STATUS AND PERSPECTIVES OF VIDEO CODING STANDARDIZATION BEYOND HEVC

Jens-Rainer Ohm and Mathias Wien

RWTH Aachen University, Institut für Nachrichtentechnik, Germany

ABSTRACT

Though High Efficiency Video Coding (HEVC) is still in its take-up phase, it is foreseeable that even better compression will be needed in the future, both in the context of traditional and new application domains. In this regard, a joint exploration activity has been started by ISO/IEC MPEG and ITU-T VCEG under the umbrella of their Joint Video Exploration Team (JVET). Test cases have been defined for various types of video, including HD, UHD, HDR, and 360-degree video, and investigations on advanced compression tools, using a common software platform, have been performed. Evidence obtained so far indicates potential for significant achievable compression efficiency improvements. As a next step towards standardization, a Call for Proposals is planned to be issued with responses expected by spring 2018.

INTRODUCTION

In past developments of video compression standard generations, it took approximately one decade to achieve a doubling in performance (i.e., half rate with same visual quality).

Though the first version of HEVC was not launched earlier than 2013, a next generation appears to be desirable earlier, potentially around 2020, specifically supporting:

- 5G applications, with the advent of ubiquitous wireless video sensors;
- Massive streaming of UHD video content;
- High dynamic range and wide colour gamut (which are emerging and are expected to become mainstream);
- Immersive formats stepping into markets, such as virtual/augmented reality, free viewpoint, light field and beyond.

At the same time, some changes in the ecosystem which has driven the sustainability of standards generations so far are happening:

- Proprietary codecs are increasingly competing against open standards (as they typically have shorter life cycles in their versions and are often running on software platforms).

- Software implementation on smart devices becomes more important than dedicated hardware and even dedicated hardware becomes increasingly programmable.

After all, this might allow for shorter development cycles than traditionally seen, in particular when potential technology exists, providing sufficient performance improvement to justify development of a new generation of standards.

In order to meet these challenges, ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11) have started a joint study about the potential for standardization of video coding technology with a compression capability that significantly exceeds the HEVC standard (Rec. ITU-T H.265 | ISO/IEC 23008-2). Such future standardization could take the form of additional extension(s) of HEVC or an entirely new standard.

To better coordinate this study, VCEG and MPEG created the Joint Video Exploration Team (JVET) as an informal collaboration activity. The scope of JVET includes consideration of a variety of video sources and video applications. Example sources include camera-captured content, screen content, consumer generated content, virtual reality/360º video, and high dynamic range content, while example applications include broadcast (with live or pre-authored content), real-time video conferencing, video chat, on-demand viewing, storage-based media replay, as well as surveillance with fixed or moving cameras.

In order to study the potential of improved compression, JVET is maintaining a software package denoted as “Joint Exploration Model” (JEM). It originated from a version of the HEVC reference software, and was launched with the purpose to investigate the benefit of additional coding tools and algorithms at the video coding layer, whereas the HEVC high-layer syntax and its signalling was mainly left untouched for simplicity of development. It should be emphasized that it is not the purpose of JEM to be considered as the potential basis of a new standard, but rather studying whether add-ons or modifications beyond HEVC would give sufficient increase of compression performance or would allow support for emerging applications, and what the impact on implementation complexity might be. Furthermore, by comparing HEVC versus JEM coding results, test cases for future standard development can be explored and systematically be defined on the basis of foreseen applications, which in the future will allow assessment and evaluation of any proposed technology on a more solid basis.

JVET has held seven meetings so far (as of September 2017). Though JVET is an exploratory group rather than a standardization body, it has well defined modes of operation, which guarantee transparency and reproducibility of results obtained:

- All JVET documents are publicly available from http://phenix.it-sudparis.eu/jvet/.
- The JEM software, which emerged from an earlier version of HEVC HM reference software and was later updated to an HM16 code base, is publicly available from https://jvet.hhi.fraunhofer.de/. A software package denoted as 360Lib, including dedicated tools e.g. for conversion and packing of 360 degree video formats is maintained, interfacing with JEM but can smoothly be used along with the HEVC HM software as well.
- Detailed descriptions of the JEM algorithm (encoder, bitstream, and decoder) [1] and of the 360Lib package [2] are available as JVET documents and are being updated with new inclusions from meeting to meeting.

