank you for allowing a resubmission of our manuscript, with an opportunity to address the reviewers' comments.

We are uploading:

(a) our point-by-point response to the comments of reviewers (below),

(b) an updated manuscript with colored highlighting that indicates changes (PDF document),

(c) an updated manuscript (LaTeX main document).

We are grateful to the Editors and the Reviewers who have given us the opportunity to improve our work. We express our deep gratitude to all the reviewers for their valid comments and constructive suggestions which have helped us improve the quality and presentation of the manuscript. Appropriate changes have been made in the revised manuscript. Please see the responses to Reviewer#1, Reviewer#2, Reviewer#3, Reviewer#4, and Reviewer#5 that are highlighted in orange (no response with the highlighted-orange text directly in the paper), highlighted in red, highlighted in green, highlighted in blue and highlighted in purple text in the manuscript, respectively. Our answers also indicate which parts of the original manuscript have been altered or added. We hope all the reviewers' comments have been addressed and resolved to their satisfaction.

Finally, we believe that our work in this manuscript is appropriate for publication in the Electronics journal. We look forward to receiving your response.

Best regards,

Miroslav Uhrina et al.

Responses to the reviewers' comments (list of changes): **electronics-2835939**

Reviewer 1:

Comment 1:

This paper is simply a comparison of codecs. There are a number of such comparisons and given that the latest codecs are designed for higher resolutions they are expected to give better results. The novelty of this work is therefore low.

Response 1:

Despite some publications dealing with codec performance comparison, only two papers examine the quality performance of the latest codecs using 8K video test sequences. Moreover, the PP8K database is the newest one, and as far as we know, no such complex performance evaluation has been conducted using this dataset.

Comment 2:

English needs considerable improvements. The paper has a lot of typos and use of incorrect tense making it sometimes difficult to read.

Response 2:

The English language in our paper has been improved.

Reviewer 2:

Comment 1:

What is shown in Table 8 is not clear. What is the anchor?

Response 1:

Since this plot is not so clear and its informative value is low, we decided to remove it from our paper.

Comment 2:

Figure 5-9 plots should be enlarged to have two in a row. There is no page limit in the journal.

Response 2:

Plots in Figures 5-9 have been enlarged to be only two next to each other.

Comment 3:

Group of Pictures is not specified in the manuscript. There is only the information that Intra frames appear every second. Frame types and profiles (e.g., low delay, random access) should be given. Results should distinguish frame types.

Response 3:

In our experiments, only the Intra frame was defined, i.e. just only the N value of the GoP was defined. The number of B frames (the M value of the GoP), as well as the GoP structure, was set to default for each of the used codecs. For better clarification, the following sentences have been added to our manuscript:

The GoP structure is referred to by two numbers - N and M, where N stands for the distance between two keyframes, i.e. I frames, also known as the length of GoP and M stands for the distance between two anchor frames (I or P). Since the GoP settings, except the I frame distance, were set to default by all codecs, the GoP structure seemed as follows: By the H.264/AVC N=60 and the M=4, by the H.265/HEVC N=60 and M=5, by the AV1 N=1 and M=1. By the H.266/VVC, the intraperiod was set to 64, perceptual optimization was enabled and the decoding refresh type was CRA, which stands for Clean Random Access. By all codecs, no certain profiles and presets were set within the encoding process.

Comment 4:

Please provide BD-PSNR apart from BD-Rate in Tables 5-9.

Response 4:

BD-PSNR apart from BD-Rate have been provided in Tables 5-12.

FHD							
	BodeMuseum	NeptuneFountain3	TiergartenParkway	Cooking	Girafe	Koi	36
H.266 vs H.264	3,50 dB	2,73 dB	3,72 dB	5,07 dB	3,60 dB	3,20 dB	3,21 dB
H.266 vs H.265	2,69 dB	1,69 dB	3,13 dB	1,88 dB	2,96 dB	1,85 dB	2,43 dB
H.266 vs AV1	1,19 dB	1,50 dB	1,84 dB	1,84 dB	1,78 dB	1,62 dB	1,22 dB
H.265 vs H.264	0,81 dB	1,05 dB	0,58 dB	3,19 dB	0,64 dB	1,35 dB	0,78 dB
AV1 vs H.264	2,31 dB	1,23 dB	1,88 dB	3,23 dB	1,82 dB	1,59 dB	1,99 dB
AV1 vs H.265	1,50 dB	0,18 dB	1,30 dB	0,04 dB	1,18 dB	0,23 dB	1,21 dB

