Recent developments in standardization of high efficiency video coding (HEVC)

Gary J. Sullivan[a] and Jens-Rainer Ohm[b]
[a]Microsoft Corporation, One Microsoft Way, Redmond, WA, USA 98052;
[b]Institute of Communications Engineering, RWTH Aachen University, D-52074 Aachen

ABSTRACT

This paper reports on recent developments in video coding standardization, particularly focusing on the Call for Proposals (CfP) on video coding technology made jointly in January 2010 by ITU-T VCEG and ISO/IEC MPEG and the April 2010 responses to that Call. The new standardization initiative is referred to as High Efficiency Video Coding (HEVC) and its development has been undertaken by a new Joint Collaborative Team on Video Coding (JCT-VC) formed by the two organizations. The HEVC standard is intended to provide significantly better compression capability than the existing AVC (ITU-T H.264 | ISO/IEC MPEG-4 Part 10) standard. The results of the CfP are summarized, and the first steps towards the definition of the HEVC standard are described.

Keywords: HEVC, MPEG, VCEG, JCT-VC, video coding, interoperability, standardization

1. INTRODUCTION

A fundamental figure of merit for a video coding design is its compression capability, which is also referred to as its *coding efficiency*. The coding efficiency relationship between two designs is typically best expressed in terms of percentage savings in bit rate for equal subjective perceptual quality. In addition to enabling service providers to deliver more content at a given quality (e.g., more television channels sent over the same data link or more video stored on the same storage medium), improved coding efficiency can alternatively be used to provide higher quality video (e.g., higher resolution or less distorted video) at a given bit rate, or to provide some other improved balance between bit rate and video quality. Improved coding efficiency can be a compelling advantage.

The last major step forward in video compression capability for world-wide use across a broad variety of applications was the Advanced Video Coding (AVC) standard[1–4], and more specifically the 2004 development of the Fidelity Range Extensions of that standard[4]. The AVC standard is published as ITU-T Rec. H.264 and ISO/IEC MPEG-4 Part 10 (ISO/IEC 14496-10). The Fidelity Range Extensions notably included the specification of a feature set known as the High Profile[4], which has rapidly emerged as the "flagship" technology for essentially all digital video applications. Although several notable further extensions were added to the AVC standard since that time – including Professional Profiles[5], Scalable Video Coding (SVC)[6], and 3D Stereo / Multiview Video Coding (MVC)[7,8] – these efforts that followed the Fidelity Range Extensions have been primarily focused on capabilities other than coding efficiency.

Meanwhile, video content has continued to become an increasing presence in our lives, with increasing diversification of usage models along with ever-increasing demands for higher quality. Consumers now expect higher resolution for their entertainment video, as standard-definition analog broadcast TV and VCR movies have given way to HDTV, DVD, and Blu-ray, and ultra-high definition video has emerged on the horizon. Similarly, interlaced CRT displays have disappeared – replaced by flat-panel displays with ever-increasing size and resolution. Moreover, video has jumped from special-purpose communication links to the Internet and from living room entertainment centers to home and office-based PCs, ubiquitous wall-mounted displays, and an expanding variety of mobile devices.

The premier video coding standardization organizations, namely the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG), have been keeping watch for emerging developments in video coding technology to help address these issues. Recently, it became clear that technology advances were beginning to emerge that could potentially form the basis of the next substantial step forward in coding efficiency.

VCEG had explored the potential for improvements relative to AVC within the framework of an exploration activity that began around 2004. Starting in January 2005, VCEG began designating certain topics as Key Technical Areas (KTA) for

