|  |  |
| --- | --- |
| **Joint Video Experts Team (JVET)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29**  23rd Meeting, by teleconference, 7–16 July 2021 | Document: JVET-W\_Notes\_d8 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Meeting Report of the 23rd Meeting of the Joint Video Experts Team (JVET), by teleconference, 7–16 July 2021** | | |
| *Status:* | Report document from the chairs of JVET | | |
| *Purpose:* | Report | | |
| *Author(s) or Contact(s):* | **Gary Sullivan** Microsoft Corp. 1 Microsoft Way Redmond, WA 98052 USA  **Jens-Rainer Ohm** Institute of Communication Engineering RWTH Aachen Melatener Straße 23 D-52074 Aachen | Tel: Email:  Tel: Email: | +1 425 703 5308 [garysull@microsoft.com](mailto:garysull@microsoft.com)  +49 241 80 27671 [ohm@ient.rwth-aachen.de](mailto:ohm@ient.rwth-aachen.de) |
| *Source:* | Chairs of JVET | | |

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

# Summary

The Joint Video Experts Team (JVET) of ITU-T WP3/16 and ISO/IEC JTC 1/‌SC 29 held its twenty-third meeting during 7–16 July 2021 as an online-only meeting. It had previously been planned to be in Prague, CZ, but this plan was changed due to the difficulties resulting from the COVID-19 pandemic. For ISO/IEC purposes, JVET is alternatively designated ISO/IEC JTC 1/‌SC 29/‌WG 5, and this was the fourth meeting as WG 5. The JVET meeting was held under the chairmanship of Dr Gary Sullivan (Microsoft/USA) and Dr Jens-Rainer Ohm (RWTH Aachen/Germany). For rapid access to particular topics in this report, a subject categorization is found (with hyperlinks) in section 2.14 of this document. It is further noted that the unabbreviated name of JVET was formerly known as “Joint Video *Exploration* Team”, but the parent bodies modified it when entering the phase of formal development of Versatile Video Coding (VVC) in April 2018. Furthermore, starting from the twentieth meeting, work items which had originally been conducted by the Joint Collaborative Team on Video Coding (JCT-VC) were continued in JVET as a single joint team, and explorations towards possible future need of standardization in the area of video coding are also conducted by JVET, as negotiated by the parent bodies.

The JVET meeting began at approximately 0500 hours UTC on Wednesday 7 July 2021. Meeting sessions were held on all days except the weekend days of Saturday and Sunday 10 and 11 July 2021, until the meeting was closed at approximately XXXX hours UTC on Friday 16 July 2021. Approximately XXX people attended the JVET meeting, and approximately XXX input documents (not counting crosschecks), 12 AHG reports, 3 CE/EE summary reports, and X BoG reports were discussed. The meeting took place in a collocated fashion with a meeting of various SC29 Working Groups – where WG 5 is representing the Joint Video Coding Team(s) and their activities from the SC 29 parent body. The subject matter of the JVET meeting activities consisted of work on further development and maintenance of the twin-text video coding technology standards *Advanced Video Coding* (AVC), *High Efficiency Video Coding* (HEVC), *Versatile Video Coding* (VVC)*, Coding-independent Code Points (Video)* (CICP), and *Versatile Supplemental Enhancement Information Messages for Coded Video Bitstreams* (VSEI), as well as related technical reports, software and conformance packages. As a primary goal, the JVET meeting reviewed the work that was performed in the interim period since the twenty-second JVET meeting in producing the following documents:

* JVET-V1000 Meeting report
* JVET-V1002 High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Encoder Description Update 15
* JVET-V1004 Errata report items for VVC, HEVC, AVC, Video CICP, and CP usage TR
* JVET-V2002 Algorithm description for Versatile Video Coding and Test Model 13 (VTM 13)
* JVET-V2005 New level and additional SEI messages for VVC (Draft 3)
* JVET-V2006 Additional SEI messages for VSEI (Draft 3)
* JVET-V2011 JVET common test conditions and evaluation procedures for HDR/WCG video
* JVET-V2016 Common Test Conditions and evaluation procedures for neural network-based video coding technology
* JVET-V2017 Common Test Conditions and evaluation procedures for enhanced compression tool testing
* JVET-V2020 VVC verification test report for HD SDR and 360° video content
* JVET-V2021 VVC verification test plan (Draft 6)
* JVET-V2022 CE on Entropy Coding for High Bit Depth and High Bit Rate Coding
* JVET-V2023 Exploration Experiment on Neural Network-based Video Coding (EE1)
* JVET-V2024 Exploration Experiment on Enhanced Compression beyond VVC capability (EE2)

Further important goals were reviewing the results of the CE on Entropy Coding for High Bit Depth and High Bit Rate Coding, of the EE on Neural Network-based Video Coding, of the EE on Enhanced Compression beyond VVC capability, of other technical input on novel aspects of video coding technology, and plan next steps for investigation of candidate technology towards further standard development.

The JVET produced 13 output documents from the current meeting:

* JVET-W1004 Errata report items for VVC, HEVC, AVC, Video CICP, and CP usage TR
* JVET-W2002 Algorithm description for Versatile Video Coding and Test Model 14 (VTM 14)
* JVET-W2005 New level and additional SEI messages for VVC (Draft 4)
* JVET-W2006 Additional SEI messages for VSEI (Draft 4)
* JVET-W2016 Common Test Conditions and evaluation procedures for neural network-based video coding technology
* JVET-W2017 Common Test Conditions and evaluation procedures for enhanced compression tool testing
* JVET-W2019 White paper on VVC
* JVET-W2020 VVC verification test report for HDR video content
* JVET-W2022 Core Experiment on film grain synthesis
* JVET-W2023 Exploration Experiment on neural network-based video coding (EE1)
* JVET-W2024 Exploration Experiment on enhanced compression beyond VVC capability (EE2)
* JVET-W2025 Algorithm description of Enhanced Compression Model 2 (ECM 2)
* JVET-W2026 Conformance testing for VVC operation range extensions (Draft 1)

For the organization and planning of its future work, the JVET established 12 “ad hoc groups” (AHGs) to progress the work on particular subject areas. At this meeting, 1 Core Experiment (CE) and 2 Exploration Experiments (EE) were defined. The next eight JVET meetings were planned for Wed. 6 – Fri. 8 October and Mon. 11 – Fri. 15 October2021 as a virtual meeting under ISO/IEC JTC 1/‌SC 29 auspices; during January 2022 under ITU-T SG16 auspices, location t.b.d.; during Fri. 22 – Fri. 29 April 2022 under ISO/IEC JTC 1/‌SC 29 auspices in Alpbach, AT; during Fri. 15 – Fri. 22 July 2022 under ISO/IEC JTC 1/‌SC 29 auspices in Cologne, DE; during October 2022 under ITU-T SG16 auspices in Geneva, CH; during January 2023 under ISO/IEC JTC 1/‌SC 29 auspices, location t.b.d.; during April 2023 under ISO/IEC JTC 1/‌SC 29 auspices, location t.b.d.; and during July 2023 under ITU-T SG16 auspices in Geneva, CH.

The document distribution site <https://jvet-experts.org/> was used for distribution of all documents. It was noted that the previous site <http://phenix.int-evry.fr/jvet/> is still accessible, but was converted to read-only.

The reflector to be used for discussions by the JVET and all its AHGs is the JVET reflector:  
[jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de) hosted at RWTH Aachen University. For subscription to this list, see <https://lists.rwth-aachen.de/postorius/lists/jvet.lists.rwth-aachen.de/>.

# Administrative topics

## Organization

The ITU-T/ISO/IEC Joint Video Experts Team (JVET) is a group of video coding experts from the ITU-T Study Group 16 Visual Coding Experts Group (VCEG) and the ISO/IEC JTC 1/‌SC 29/‌WG 5. The parent bodies of the JVET are ITU-T WP3/16 and ISO/IEC JTC 1/‌SC 29.