- An Exploration Experiment (EE) procedure was defined, where newly proposed ideas are validated and studied by a larger group of independent parties, such that a common understanding is achieved before inclusion into JEM.
- Common testing conditions are defined to explore the impact of tools with up-to-date video material. Naturally, these are different for the cases of conventional video [3], HDR [4] and 360 degree video [5].

A Call for Evidence (CfE) has been produced by JVET and was issued jointly by VCEG and MPEG as part of this study [6]. In the CfE, companies and organizations that have developed technology that they believe to have compression capability better than that of the Main 10 Profile of the HEVC standard were kindly invited to bring such information to the attention of JVET. Additionally, contributions were also sought addressing technology that better supports newly emerging application areas of video coding. Responses to the CfE were analysed during the 7th JVET meeting in July 2017, and will be reported in the presentation at IBC.

The following sections give an overview about technology elements contained in JEM, and further report on the comparison to HEVC, both in terms of objective and subjective quality. In the final section, conclusions are drawn and next steps of planned actions are described.

OVERVIEW OF JEM

The overall architecture of the JEM codec is very similar to HEVC. It operates by employing closed-loop prediction with motion compensation from previously decoded reference pictures, or intra prediction from previously decoded areas of the current picture. Also high-level syntax signalling, reference picture buffer management and other high-level elements have been simply copied from HEVC, as applicable. Modifications and add-ons are primarily targeting the coding layer. In this context, the following aspects can be considered as relevant for achieving improvements of compression:

- Enhanced inter prediction, in particular higher precision of motion compensation, e.g., more accurate segmentation/partitioning of motion blocks, usage of non-translational motion models, including improved coding of motion information and partitioning information;
- Improved intra prediction, as particularly relevant in cases where motion compensation fails, e.g. under presence of occlusions; this also benefits from more accurate partitioning;
- Improved coding of the prediction residual, including invocation of adaptive transform bases, and enhancements of entropy coding;
- Enhanced loop filtering, targeting the removal of coding artefacts as well as noise from the prediction signal.

Generally, in the items listed above, the decoder may be designed to make more use of information which is implicitly available from previously decoded parts of the video signal. This may however imply that the computational complexity is increased.

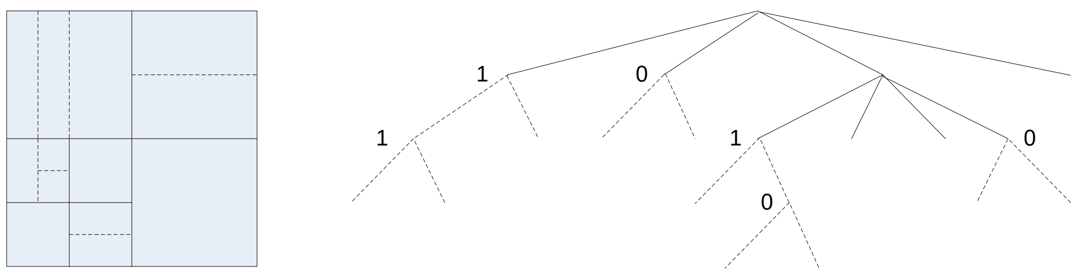

Figure 1 – QTBT block structure (from [1])

Subsequently, some elements of JEM (version 6.0, [1]) are described in more detail:

- Block partitioning: The highest level is denoted as Coding Tree Unit (CTU, square shaped) as in HEVC, however its maximum size is increased to a value of up to 256x256 luma samples. 2D transforms can be applied over areas of up to 64x64 samples. As a major change, block splitting below the CTU level can be performed either by quad or binary split steps, by a method entitled as quad tree binary tree (QTBT), as illustrated in Fig. 1. Finally, distinction between coding units, prediction units and transform units has been abandoned, at the same time allowing for simplified syntax and decoding operations.
- Intra prediction: The number of intra prediction directions is increased from 33 to 65, and a linear model is operated to predict chroma components from luma. Further filtering can be applied, such as a 4-tap interpolation filter, and an additional boundary filter used for more directions than only horizontal and vertical. Intra mode coding is also modified with two additional tools, namely position dependent intra prediction combination (PDPC), and adaptive reference sample smoothing (ARSS).
- Transforms: As a primary transform, different types of DCT and DST can be used for the case of intra coding, indicated by explicit signalling. Furthermore, an intra mode dependent non-separable secondary transform (NSST) is defined. JEM also implements a specific signal dependent transform (SDT), which determines the best transform basis from already decoded neighbour samples; this tool is however disabled by default as it is overly complex.
- Inter prediction: Generally, representation of motion vectors for smaller blocks is made more efficient by sub-block level motion vector prediction and affine motion prediction. Temporal motion vector prediction is also supplemented by more advanced mechanisms and by increasing the resolution of the reference vectors. Furthermore, overlapped block motion compensation (OBMC) and local illumination compensation (LIC) are used. There are also two tools for decoder-side refinement/derivation of motion vectors, including template matched motion vector deriva-

tion and bi-directional optical flow (BIO), both quite effective for improving motion compensation in B picture coding (see Figure 2).

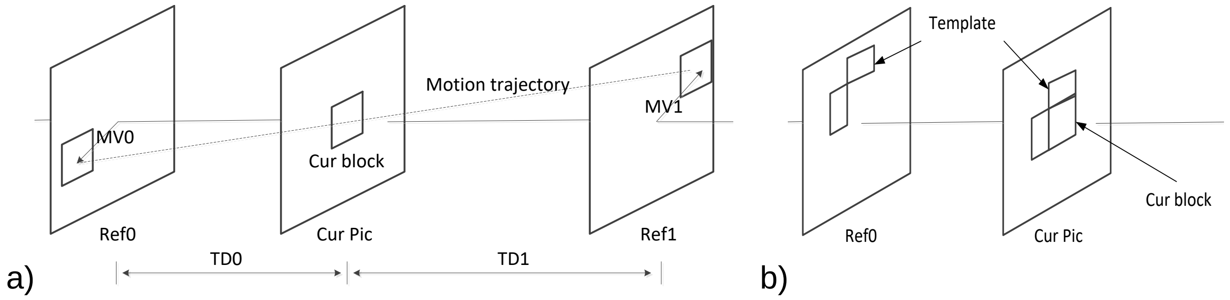

Figure 2 – Decoder-side motion vector derivation:
a) General principle b) Template matching (from [1])

- Loop filter: An adaptive loop filter (ALF) is operated in the prediction loop, which uses a set of explicitly signalled filters and performs a local classification of the picture content to decide which filter from the set to use. Furthermore, a bilateral filter is operated directly following the inverse transform, having the property of denoising of flat areas in combination with edge preservation.
- Entropy coding: An enhanced design of context-adaptive binary arithmetic coding (CABAC) is used, performing context model selection for dedicated transform coefficient levels, using a multi-hypothesis probability estimation, and an improved initialization for context models.

Particularly in the case of (motion compensated) inter prediction, the decoding operation performs at significantly higher computational complexity compared to HEVC (see subsequent section). However, as usual in video coding, the encoder requires an even higher amount of computational power than the decoder, for the purpose of motion estimation, mode decision, decision about partitioning, transforms, etc.

RESULTS OF JEM CODING

As of the current status, the bit rate reduction of JEM6, when compared to an HM16 software encoder implementing the HEVC Main 10 Profile, is around 30 % in Random Access (RA) configuration using motion compensation in a hierarchical B picture structure, and around 20% in All Intra (AI) configuration (without motion compensation). This result was obtained when averaging the bit rate reduction comparing at same PSNR by the so-called Bjøntegaard Delta criterion [7] [8] over a set of more than 20 sequences with resolutions ranging from QVGA, WVGA, HD720p, HD1080p up to 4K UHD [3]. Similar objective gains are observed for cases of 360 degree (4K panorama resolution) and HDR video (HD resolution), for which test cases are described in [4] and [5], respectively.

