

Video Quality Evaluation Methodology and Verification Testing of HEVC Compression Performance

Thiow Keng Tan, *Senior Member, IEEE*, Rajitha Weerakkody, Marta Mrak, *Senior Member, IEEE*, Naeem Ramzan, *Senior Member, IEEE*, Vittorio Baroncini, Jens-Rainer Ohm, *Member, IEEE*, and Gary J. Sullivan, *Fellow, IEEE*

Abstract—The High Efficiency Video Coding (HEVC) standard (ITU-T H.265 and ISO/IEC 23008-2) has been developed with the main goal of providing significantly improved video compression compared with its predecessors. In order to evaluate this goal, verification tests were conducted by the Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29. This paper presents the subjective and objective results of a verification test in which the performance of the new standard is compared with its highly successful predecessor, the Advanced Video Coding (AVC) video compression standard (ITU-T H.264 and ISO/IEC 14496-10). The test used video sequences with resolutions ranging from 480p up to ultra-high definition, encoded at various quality levels using the HEVC Main profile and the AVC High profile. In order to provide a clear evaluation, this paper also discusses various aspects for the analysis of the test results. The tests showed that bit rate savings of 59% on average can be achieved by HEVC for the same perceived video quality, which is higher than a bit rate saving of 44% demonstrated with the PSNR objective quality metric. However, it has been shown that the bit rates required to achieve good quality of compressed content, as well as the bit rate savings relative to AVC, are highly dependent on the characteristics of the tested content.

Index Terms—Advanced Video Coding (AVC), H.265, High Efficiency Video Coding (HEVC), Joint Collaborative Team on Video Coding (JCT-VC), Moving Picture Experts Group (MPEG), MPEG-H Part 2, standards, Video Coding Experts Group (VCEG), video compression.

I. INTRODUCTION

WE ARE currently witnessing something that has become a once-in-a-decade event in the world of

Manuscript received December 16, 2014; revised June 3, 2015; accepted September 1, 2015. Date of publication September 10, 2015; date of current version January 6, 2015. This paper was recommended by Associate Editor T. Wiegand.

T. K. Tan is with M-Sphere Consulting Pte. Ltd., Singapore 808379, and also with NTT DOCOMO, Inc., Tokyo 100-6150, Japan (e-mail: tktan@m-sph.com).

R. Weerakkody and M. Mrak are with the Research and Development Department, British Broadcasting Corporation, London W12 7SB, U.K. (e-mail: rajitha.weerakkody@bbc.co.uk; marta.mrak@bbc.co.uk).

N. Ramzan is with the University of the West of Scotland, Paisley PA1 2BE, U.K. (e-mail: naeem.ramzan@uws.ac.uk).

V. Baroncini is with Fondazione Ugo Bordoni, Rome 00144, Italy (e-mail: vittorio@fub.it).

J.-R. Ohm is with RWTH Aachen University, Aachen 52062, Germany (e-mail: ohm@ient.rwth-aachen.de).

G. J. Sullivan is with Microsoft Corporation, Redmond, WA 98052 USA (e-mail: garys@ieee.org).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSVT.2015.2477916

video compression: the emergence of a major new family of video compression standards. The mid-1990s saw the introduction of the Moving Picture Experts Group (MPEG)-2 video coding standard (ITU-T Rec. H.262 and ISO/IEC 13818-2 [1]), the first compression standard to be widely adopted in broadcasting and entertainment applications. Advanced Video Coding (AVC) (ITU-T Rec. H.264 and ISO/IEC 14496-10 [2]) appeared in the mid-2000s, offering the same subjective quality at approximately half the bit rate. Now, a new standard, High Efficiency Video Coding (HEVC) (ITU-T Rec. H.265 and ISO/IEC 23008-2), has been developed that promises a further factor of two improvement in compression efficiency for the mid-2010s [3].

The HEVC standard has been jointly developed by the same two standardization organizations whose previous collaboration resulted in both MPEG-2 and AVC: 1) the ISO/IEC MPEG and 2) the ITU-T Video Coding Experts Group (VCEG), through the Joint Collaborative Team on Video Coding (JCT-VC) [4]. HEVC version 1 was ratified in 2013 as H.265 by the ITU-T and as MPEG-H Part 2 by ISO/IEC [5]. This first version supports applications that use conventional (single-layer) encoding of 4:2:0-sampled video with 8- or 10-bit precision. A second edition was completed in July 2014, and an additional extension was completed in February 2015 [6]. These extend the standard to support contribution applications with tools that enable 4:2:2- and 4:4:4-sampled video formats as well as 12- and 16-bit precision [7], and multilayer coding enhancements for efficient scalability [8] and stereo/multiview and depth-enhanced 3D compression [9]. A further amendment is currently being developed to enable more efficient coding of screen-captured graphics and text content and mixed-source content [10].

A few evaluations have previously been reported comparing the compression performance of HEVC with AVC and also demonstrating the suitability of HEVC for various applications, particularly including some evaluations for high-resolution video content [11]. The subjective test results for a number of test sequences and some analysis of the bit rate savings at different quality levels were presented in [12]. In [13], a study on the suitability of HEVC for beyond-HDTV broadcast services was presented, but with no comparison with previous standards. A comparison of HEVC and AVC performance for frame rates up to 30 Hz on a small number of ultra high definition (UHD)

video sequences was presented in [14] and [15], and an informal study focused on low-delay (LD) applications and real-time encoding with HD resolutions was presented in [16]. All of these, including some that were carried out at the early stages of HEVC development, have provided very consistent evidence of the substantial coding efficiency improvements enabled by HEVC.

For its formal evaluation of HEVC performance, the JCT-VC performed a subjective evaluation on a wider range of content of resolutions varying from 480p (832 × 480) up to UHD with frame rates of up to 60 Hz. The test sequences used for the verification testing were deliberately chosen to be different from those that had been used during the development of the standard, to avoid any possible bias that the standard could have toward those sequences. A test report was produced for the JCT-VC itself [17], and some additional analysis of HEVC performance for UHD content has been presented in [18] (using cropped 2560 × 1600 regions) and [19] (a recent brief conference publication about the testing).

As video coding standards generally specify only the format of the data and the associated decoding process without specifying how encoding is to be performed, it is not possible in general to test the compression performance of a standard. Some particular encoding methods must be used as a proxy to represent the capability of the standard instead. The outcome of such a comparison is generally more reliable when similar encoding techniques with similar configurations are applied in the compared encoders rather than simply comparing unknown technologies as black boxes. The verification tests were therefore performed using reference software encoders that had been developed in the standardization work and used very similar encoding algorithms and configurations that were selected to represent important applications. These publicly available reference software codebases are known as the HEVC model (HM) for HEVC [20] and the joint model (JM) for AVC [21].

While this paper reports an extended set of results of HEVC verification tests, it also summarizes the details of tools that can be used in the analysis of the results, pointing to a number of factors an evaluation should consider. Compared with the initial results [17], those presented in this paper are based on the use of more viewers for subjective tests, the objective results are presented and compared with the subjective results, and additional analysis of the coding gains versus the bit rate is provided. Ultimately, the verification test showed that the key objective of HEVC had been achieved—i.e., providing a substantial improvement in compression efficiency relative to its predecessor AVC.

This paper is organized as follows. An overview of video quality evaluation and statistical analysis methodology is presented in Section II, and the test settings used in the subjective evaluations are detailed in Section III. Section IV presents the test results and detailed analysis. Finally, the conclusion is given in Section V.

II. VIDEO QUALITY EVALUATION

This section provides an overview of the objective and subjective video quality metrics and the related analysis used in this study.

For the convenience of video coding performance assessment, the most commonly used objective metric is peak signal-to-noise ratio (PSNR). However, it is commonly acknowledged that PSNR has the disadvantages of disregarding the viewing conditions and the characteristics of human visual system perception. In addition, the PSNR for a given video sequence can be computed in different ways, depending on how the picture components (e.g., luma and chroma) or individual picture PSNR values are combined. Nevertheless, for a particular content item and small variations of coding conditions, the changes in PSNR values for an overall video sequence can typically be reliably interpreted.

Other objective video quality metrics, such as the structural similarity index (SSIM) and video quality metric (VQM), have been proposed, but are not used nearly as frequently as PSNR [22]. VQM is not often used—primarily due to its computational complexity—and for both metrics, the interpretation of the values they provide has not yet become a common practice in video coding community. Therefore, in the context of HEVC and AVC compression, this paper provides comparisons and analysis using the PSNR objective measure and subjective quality evaluation results.

Subjective quality evaluation is the process of employing human viewers for grading video quality based on individual perception. Formal methods and guidelines for subjective quality assessments are specified in various ITU recommendations. Among the many of these, the most relevant to this context are ITU-T Rec. P.910 [23], which defines subjective video quality assessment methods for multimedia applications, and ITU-R Rec. BT.500 [24], which defines a methodology for the subjective assessment of the quality of television pictures. These specifications describe a number of test methods with distinct presentation and scoring schemes, along with the recommended viewing conditions.

Explanations of the quality metrics and data analysis methods are provided in the following sections.