The Joint Video Experts Team (JVET) of ITU-T WP3/16 and ISO/IEC JTC 1/‌SC 29 held its twenty-second meeting during 7–16 July 2021 as an online-only meeting, using Zoom teleconferencing tools. For ISO/IEC purposes, JVET is alternatively designated ISO/IEC JTC 1/‌SC 29/‌WG 5, and this was the fourth meeting as WG 5. The JVET meeting was held under the chairmanship of Dr Gary Sullivan (Microsoft/USA) and Dr Jens-Rainer Ohm (RWTH Aachen/Germany).

It is further noted that the unabbreviated name of JVET was formerly known as “Joint Video *Exploration* Team”, but the parent bodies modified it when entering the phase of formal development of the *Versatile Video Coding* (VVC) and *Versatile Supplemental Enhancement Information Messages for Coded Video Bitstreams* (VSEI) standards. Furthermore, starting from the twentieth meeting, work items which had originally been conducted by the Joint Collaborative Team on Video Coding (JCT-VC) were continued to be conducted in JVET as a single joint team, as negotiated by the parent bodies. This particularly consists of work on:

* *High Efficiency Video Coding* (HEVC) and its extensions, the development of associated conformance test sets, reference software, verification testing, and non-normative guidance information,
* Specification of *Coding-independent Code Points (Video)* (CICP), and associated technical report(s),
* Maintenance and minor enhancement work on the *Advanced Video Coding* (AVC) standard, associated conformance test sets and reference software.

Furthermore, explorations towards possible future need of standardization in the area of video coding are also conducted by JVET. Currently, the following topics are under investigation:

* Exploration on Neural Network-based Video Coding
* Exploration on Enhanced Compression beyond VVC capability

This report contains three important annexes, as follows:

* Annex A contains a list of the documents of the JVET meeting
* Annex B contains a list of the meeting participants, as recorded by the teleconferencing tool used for the meeting
* Annex C contains the meeting recommendations of ISO/IEC JTC 1/‌SC 29/‌WG 5 for purposes of results reporting to ISO/IEC.

## Meeting logistics

Information regarding logistics arrangements for the meeting had been provided via the email reflector [jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de) and at <http://wftp3.itu.int/av-arch/jvet-site/2022_07_W_Virtual/>.

## Primary goals

As a primary goal, the JVET meeting reviewed the work that was performed in the interim period since the twenty-second JVET meeting in producing the following documents:

* JVET-V1000 Meeting report
* JVET-V1002 High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Encoder Description Update 15
* JVET-V1004 Errata report items for VVC, HEVC, AVC, Video CICP, and CP usage TR
* JVET-V2002 Algorithm description for Versatile Video Coding and Test Model 13 (VTM 13)
* JVET-V2005 New level and additional SEI messages for VVC (Draft 3)
* JVET-V2006 Additional SEI messages for VSEI (Draft 3)
* JVET-V2011 JVET common test conditions and evaluation procedures for HDR/WCG video
* JVET-V2016 Common Test Conditions and evaluation procedures for neural network-based video coding technology
* JVET-V2017 Common Test Conditions and evaluation procedures for enhanced compression tool testing
* JVET-V2020 VVC verification test report for HD SDR and 360° video content
* JVET-V2021 VVC verification test plan (Draft 6)
* JVET-V2022 CE on Entropy Coding for High Bit Depth and High Bit Rate Coding
* JVET-V2023 Exploration Experiment on Neural Network-based Video Coding (EE1)
* JVET-V2024 Exploration Experiment on Enhanced Compression beyond VVC capability (EE2)

Further important goals were reviewing the results of the CE on Entropy Coding for High Bit Depth and High Bit Rate Coding, of the EE on Neural Network-based Video Coding, of the EE on Enhanced Compression beyond VVC capability, of other technical input on novel aspects of video coding technology, and plan next steps for investigation of candidate technology towards further standard development.