The increase of computational complexity compared to HEVC is reflected by the fact that the encoder software run time increases by factors of approximately 12 and 60 in RA and AI configurations, respectively. Correspondingly, the decoder run time increases by factors of approximately 10 and 2.5 in RA and AI, respectively. These are average numbers over the entire set of sequences from [3]. The worst case complexity may even be more dramatically higher compared to HEVC.

Measuring PSNR can be misleading as a criterion for judging quality. In order to assess the subjective visual benefit, expert viewing tests were performed during the 7th JVET meeting in the context of evaluating responses to the Call for Evidence [6]. Herein, JEM and HEVC HM, as well as some other codecs that were modified versions of JEM were investigated for subjective quality. Four classes of test sequences were used, where it was required to perform decoding from bitstreams with four matching rate points of all codecs under test:

- Standard Dynamic Range with 4K resolution.
- Standard Dynamic Range with HD resolution.
- High Dynamic Range with HD resolution.
- 360° Video with panorama encoded as 4K resolution, from which dynamic viewport projections of size 1816x1816 were shown on an UHD.

In all four classes, identical Random Access configuration settings were used. For full details, the reader is referred to [9]. An expert viewing protocol (EVP) [10] was used with a five grade impairment scale. Representative results of the four sequence classes are shown in Figures 3 to 6, where JEM and HM are explicitly identified, whereas the other codecs are indicated as “Px”, x being some letter. It becomes evident that in particular at low rates, where the quality of HEVC HM would dramatically break down, significant improvements in visual quality are achieved by JEM and the other new codecs. Typically, bit rate savings at same visual quality of between approximately 35 and 60% were observed.

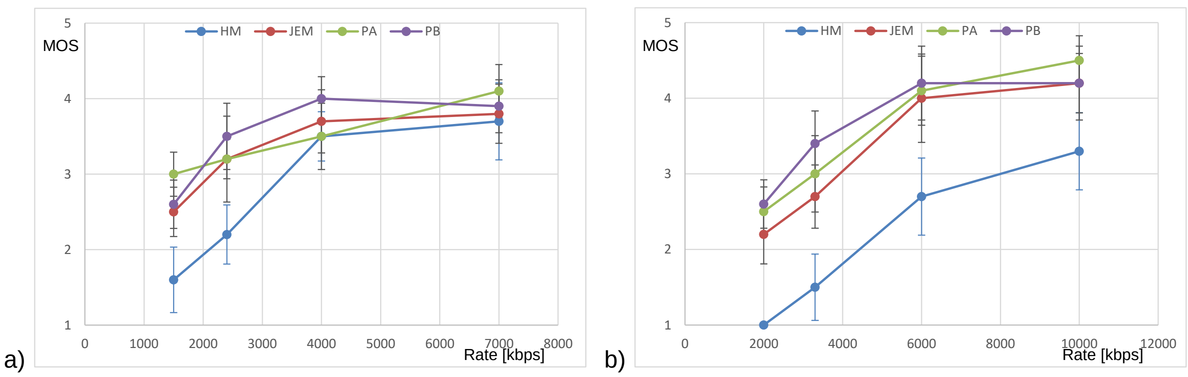

Figure 3 – Results of expert viewing test, SDR 4K video sequences, Random Access configuration: a) “Daylight Road” b) “Campfire Party” (from [9])