UHD							
	BodeMuseum	NeptuneFountain3	TiergartenParkway	Cooking	Girafe	Koi	36
H.266 vs H.264	4,49 dB	5,14 dB	5,40 dB	7,99 dB	3,82 dB	4,92 dB	4,24 dB
H.266 vs H.265	3,15 dB	2,06 dB	3,88 dB	2,45 dB	2,73 dB	1,65 dB	2,24 dB
H.266 vs AV1	1,62 dB	1,37 dB	1,61 dB	1,51 dB	1,53 dB	1,32 dB	0,83 dB
H.265 vs H.264	1,34 dB	3,09 dB	1,53 dB	5,54 dB	1,10 dB	3,27 dB	2,00 dB
AV1 vs H.264	2,87 dB	3,77 dB	3,79 dB	6,49 dB	2,29 dB	3,59 dB	3,41 dB
AV1 vs H.265	1,53 dB	0,68 dB	2,27 dB	0,95 dB	1,19 dB	0,32 dB	1,41 dB

8K							
	BodeMuseum	NeptuneFountain3	TiergartenParkway	Cooking	Girafe	Koi	36
H.266 vs H.264	3,39 dB	4,48 dB	5,57 dB	5,01 dB	2,67 dB	2,69 dB	3,03 dB
H.266 vs H.265	1,78 dB	1,59 dB	2,69 dB	1,24 dB	1,38 dB	0,76 dB	1,54 dB
H.266 vs AV1	1,31 dB	1,30 dB	1,66 dB	0,74 dB	1,06 dB	0,61 dB	0,36 dB
H.265 vs H.264	1,61 dB	2,89 dB	2,88 dB	3,77 dB	1,29 dB	1,93 dB	1,49 dB
AV1 vs H.264	2,08 dB	3,18 dB	3,91 dB	4,27 dB	1,61 dB	2,08 dB	2,67 dB
AV1 vs H.265	0,46 dB	0,29 dB	1,03 dB	0,50 dB	0,32 dB	0,15 dB	1,17 dB

	FHD	UHD	8K
H.266 vs H.264	3,58 dB	5,14 dB	3,83 dB
H.266 vs H.265	2,37 dB	2,59 dB	1,57 dB
H.266 vs AV1	1,57 dB	1,40 dB	1,01 dB
H.265 vs H.264	1,20 dB	2,55 dB	2,27 dB
AV1 vs H.264	2,01 dB	3,75 dB	2,83 dB
AV1 vs H.265	0,81 dB	1,19 dB	0,56 dB

Comment 5:

There is a typo in line 36.

Response 5:

The typo has been corrected.

Reviewer 3:

Comment 1:

1. Does this test still have some limitations or factors that may influence the results, such as the selection of the test set, adjustments to encoding parameters, etc.? Addressing these aspects would help readers better understand the reliability and applicability of the study.

Response 1:

Of course, the whole experimental setup which involves the selection of test materials as well as the encoding process may influence the final results. To better understand, we have added a few sentences to describe these aspects:

There are many factors that can influence the results. Firstly, the selection of test set, which means test sequences, plays an important role. The more complex the test sequence, in the meaning of SI-TI parameters, the higher the difficulty of the coding process. On the contrary, sequences with slow motion or small amount of spatial details can be coded with higher efficiency. The resolution, bitdepth, framerate, i.e. number of frames per second, and color space are input factors that can also affect the final results. Last but not least, the experimental setup, in the meaning of adjustment of the encoding parameters, can influence the results. This includes, for instance, the Group of Pictures (GoP) setting, encoding quality choice by either Quantization Parameter (QP) or bitrate (BR) constrain, selection of encoding modes as Constant Bitrate (CBR), Variable Bitrate (VBR) or Adaptive Bitrate (ABR), choice of rate control modes as 1-pass, 2-pass or CRF encoding, selection of used presets, tunes or profiles.

Comment 2:

2. In the paper, we observe that as the resolution increases, newly developed codecs such as H.266/VVC and AV1 exhibit superior performance. If necessary, a more detailed explanation could be provided on why H.266/VVC performs better at different resolutions, including the presence of specific technological features or optimization strategies.