Applications of Digital Image Processing XXXIII, edited by Andrew G. Tescher,
Proc. of SPIE Vol. 7798, 77980V · © 2010 SPIE · CCC code: 0277-786X/10/$18
doi: 10.1117/12.863486

focused investigation. A decision was made in April 2005 to establish a group software codebase for KTA work, based on other group-maintained AVC reference software. A close relationship was maintained between the KTA software and the AVC reference software as the codebases were further refined. In related efforts, MPEG organized several workshops on the topic of future video coding standardization during 2006-2008, inviting presentations of developers of potential technology, and subsequently organized a Call for Evidence on High Performance Video Coding in 2009, where expert viewing tests were conducted in comparison against AVC results. After their individual investigations, both organizations concluded that the time had come to initiate efforts towards the definition of a new generation of video coding standard, and it was considered to perform such work jointly. After reaching a consensus in both groups and establishing the necessary arrangements for joint work, an agreement was reached in January 2010 to establish a Joint Collaborative Team on Video Coding[9] (JCT-VC) and to issue a joint Call for Proposals[10] (CfP). The JCT-VC held its first meeting[11] during 15–23 April 2010 to evaluate the responses to the CfP.

1.1 Design of the Call

Respondents to the CfP were tasked with submitting a set of encodings of 18 source video sequences, which were grouped into five classes of video resolution, ranging from quarter WVGA (416×240) at the low end up to areas of size 2560×1600 cropped from 4K×2K Ultra HD (UHD) material at the high end. All source video test material was progressively scanned and used 4:2:0 YCbCr color sampling with 8 bits per sample. The video sequence encodings that were used in the subjective testing each had a duration of 10 seconds. (The UHD video sequence encodings, which were not used in the subjective testing due to logistic reasons, each had a duration of 5 seconds.) Proponents were required to submit complete results for all test cases – amounting to 145 subjectively-assessed encodings for each proposal and an additional 10 UHD encodings for which only objective analysis was performed. This included encodings for two different constraint conditions (representing *random access* and *low delay* application scenarios as further discussed below) and five rate points as test cases per sequence and constraint condition. Target rate points, which were not to be exceeded by submissions, were as shown in Table 1.

Table 1. Classes of video resolutions and bit rate points used in the CfP.

Class	Rate 1	Rate 2	Rate 3	Rate 4	Rate 5
A: 2560×1600p30	2.5 Mbit/s	3.5 Mbit/s	5 Mbit/s	8 Mbit/s	14 Mbit/s
B1: 1080p24	1 Mbit/s	1.6 Mbit/s	2.5 Mbit/s	4 Mbit/s	6 Mbit/s
B2: 1080p50-60	2 Mbit/s	3 Mbit/s	4.5 Mbit/s	7 Mbit/s	10 Mbit/s
C: WVGAp30-60	384 kbit/s	512 kbit/s	768 kbit/s	1.2 Mbit/s	2 Mbit/s
D: WQVGAp30-60	256 kbit/s	384 kbit/s	512 kbit/s	850 kbit/s	1.5 Mbit/s
E: 720p60	256 kbit/s	384 kbit/s	512 kbit/s	850 kbit/s	1.5 Mbit/s

For each test case, two anchor encodings were generated and their decoded results were included in the formal subjective tests in the same way if they were proposal submissions. The anchors were generated by encoding the above source sequences using a reference AVC encoder (based on the "JM16.2" reference software developed by VCEG and MPEG). The purpose of the anchors was to facilitate the analysis of the test results, providing two reference points to demonstrate the behaviour of well-understood configurations of current technology obeying the same constraints as imposed on the proposals.

The tests were conducted for two types of coding constraint conditions:

- **Random access**: A set of conditions requiring relatively frequent (approximately 1 second) random access refreshes (representing such applications as broadcast entertainment video)
- **Low delay**: A set of conditions requiring low algorithmic delay (representing video usage for real-time communication – with no picture reordering between decoder processing and output).

In the low delay case, the anchor encodings were prepared for two different example configurations of the AVC reference encoder – one having lower encoding and decoding complexity than the other. In the random access case, the two tested anchor encodings were actually identical (which enabled statistical comparison of the resulting measurements).

1.2 Testing Approach

The selection of the test method took into account the quality of the video across the various Classes.

The tests were conducted during March 2010 at three testing laboratories, under the overall coordination of Vittorio Baroncini of Fondazione Ugo Bordoni (FUB):

- FUB (Fondazione Ugo Bordoni, Rome, Italy)
- EPFL (École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland), and
- EBU (European Broadcasting Union, Geneva, Switzerland).