## Documents and document handling considerations

### General

The document distribution site <https://jvet-experts.org/> was used for distribution of all documents. It was noted that the previous site <http://phenix.int-evry.fr/jvet/> is still accessible, but was converted to read-only.

Registration timestamps, initial upload timestamps, and final upload timestamps are listed in Annex A of this report.

The document registration and upload times and dates listed in Annex A and in headings for documents in this report are in Paris/Geneva time. Dates mentioned for purposes of describing events at the meeting (other than as contribution registration and upload times) follow the local time at the meeting facility.

Highlighting of recorded decisions in this report is practised as follows:

* Decisions made by the group that might affect the normative content of a future standard are identified in this report by prefixing the description of the decision with the string “Decision:”.
* Decisions that affect one of the various software packages but have no normative effect are marked by the string “Decision (SW):”.
* Decisions that fix a “bug” in one of the test model descriptions such as VTM, HM, etc. (an error, oversight, or messiness) or in the associated software package are marked by the string “Decision (BF):”.
* Decisions that are merely editorial without effect on the technical content of a draft standard are marked by the string "Decision (Ed.):". Such editorial decisions are merely suggestions to the editor, who has the discretion to determine the final action taken if their judgment differs.

This meeting report is based primarily on notes taken by the JVET chairs. The preliminary notes were also circulated publicly by ftp and http during the meeting on a daily basis. It should be understood by the reader that 1) some notes may appear in abbreviated form, 2) summaries of the content of contributions are often based on abstracts provided by contributing proponents without an intent to imply endorsement of the views expressed therein, and 3) the depth of discussion of the content of the various contributions in this report is not uniform. Generally, the report is written to include as much information about the contributions and discussions as is feasible (in the interest of aiding study), although this approach may not result in the most polished output report. Expressions such as “X.XX%” indicate that the desired results were not available at the time the information was recorded.

### Late and incomplete document considerations

The formal deadline for registering and uploading non-administrative contributions had been announced as Wednesday, 30 June 2021. Any documents uploaded after 1159 hours Paris/Geneva time on Thursday 1 July 2021 were considered “officially late”, giving a grace period of 12 hours to accommodate those living in different time zones of the world. The deadline does not apply to AHG reports, and other such reports which can only be produced after the availability of other input documents.

All contribution documents with registration numbers higher than JVET-W0132 were registered after the “officially late” deadline (and therefore were also uploaded late). However, some documents in the “late” range might include break-out activity reports that were generated during the meetings, and are therefore better considered as report documents rather than as late contributions.

In many cases, contributions were also revised after the initial version was uploaded. The contribution document archive website retains publicly accessible prior versions in such cases. The timing of late document availability for contributions is generally noted in the section discussing each contribution in this report.

One suggestion to assist with the issue of late submissions was to require the submitters of late contributions and late revisions to describe the characteristics of the late or revised (or missing) material at the beginning of discussion of the contribution. This was agreed to be a helpful approach to be followed at the meeting.

The following technical design proposal contributions were registered and/or uploaded late:

* JVET-W0XXX (a proposal on …), uploaded 07-XX.
* …

It may be observed that some of the above-listed contributions were submissions made in response to issues that arose in discussions during the meeting or from the study of other contributions, and thus could not have been submitted by the ordinary deadline.

The following other document not proposing normative technical content, but with some need for consideration, were registered and/or uploaded late:

* JVET-W0XXX (a document on …), uploaded 07-XX.
* …

All cross-verification reports at this meeting were registered late, and/or uploaded late. In the interest of brevity, these are not specifically identified here. Initial upload times for each document are recorded in Annex A of this report.

The following contribution registrations were noted that were later cancelled, withdrawn, never provided, were cross-checks of a withdrawn contribution, or were registered in error: JVET-W0XXX.