Response 2:

More detailed explanations discussing why the newly developed codecs achieve better quality in high resolutions than the older ones have been added to our manuscript:

The reasons why the newly developed codecs AV1 and H.266/VVC achieve better video quality at high resolutions are given in the following lines.

Since the H.266/VVC codec has been developed to be versatile, it comes along with several new functionalities. One of them is random access capability, which refers to the ability to start consuming video content from positions other than at the very beginning of the bitstream. VVC also allows the spatial resolution to change at inter-coded pictures through the support of the feature referred to as reference picture resampling (RPR). In VVC, the Coding Tree Units (CTUs), which are basic processing units within a frame, can be larger than in HEVC, but the concept is the same and is similar to the approach of a macroblock in AVC. While the basic well-known block-based hybrid video coding scheme used in all previous MPEG standards has been retained in VVC, core compression technologies have been improved in some ways. First, we can mention the quadtree partitioning of a CTU which has been extended by enabling more flexible partitioning and supporting larger block sizes. In intra-picture prediction, VVC contains also finer-granularity angular prediction with 93 angle modes except the DC and planar modes similar to HEVC and othermatrix-based prediction modes for luma as well as cross-component prediction modes for chroma samples. In inter-frame prediction, VVC uses either single motion vector (MV) unprediction referencing a frame in a list of previously decoded reference frames or bi-prediction using two MVs. Besides that, VVC offers a variety of new coding tools for more efficient representation, prediction, and coding of motion compensation control information, as well as for enhancing the motion compensation processing itself. These techniques can be categorized into advances in coding motion information, advances in CU-level motion compensation, improved motion compensation processes using subblock based motion derivation and prediction refinement at the decoder, and horizontal wrap-around motion compensation. In transform and quantization, VVC uses the same concept as by the HEVC, but VVC achieves better energy compaction of the prediction residual by extended transforms complemented by improved quantization and residual coding. The entropy coding in VVC is achieved by the CABAC coding as in HEVC, but the efficiency is refined by some changes in the coefficient coding and probability estimation. In VVC, refined and new in-loop filters are used to reach better visual quality. Last but not least, VVC contains special coding tools which increase coding efficiency.

AV1 has larger superblock partitioning up to 128x128 luma samples. It can be partitioned into smaller block sizes, where the minimum block size is extended to 4x4 luma samples, which provides more coding flexibility. AV1 also enables a two-stage block partitioning search, where the first pass starts from the largest block size. In intraframe prediction, AV1 extends the directional Intra prediction options to support higher granularity and adds a new smooth prediction mode in nondirectional smooth Intra prediction. AV1 also allows intraframe motion-compensated prediction, which uses the previously coded pixels within the same frame, namely Intrablock Copy (IntraBC). Also, other features increase coding efficiency as chroma from luma prediction models and color palette mode. In interframe prediction, AV1 supports many toolsets to exploit the temporal correlation in video signals, which include adaptive filtering in translational motion compensation, affine motion compensation, and highly flexible compound prediction modes. AV1, unlike the older codecs, employs a dynamic motion vector referencing scheme that obtains candidate motion vectors from the spatial and temporal neighbors and ranks them for efficient entropy coding. In transform coding, AV1 extends its flexibility in terms of both the transform block sizes and kernels, where expends the maximum transform block size to 64x64. In kernels, AV1 allows each transform block to choose its own transform kernel independently. In the quantization step, where the transform coefficients are quantized, the quantization parameter (QP) ranges between 0 and 255. In the entropy coding system, AV1 employs the M-ary symbol arithmetic coding method which was basically developed for the Daala codec. In postprocessing, AV1 allows three optional in-loop filter stages: a deblocking filter, a constrained directional enhancement filter (CDEF), and a loop restoration filter.

All above-mentioned features help both codecs achieve better coding efficiency and quality, especially for high resolution video content.

Comment 3:

3. If necessary, the "Conclusion" section could benefit from further elaboration, encouraging the authors to highlight key findings and underscore the significance of their work more effectively. Adding more comprehensive content to this section will enhance the overall conclusion, ensuring that readers gain a clear understanding of the critical contributions and implications of the study.

