|  |  |
| --- | --- |
| **ITU – Telecommunications Standardization Sector**STUDY GROUP 16 Question 6**Video Coding Experts Group (VCEG)**75th Meeting: 2-8 November 2024, Kemer, TR | Document VCEG-BW08-v1 |

|  |  |
| --- | --- |
| Question: | Q.6/SG16 (VCEG) |
| Source: | **Yan Ye (Alibaba)****Marta Karczewicz (Qualcomm)****Mary-Luc Champel (Xiaomi)****Patrice Onno (Canon)****Li Zhang (Bytedance)****Xianglin Wang (Kwai)****Yue Yu (OPPO)****Jie Yuan (vivo)****Yongkai Huo (Transsion)****Ajay Luthra (Picsel Labs)****Edouard François (InterDigital)****Han Boon Teo (Panasonic)****Yoshitaka Kidani (KDDI)****Sung-Chang Lim (ETRI)****Damian Ruiz (Ofinno)****Fangdong Chen (Hikvision)****Walt Husak (Dolby)****Philippe Wetzel (VITEC)****Jay Shingala (Ittiam)****John Juhyung Son (WILUS)** | Email: | yan.ye@alibaba-inc.commartak@qti.qualcomm.comchampelmaryluc@xiaomi.comPatrice.Onno@crf.canon.frlizhang.idm@bytedance.comxianglinwang@kwai.comyue.yu@oppo.comjie.yuan@vivo.comyongkai.huo@transsion.comajay@picsellabs.comEdouard.Francois@interdigital.comhanboon.teo@sg.panasonic.comyo-kidani@kddi.comsclim@etri.re.krdruizcoll@ofinno.comchenfangdong@hikvision.comWJH@dolby.comphilippe.wetzel@vitec.comjay.shingala@ittiam.comjohn.son@wilusgroup.com |
| Title: | **Proposed timeline and requirements for the next generation video coding standard**  |
| Purpose: | Proposal |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

In the July 2024 meeting, WG 2 approved “Preliminary draft of use cases and requirements for potential next-generation video coding beyond VVC capability” in output document N00394 based on input contribution m68403 from fourteen companies. Besides requirements, m68403 also proposed timeline for the next generation video coding standard. This contribution is a follow-up contribution to m68403: the same timeline is proposed, more requirements have been raised for consideration and discussion, and compared to m68403, the list of proponents has grown from fourteen to twenty.

Mirror documents of this contribution will be submitted to MPEG Technical Requirements as m69875 and JVET as JVET-AJ0251.

# Introduction

JVET has been exploring next generation video coding technologies beyond the capability of VVC in the form of exploration experiments (EE) and Ad hoc Groups (AhG) following two main directions: enhanced compression model (ECM) and neural network video coding (NNVC). For the ECM direction, Figure 1 plots the evolution of ECM’s performance (BD rate savings in terms of luma PSNR) in the random access (RA) configuration of the ECM common test conditions [1]. As shown, since ECM-2.0, ECM’s performance has been steadily increasing for every version of ECM. After integration of all tools adopted at the 35th JVET meeting in Sapporo, the latest ECM-14.0 achieves average coding performance gain of 25.9% for camera-captured content and 32-41% for screen content (class F and class TGM) in terms of luma PSNR in the RA configuration. Detailed per-class performance of ECM-14.0 can be obtained from JVET-AJ0006 “JVET AHG report: ECM software development (AHG6)” [2].

****

Figure 1. ECM's compression performance vs. VVC in terms of luma PSNR in the RA configuration

In April 2024, subjective assessments of ECM’s compression capability were conducted in the form of expert viewings and formal viewing using naïve subjects [3]. These assessments showed that, for most test cases, ECM delivers higher BD-rate gain in terms of MOS than in terms of PSNR.

In April 2024, JVET held a joint session with AG5 MPEG visual quality assessment, WG2 MPEG technical requirements, and ITU-T VCEG to discuss these subjective assessment results. As an outcome of the joint discussion, WG2 issued a recommendation in Section 2.1.2 of its output document no. 355 “Recommendations of the 15th MPEG WG2 meeting (MPEG 146)” [3] as follows:

*WG2 recommends MPEG members to submit use cases and requirements in the area of future video standardization beyond VVC.*

In response to this recommendation from WG2, in July 2024, fourteen companies active in JVET’s technical work submitted a contribution [5] (m68403, VCEG-BV02 and JVET-AI0247) to propose timeline and requirements for the next generation video coding standard beyond VVC. The contribution was discussed in multiple joint meeting that were highly attended by many experts. As a result, WG2 MPEG technical requirements issued an output document (N00394) to capture preliminary use cases and requirements for the potential next generation video coding standard.