“Placeholder” contribution documents that were basically empty of content, or lacking any results showing benefit for the proposed technology, and obviously uploaded with an intent to provide a more complete submission as a revision, had been agreed to be considered unacceptable and to be rejected in the document management system until a more complete version was available (which would then typically be counted as a late contribution). At the current meeting, this situation did not apply.

Contributions that had significant problems with uploaded versions were not observed.

As a general policy, missing documents were not to be presented, and late documents (and substantial revisions) could only be presented when there was a consensus to consider them and there was sufficient time available for their review. Again, an exception is applied for AHG reports, CE and HLS topic summaries, and other such reports which can only be produced after the availability of other input documents. There were no objections raised by the group regarding presentation of late contributions, although there was some expression of annoyance and remarks on the difficulty of dealing with late contributions and late revisions.

It was remarked that documents that are substantially revised after the initial upload can also be a problem, as this becomes confusing, interferes with study, and puts an extra burden on synchronization of the discussion. This can especially be a problem in cases where the initial upload is clearly incomplete, and in cases where it is difficult to figure out what parts were changed in a revision. For document contributions, revision marking is very helpful to indicate what has been changed. Also, the “comments” field on the web site can be used to indicate what is different in a revision although participants tend to seldom notice what is recorded there.

A few contributions may have had some problems relating to IPR declarations in the initial uploaded versions (missing declarations, declarations saying they were from the wrong companies, etc.). These issues were corrected by later uploaded versions in a reasonably timely fashion in all cases (to the extent of the awareness of the responsible coordinators).

Some other errors were noticed in other initial document uploads (wrong document numbers or meeting dates or meeting locations in headers, etc.) which were generally sorted out in a reasonably timely fashion. The document web site contains an archive of each upload.

### Outputs of the preceding meeting

All output documents of the previous meeting, particularly the meeting report JVET-V1000, the High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Encoder Description Update 15 JVET-V1002, the Errata report items for VVC, VSEI, HEVC, AVC, Video CICP, and CP usage TR JVET-V1004, the Algorithm description for Versatile Video Coding and Test Model 13 (VTM 13) JVET-V2002, the Operation range extensions for VVC (Draft 3) JVET-V2005, the Additional SEI messages for VSEI (Draft 3) JVET-V2006, the JVET common test conditions and evaluation procedures for HDR/WCG video JVET-V2011, the Common Test Conditions and evaluation procedures for neural network-based video coding technology JVET-V2016, the Common Test Conditions and evaluation procedures for enhanced compression tool testing JVET-V2017, the VVC verification test report for HD SDR and 360° video content JVET-V2020, the VVC verification test plan (Draft 6) JVET-V2021, the Description of the CE on Entropy Coding for High Bit Depth and High Bit Rate Coding JVET-V2022, the Description of the EE on Neural Network-based Video Coding JVET-V2023, and the Description of the EE on Enhanced Compression beyond VVC capability JVET-V2024 had been completed and were approved. The software implementations of VTM (versions 13.0), and HDRTools (version 0.22) were also approved.

Only minor editorial issues in meeting report – no need to produce an update was identified (see section 2.12 for details).

It is reported that there was a modification in the AVC extension text (JVET-T1006) before submitting to ITU consent and ISO/IEC ballot, due to an obvious error in referencing VSEI.

The available output documents of the previous meeting and the software had been made available in a reasonably timely fashion.

## Attendance

The list of participants in the JVET meeting can be found in Annex B of this report.

The meeting was open to those qualified to participate either in ITU-T WP3/16 or ISO/IEC JTC 1/‌SC 29/‌WG 5 (including experts who had been personally invited as permitted by ITU-T or ISO/IEC policies).

Participants had been reminded of the need to be properly qualified to attend. Those seeking further information regarding qualifications to attend future meetings may contact the responsible coordinators.

It was further announced that it is necessary to register for the meeting through the ISO Meetings website for ISO/IEC experts or through the Q6/16 rapporteur for ITU-T experts. The password had been sent to registered participants via these channels. Links to the Zoom sessions (without password) were available in the posted meeting logistics information and the calendar of meeting sessions in the JVET web site. No particular problems were observed that resulted in interference with the meeting, nor was anybody identified who would have attended sessions without being authorized.