A. Objective Quality Evaluation Using PSNR

PSNR is defined as the ratio between the maximum possible power of the signal (the original image) and the power of noise, which in the considered scenario is introduced by lossy compression. For a decoded image component I_d , the mean square error (MSE) with reference to an original image component I is computed as

$$\text{MSE} = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I(i, j) - I_d(i, j))^2}{M \cdot N} \quad (1)$$

where M and N are the width and height of the image component, and the image component is, for example, an array of luma samples or C_B or C_R chroma samples. The PSNR value is then computed as

$$\text{PSNR} = 10 \cdot \log \frac{(2^B - 1)^2}{\text{MSE}} \quad (2)$$

where B is the bit depth of image samples. This is typically calculated for each frame separately, and then averaged for the frames of a video sequence. Due to the logarithmic transformation, this corresponds to using the geometric average

TABLE I
SUBJECTIVE QUALITY SCALE WITH 11 POINTS
USED IN AN HEVC VERIFICATION TEST

Score	Description of scores
10	Denotes a quality of reproduction that is perfectly faithful to the original. No further improvement is possible.
9	If some difference is “seen or even only thought to be seen” (8 to be used when this happens for more than one part of the image).
8	
7	If the viewer is sure to have seen some difference (6 to be used when this happens for more than one part of the image).
6	
5	When the differences are evident and visible with no particular effort (4 to be used when this happens for more than one part of the image).
4	
3	When differences are many and annoying (2 to be used when this is severely impairing the image).
2	
1	When the image is severely impaired and very far from the original (0 to be used when this happens all over the image making it hardly understandable). Score 0 denotes a quality of reproduction that has no similarity to the original. A worse quality cannot be imagined.
0	

of frame MSEs, and the impact of this should be critically considered when a high fluctuation over frames is present.

For video sequences, which ordinarily consist of three color components, either the luma PSNR value ($PSNR_Y$), calculated using only luma component values, may be reported or a weighted PSNR value ($PSNR_W$) using all three components can be computed using some weighting criteria. An example of a popular weighting for content with 4:2:0 sampling is

$$PSNR_W = \frac{6 \cdot PSNR_Y + PSNR_{C_B} + PSNR_{C_R}}{8}. \quad (3)$$

The most accurate interpretation of the objective results is obtained by looking at the frame-by-frame results for each component. However, this may not be practical for the final presentation of the results for a large data set and a large number of test points.

B. Subjective Quality Evaluation

For the HEVC verification test that includes a wide range of visual quality points, a degradation category rating (DCR) [23] test method was selected. For this purpose, it was used to evaluate the quality (and not the impairment) with a quality rating scale made of 11 levels [23], ranging from 0 (lowest quality) to 10 (highest quality), which may be interpreted as in Table I. The numerical scale helps avoid misinterpretations associated with the use of category adjectives (e.g., excellent or good), especially in cases where the tests are performed across different countries and including nonnative English speakers.

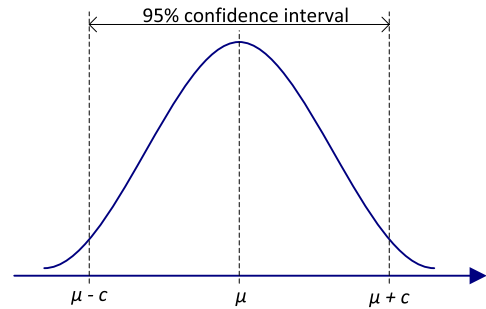


Fig. 1. Confidence interval for normal distribution and 95% probability.

The basic results of the subjective test are evaluated in terms of the average rating, which is called the mean opinion score (MOS), and the associated confidence interval values that are computed for each coding point, after having verified the reliability of each viewer. For the DCR method, it is recommended to hire more than 15 naïve viewers that have been properly screened for visual acuity and color blindness, to allow for an accurate statistical analysis of the subjective scores [24].

From the raw data, i.e., the individual subjective scores, the reliability of each viewer is calculated. The individual reliability is evaluated using the correlation coefficient r computed between each score x_i provided by a viewer and the overall MOS value y_i assigned for that test point i as

$$r = \frac{\sum_{i=1}^T (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^T (x_i - \bar{x})^2 \cdot \sum_{i=1}^T (y_i - \bar{y})^2}} \quad (4)$$

where T is the total number of test points for a viewing session, y_i is the average of all scores for the test point i , and \bar{x} and \bar{y} are the average values of x_i and y_i for all test points, respectively. In this HEVC verification test, a correlation index greater than or equal to 0.75 is considered as valid for the acceptance of the viewer’s scorings; otherwise, the viewer is considered as an outlier. Once the results for outliers are discarded, the MOS for each test point is computed using the arithmetic average of scores of the remaining viewers. In addition, the confidence interval is computed for each test point to estimate a range of values covered by a certain probability.

Assuming a Gaussian (normal) distribution for the population of subjective scores with sample size n , mean (MOS) μ , and sample-based standard deviation measurement s , the confidence interval is defined as $(\mu - c, \mu + c)$, where c is computed as

$$c = z \cdot \frac{s}{\sqrt{n}}. \quad (5)$$

In the analysis of the subjective test results, the 95% confidence interval, as shown in Fig. 1, is calculated for each test point. For a 95% confidence interval with a Gaussian distribution, the value of z in (5) is 1.96. For the results presented in Section IV, the confidence interval is plotted alongside the MOS, as shown in Fig. 2, with an interpolated curve from MOS values.

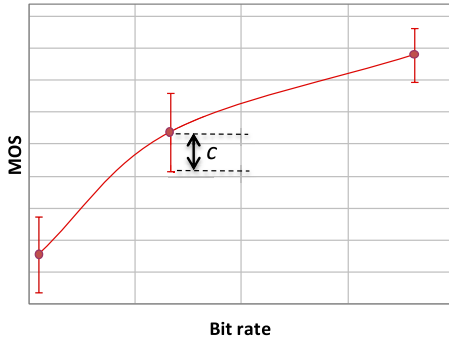


Fig. 2. Example of a confidence interval related to a test point and an associated MOS versus bit rate curve.

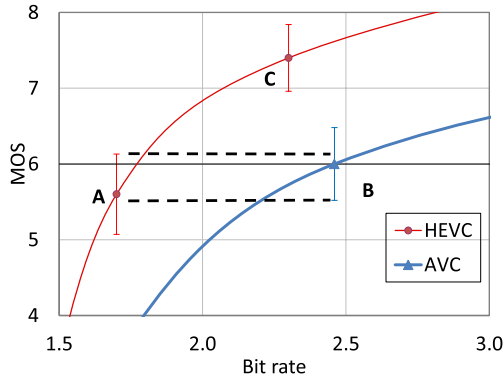


Fig. 3. Overlapping confidence interval of test points.

C. Interpretation of Bit Rate Savings From Subjective Quality Comparison

The objective of the verification test is to gauge the bit rate savings of HEVC over AVC when the AVC and HEVC test points have the same subjective quality.

Fig. 3 shows an example of a plot comparing the AVC and HEVC MOS versus bit rate curves. There is no overlap in the MOS confidence intervals of the HEVC test point C and AVC test point B, and hence, there is sufficient statistical significance to conclude that the HEVC test point C has a better quality than the AVC test point B. There is, however, an overlap in the MOS confidence intervals of the HEVC test point A and AVC test point B. This means that it is highly likely that the HEVC test point A and AVC test point B have subjective quality that cannot be distinguished. However, there is still a chance that the subjective qualities of HEVC test point A and AVC test point B are not the same.

A more rigorous analysis is to perform a two-sample unequal variance (heteroscedastic) student's t -test test using the two-tailed distribution to determine if indeed the subjective qualities given by the sample mean values of the pair of test points are not the same. The null hypothesis, H_0 , in this case would be that the HEVC test points have the same quality as the AVC test point, and the alternate hypothesis, H_a , is that the HEVC test points do not have the same quality as the AVC test point.

To compare the means of two populations, the t -statistic can be used, which is expressed as

$$t = (\bar{X}_1 - \bar{X}_2) / \sqrt{s_1^2/n_1 + s_2^2/n_2} \quad (6)$$

where \bar{X}_i , s_i^2 , and n_i denote the sample mean, the sample variance, and the size of the i th sample, $i \in \{1, 2\}$. By computing the t -statistic in this way and approximating it with a student's t -distribution whose degree of freedom (DF) is specified as

$$DF = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)} \quad (7)$$

a probability value p can be computed from the t -statistic that indicates the extent to which the means of the two populations are considered to be different. The smaller the p -value is, the more significant the difference between the distributions of the two populations is.

A p -value less than 0.05 indicates a very low probability of committing a type-I error (i.e., rejecting the null hypothesis when it is true). In such a case, the null hypothesis can thus be safely rejected, and it can be concluded that there is statistical significance that the HEVC test point does not have the same quality as the AVC test point. A p -value greater than or equal to 0.05 means that the null hypothesis cannot be confidently rejected. For the purpose of this paper, the HEVC test point is considered to have the same quality as the corresponding AVC test point in such a case. However, there is still a possibility of committing a type-II error (i.e., failure to reject the null hypothesis when in fact the alternate hypothesis is true). The power or sensitivity of a statistical test is the probability of correctly rejecting the null hypothesis (H_0) when it is false—i.e., the probability of correctly accepting the alternative hypothesis (H_a) when it is true [25]. A statistical power test of the data has shown that if in fact the true population mean for the difference in the HEVC MOS and AVC MOS is greater than or equal to 0.8, then the mean probability of committing a type-II error, β , is less than or equal to 0.14, and hence the mean power of the test (defined as $1 - \beta$) is 0.86. By convention, a test with a power greater than 0.8 (or $\beta \leq 0.2$) is considered statistically powerful [26].

In the design of the verification test, four bit rates per codec, R_{HEVC} and R_{AVC} , were used. The bit rates were carefully selected for each of 20 test sequences so that each R_{HEVC} is approximately half of the corresponding R_{AVC} . These gave 80 pairs of test points on which the t -test described above was applied. The results of the test determine, for each pair of test points, whether the HEVC test point has a quality better than, the same as, or less than the AVC test point, and give a rough estimate of the bit rate savings of HEVC compared with AVC. The following are the possible outcomes for each pair of test points.