Response 3:

The "Conclusion" section has been enlarged and rewritten by highlighting key findings and conclusions proceeding from our research:

In our experiments, we worked with 63 test sequences from three different databases with large diversity. Firstly, we had to choose some of them according to SI-TI analysis. We have selected seven sequences, one sequence from each of the four corners and three ones from the middle of the SI-TI plot. Subsequently, we encoded them to particular codecs, namely H.264/AVC, H.265/GEVC, H.266/VVC, and AV1 at Full HD (FHD), Ultra HD (UHD), and 8K resolution. The bitrates were set to 1, 3, 5, 7, 10, 15 Mbps for FHD and UHD resolution and 5, 7, 10, 15, 30, 50 Mbps for 8K resolution, respectively. Altogether, 420 test sequences were encoded. For encoding, we decided to use Average Bitrate Coding (ABR) mode and 1-pass rate control mode except encoding sequences to AVC and HEVC codecs at 8K resolution, where 2-pass encoding mode was used. For quality assessment, we used objective metrics, namely Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) and Video Multi-Method Assessment Fusion (VMAF), which belong to full-reference methods.

Beside the bitrate savings, we also compared the processing time of all codecs at all resolutions. For this analysis, we decided to choose the "NeptuneFountain3" test sequence, which is situated in the middle of the SI-TI plot. From the results is it clear that H.264/AVC reaches a very similar coding time as AV1 and H.265/HEVC achieves about two times longer computational time compared to AV1. The longest coding time, as expected, reaches H.266/VVC codec, starting from 27 times longer at 1 Mbps bitrate at UHD resolution and ending at 174 times longer at 15 Mbps bitrate at FHD resolution than AV1 codec.

Comment 4:

4. The authors have conducted a commendable comparison of the performance of H.266/VVC, AV1, H.265/HEVC, and H.264/AVC. However, I am curious whether differences in the experimental configurations may also impact the performance comparison. Therefore, I believe it is essential for the authors to showcase the experimental setup.

Response 4:

We are not sure what the reviewer means by this comment, but we have included the time of the encoding process for each codec. For this analysis, we decided to choose a test sequence called "NeptuneFountain3", which is situated in the middle of the SI-TI plot.

The following sentences and tables have been added to our manuscript:

Apart from the bitrate savings, we have also noticed and compared the processing time of all codecs at all resolutions. For this analysis, we decided to choose the "NeptuneFountain3" test sequence, which is situated in the middle of the SI-TI plot. Table \ref{tab:Testing PC} describes the parameters of the PC where the experimental setup was done. Table \ref{tab:Coding time} shows coding time, in seconds, and Table \ref{tab:Coding time comparison} comparison of coding time relative to AV1 codec, which achieved the shortest time. As can be seen, H.264/AVC reaches a very similar coding time as AV1, and H.265/HEVC achieves about two times longer computational time compared to AV1. The longest coding time, as expected, reaches H.266/VVC codec, starting from 27 times longer at 1 Mbps bitrate at UHD resolution and ending at 174 times longer at 15 Mbps bitrate at FHD resolution than AV1 codec.

FHD						
Codec / Bitrate	**1 Mbps**	**3 Mbps**	**5 Mbps**	**7 Mbps**	**10 Mbps**	**15 Mbps**
H.264/AVC	3.933 s	4.284 s	4.403 s	4.388 s	4.446 s	4.813 s
H.265/HEVC	6.284 s	7.462 s	8.457 s	9.263 s	10.042 s	11.245 s
H.266/VVC	185.142 s	328.863 s	432.875 s	514.937 s	625.180 s	765.770 s
AV1	3.079 s	3.572 s	3.676 s	3.802 s	3.952 s	4.387 s
UHD						
Codec / Bitrate	**1 Mbps**	**3 Mbps**	**5 Mbps**	**7 Mbps**	**10 Mbps**	**15 Mbps**
H.264/AVC	15.719 s	16.177 s	16.424 s	16.578 s	16.864 s	17.218 s
H.265/HEVC	23.463 s	27.036 s	34.500 s	35.183 s	36.568 s	32.863 s
H.266/VVC	312.819 s	519.568 s	674.671 s	807.677 s	973.196 s	1229.545 s
AV1	11.467 s	11.826 s	12.355 s	12.664 s	12.889 s	13.282 s
8K						
Codec / Bitrate	**5 Mbps**	**7 Mbps**	**10 Mbps**	**15 Mbps**	**30 Mbps**	**50 Mbps**
H.264/AVC	48.772 s	49.057 s	48.811 s	49.487 s	49.916 s	49.632 s
H.265/HEVC	96.779 s	102.943 s	157.587 s	168.960 s	183.300 s	243.093 s
H.266/VVC	1693.086 s	1921.788 s	2271.042 s	2815.820 s	4286.146 s	6162.371 s
AV1	39.473 s	39.268 s	40.079 s	41.569 s	43.875 s	52.724 s