The specifics of the test methodology are described in the test report, which is publicly available[12]. A total of 27 responses were received, which resulted in a total of approximately 23 000 video clips that needed to be tested. For that purpose, more than 130 test sessions (of approximately 20 minutes each) were organized, and a total of 850 test subjects were employed in the viewing – resulting in the collection of approximately 300 000 quality scores. To the extent of our awareness, this was the largest subjective video quality testing effort ever conducted.

A total of 4205 Mean Opinion Score (MOS) values were obtained, which were analyzed and represented on tables and graphs with associated Confidence Interval (CI) values, so that the performance of the proposals could be understood reasonably easily – both in relation to the other proposals and in relation to the performance of the Anchors.

1.3 Results of the Call

The subsequent graphs (Figures 1 and 2) shows results averaged over all of the test sequences; in which the first graph (Figure 1) shows the average results for the Random Access constraint conditions, and the second graph (Figure 2) shows the average results for the Low Delay constraint conditions.

The results were based on an 11 grade scale, where 0 represents the worst and 10 represents the best individual quality measurements. Along with each mean opinion score (MOS) data point in the figures, a 95% confidence interval (CI) is shown.

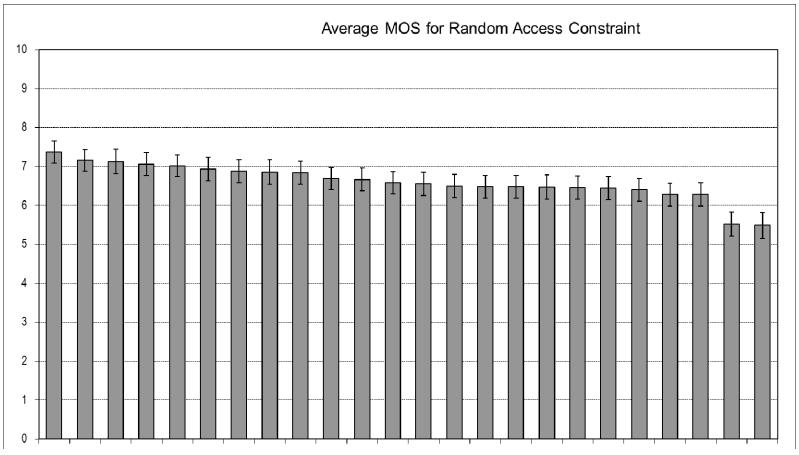

Figure 1. Overall average MOS results over all Classes for Random Access coding conditions.

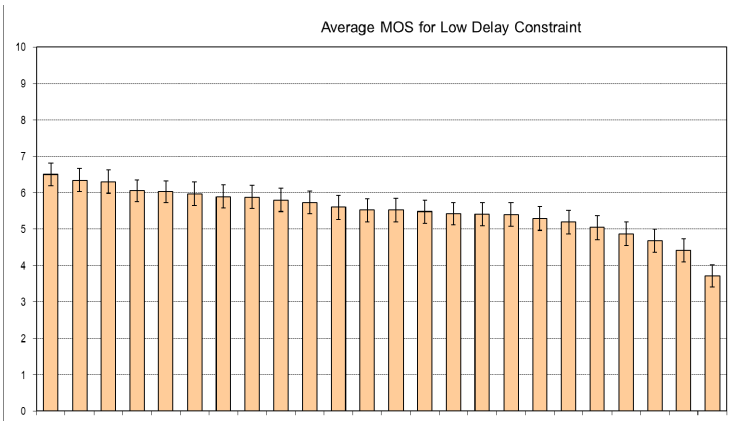

Figure 2. Overall average MOS results over all Classes for Low Delay coding conditions.

For Random Access cases, the same "Alpha" anchor was tested twice and the two results are indicated by the two right-most bars. For Low Delay cases the higher quality "Beta" anchor relating to AVC High Profile is shown second from the right and the lower-complexity "Gamma" anchor relating to AVC Baseline Profile is the right-most case. A significant quality gap can be observed between the AVC anchors and most proposals. From a more detailed analysis performed after the tests and provided in the test report[12], it could be concluded that the best-performing proposals in a significant number of cases showed similar quality as the AVC anchors when encoded at roughly half of the anchor bit rate.