The first case is when the null hypothesis is rejected and there is statistical significance that the HEVC MOS at R_{HEVC} is greater than the AVC MOS at R_{AVC} . This means that one can reasonably conclude that the HEVC test point is achieving a better quality than the AVC test point at half the bit rate of AVC. Note that by the design of the test, the bit rate saving when R_{HEVC} is half R_{AVC} is 50%. Since the bit rate for an HEVC test point could be further reduced to achieve the same quality, the bit rate saving of HEVC compared with AVC is therefore greater than 50% for this case.

The second case is when the null hypothesis was failed to be rejected. This means that the HEVC test point has about the same quality as the AVC test point at half the AVC bit rate, since R_{HEVC} is approximately half of R_{AVC} . Therefore, the bit rate saving of HEVC compared with AVC is approximately 50% for this case.

The third case is when the null hypothesis is rejected and there is statistical significance to conclude that the HEVC MOS at R_{HEVC} is less than the AVC MOS at R_{AVC} . This means that the HEVC test point is not achieving equal or better quality than the AVC test point at half the bit rate of AVC. More bits would need to be allocated to the HEVC test point before the same quality would be achieved. Therefore, the bit rate saving of HEVC compared with AVC is less than 50% for this case.

D. Bjøntegaard Model

The Bjøntegaard model [27], [28] has become a popular tool for evaluating the coding efficiency of a given video codec in comparison with a reference codec over a range of quality points or bit rates. Bjøntegaard delta (BD) metrics are typically computed as a difference in bit rate or a difference in quality based on interpolating curves from the tested data points. In this paper, the focus is on the difference in bit rate, expressed as a percentage of a reference bit rate, as this is easily interpreted as the bit rate saving benefit for equal measured quality.

The BD-rate represents the average bit rate savings for the same video quality (e.g., PSNR or MOS) and is calculated between two rate-distortion curves, such as AVC and HEVC MOS curves in Fig. 3. The bit rate saving difference between the two rate-distortion curves at a given level of quality is

$$\Delta R(D) = \frac{R_B(D) - R_A(D)}{R_A(D)} \quad (8)$$

where $R_A(D)$ and $R_B(D)$ are the bit rate of the interpolated reference and tested bit rate curves, respectively, at the given level of quality/distortion D . $\Delta R(D)$ is typically represented as a percentage of the reference bit rate $R_A(D)$ so that a negative value represents compression gain, while a positive value represents compression loss.

The Bjøntegaard model uses a logarithmic scale for the domain of the bit rate interpolation, so by defining $r = \log R$, the bit rate savings can be expressed as

$$\Delta R(D) = 10^{r_B(D) - r_A(D)} - 1. \quad (9)$$

Taking into account the actual measured rate-distortion points $(R(i), D(i))$, the fitted rate-distortion curves $\hat{r}(D)$ are used in BD-rate computation. Over a range of quality levels, the BD-rate is approximated as

$$\Delta R_{\text{Overall}} \approx 10^{\frac{1}{D_H - D_L} \int_{D_L}^{D_H} [\hat{r}_B(D) - \hat{r}_A(D)] dD} - 1 \quad (10)$$

where the lower D_L and higher D_H integration bounds are computed from the range of the interpolated distortion values D_A and D_B for the reference and tested data sets,

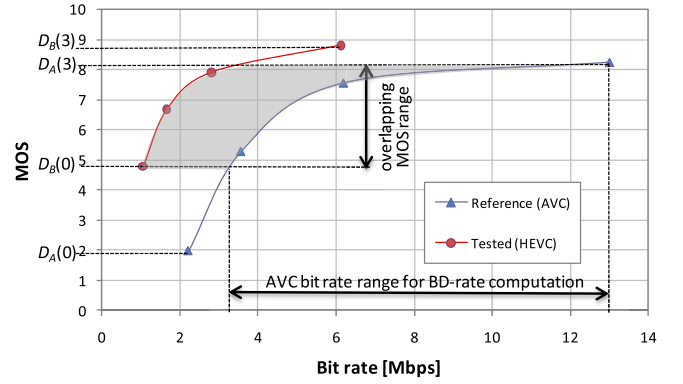


Fig. 4. Illustration of BD-rate computation.

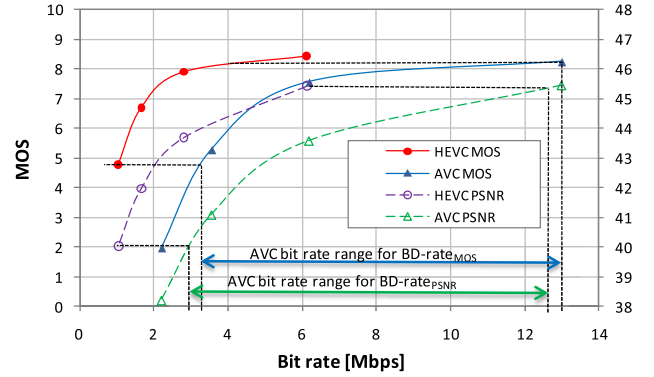


Fig. 5. BD-rate for PSNR and MOS ratings. Actual bit rate ranges and test points taken into account do not necessarily overlap.

respectively, as

$$\begin{aligned} D_L &= \max\{\min(D_A(0), \dots, D_A(N_A - 1)), \\ &\quad \min(D_B(0), \dots, D_B(N_B - 1))\} \\ D_H &= \min\{\max(D_A(0), \dots, D_A(N_A - 1)), \\ &\quad \max(D_B(0), \dots, D_B(N_B - 1))\} \end{aligned} \quad (11)$$

where $D(0)$ is the lowest and $D(N-1)$ is the highest measured quality point, for either the tested or reference sets, as shown in Fig. 4.

In the HEVC verification test, the number of test points for both the reference and evaluated sets is four (i.e., $N_A = N_B = 4$) and the curve fitting uses cubic spline interpolation.

As can be observed from Fig. 4, in some cases, the overall BD-rate measure may be computed over a relatively small interval of overlapping distortion regions. In such a case, the BD-rate metric does not necessarily represent average coding efficiency for all test points involved in the actual test. Therefore, it is important to design the test in a way that the distortion overlap between the two tested codecs covers a range of qualities of interest for specific application.

As the metrics derived from the Bjøntegaard model can be applied to different evaluation criteria, it is important to understand the range on which they are computed. For example, BD-rate can be computed for MOS and PSNR, for the same test material. However, as demonstrated in Fig. 5, the actual bit rates on which the two are computed may not

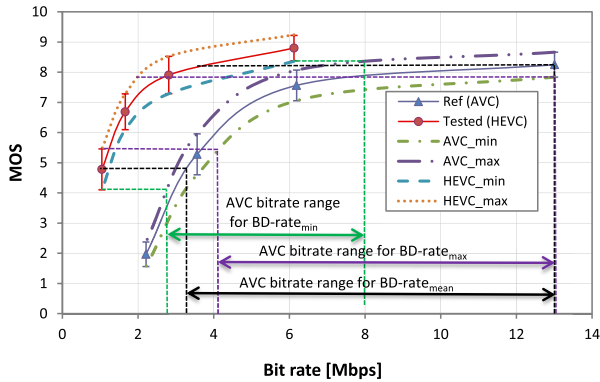


Fig. 6. BD-rate with MOS confidence intervals.

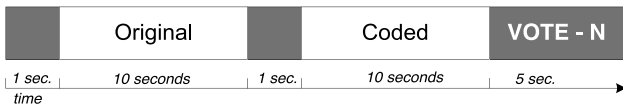


Fig. 7. BTCs for subjective evaluation.

be the same. Therefore, in addition to providing BD-rates for both PSNR and MOS in our evaluation reported in Section IV, we also compute BD-rate on the bit rate interval common for both criteria (MOS and PSNR).

Newer studies [29] have shown how the Bjøntegaard model can further be extended to compute BD-rate intervals considering the confidence intervals of the MOS ratings for each test point, as shown in Fig. 6. The dotted curves show the boundaries of confidence intervals for each curve, and two new BD-rate values are computed comparing $D_{B,\min}$ with $D_{A,\max}$ (labeled $BD\text{-rate}_{\min}$) and $D_{B,\max}$ with $D_{A,\min}$ (labeled $BD\text{-rate}_{\max}$), where $[D_{\min}, D_{\max}]$ represents 95% confidence intervals of MOS. The new BD-rates thus provide lower and upper limits for the BD-rate. However, it is noted that these three values of BD-rate are based on different reference (AVC) bit rate ranges as shown in Fig. 6. Although in the results reported in Section IV, these intervals are reported, they have to be carefully interpreted, as the limits of the intervals are defined for significantly different bit rate ranges. However, for relatively small differences between rate-distortion curves, it can be useful to evaluate BD-rate confidence intervals.

III. TEST SETTINGS

This section provides information regarding the test material used, test settings, and logistics.