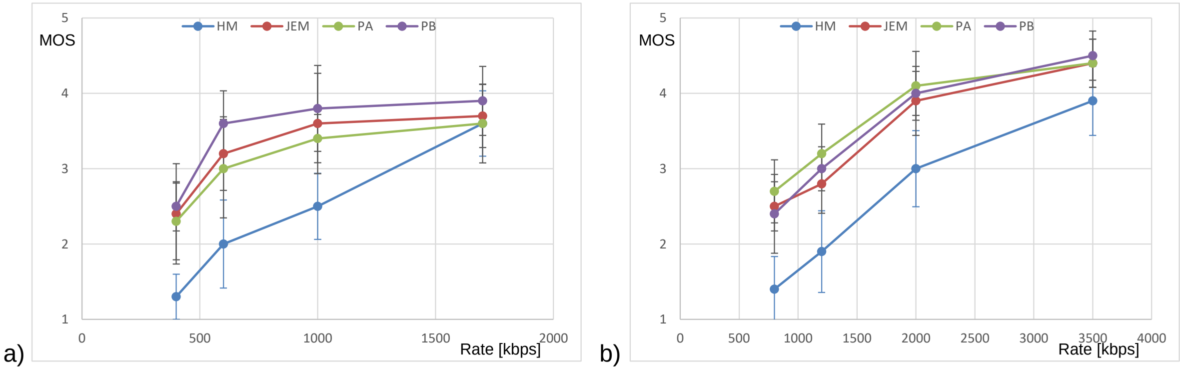

Figure 4 – Results of expert viewing test, SDR HD video sequences, Random Access configuration: a) “BQ Terrace” b) “Cactus” (from [9])

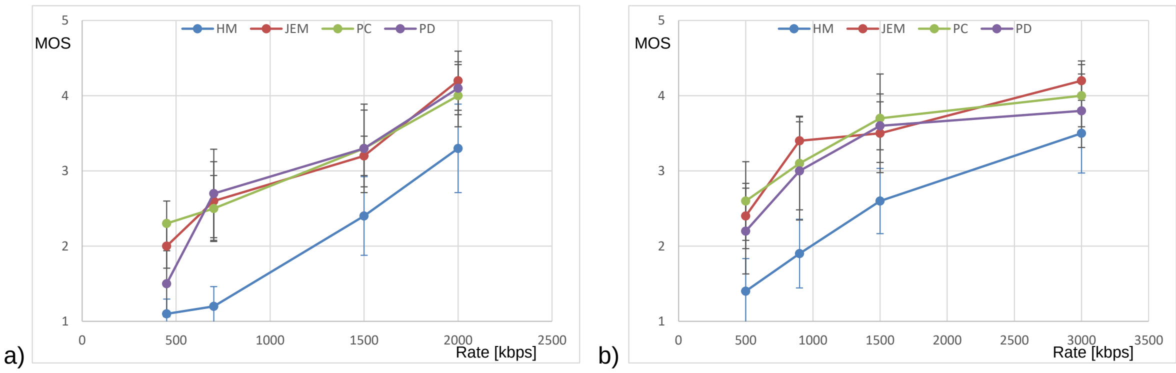

Figure 5 – Results of expert viewing test, HDR HD video sequences, Random Access configuration: a) “Hurdles” b) “Cosmos” (from [9])