2. DECISIONS AND DEVELOPMENTS FROM THE CFP TESTING

Despite the Eyjafjallajökull volcano eruption, which effectively stopped air travel to Europe as the meeting began, approximately 200 experts participated in the first JCT-VC meeting in Dresden Germany. The subjective test results had indicated that a clear quality improvement has been achieved by many proposals, as compared to the quality of the AVC anchors, for both constraint conditions (Random Access and Low Delay). For a considerable number of test points, the subjective quality of the proposal encoding was as good, for the best performing proposals, as the quality of the anchors when using only about half of the bit rate of the anchors. Even when considering the fact that some proposals certainly used more advanced encoder optimization than the AVC anchors, a substantial gain could be identified for a prospective starting point of the new generation of video coding standard.

The technical assessment of the proposed technology, as it was performed at the first JCT-VC meeting held in Dresden, Germany, between 15 and 23 April 2010, revealed that all proposed algorithms were based on the traditional hybrid coding approach, combining motion-compensated prediction between video frames with intra-picture prediction, closed-loop operation with in-loop filtering, 2D transformation of the spatial residual signals, and advanced adaptive entropy coding. Many specific candidate technology improvements were identified from the proposal responses, as was summarized in technology survey documents issued at the meeting[13,14].

After reviewing the state of the effort, the JCT-VC settled on the project name of "High Efficiency Video Coding" (HEVC) for the new initiative. As an initial step toward moving forward into collaborative work, an initial *Test Model under Consideration* (TMuC) document[15] was produced, combining identified key elements from a group of seven well-performing proposals. This first TMuC became the basis of a first software implementation, which after its development has begun to enable more rigorous assessment of the coding tools that it contains as well as additional tools to be investigated within a process of "Tool Experiments" as planned at the first JCT-VC meeting[11].

2.1 Elements of the Test Model under Consideration

The first TMuC[15] has become the intitial framework in which to study and evaluate a number of key design elements that could likely be part of the HEVC standard. Although this design is still a moving target, which will certainly be somewhat outdated quite soon, we review the initial TMuC design in this section as an initial glimpse at some potential features of that standard.

The first TMuC embraces similar concepts of network abstraction layer (NAL) and high-level syntax (such as frame buffer management, sequence and picture parameter sets) as used in AVC. For the slice layer, some new elements are added as appropriate for additional adaptive elements.

One of the most beneficial elements for higher compression performance in high-resolution video comes due to introduction of larger block structures with flexible mechanisms of sub-partitioning. For this, the TMuC defines *coding units* (CUs) which define a sub-partitioning of a picture into rectangular regions of equal or (typically) variable size. The coding unit replaces the macroblock structure as known from previous video coding standards, and contains one or several *prediction unit(s)* (PUs) and *transform units* (TUs). The basic partition geometry of all these elements is encoded by a scheme similar to the well-known quad-tree segmentation structure. At the level of PU, either intra-picture or inter-picture prediction is selected:

- Intra-picture prediction is performed from samples of already decoded adjacent PUs, where the different modes are DC (flat average), horizontal, vertical, or one of up to 28 angular directions (number depending on block size), plane (amplitude surface) prediction, and bilinear prediction. The signaling of the mode is derived from the modes of adjacent PUs and syntax indicators.
- Inter-picture prediction is performed from region(s) of already decoded pictures stored in a reference picture buffer (with a prediction dependency independent of display order, as in AVC). This allows selection among multiple reference pictures, as well as bi-prediction (including weighted averaging) from two reference pictures or two positions in the same reference picture. The reference area is selected by specifying a motion vector displacement and a reference picture index. In terms of the usage of the motion vector, merging of adjacent PUs is possible, and non-rectangular sub-partitions are also possible in this context. For efficient encoding, skip and direct modes similar to the ones of AVC are defined, and derivation of motion vectors from those of adjacent PUs is performed by various means such as median computation or a new scheme referred to as *motion vector competition*. Motion compensation is performed with a motion vector precision up to quarter-sample precision.