A. Selection of Test Material and Test Points

The HEVC verification tests were carried out for four categories of spatial resolutions: UHD (3840×2160), except for the *Traffic* sequence, which is 4096×2048), 1080p (1920×1080), 720p (1280×720), and 480p (832×480). The details of the test sequences are provided in Tables II and III. Screenshots are given in Figs. 8–11. The sequences are selected from different sources and have different spatiotemporal characteristics, which leads to different behavior of compression algorithms. This is the first formal test of video compression standards where a wide range of

TABLE II
TEST SEQUENCES

Sequence category	Sequence	Source / Copyright
UHD	BT709Birthday	Technicolor
	Book	BBC
	Manege	4EVER ¹
	HomelessSleeping	Kamerawerk
	Traffic	Plannet, Inc.
1080p	JohnnyLobby	Vidyo
	Calendar	BBC
	SVT15	SVT
	SedofCropped	4EVER ¹
	UnderBoat1	NTIA
720p	ThreePeople	Vidyo
	BT709Parakeets	Technicolor
	QuarterBackSneak1	NTIA
	SVT01a	SVT
	SVT04a	SVT
480p	Cubicle	Vidyo
	Anemone	NTIA
	BT709BirthdayFlash	Technicolor
	Ducks	Plannet, Inc.
	WheelAndCalendar	BBC

TABLE III
PARAMETERS OF USED TEST SEQUENCES

Sequence	Resolution (width \times height)	Frame rate (fps)	Frames	Cfg
BT709Birthday	3840×2160	50	500	RA
Book	3840×2160	50	500	RA
Manege	3840×2160	60	600	RA
HomelessSleeping	3840×2160	60	600	RA
Traffic	4096×2048	30	300	RA
JohnnyLobby	1920×1080	60	600	LD
Calendar	1920×1080	50	500	RA
SVT15	1920×1080	50	500	RA
SedofCropped	1920×1080	60	600	RA
UnderBoat1	1920×1080	30	300	RA
ThreePeople	1280×720	60	600	LD
BT709Parakeets	1280×720	50	500	RA
QuarterBackSneak1	1280×720	30	300	RA
SVT01a	1280×720	50	500	RA
SVT04a	1280×720	50	500	RA
Cubicle	832×480	30	300	LD
Anemone	832×480	30	300	RA
BT709BirthdayFlash	832×480	50	500	RA
Ducks	832×480	60	600	RA
WheelAndCalendar	832×480	50	500	RA

resolutions including content with UHD resolution with high frame rate has been evaluated. The format, as specified in ITU-R Rec. BT.2020 [30], has a number of extended features compared with ordinary HD video. In addition to containing more pixels per frame, it specifies support for higher frame rates, wider color gamut, and higher bit depths [31]. However, for compatibility with available playout and display systems,

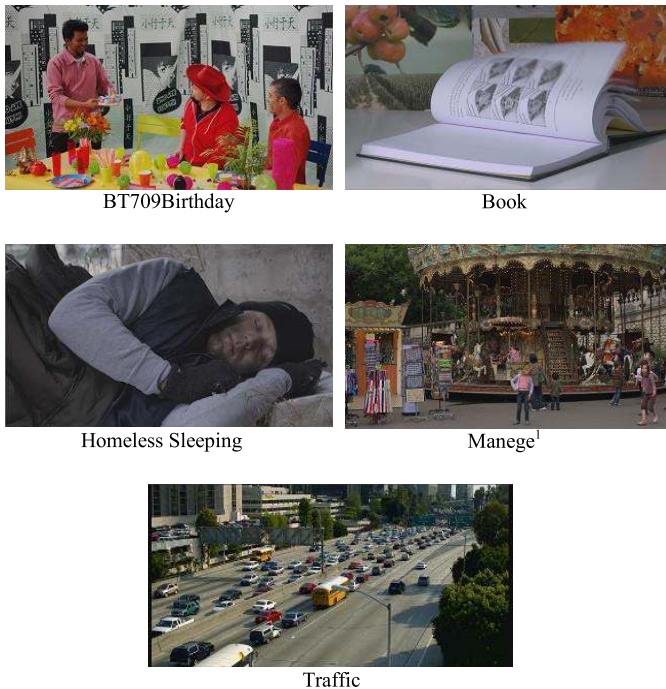


Fig. 8. UHD test sequences.

all tested video sequences have 8 bits per component per sample and are in the $Y'CBC_R$ color space defined by ITU-R Rec. BT.709 [32].

The test sequences were compressed using HEVC (HM-12.1, Main profile [20]) and AVC (JM-18.5, High profile [21]) encoding. Either a random access (RA) or low delay (LD) configuration (Cfg) was used (similarly configured for both HEVC and AVC, with a refresh period of approximately 1 s and hierarchical referencing for RA and with no periodic refresh and no reordered referencing for LD). For each test sequence, four test points using different fixed quantization parameter (QP) settings were selected so that the tested HEVC bit rates are approximately half of the AVC bit rates. Also, the ranges of the QP values were selected so that the subjective quality of the encoded sequences spans a large range of MOS values. This bit rate ratio was chosen because it is already well established that the quality of HEVC is much better than the quality of AVC at the same bit rate. These test conditions can identify whether a bit rate saving of 50% or more is achieved for the majority of the tested video sequences. The full details of the QP values and bit rates can be found in [17].

B. Subjective Test Structure

Subjective tests for different categories of video sequences were conducted in separate sessions. Each subjective test session of the DCR method consisted of a series of basic test cells (BTCs). Each BTC was made of two consecutive presentations of the video clip under test, as shown in Fig. 7. First, the original version of the video sequence was displayed, followed by the coded version of the video sequence, with a gap of 1 s. Then, a message asking the viewers to vote was displayed for 5 s.

Each test session was designed with 45 BTCs in total: the first three BTCs represented the stabilization phase and were

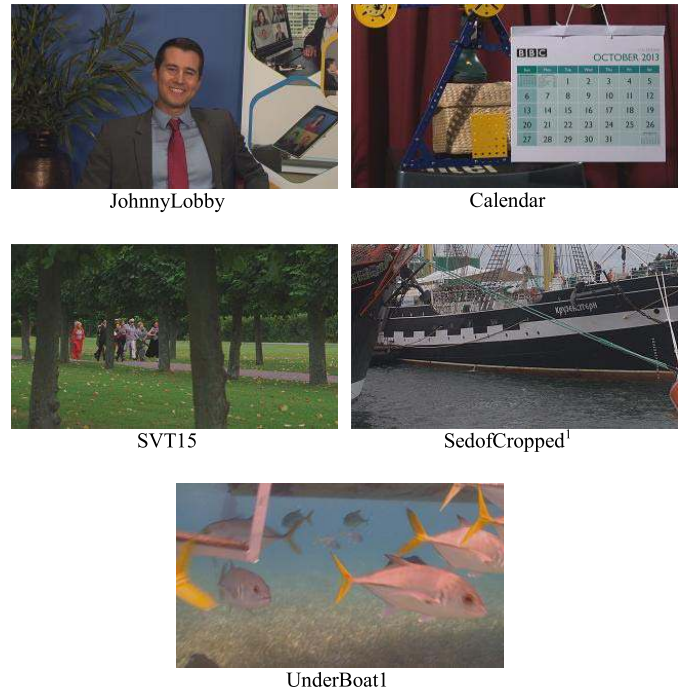


Fig. 9. 1080p test sequences.

selected to show to the viewers the whole range of quality they would see during the test. Two BTCs showing original versus original were also used as a sanity check for the range of ratings made by the viewers. The scores coming from the original BTCs and from the stabilization phase were excluded from further analysis. All the BTCs were randomly ordered to avoid the same content being seen repeatedly and to spread the quality as much as possible in a uniform way across the whole test.

C. Subjective Test Logistics

The subjective tests were performed at two sites, under a controlled laboratory environment, adhering to the recommendations in ITU-R Rec. BT.500 and ITU-T Rec. P.910. The tests for UHD and 720p resolutions were done at the BBC R&D labs in London and for the other resolutions at the University of the West of Scotland (UWS). The equipment and session details for each test are given in Tables IV and V, respectively. Additional analysis of the influence of viewing distances (front and back rows in the seating arrangement for viewers) on the subjective rating has been provided in [19].

IV. RESULTS

This section summarizes the subjective test results. It also provides an analysis with a focus on a comparison with the objective test results, which are easier to obtain in practice.

¹The test sequences *Manege* and *SedofCropped* are “Copyright © 2012-2013, all rights reserved to the 4EVER participants and their licensors. 4EVER consortium: Orange, Technicolor, ATEME, France Télévisions, INSA-IETR, Globecast, TeamCast, Telecom ParisTech, HighlandsTechnologies Solutions, www.4ever-project.com, contact: maryline.clare@orange.com. The 4EVER research Project is coordinated by Orange and has received funding from the French State (FUI/Oseo) and French local Authorities (Région Bretagne) associated with the European funds FEDER.” Copyright holders for other test sequences are single organizations identified in Table II.



Fig. 10. 720p test sequences.

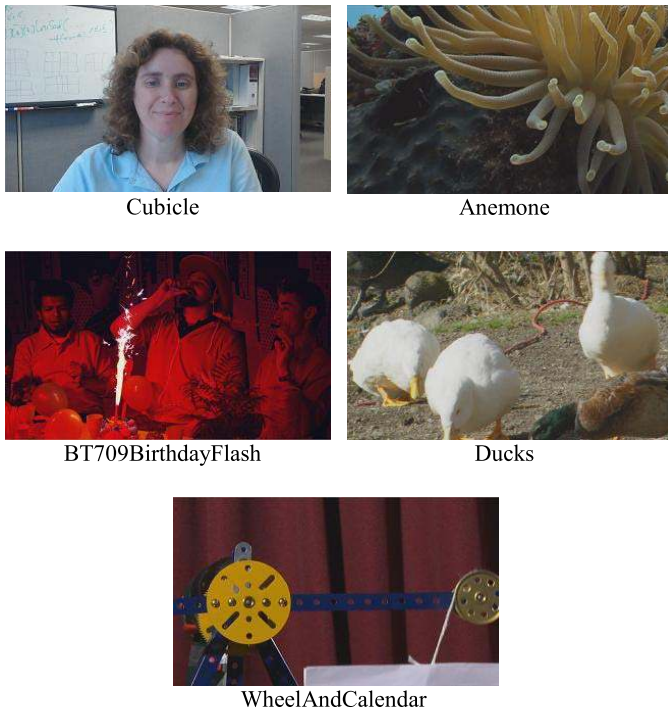


Fig. 11. 480p test sequences.