No agreement on the timeline had been reached at the last JVET meeting. In this contribution, the proponents propose the same timeline as in m68403, VCEG-BV02 and JVET-AI0247. Since the last meeting, the number of proposing companies has grown from fourteen to twenty. It is particularly noteworthy that companies in support of this contribution encompass major video product and service providers including internet companies, chipset vendors, handset manufacturers, camera vendors, network operators, and encoding product vendors; further, these companies cover major economies in the world.

# Proposed timeline for the next generation coding standard

As a starting point and to facilitate effective discussions, the following timeline is proposed:

* Call for Proposals (CfP) to be issued: April or July, 2025
* Responses to the Call to be evaluated: January or April, 2026
* Version 1 finalization: July or October, 2028

This is the same as in m68403/VCEG-BV02/JVET-AI0247 [5]. Further remarks regarding this proposed timeline are provided below:

* Though the proposed timeline does not include a Call for Evidence (CfE) stage, CfE could be planned according to the agreed-upon CfP date;
* The proposed timeline provides at least 6 months of preparation time to interested parties to prepare for their responses to the CfP;
* The proposed finalization date (late 2028) would create an 8-plus-year gap between version 1 of VVC and that of the next generation standard, which we assert to be inline with our usual practice (the gap between AVC and HEVC was ~10 years and the gap between HEVC and VVC was ~7 years);
* VCEG, MPEG, and JVET are not the only “player in town” who develops video codec standards. The proposed timeline is aimed at maintaining an appropriate cycle such that the joint standards of ITU-T and ISO/IEC continue to fulfill industry’s needs in a timely manner. The proponents consider it of great importance for VCEG, MPEG and JVET to remain competitive and relevant.

The proponents welcome constructive feedback around the timeline with the goal of establishing a clear path for progressing JVET’s work toward a next generation video coding standard.

# Additional requirements

Based on the preliminary use cases and requirements captured in [6], this contribution proposes adjustments and refinements to requirements to better satisfy industry needs.

## Use cases and applications

Though [6] includes “Content with film grain” in “content types to be addressed”, it is remarked that special care must be taken to provide a robust and clear solution to film grain content coding. It is noted that the non-normative solutions provided in HEVC and VVC have sometimes been met with confusion by practitioners outside of JVET, VCEG and MPEG. A more preferable solution to film grain content coding may be normative. Note that the intention of this remark is not to discuss what technical solution should be eventually included, but rather to emphasize the importance of sending more clear and definitive marketing messages to industry practitioners.

## Encoder complexity

The following is encoder complexity-related requirement in [6]:

* The encoder complexity shall be capable of trading off complexity and coding efficiency. Real-time encoding with adequate coding efficiency advantage over existing standards shall be feasible at the expected time of usage. Non-real-time encoding with further improvement of coding efficiency should be feasible at the expected time of usage.

As newer standards continue to incorporate more coding tools to achieve higher coding efficiency, an encoder needs to “work harder” to select from a larger set of tools, which often results in rapidly increasing encoder complexity in terms of encoding time of a software encoder and silicon area of a hardware encoder. Further, more applications nowadays are ultra-low-latency driven. Therefore, it is proposed to amend the above encoder complexity requirement as follows to better satisfy an encoder vendor’s needs:

* An encoder implementer shall be capable of trading off encoding complexity (including encoding time on a general processor and hardware encoder’s silicon area) with coding efficiency in an economically viable way. Real-time encoding with substantial coding efficiency advantage over existing standards or with substantial complexity reduction while maintaining the quality of existing standards, non-real-time encoding with further improvement of coding efficiency, and ultra-low-latency encoding (e.g., one picture “glass-to-glass” delay) shall be feasible with meaningfully better coding efficiency than VVC upon the completion of the next generation coding standard.