At the level of the TU (which typically would not be larger than the PU), an integer spatial transform similar in concept to the DCT is used, with a selectable block size ranging from 4×4 to 64×64. For the directional intra modes, which usually exhibit directional structures in the prediction residual, special *mode-dependent directional transforms* (MDDT) are employed for block sizes 4×4 and 8×8. Additionally, a *rotational transform* can be used for the cases of block sizes larger than 8×8. The rotational transform adds an additional processing step for spatial transform coefficients, applied only to lower frequency components, such that the related set of basis vectors in the transform matrix is rotated to better adapt for directional 2D structures. Scaling, quantization and scanning of transform coefficient values are performed in a similar manner as for the design in AVC.

At the level of CU, it is possible to switch on an *adaptive loop filter* (ALF) which is applied in the prediction loop prior to copying the frame into the reference picture buffer. This is an FIR filter which is designed with the goal to minimize distortion relative to the original picture (e.g., with a least-squares or Wiener filter optimization). Filter coefficients are encoded at the slice level. In addition, a deblocking filter (similar to the deblocking filter design in AVC) is operated within the prediction loop. The display output of the decoder is written to the decoded picture buffer after applying these two filters.

The TMuC defines two context-adaptive entropy coding schemes, one for operation in a lower-complexity mode, and one for higher-complexity mode:

- The lower-complexity scheme is based on a variable length code (VLC) table selection for all syntax elements (based on either fixed-length code or exponential Golomb code as appropriate), with a particular code table that is selected in a context-dependent fashion based on previous decoded values. This is similar in concept to the CAVLC scheme from AVC, but allows even simpler implementation due to the more systematic structure. As an additional element to potentially improve compression performance, a re-sorting of code table elements can be used (with signaling of the sorting selection).
- The higher-complexity scheme uses a binarization and context adaptation mechanism similar to the CABAC entropy coder of AVC, but uses a set of variable-length-to-variable-length codes (mapping a variable number of bins into a variable number of encoded bits) instead of an arithmetic coding engine. This is performed by employing a bank of parallel VLC coders – each of which is responsible for a certain range of probabilities of binary events (which are referred to as bins). While the coding performance is very similar to CABAC, it can be better

parallelized and has higher throughput per processing cycle in a software or hardware implementation. The compression performance of this scheme is significantly higher than for the low-complexity VLC.

2.2 Experiments and ad hoc groups

A software implementation of the TMuC has been developed. On this basis, the JCT-VC is currently performing a detailed investigation of the performance of the coding tools contained in the TMuC package, as well as other tools that have been proposed in addition to those. Based on the results of such *Tool Experiments* (TE), it is anticipated that the group will define a more well-validated design referred to as a *Test Model* (TM) as the next significant step in HEVC standardization. Specific experiments have been planned relating to tool-by-tool evaluation of the elements of the current TMuC, as well as evaluation of other tools that could give additional benefit in terms of compression capability or complexity reduction in areas such as intra-frame and inter-frame prediction, transforms, entropy coding and motion vector coding. Various *ad hoc groups* (AHGs) have also been set up to perform additional studies on issues such as complexity analysis.

3. CONCLUSIONS

The results of the CfP show that technology exists that can deliver significantly improved compression performance relative to that of the AVC (ITU-T H.264 | ISO/IEC 14496-10), which will soon lead to the definition of a new video compression standard developed jointly by ITU-T and ISO/IEC. The new standardization initiative is now known as High Efficiency Video Coding (HEVC). A Test Model under Consideration (TMuC) has been established, containing a preliminary selection of coding tools that show promise as candidate technologies for inclusion in the standard. The anticipated plan is to have the first version of a well-validated group Test Model (TM) for this standard defined in October 2010. The TM design will then be further developed over various draft standard versions, such that the first version of the new HEVC standard can be expected to be completed in 2012 (or perhaps 2013, depending on the progress of the further work and the eventual final scope of target applications selected for the first version).