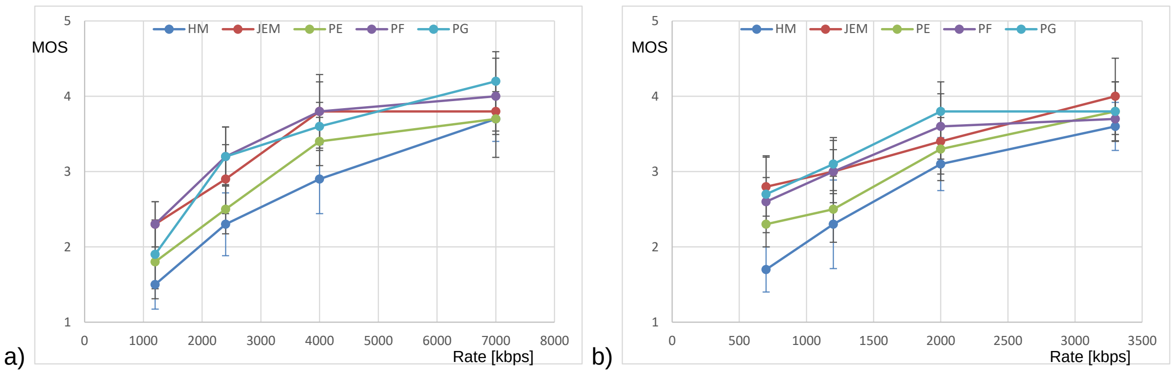

Figure 6 – Results of expert viewing test, 360° video sequences, Random Access configuration: a) “Kite Flite” b) “Harbour” (from [9])

From the findings, two important aspects should be considered in the context of developing a new video compression standard: Firstly, already the JEM development indicates that significant compression gain can only be achieved when modifying core parts of HEVC, such as the block structure, or add new and potentially more complex building blocks to the decoder loop. Secondly, due to the advent of new applications such as 360° and HDR video, it can be expected that further coding tools would be proposed which might have dedicated benefits in such emerging domains. The likelihood that the next generation compression standard would be a simple extension of HEVC can therefore already today be expected of being rather low.

CONCLUSIONS AND NEXT STEPS

Initial results obtained by JVET show clear evidence that video coding tools exist which can significantly improve the compression performance compared to HEVC. These findings look extremely promising, even though, due to complexity, not all elements included currently in JEM might be of practical importance in the context of real-world products. Nevertheless, based on these findings, a formal Call for Proposals (CfP) is planned to be issued by JVET's parent bodies by October 2017 in preparation for starting a formal standardization project. A draft version of this Call is already available [11]. The currently anticipated tentative timeline is as follows:

- Submission deadline: February 2018.
- Evaluation of responses: April 2018.
- First test model: October 2018.
- First version of new video compression standard: October 2020.

REFERENCES

Note: All JVET documents are available from http://phenix.it-sudparis.eu/jvet/.

[1] "Algorithm Description of Joint Exploration Test Model 6 (JEM6)", Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11), 6th Meeting, Hobart, April 2017, Doc. JVET-F1001.

[2] "Algorithm descriptions of projection format conversion and video quality metrics in 360Lib", Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11), 6th Meeting, Hobart, April 2017, Doc. JVET-F1003.

[3] "JVET common test conditions and software reference configurations", Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11), 2nd Meeting, San Diego, February 2016, Doc. JVET-B1010.

[4] “JVET common test conditions and evaluation procedures for HDR/WCG video”, Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11), 6th Meeting, Hobart, April 2017, Doc. JVET-F1020.

[5] “JVET common test conditions and evaluation procedures for 360^0 video”, Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11), 6th Meeting, Hobart, April 2017, Doc. JVET-F1030.

[6] “Joint Call for Evidence on Video Compression with Capability beyond HEVC”, Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11), 6th Meeting, Hobart, April 2017, Doc. JVET-F1002.

[7] Gisle Bjøntegaard, "Calculation of Average PSNR Differences between RD curves", ITU-T SG16/Q6, 13th VCEG Meeting, Austin, Texas, USA, April 2001, Doc. VCEG-M33.

[8] Gisle Bjøntegaard, "Improvements of the BD-PSNR model", ITU-T SG16/Q6, 35th VCEG Meeting, Berlin, Germany, 16–18 July 2008, Doc. VCEG-AI11.

[9] Vittorio Baroncini, Philippe Hanhart, Mathias Wien, Jill Boyce, Andrew Segall, and Teruhiko Suzuki: “Results of the Joint Call for Evidence on Video Compression with Capability beyond HEVC”, Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11), 7th Meeting, Torino, July 2017, Doc. JVET-G1004.

[10] “Subjective assessment of video quality using Expert Viewing Protocol”, ITU-R Recommendation BT.2095-0, Geneva, April 2016.

[11] “Draft Joint Call for Proposals on Video Compression with Capability beyond HEVC”, Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11), 7th Meeting, Torino, July 2017, Doc. JVET-G1002.