## Multi-view and 3D video for AR/VR applications

With more 3D video capturing and display devices coming to the market in recent years, multi-view and 3D video is expected to become more widespread and enable more AR/VR applications. For example, starting from iPhone 15 Pro and iPhone 15 Pro Max, users can capture spatial video and display it on the Apple Vision Pro, other VR devices such as the Meta Quest, and 3D TV. The spatial video is captured using the main camera and the ultra wide camera, which are two cameras with different features and capabilities.

In [6], requirements regarding multi-view content coding are only captured as “Stereoscopic and multi-view content” as “content types to be addressed” (Section 1) and as “Stereoscopic 3D, multi-view, depth, alpha, and auxiliary components” as “formats that can be readily supported without major architectural modifications” (Section 4). It is suggested to expand the requirements to include a separate section on multi-view and 3D video, and to include considerations for such video being captured using “asymmetric” cameras. It is remarked that setting aside a separate section will serve to better address the increasing importance and prevalence of AR/VR applications.

## Proposed expanded use cases

In section 1, the expanded use cases for stereoscopic and multiview content should include:

* Stereoscopic equirectangular content for playback in large venues, domes or HMD
* Low-delay 3D representation for remote communication
* Representation of content to allow the rendering of a viewpoint from a limited specified location
* Medical applications for multiview representation

Additional content types should also include:

* Representations for assets created from multiple asymmetric cameras

## Other requirements

The following requirements were proposed in m68403/VCEG-BV03/JVET-AI0247 [5] but not discussed due to lack of time. The proponents suggest to further discuss them at this meeting:

* More and more devices are already able to capture RGBa rather than just RGB in order to enable use cases with transparency such as video overlay. Similarly, many applications are using RGBd format where depth can be either captured or generated, e.g. via AI-based algorithms. Existing coding standards use auxiliary pictures to code these additional video channels. It is suggested that the next generation coding standard should be broadened to a “4-channel” concept so as to code alpha or depth together with the YUV channels in a more coherent manner.
* For some applications, the alpha information would need to be coded with different parameters (for instance, with a lower maximum QP or even with lossless coding) than those used to code the video channels. This is a capability that the next generation coding standard should offer.

# References

1. M. Karczewicz, Y. Ye, “Common test conditions and evaluation procedures for enhanced compression tool testing”, JVET 35th meeting, Sapporo, Japan, July 2024, Doc. JVET-AI2017 (available <https://jvet-experts.org/doc_end_user/documents/35_Sapporo/wg11/JVET-AI2017-v1.zip>).
2. V. Seregin, J. Chen, R. Chernyak, F. Le Léannec, K. Zhang, “JVET AHG report: ECM software development (AHG6)”, 36th meeting of JVET, Kemer, Türkiye, November 2024, Doc. JVET-AJ0006.
3. M. Wien, V. Baroncini, G. Baroncini, “Preliminary report on subjective performance evaluation of the ECM”, JVET 34th meeting, Rennes, France, April 2024, Doc. JVET-AH0344 (available at <https://jvet-experts.org/doc_end_user/documents/34_Rennes/wg11/JVET-AH0344-v1.zip>).
4. “Recommendations of the 15th MPEG WG2 meeting (MPEG 146)”, 15th meeting of ISO/IEC JTC1/SC29/WG2 MPEG technical requirements, Doc. N00355, Rennes, France, April 2024 (available at <https://dms.mpeg.expert/doc_end_user/documents/146_Rennes/wg11/MDS23862_WG02_N00355.zip>).
5. Y. Ye, M. Karczewicz, M.-L. Champel, P. Onno, L. Zhang, X. Wang, D. Wang, Z. Lyu, Y. Huo, A. Luthra, E. François, H.-B. Teo, Y. Kidani, S.-C. Lim, “Proposed timeline and requirements for the next generation video coding standard”, MPEG technical requirements Doc. m68403, VCEG Doc. VCEG-BV03, and JVET Doc. JVET-AI0247, Sapporo, Japan, July 2024.
6. “Preliminary draft of use cases and requirements for potential next-generation video coding beyond VVC capability”, 16th meeting of ISO/IEC JTC1/SC29/WG2 MPEG technical requirements, Doc. N00394, Sapporo, Japan, July 2024 (available at <https://dms.mpeg.expert/doc_end_user/documents/147_Sapporo/wg11/MDS24308_WG02_N00394.zip>).

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_