A. Objective and Subjective Test Results

The subjective evaluation results for each category of test sequences are shown in Figs. 12–15 in the form of MOS versus bit rate plots. The objective quality metric (PSNR) values for the same test points are also plotted on the same graph using a second vertical axis. Note that the scales for the two vertical axes in these plots are independently selected, and thus no direct connection between the subjective (i.e., MOS) and objective (i.e., PSNR) plots is demonstrated. The legend

TABLE IV
TEST LOGISTICS (BBC)

Test performed	UHD test set	720p test set
Display (resolution setting)	TVLogic (56" LUM-560W (3840 × 2160))	Pioneer (50" PDP-5000EX (1920 × 1080))
Sitting position and viewing distance (H is the screen height)	3 viewers at 1.5H (front row), 2 viewers at 2H (back row)	3 viewers at 2H (front row), 2 viewers at 3H (back row)
Total number of viewers	19 (front row), 13 (back row)	19 (front row), 14 (back row)

TABLE V
TEST LOGISTICS (UWS)

Test performed	1080p test set	480p test set
Display (resolution setting)	HP (30" ZR30W (1920 × 1080))	HP (30" ZR30W (1920 × 1080))
Sitting position and viewing distance (H is the screen height)	2 simultaneous displays, 1 viewer at 3H per display	2 simultaneous displays, 1 viewer at 3H per display
Total number of viewers	31	31

in all plots shows that circle and triangle markers represent the results for actual test points, while the curves between them were calculated using cubic spline interpolation with the bit rate in a logarithmic scale, as in typical BD-rate computation. Only the parts of the curves related to BD-rate calculation, either for MOS or PSNR BD-rate computation, are displayed. The solid and dotted lines represent MOS and PSNR curves, respectively. Confidence intervals are displayed for each MOS test point. The PSNR results presented are for the luma color component only. Because of space limitation, the chroma results are not presented. However, the authors note that the weighted PSNR results as in (3) are highly correlated to luma-only PSNR results. Typically, the weighted PSNR has a somewhat higher value than luma PSNR, but the values of BD-rate for weighted PSNR are typically close to the values of BD-rate for luma PSNR.

By considering the positions of MOS points for AVC and HEVC, it can be observed that HEVC achieves the same subjective quality as AVC while typically requiring substantially lower bit rates. Table VI shows the results of the student's *t*-test on the 80 pairs of HEVC and AVC test points. These pairs of test points were classified into the three categories of bit rate savings as described in Section II-C. The first four rows show the distribution of bit rate savings achieved for each resolution. The last row summarizes the distribution of bit rate saving statistics for all resolutions. This shows that for 74 out of the 80 pairs of test points (or 92.5%), HEVC has a bit rate saving compared with AVC that is greater

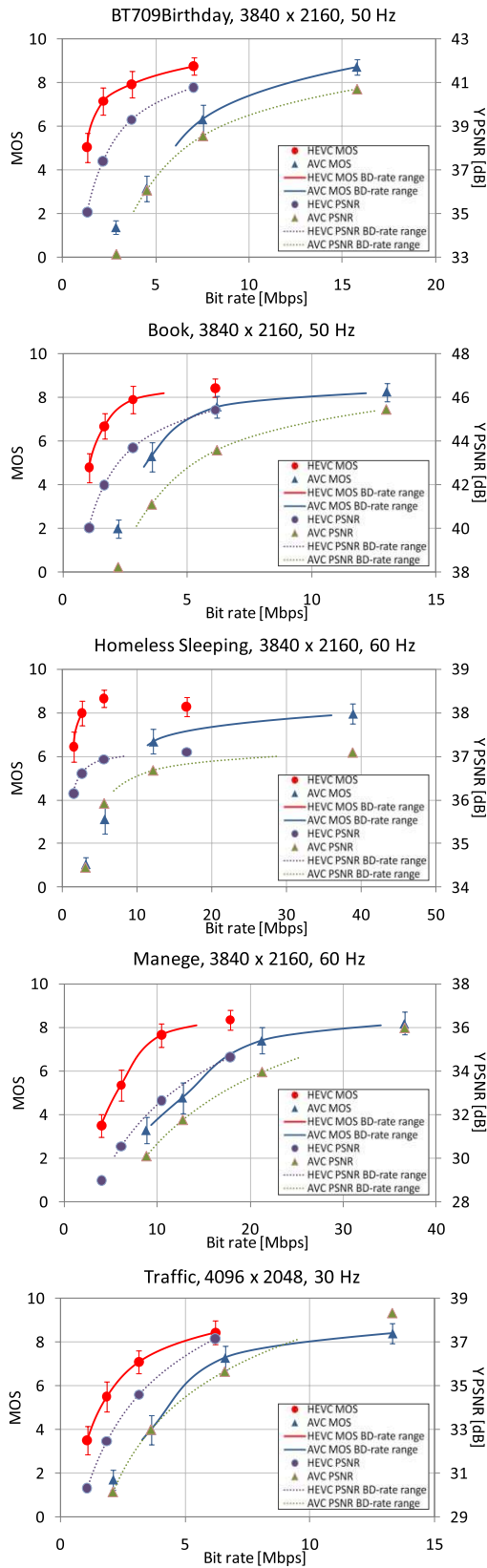


Fig. 12. Subjective and objective evaluation results for UHD content; subjective results, with associated 95% confidence intervals.

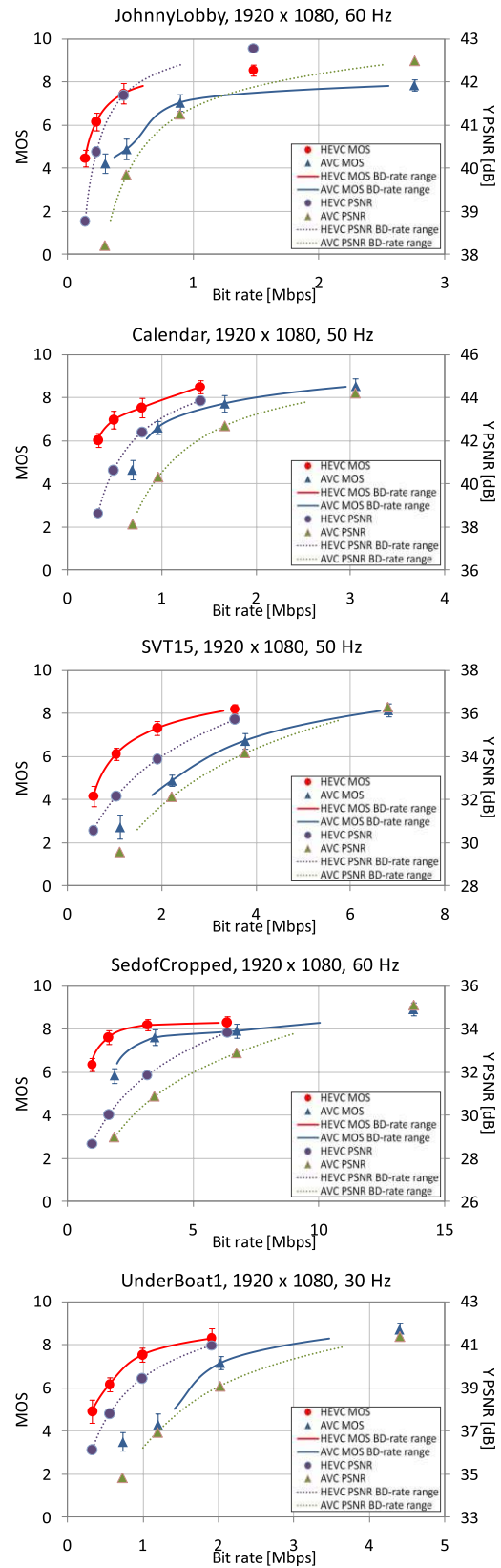


Fig. 13. Subjective and objective evaluation results for 1080p content; subjective results, with associated 95% confidence intervals.

than or equal to 50%. The amount of bit rate saving is similar for both the RA and LD test cases. Only six pairs of test points (or 7.5%) show a bit rate saving of less than 50%. Four of the

six pairs of test points were contributed by one video sequence (*SVT04a*), where the HEVC encoder did not perform as well as in the other cases.

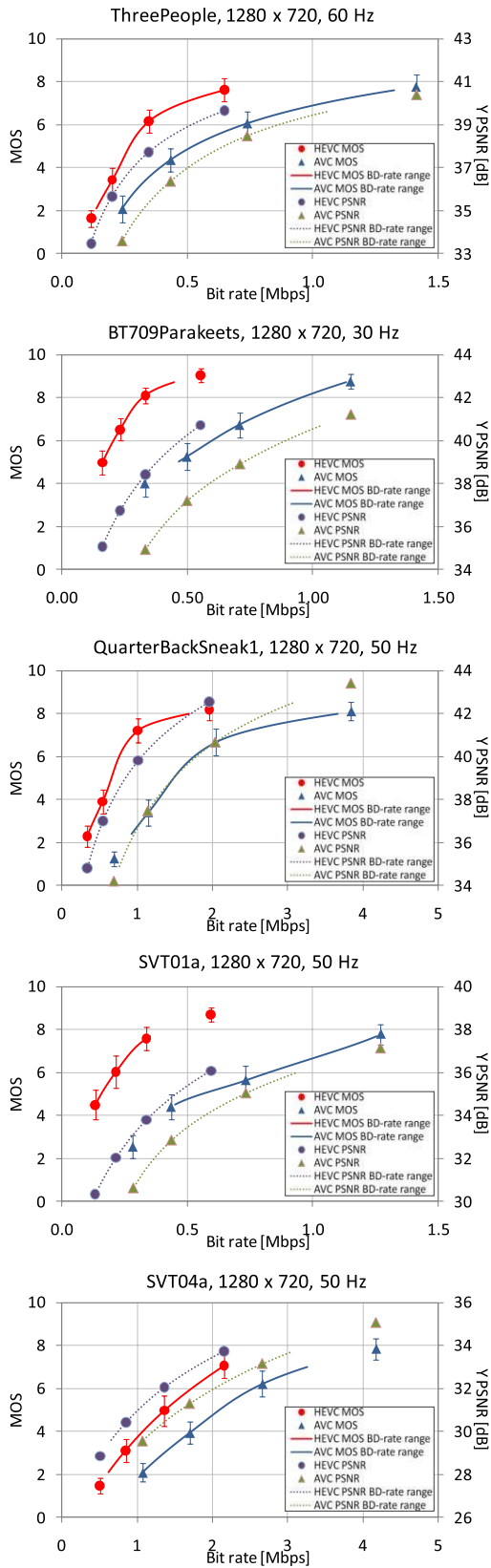


Fig. 14. Subjective and objective evaluation results for 720p content; subjective results, with associated 95% confidence intervals.