FHD						
Codec / Bitrate	**1 Mbps**	**3 Mbps**	**5 Mbps**	**7 Mbps**	**10 Mbps**	**15 Mbps**
H.264/AVC	1,28x	1,20x	1,20x	1,15x	1,13x	1,10x
H.265/HEVC	2,04x	2,09x	2,30x	2,44x	2,54x	2,56x
H.266/VVC	60,13x	92,07x	117,76x	135,44x	158,19x	174,55x
AV1	1,00x	1,00x	1,00x	1,00x	1,00x	1,00x

UHD						
Codec / Bitrate	**1 Mbps**	**3 Mbps**	**5 Mbps**	**7 Mbps**	**10 Mbps**	**15 Mbps**
H.264/AVC	1,37x	1,37x	1,33x	1,31x	1,31x	1,30x
H.265/HEVC	2,05x	2,29x	2,79x	2,78x	2,84x	2,47x
H.266/VVC	27,28x	43,93x	54,61x	63,78x	75,51x	92,57x
AV1	1,00x	1,00x	1,00x	1,00x	1,00x	1,00x

8K						
Codec / Bitrate	**5 Mbps**	**7 Mbps**	**10 Mbps**	**15 Mbps**	**30 Mbps**	**50 Mbps**
H.264/AVC	1,24x	1,25x	1,22x	1,19x	1,14x	0,94x
H.265/HEVC	2,45x	2,62x	3,93x	4,06x	4,18x	4,61x
H.266/VVC	42,89x	48,94x	56,66x	67,74x	97,69x	116,88x
AV1	1,00x	1,00x	1,00x	1,00x	1,00x	1,00x

Processor	AMD Ryzen 9 5950X 16-Core 4000.0 MHz
SSD	Samsung SSD 980 Pro 1TB
RAM	64 GB DDR4 SDRAM
Graphic card	NVIDIA GeForce RTX 3080 Founders Edition LHR
Operating system	Microsoft Windows 10 Education 64-bit

Comment 5:

5.In addition to PSNR, SSIM, and VMAF, other evaluation metrics should be considered to comprehensively evaluate the performance of the codec, especially the subjective experience in different application scenarios.

Response 5:

For the performance comparison, we have decided to choose PSNR, SSIM, and VMAF objective quality metrics. Of course, the quality could be assessed by any other objective metrics, but we are convinced that our selection is sufficient for this purpose. The idea of evaluating the quality by subjective methods is very relevant and we appreciate it, but it is behind the scope of this paper. Of course, we plan to realize it in the near future.

Comment 6:

6.Since a subset of test sequences is selected from the entire data set, it is recommended to gain in-depth knowledge of the characteristics of each selected sequence. This can help understand why these sequences were chosen and how they represent the entire data set.

Response 6:

The reason why we have decided to choose these test sequences is mentioned in the following sentences: "To cover all SI-TI diagrams, we have decided to choose sequences with a wide variety of SI and TI values. We have picked one sequence from each of the four corners, namely "Cooking", "TiergartenParkway", "BodeMuseum", "Koi" and three sequences from the middle of the SI-TI plot, namely "NeptuneFountain3", "Giraffe" and "36"."