The following rules were initially set up for the Zoom teleconference meeting:

* Use the “hand-raising” function to enter yourself in the queue to speak (unless otherwise instructed by the session chair). If you are dialed in by phone, request your queue position verbally.
* Stay muted unless you have something to say. People were muted by default when they join and would need to unmute themselves to speak. The chair may mute anyone who is disrupting the proceedings (e.g. by forgetting they have a live microphone while chatting with their family or by causing bad noise or echo).
* Identify who you are and your affiliation when you begin speaking.
* Use your full name and company/organization affiliation in your joining information, as the participation list of Zoom would also be used to compile attendance records.
* Turn on the chat window and watch for chair communication and side commentary there as well as by audio.
* Avoid overloading people’s internet connections by not using video for the teleconferencing calls – only voice and screen sharing. Extensive use of screen sharing is encouraged.

## Agenda

The agenda for the meeting, for the further development and maintenance of the twin-text video coding technology standards *Advanced Video Coding* (AVC), *High Efficiency Video Coding* (HEVC), *Versatile Video Coding* (VVC)*, Coding-independent Code Points (Video)* (CICP), and *Versatile Supplemental Enhancement Information Messages for Coded Video Bitstreams* (VSEI), as well as related technical reports, software and conformance packages, was as follows:

* Opening remarks and review of meeting logistics and communication practices
* Code of conduct policy reminder
* IPR policy reminder and declarations
* Contribution document allocation
* Review of results of the previous meeting
* Reports of ad hoc group (AHG) activities
* Report of core experiment on entropy coding for high bit depth and high bit rate coding
* Report of exploration experiments on neural-network-based video coding
* Report of exploration experiments on enhanced compression beyond VVC capability
* Consideration of contributions on high-level syntax
* Consideration of contributions and communications on project guidance
* Consideration of video coding technology contributions
* Consideration of contributions on conformance and reference software development
* Consideration of contributions on coding-independent code points for video signal type identification
* Consideration of contributions on errata relating to standards in the domain of JVET
* Consideration of contributions on technical reports relating to standards and exploration study activities in the domain of JVET
* Consideration of contributions providing non-normative guidance relating to standards and exploration study activities in the domain of JVET
* Consideration of information contributions
* Coordination of visual quality testing
* Coordination activities with other organizations
* Approval of output documents and associated editing periods
* Future planning: Determination of next steps, discussion of working methods, communication practices, establishment of coordinated experiments (if any), establishment of AHGs, meeting planning, other planning issues
* Other business as appropriate for consideration

Other business as appropriate for considerationThe plans for the times of meeting sessions were established as follows, in UTC (2 hours behind the time in Geneva, Paris; 7 hours ahead of the time in Los Angeles, etc.). No session should last longer than 2 hrs.

* 0500–0700 1st “morning” session [break after 2 hours]
* 0720–0920 2nd “morning” session
* [“overday” break – nearly 12 hours]
* 2100–2300 1st “night” session [break after 2 hours]
* 2320–0120+1 2nd “night” session

It was also pointed out that the session times had been changed from meeting to meeting, such that different time zones of the world might be treated approximately equally fair either in one meeting or another. For the current meeting, the same session times were used as in the 19th JVET meeting (which used to be the second meeting conducted as online meeting)

* 1. ***ISO and IEC Code of Conduct reminders***

Participants were reminded of the ISO and IEC Codes of Conduct, found at

<https://www.iso.org/publication/PUB100397.html>.

<https://www.iecapc.jp/F/IEC_Code_of_Conduct.pdf>

This includes points relating to:

* Respecting others
* Behaving ethically
* Escalating and resolving disputes
* Working for the net benefit of the international community
* Upholding consensus and governance
* Agreeing to a clear purpose and scope
* Participating actively and managing effective representation

## IPR policy reminder

Participants were reminded of the IPR policy established by the parent organizations of the JVET and were referred to the parent body websites for further information. The IPR policy was summarized for the participants.