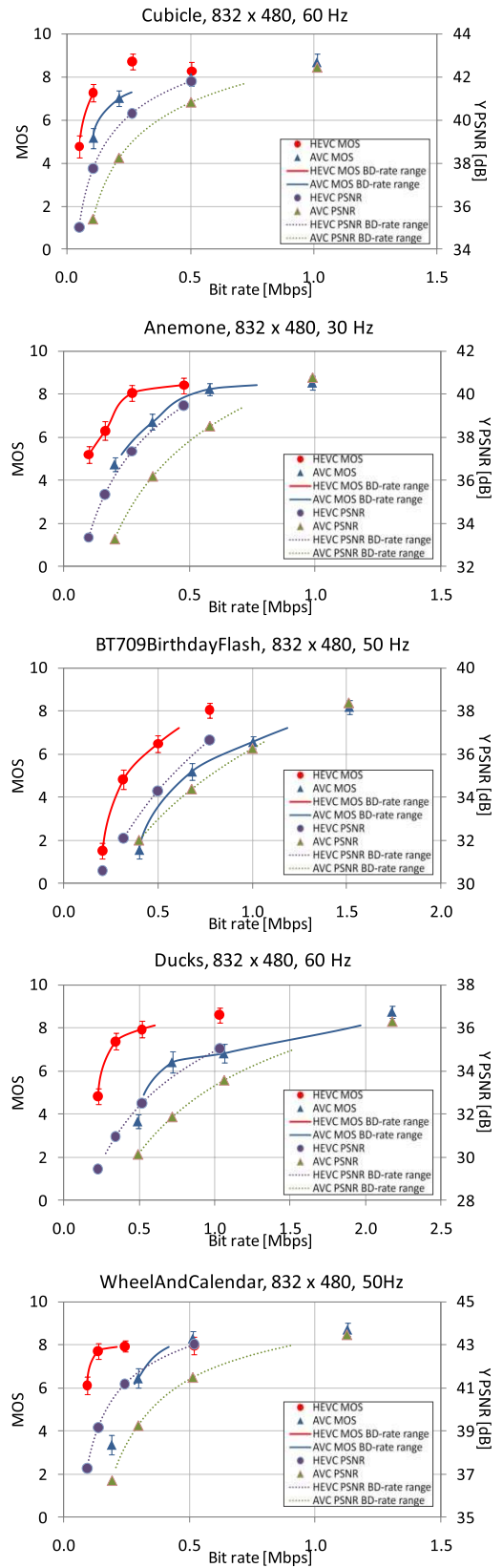


Fig. 15. Subjective and objective evaluation results for 480p content; subjective results, with associated 95% confidence intervals.

The data convincingly show that the HEVC is achieving bit rate saving that is at least 50%. However, the granularity of the above bit rate saving estimation was limited by the

test design, where in each pair of test points, the HEVC bit rate was selected as approximately half the AVC bit rate. In order to get a more precise quantification of the

TABLE VI
CODING GAIN ESTIMATES USING t -TEST ON MOS SCORE

	Bit rate savings of HEVC over AVC is		
	greater than 50%	approximately 50%	less than 50%
UHD	10	10	0
1080p	9	10	1
720p	7	9	4
480p	6	13	1
Total	32	42	6
(%)	(40.0%)	(52.5%)	(7.5%)

TABLE VII
BD-RATES FOR SUBJECTIVE AND OBJECTIVE EVALUATION RESULTS

Sequence	MOS (full range) BD-rate with upper and lower limits	MOS ≥ 7 BD-rate	PSNR BD-rate
UHD			
BT709Birthday	-75 % [-83 %, -66 %]	-70 %	-62 %
Book	-66 % [-76 %, -50 %]	-65 %	-59 %
HomelessSleeping	*	*	-76 %
Manege	-56 % [-68 %, -39 %]	-56 %	-33 %
Traffic	-58 % [-70 %, -44 %]	-50 %	-41 %
1080p			
JohnnyLobby (LD)	-67 % [-76 %, -49 %]	-69 %	-62 %
Calendar	-53 % [-70 %, -34 %]	-49 %	-51 %
SVT15	-65 % [-73 %, -57 %]	-57 %	-50 %
SedofCropped	-57 % [-75 %, -32 %]	-59 %	-35 %
UnderBoat1	-64 % [-71 %, -58 %]	-55 %	-59 %
720p			
ThreePeople (LD)	-48 % [-61 %, -28 %]	-53 %	-44 %
BT709Parakeets	-66 % [-73 %, -55 %]	-65 %	-49 %
QuarterBackSneak1	-58 % [-68 %, -45 %]	-58 %	-45 %
SVT01a	-73 % [-82 %, -59 %]	-73 %	-40 %
SVT04a	-36 % [-52 %, -16 %]	-35 %	-35 %
480p			
Cubicle (LD)	*	*	-62 %
Anemone	-49 % [-60 %, -35 %]	-47 %	-51 %
BT709BirthdayFlash	-49 % [-57 %, -40 %]	-48 %	-50 %
Ducks	-67 % [-76 %, -55 %]	-71 %	-35 %
WheelAndCalender	-66 % [-72 %, -54 %]	-64 %	-59 %
Averages			
Average UHD	-64 %	-60 %	-54 %
Average 1080p	-61 %	-58 %	-51 %
Average 720p	-56 %	-57 %	-43 %
Average 480p	-58 %	-58 %	-51 %

estimated bit rate savings, a different method is used, where the MOS BD-rate values for the subjective ratings for each test sequence, as shown in Table VII, are computed. The upper and lower limits of BD-rates corresponding to the 95% confidence interval of MOS, as discussed in Section II-D, are also indicated.

In addition to the BD-rates for the available MOS range of each video sequence, an additional measurement was also calculated for the range of MOS scores greater than or equal to seven. This range (MOS ≥ 7) is typically expected for a number of services, such as broadcasting, where targeted

TABLE VIII
AVERAGE DIFFERENCES BETWEEN MOS BD-RATE AND PSNR
BD-RATE FOR THE SAME REFERENCE BIT RATES

Sequences	MOS BD-rate	PSNR BD-rate	BD-rate difference
UHD	63 %	47 %	16 %
1080p	61 %	50 %	11 %
720p	56 %	42 %	14 %
480p	57 %	39 %	18 %
Average	59 %	44 %	15 %

quality levels are good to excellent. Negative BD-rate values in Table VII indicate the bit rate savings measured for HEVC relative to the bit rate used for AVC.

The averages in Table VII are computed only for the BD-rate values displayed in the table. Note that the results for MOS BD-rates for the *HomelessSleeping* and *Cubicle* sequences are excluded. In order for the BD-rate interpolation to work correctly, the MOS values should exhibit a smooth curve and the interval of the averaging (shown as curves in Figs. 12–15) should be interpolated from at least three points that are monotonically increasing with bit rate. This condition was not satisfied for the omitted test sequences (although a very substantial gain is evident for HEVC in both omitted cases).

It is noted that for most video sequences, the MOS-based BD-rate benefit is substantially higher than what is measured by the PSNR-based BD-rate. MOS BD-rates indicate that HEVC could provide the same visual quality as AVC for most tested content categories at well below half the bit rate of the latter, surpassing the performance expected at the launch of the HEVC standard development process.

The fact that BD-rates for MOS, confidence intervals, and PSNR in Table VII have not been computed over the same bit rate range for a given test sequence, which has been discussed in Section II, is further addressed in the following section.

B. Compression Efficiency Results for Specific Bit Rates

Although different BD-rate measures presented in Table VII are useful indicators of the compression performance of HEVC, comparing them with each other raises validity concerns, as they are computed for different reference bit rate ranges, as discussed in Section II. To partly address this problem, additional analysis of the results has been conducted to evaluate compression efficiency for bit rates that are common to both PSNR and MOS BD-rate computation.