Moreover, to gain in-depth knowledge of the characteristics of selected sequences some new information has been added to Table 2 of the paper:

Dataset	Test sequence	Description	8K		UHD		FHD	
			SI value	TI value	SI value	TI value	SI value	TI value
Fraunhofer	BodeMuseum	A view of the Bode Museum in Berlin. Only a train and a boat move in the background. The camera is fixed.	50,83	6,78	71,77	6,02	88,57	4,39
	NeptuneFountain3	A view of the Neptune Fountain in Berlin. Water splashes from the fountain. The camera moves around the fountain from left to right.	26,55	21,23	39,04	21,16	55,23	20,75
	TiergartenParkway	A view of a park way close to the Tiergarten in Berlin. The camera moves forward as if a person is walking with a camera.	54,13	33,97	79,56	33,91	101,25	32,98
PP8K	Cooking	A chef is cooking and the dancing flame is in the pan. The camera is fixed.	11,60	38,42	13,98	38,36	23,65	37,93
	Giraffe	Three giraffes are walking in the zoo. The camera is fixed.	34,62	8,92	52,70	8,07	63,72	7,47
	Koi	The colorful koi are swimming in the fish tank. A lot of bubbles, produced by the working oxygen generator, are floating upward in the water. The camera moves slowly from left to right.	12,91	7,70	12,70	6,55	18,84	5,89
SEPE	36	Many people are skating on the ice rink. The camera is fixed.	36,89	14,53	59,42	14,32	78,11	13,45

Comment 7:

7.For metrics such as PSNR, SSIM and VMAF, it is recommended to emphasize their value ranges and indicate the units of the indicators in the text to help readers better understand the significance of the evaluation results.

Response 7:

New sentences to emphasize objective metric value ranges and indicate the units of the indicators have been added to our manuscript:

The output value of this metric is given in decibels [dB]. Based on the measurements the maximum value that can PSNR achieve is 100 dB which essentially equals to quality of a reference video.

The results obtained using the SSIM metrics fall within the interval [0, 1], where 1 represents the best quality that can be achieved only if all the compared images or videos are identical.

All features are concatenated using SVM-based regression to determine an output score ranging from 0 to 100 per video frame, where 100 represents identical quality compared to a reference video.

Comment 8:

8.In the conclusion, possible future research directions can be briefly mentioned, such as further optimizing the codec, considering performance in different application scenarios, etc. This helps guide the reader in thinking about possible extensions and further research directions.

Response 8:

The conclusion has been extended with possible future research plans and directions.

In the near future, we would like to assess other test sequences from 8K datasets using objective metrics. Subsequently, we plan to select appropriate sequences to evaluate them by subjective methods such as Absolute Category Rating (ACR), Absolute Category Rating with Hidden Reference (ACR-HR), or Double Stimulus Impairment Scale (DSIS). From the results, we intend to calculate the correlation between objective and subjective results by Pearson and Spearman's correlation coefficient. Moreover, the results will be used as inputs to neural networks to refine our proposed model, which can predict quality based on objective metrics.

Reviewer 4:

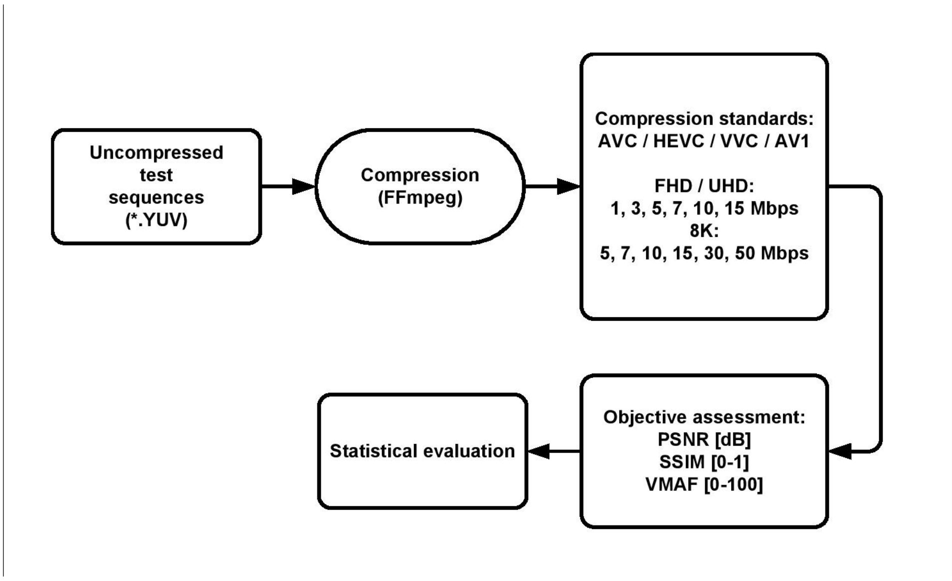
Comment 1:

- To enrich the context and provide a clearer understanding of each codec, an introduction by flow charts highlighting the encoding process of each standard would be highly beneficial. These visual aids would help in overviewing the technical complexities and make the paper more accessible to readers who may not be deeply familiar with video encoding technologies.