ACKNOWLEDGEMENTS

The authors would like to thank the contributors to VCEG, MPEG, and the JCT-VC for their excellent work that is reviewed herein, and the following persons in particular for their great contributions in making the subjective testing of the HEVC CfP results successful: Vittorio Baroncini, Licia Capodiferro, Luca Costantini, Francesca De Simone, Cristina Delogu, Touradj Ebrahimi, Ulrich Engelke, Christine Gabriel, Lutz Goldmann, Ivan Ivanov, Adi Kouadio, Jong-Seok Lee, Federica Mangiatordi, Eugene Myakotnykh, Emiliano Pallotti, Stéphane Pateux, Andrew Perkis, Fitri Rahayu, Ulrich Reiter, Lionel Rhyn, Hamidreza Shirazi, Paolo Sità, Rickard Sjöberg, Christoph Steindl, Peter Vajda, Liyuan Xing, Ashkan Yazdani, and Jungyong You.

REFERENCES

(Note: The referenced JCT-VC documents are publicly available at http://ftp3.itu.ch/av-arch/jctvc-site.)

[1] ITU-T and ISO/IEC, ITU-T Rec. H.264 | ISO/IEC 14496-10 *Advanced Video Coding (AVC)*, May 2003 (with subsequent editions and extensions).

[2] Sullivan, G. J., and Wiegand, T., "Video Compression – From Concepts to the H.264/AVC Standard", *Proc. IEEE* 93(1), 18-31 (2005).

[3] Wiegand, T., Sullivan, G. J., Bjøntegaard, G., and Luthra, A., "Overview of the H.264/AVC Video Coding Standard", *IEEE Trans. Circuits and Systems for Video Tech.*, 13(7), 560–576 (2003).

[4] Sullivan, G. J., Topiwala, P., and Luthra, A., "The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions", in *SPIE Conference on Applications of Digital Image Processing*

XXVII (Special Session on Advances in the Emerging H.264/AVC Video Coding Standard), 5558(1), 454–474 (2004).

[5] Sullivan, G. J., Yu, H., Sekiguchi, S., Sun, H., Wedi, T., Wittmann, S., Lee, Y.-L., Segall, A., and Suzuki, T., "New Standardized Extensions of MPEG4-AVC/H.264 for Professional-Quality Video Applications", *Proc. IEEE Intl. Conf. Image Proc. (ICIP)* I, 13–16 (2007).

[6] Schwarz, H., Marpe, D., and Wiegand, T., "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", *IEEE Trans. Circuits and Systems for Video Tech.* 17(9), 110–1120 (2007).

[7] Sullivan, G. J., "Standards-based approaches to 3D and multiview video coding", *SPIE Applications of Digital Image Processing XXXII* 7443, (2009).

[8] Chen, Y., Wang, Y.-K., Ugur, K., Hannuksela, M. M., and Lainema, J., and Gabbouj, M., "3D video services with the emerging MVC standard", *EURASIP Journal on Advances in Signal Processing,* (2009).

[9] ITU-T VCEG and ISO/IEC MPEG, "Terms of Reference of the Joint Collaborative Team on Video Coding Standard Development", document VCEG-AM90 of VCEG and N11112 of MPEG, (2010).

[10] ITU-T VCEG and ISO/IEC MPEG, "Joint Call for Proposals on Video Compression Technology", document VCEG-AM91 of VCEG and N11113 of MPEG, (2010).

[11] Sullivan, G. J., and Ohm, J.-R., "Meeting report of the first meeting of the Joint Collaborative Team on Video Coding (JCT-VC), Dresden, DE, 15–23 April, 2010", document JCTVC-A200 of JCT-VC, (2010).

[12] Baroncini, V., Sullivan, G. J., and Ohm, J.-R., "Report of subjective testing of responses to Joint Call for Proposals (CfP) on video coding technology for High Efficiency Video Coding (HEVC)", document JCTVC-A204 of JCT-VC, (2010).

[13] JCT-VC, "Architectural Outline of Proposed High Efficiency Video Coding Design Elements", document JCTVC-A202 of JCT-VC, (2010).

[14] JCT-VC, "Table of Proposal Design Elements for High Efficiency Video Coding (HEVC)", document JCTVC-A203 of JCT-VC, (2010).

[15] JCT-VC, "Test Model under Consideration", document JCTVC-A205 of JCT-VC, (2010).