The bit rate savings achieved by HEVC with reference to the associated AVC bit rate, computed for the continuous bit rate range using cubic spline interpolation as discussed in Section II-D, are shown in Fig. 16, considering both subjective and objective quality assessments. In the majority of cases, the bit rate savings for equal MOS are higher than the bit rate savings for equal PSNR. The savings vary across different bit rates and different video source content. Considering only the parts of the curves from Fig. 16 that are defined both for MOS and PSNR for a given AVC bit rate, we have measured the average BD-rates and the average difference between the PSNR and MOS curves, which are presented

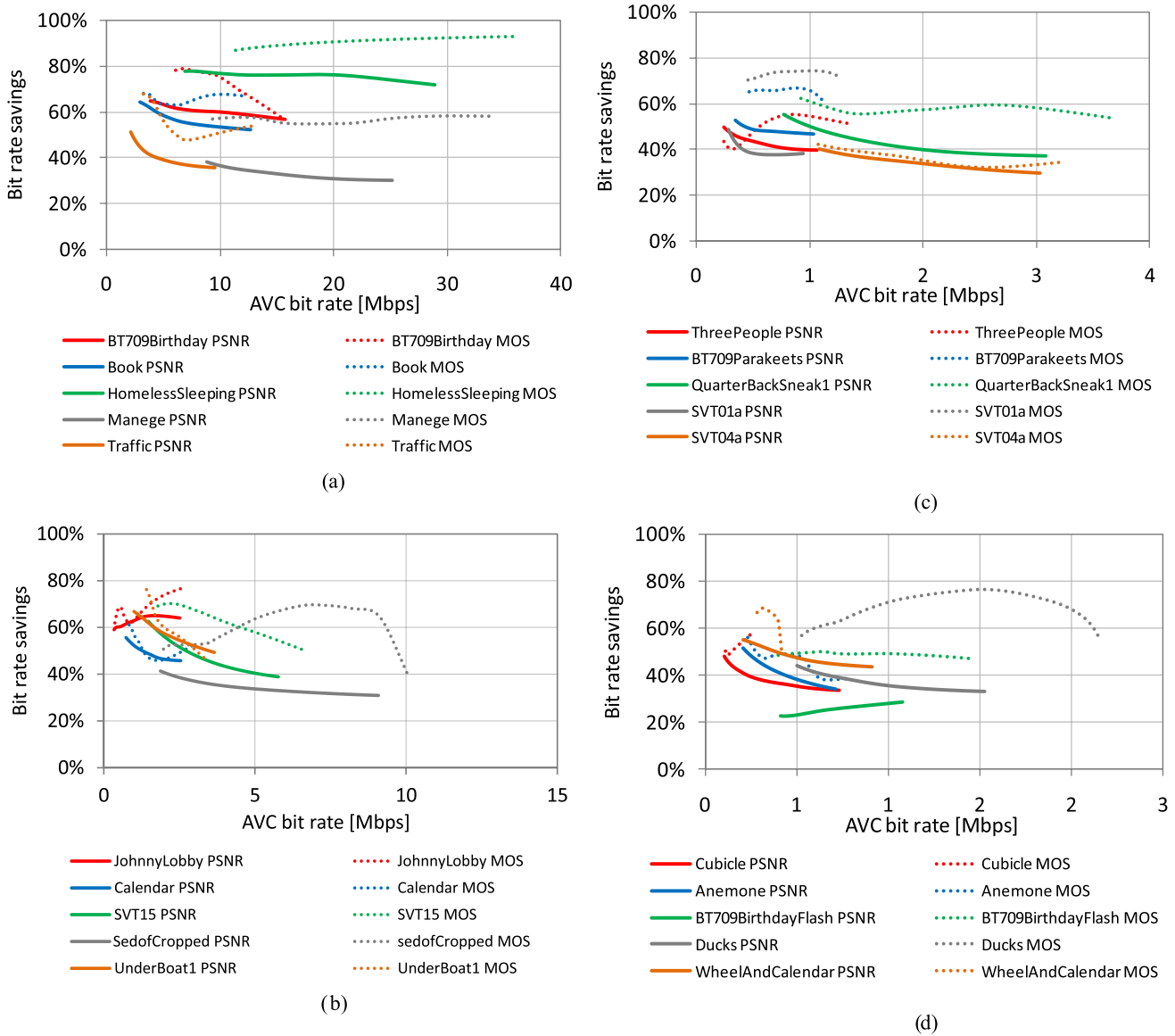


Fig. 16. Bit rate savings provided by HEVC at various reference bit rates. (a) UHD content. (b) 1080p content. (c) 720p content. (d) 480p content.

in Table VIII. The results in Table VIII do not take into account the test sequences that were discarded according to MOS BD-rate computation problems (*HomelessSleeping* and *Cubicle*) described in Section IV-A.

For the same bit rate range, the MOS-based BD-rate saving is 59% and the PSNR-based BD-rate saving is 44%, averaged over all test sequences. Depending on the content category, the average differences between the two measurement methods are between 11% and 18%. In other words, the bit rate savings measured for equal PSNR are lower than the bit rate savings measured for equal MOS between AVC and HEVC by roughly 15%. This is consistent with the difference between the average values in Table VII, with the difference being that in this case the bit rate range for PSNR and MOS rate saving measurement is equal.

V. CONCLUSION

This paper presents the results of a formal subjective verification test that was carried out by the JCT-VC for the new

HEVC video coding standard. In this paper, a more rigorous analysis of the subjective test results using student's *t*-test shows that the HEVC test points at half or less than half the bit rate of the AVC reference were found to achieve a comparable quality in 92.5% of the test cases. In addition, it provides a summary of evaluation metrics and an analysis of the results that were obtained, with a focus on comparison between the subjective and objective evaluation results and the performance across different bit rate ranges.

A more precise quantitative estimate of the bit rate savings was obtained by applying the MOS-based BD-rate measurement on the results of the subjective test. It was found for the investigated test cases that the HEVC Main profile can achieve the same subjective quality as the AVC High profile while requiring on average approximately 59% fewer bits. The PSNR-based BD-rate average over the same sequences was calculated to be 44%. This confirms that the subjective quality improvements of HEVC are typically greater than the objective quality improvements measured by the method

that was primarily used during the standardization process of HEVC.

It can therefore be concluded that the HEVC standard is able to deliver the same subjective quality as AVC, while on average (and in the vast majority of typical sequences) requiring only half or even less than half of the bit rate used by AVC. This means that the initial objective of the HEVC development (substantial improvement in compression compared with the previous state of the art) has been successfully achieved.

ACKNOWLEDGMENT

The authors would like to thank the contributors of the video test sequence material for permitting its use for the tests and this publication, and their collaborators in the JCT-VC for their assistance in developing the test plan and analyzing its results.

REFERENCES

- [1] *Generic Coding of Moving Pictures and Associated Audio Information—Part 2: Video*, ITU-T Rec. H.262 and ISO/IEC 13818-2 (MPEG-2 Video), ITU-T and ISO/IEC JTC 1, 1994.
- [2] *Advanced Video Coding for Generic Audio-Visual Services*, ITU-T Rec. H.264 and ISO/IEC 14496-10, ITU-T and ISO/IEC JTC 1, May 2003.
- [3] *Joint Call for Proposals on Video Compression Technology*, document ITU-T SG 16/Q6 document VCEG-AM91 and WG 11 document N11113, ITU-T SG 16/Q6 and ISO/IEC JTC 1/SC 29/WG 11, 39th VCEG and 91st MPEG meeting, Kyoto, Japan, Jan. 2010.
- [4] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
- [5] *High Efficiency Video Coding*, ITU-T Rec. H.265 and ISO/IEC 23008-2, ITU-T and ISO/IEC JTC 1, Apr. 2013.
- [6] G. J. Sullivan, J. M. Boyce, Y. Chen, J.-R. Ohm, C. A. Segall, and A. Vetro, "Standardized extensions of High Efficiency Video Coding (HEVC)," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1001–1016, Dec. 2013.
- [7] D. Flynn *et al.*, "Overview of the range extensions for the HEVC standard: Tools, profiles and performance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1.
- [8] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of the SHVC scalable extension to the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1.
- [9] G. Tech, K. Müller, J.-R. Ohm, and A. Vetro, "Overview of the multiview and 3D extensions of High Efficiency Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1.
- [10] J. Xu, R. Joshi, and R. A. Cohen, "Overview of the emerging HEVC screen content coding extension," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1.
- [11] R. Weerakkody and M. Mrak, "High Efficiency Video Coding for ultra high definition television," in *Proc. NEM Summit*, Nantes, France, Oct. 2013, pp. 9–14.
- [12] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards—Including High Efficiency Video Coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012.
- [13] S.-H. Bae, J. Kim, M. Kim, S. Cho, and J. S. Choi, "Assessments of subjective video quality on HEVC-encoded 4K-UHD video for beyond-HDTV broadcasting services," *IEEE Trans. Broadcast.*, vol. 59, no. 2, pp. 209–222, Jun. 2013.
- [14] P. Hanhart, M. Reřábek, F. De Simone, and T. Ebrahimi, "Subjective quality evaluation of the upcoming HEVC video compression standard," *Proc. SPIE, Appl. Digit. Image Process.* XXXV, vol. 8499, p. 84990V, Aug. 2012.
- [15] M. Reřábek and T. Ebrahimi, "Comparison of compression efficiency between HEVC/H.265 and VP9 based on subjective assessments," *Proc. SPIE, Appl. Digit. Image Process.* XXXVII, vol. 9217, p. 92170U, Aug. 2014.
- [16] M. Horowitz *et al.*, "Informal subjective quality comparison of video compression performance of the HEVC and H.264/MPEG-4 AVC standards for low-delay applications," *Proc. SPIE, Appl. Digit. Image Process.* XXXV, vol. 8499, p. 84990W, Oct. 2012.
- [17] T. K. Tan, M. Mrak, V. Baroncini, and N. Ramzan, *HEVC Verification Test Report*, document JCTVC-Q1011, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Valencia, Spain, Mar./Apr. 2014.
- [18] A. Tabatabai *et al.*, "Compression performance analysis in HEVC," *High Efficiency Video Coding: Algorithms and Architectures*, V. Sze, M. Budagavi, and G. J. Sullivan, Eds. New York, NY, USA: Springer, Sep. 2014, pp. 275–302, ch. 9.
- [19] R. Weerakkody, M. Mrak, T. K. Tan, V. Baroncini, G. J. Sullivan, and J.-R. Ohm, "Verification testing of HEVC compression performance for UHD video," in *Proc. IEEE GlobalSIP*, Dec. 2014, pp. 1083–1087.
- [20] JCT-VC of ITU-T SG 16/Q6 and ISO/IEC JTC 1/SC 29/WG 11. (Oct. 2013). *HEVC HM Reference Software Codebase, Version 12.1*. [Online]. Available: http://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/
- [21] JVT of ITU-T SG 16/Q6 and ISO/IEC JTC 1/SC 29/WG 11. (May 2013). *AVC JM Reference Software Codebase, Version 18.5*. [Online]. Available: <http://iphome.hhi.de/suehring/tml/>
- [22] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.
- [23] *Subjective Video Quality Assessment Methods for Multimedia Applications*, ITU-T Rec. P.910, ITU-T, Apr. 2008.
- [24] *Methology for the Subjective Assessment of the Quality of Television Picture*, ITU-R Rec. BT.500, ITU-R, Jan. 2012.
- [25] Wikimedia Foundation. (Apr. 23, 2015). *Statistical Power*. [Online]. Available: http://en.wikipedia.org/w/index.php?title=Statistical_power&oldid=658857799
- [26] A. M. M. Mazen, M. Hemmasi, and M. F. Lewis, "In search of power: A statistical power analysis of contemporary research in strategic management," in *Proc. Acad. Manage.*, 1985, pp. 30–34.
- [27] G. Bjøntegaard, *Calculation of Average PSNR Differences Between RD-Curves*, document VCEG-M33, ITU-T SG 16/Q6, 13th VCEG Meeting, Austin, TX, USA, Apr. 2001.
- [28] G. Bjøntegaard, *Improvements of the BD-PSNR Model*, document VCEG-AI11, ITU-T SG 16/Q6, 35th VCEG Meeting, Berlin, Germany, Jul. 2008.
- [29] P. Hanhart and T. Ebrahimi, "Calculation of average coding efficiency based on subjective quality scores," *J. Vis. Commun. Image Represent.*, vol. 25, no. 3, pp. 555–564, Apr. 2014.
- [30] *Parameter Values for Ultra-High Definition Television Systems for Production and International Programme Exchange*, ITU-R Rec. BT.2020, ITU-R, Aug. 2012.
- [31] M. Sugawara, S.-Y. Choi, and D. Wood, "Ultra-high-definition television (Rec. ITU-R BT.2020): A generational leap in the evolution of television," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 170–174, May 2014.
- [32] *Parameter Values for the HDTV Standards for Production and International Programme Exchange*, ITU-R Rec. BT.709, ITU-R, Apr. 2002.