Response 1:

A flowchart that highlights the coding and evaluating process has been added to our manuscript with the following sentence:

A flowchart highlighting the coding and evaluating process is depicted in Figure 3.



Comment 2:

- While the paper excels in comparing the quality and bitrate efficiency of different codecs, a deeper analysis of their computational efficiency would be beneficial. This could include evaluating the processing time and resource utilisation of each codec, providing a more holistic view of their practical applicability in various scenarios.

Response 2:

The processing time for each codec has been added to our paper. For this analysis, we decided to choose a test sequence called "NeptuneFountain3", which is situated in the middle of the SI-TI plot. The following sentences have been added to our manuscript:

Apart from the bitrate savings, we have also noticed and compared the processing time of all codecs at all resolutions. For this analysis, we decided to choose the "NeptuneFountain3" test sequence, which is situated in the middle of the SI-TI plot. Table \ref{tab:Testing PC} describes the parameters of the PC where the experimental setup was done. Table \ref{tab:Coding time} shows coding time, in seconds, and Table \ref{tab:Coding time comparison} comparison of coding time relative to AV1 codec, which achieved the shortest time. As can be seen, H.264/AVC reaches a very similar coding time as AV1, and H.265/HEVC achieves about two times longer computational time compared to AV1. The longest coding time, as expected, reaches H.266/VVC codec, starting from 27 times longer at 1 Mbps bitrate at UHD resolution and ending at 174 times longer at 15 Mbps bitrate at FHD resolution than AV1 codec.

FHD						
Codec / Bitrate	**1 Mbps**	**3 Mbps**	**5 Mbps**	**7 Mbps**	**10 Mbps**	**15 Mbps**
H.264/AVC	3.933 s	4.284 s	4.403 s	4.388 s	4.446 s	4.813 s
H.265/HEVC	6.284 s	7.462 s	8.457 s	9.263 s	10.042 s	11.245 s
H.266/VVC	185.142 s	328.863 s	432.875 s	514.937 s	625.180 s	765.770 s
AV1	3.079 s	3.572 s	3.676 s	3.802 s	3.952 s	4.387 s
UHD						
Codec / Bitrate	**1 Mbps**	**3 Mbps**	**5 Mbps**	**7 Mbps**	**10 Mbps**	**15 Mbps**
H.264/AVC	15.719 s	16.177 s	16.424 s	16.578 s	16.864 s	17.218 s
H.265/HEVC	23.463 s	27.036 s	34.500 s	35.183 s	36.568 s	32.863 s
H.266/VVC	312.819 s	519.568 s	674.671 s	807.677 s	973.196 s	1229.545 s
AV1	11.467 s	11.826 s	12.355 s	12.664 s	12.889 s	13.282 s
8K						
Codec / Bitrate	**5 Mbps**	**7 Mbps**	**10 Mbps**	**15 Mbps**	**30 Mbps**	**50 Mbps**
H.264/AVC	48.772 s	49.057 s	48.811 s	49.487 s	49.916 s	49.632 s
H.265/HEVC	96.779 s	102.943 s	157.587 s	168.960 s	183.300 s	243.093 s
H.266/VVC	1693.086 s	1921.788 s	2271.042 s	2815.820 s	4286.146 s	6162.371 s
AV1	39.473 s	39.268 s	40.079 s	41.569 s	43.875 s	52.724 s

FHD						
Codec / Bitrate	**1 Mbps**	**3 Mbps**	**5 Mbps**	**7 Mbps**	**10 Mbps**	**15 Mbps**
H.264/AVC	1,28x	1,20x	1,20x	1,15x	1,13x	1,10x
H.265/HEVC	2,04x	2,09x	2,30x	2,44x	2,54x	2,56x
H.266/VVC	60,13x	92,07x	117,76x	135,44x	158,19x	174,55x
AV1	1,00x	1,00x	1,00x	1,00x	1,00x	1,00x

UHD						
Codec / Bitrate	**1 Mbps**	**3 Mbps**	**5 Mbps**	**7 Mbps**	**10 Mbps**	**15 Mbps**
H.264/AVC	1,37x	1,37x	1,33x	1,31x	1,31x	1,30x
H.265/HEVC	2,05x	2,29x	2,79x	2,78x	2,84x	2,47x
H.266/VVC	27,28x	43,93x	54,61x	63,78x	75,51x	92,57x
AV1	1,00x	1,00x	1,00x	1,00x	1,00x	1,00x

8K						
Codec / Bitrate	**5 Mbps**	**7 Mbps**	**10 Mbps**	**15 Mbps**	**30 Mbps**	**50 Mbps**
H.264/AVC	1,24x	1,25x	1,22x	1,19x	1,14x	0,94x
H.265/HEVC	2,45x	2,62x	3,93x	4,06x	4,18x	4,61x
H.266/VVC	42,89x	48,94x	56,66x	67,74x	97,69x	116,88x
AV1	1,00x	1,00x	1,00x	1,00x	1,00x	1,00x

Processor	AMD Ryzen 9 5950X 16-Core 4000.0 MHz
SSD	Samsung SSD 980 Pro 1TB
RAM	64 GB DDR4 SDRAM
Graphic card	NVIDIA GeForce RTX 3080 Founders Edition LHR
Operating system	Microsoft Windows 10 Education 64-bit

Reviewer 5:

Comment 1:

In my opinion, this paper lacks definitions and explanations for terminology, approaches, and quality measures used in the area. In the first five pages, the authors use a large number of abbreviations and terms which are partially explained on page 6 and later.

A non-full list of examples is given below.

- For example, the quality measures: PSNR, SSIM, and VMAF appear already in the abstract, but they are explained only on page 6.
- For resolutions: 240p, 480p, Full HD, Ultra HD, and 8K, frame sizes, frame rates, and number of bits per pixel have to be specified.
- 4:2:0 YUV is not defined
- Abbreviation SI-TI values appears at the beginning of page 3 and is explained only at the bottom of the page.
- Page 2, "…in terms of coding gains and complexity…." Coding gains are not defined.

Response 1:

All mentioned abbreviations have been fully explained.

A non-full list of examples is given below.

Comment 1a:

For example, the quality measures: PSNR, SSIM, and VMAF appear already in the abstract, but they are explained only on page 6.

Response 1a:

All mentioned abbreviations of used objective metrics have been explained also in the abstract.

Comment 1b:

For resolutions: 240p, 480p, Full HD, Ultra HD, and 8K, frame sizes, frame rates, and number of bits per pixel have to be specified.

Response 1b:

Not in all articles that we have cited in the Introduction of our manuscript, the authors noticed all technical parameters including frame sizes, frame rates, and bit depths. We tried to do our best to describe all the parameters that were mentioned in the cited papers.

Comment 1c:

4:2:0 YUV is not defined

Response 1c:

We have defined the 4:2:0 YUV color format by adding following sentences:

The YUV, also known as YPbPr in analog or YCbCr in digital video, is a color model, composed of luma (Y) and two chroma (UV) components. In the past, it was primarily used on an analog television. Nowadays, it is used for chroma subsampling which is a type of compression that reduces the color information in a video signal in favor of luminance (Y) data in order to reduce bandwidth usage without significantly affecting picture quality. There are many types of subsampling modes for instance 4:2:2, 4:2:0, or 4:1:1. In our experiments, we have used the 4:2:0 subsampling format, which means the bandwidth of a video signal is reduced by half compared to no chroma subsampling signal.

Comment 1d:

Abbreviation SI-TI values appears at the beginning of page 3 and is explained only at the bottom of the page.

Response 1d:

We have explained the abbreviations SI-TI at the beginning of page 3, where these parameters were first mentioned.

Comment 1e:

Page 2, "…in terms of coding gains and complexity…." Coding gains are not defined.

Response 1e:

Given that the "coding gains" were not mentioned on page 2, we have erased it from the manuscript.

Comment 2:

Typos

1. Page 1, row 36, igh

2. Page 2, row 89 From this reason should be for this reason

3. Page 5, row 156 parameteres should be parameters

4. Page 6, row 168 Bjontegaard Delta should be Bjøntegaard Delta

Response 2:

The mentioned typos have been corrected.