Thio Keng Tan (S'89–M'94–SM'03) received the B.Sc. degree, the bachelor's degree in electrical and electronics engineering, and the Ph.D. degree in electrical engineering from Monash University, Melbourne, VIC, Australia, in 1987, 1989, and 1994, respectively.

He is currently a Consultant with NTT DOCOMO, Inc., Tokyo, Japan. He is an active participant at the video subgroup of the ISO/IEC JTC 1/SC 29/WG 11 Moving Picture Experts Group (MPEG), the ITU-T SG 16/Q6 Video Coding Experts Group, and the ITU-T/ISO/IEC Joint Video Team and the ITU-T/ISO/IEC Joint Collaborative Team for Video Coding standardization activities. He holds over 60 granted U.S. patents. His research interests include image and video coding, analysis, and processing.

Dr. Tan received the Douglas Lampard Electrical Engineering Medal for his Ph.D. thesis and the first prize IEEE Region 10 Student Paper Award for his final year undergraduate project. He also received three ISO certificates for outstanding contributions to the development of the MPEG-4 standard. He has served on the Editorial Board of IEEE TRANSACTIONS ON IMAGE PROCESSING.



Rajitha Weerakkody received the B.Sc. degree in electronic and telecommunication engineering from University of Moratuwa, Moratuwa, Sri Lanka, in 2000 and the Ph.D. degree in electronics engineering from University of Surrey, Guildford, U.K., in 2008.

He was a Network Engineer and Manager with the wireless telecommunication industry in Sri Lanka from 2000 to 2005. He joined British Broadcasting Corporation, London, U.K., in 2008, for research on video coding for audio-visual archiving applications, and managed the technical design and build of the U.K. distribution and public viewing network for the ultra-high definition–super hi-vision showcase at the London 2012 Olympic Games. He is currently a Research Technologist with British Broadcasting Corporation, where he is involved in video compression technology. He has authored or co-authored over 50 research publications.



Marta Mrak (SM'13) received the Dipl.-Ing. and M.Sc. degrees in electronics engineering from University of Zagreb, Zagreb, Croatia, and the Ph.D. degree from Queen Mary University of London, London, U.K.

She was a Post-Doctoral Researcher with University of Surrey, Guildford, U.K., and the Queen Mary University of London. She joined the Research and Development Department, British Broadcasting Corporation, London, in 2010, to work on video compression research and the H.265/High Efficiency Video Coding standardization. She has co-authored over 100 papers, book chapters, and standardization contributions, and also co-edited a book entitled *High-Quality Visual Experience* (Springer, 2010).

Dr. Mrak is a member of the IEEE Multimedia Signal Processing Technical Committee, an Area Editor of *Signal Processing: Image Communication* (Elsevier), and a Guest Editor of several special issues in relevant journals. She received the German DAAD Scholarship for video compression research from the Heinrich Hertz Institute, Berlin, Germany, in 2002. She has been involved in several projects funded by European and U.K. research councils in roles ranging from Researcher to Scientific Coordinator.



Naem Ramzan (S'04–M'08–SM'13) received the M.Sc. degree in telecommunication from University of Brest, Brest, France, in 2004 and the Ph.D. degree in electronics engineering from Queen Mary University of London, London, U.K., in 2008.

He was a Senior Fellow Researcher and Lecturer with Queen Mary University of London. He joined University of the West of Scotland, Paisley, U.K., where he is currently a Reader in Visual Communication. He has authored or co-authored over 70 research publications, including journals, book chapters, and standardization contributions. He co-edited a book entitled *Social Media Retrieval* (Springer, 2013).

Dr. Ramzan is a fellow of the Higher Education Academy and a member of Institution of Engineering and Technology. He served as a Guest Editor for a number of special issues in technical journals. He has organized and co-chaired three ACM Multimedia Workshops, and served as the Session Chair/Co-Chair for a number of conferences. He is the Co-Chair of the Ultra HD Group of the Video Quality Experts Group (VQEG) and the Co-Editor-in-Chief of *VQEG E-Letter*. He has participated in several projects funded by European and U.K. research councils.



Vittorio Baroncini received the bachelor's degree in physics from Rome University "La Sapienza."

He was a Hardware Designer and Group Leader with ITT R&D Laboratory, beginning in 1976, and he was Counsellor with the Italian Telecommunication Ministry in 1984. In 1986, he joined Fondazione Ugo Bordoni (FUB), Rome, Italy, in the Television Group, as a Hardware Designer and Visual Quality Expert. In 1992, he began his activity in the CCIR (ITU-R) and ISO/IEC MPEG standardization organizations. He designed a metric to measure the Quality of Service of TV services with a reduced reference approach. He is a Co-Founder of the Video Quality Experts Group (1997) that led to the standardization of the first world-wide International Standard for an objective quality metric for digital TV (ITU-R Rec. BT.1683). Among his many activities within MPEG, some of the more notable are the tests for Calls for Proposals for Digital Cinema, High Efficiency Video Coding (HEVC), 3D Video (3DV), and HEVC Screen Content Coding extensions, and the recent the Call for Evidence of new technologies for high-dynamic range and wide color gamut video. He is the Technical Leader responsible for video quality assessment and the HDTV, 3DV, and Digital Cinema projects with FUB. He has authored many papers for conferences and scientific journals, and two books on MPEG technologies.



Jens-Rainer Ohm (M'92) received the Dipl.-Ing., Dr. Ing., and Habilitation degrees from Technical University of Berlin, Berlin, Germany, in 1985, 1990, and 1997, respectively.

He has been participating in the Moving Picture Experts Group (MPEG) since 1998. He has been a Full Professor and the Chair of Institute of Communication Engineering with RWTH Aachen University, Aachen, Germany, since 2000. He has authored numerous papers and German and English textbooks in multimedia signal processing, analysis, and coding, and basics of communication engineering and signal transmission. His research interests include motion-compensated, stereoscopic and 3D image processing, multimedia signal coding, transmission and content description, audio signal analysis, and fundamental topics of signal processing and digital communication systems.

Dr. Ohm has been the Chair and Co-Chair of various standardization activities in video coding, namely, the MPEG Video Subgroup since 2002, the Joint Video Team of MPEG and ITU-T SG16 Video Coding Experts Group from 2005 to 2009, and currently the Joint Collaborative Team on Video Coding (JCT-VC) and the Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V).



Gary J. Sullivan (S'83–M'91–SM'01–F'06) received the B.S. and M.Eng. degrees in electrical engineering from University of Louisville, Louisville, KY, USA, in 1982 and 1983, respectively, and the Ph.D. and Engineering degrees in electrical engineering from University of California at Los Angeles, Los Angeles, CA, USA, in 1991.

He has been a longstanding Chairman or Co-Chairman of various video and image coding standardization activities in ITU-T Video Coding Experts Group, ISO/IEC Moving Picture Experts Group, ISO/IEC JPEG, and their joint collaborative teams, since 1996. He was the originator and a Lead Designer of the DirectX Video Acceleration video decoding feature of the Microsoft Windows operating system. He is currently a Video/Image

Technology Architect with the Corporate Standardization Group, Microsoft Corporation, Redmond, WA, USA. His research interests include image and video compression, rate-distortion optimization, motion estimation and compensation, scalar and vector quantization, and loss-resilient video coding.

Dr. Sullivan is a fellow of the International Society for Optics and Photonics. He has received the IEEE Masaru Ibuka Consumer Electronics Award, the IEEE Consumer Electronics Engineering Excellence Award, the INCITS Technical Excellence Award, the IMTC Leadership Award, and the University of Louisville J. B. Speed Professional Award in Engineering. The team efforts he has led have been recognized by the ATAS Primetime Emmy Engineering Award and the NATAS Technology and Engineering Emmy Award. He has been a Guest Editor of several previous special issues and special sections of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.