The ITU-T/ITU-R/ISO/IEC common patent policy shall apply. Participants were particularly reminded that contributions proposing normative technical content shall contain a non-binding informal notice of whether the submitter may have patent rights that would be necessary for implementation of the resulting standard. The notice shall indicate the category of anticipated licensing terms according to the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form.

This obligation is supplemental to, and does not replace, any existing obligations of parties to submit formal IPR declarations to ITU-T/ITU-R/ISO/IEC.

Participants were also reminded of the need to formally report patent rights to the top-level parent bodies (using the common reporting form found on the database listed below) and to make verbal and/or document IPR reports within the JVET necessary in the event that they are aware of unreported patents that are essential to implementation of a standard or of a draft standard under development.

Some relevant links for organizational and IPR policy information are provided below:

* <http://www.itu.int/ITU-T/ipr/index.html> (common patent policy for ITU-T, ITU-R, ISO, and IEC, and guidelines and forms for formal reporting to the parent bodies)
* <http://ftp3.itu.int/av-arch/jvet-site> (JVET contribution templates)
* <http://www.itu.int/ITU-T/dbase/patent/index.html> (ITU-T IPR database)

It is noted that the ITU TSB director’s AHG on IPR had issued a clarification of the IPR reporting process for ITU-T standards, as follows, per SG 16 TD 327 (GEN/16):

“TSB has reported to the TSB Director’s IPR Ad Hoc Group that they are receiving Patent Statement and Licensing Declaration forms regarding technology submitted in Contributions that may not yet be incorporated in a draft new or revised Recommendation. The IPR Ad Hoc Group observes that, while disclosure of patent information is strongly encouraged as early as possible, the premature submission of Patent Statement and Licensing Declaration forms is not an appropriate tool for such purpose.

In cases where a contributor wishes to disclose patents related to technology in Contributions, this can be done in the Contributions themselves, or informed verbally or otherwise in written form to the technical group (e.g. a Rapporteur’s group), disclosure which should then be duly noted in the meeting report for future reference and record keeping.

It should be noted that the TSB may not be able to meaningfully classify Patent Statement and Licensing Declaration forms for technology in Contributions, since sometimes there are no means to identify the exact work item to which the disclosure applies, or there is no way to ascertain whether the proposal in a Contribution would be adopted into a draft Recommendation.

Therefore, patent holders should submit the Patent Statement and Licensing Declaration form at the time the patent holder believes that the patent is essential to the implementation of a draft or approved Recommendation.”

The responsible coordinators invited participants to make any necessary verbal reports of previously-unreported IPR in technology that might be considered as prospective candidate for inclusion in future standards, and opened the floor for such reports: No such verbal reports were made.

## Software copyright disclaimer header reminder

It was noted that the VTM software implementation package uses the same software copyright license header as the HEVC reference software, where the latter had been agreed at the 5th meeting of the JCT-VC and approved by both parent bodies at their collocated meetings at that time. This license header language is based on the BSD license with a preceding sentence declaring that other contributor or third party rights, including patent rights, are not granted by the license, as recorded in [N 10791](http://phenix.it-sudparis.eu/mpeg/doc_end_user/current_document.php?id=27881&id_meeting=16) of the 89th meeting of ISO/IEC JTC 1/‌SC 29/‌WG 11. Both ITU and ISO/IEC will be identified in the <OWNER> and <ORGANIZATION> tags in the header. This software is used in the process of designing the VTM software, and for evaluating proposals for technology to be potentially included in the design. This software or parts thereof might be published by ITU-T and ISO/IEC as an example implementation of a future video coding standard and for use as the basis of products to promote adoption of such technology.

Different copyright statements shall not be committed to the committee software repository (in the absence of subsequent review and approval of any such actions). As noted previously, it must be further understood that any initially-adopted such copyright header statement language could further change in response to new information and guidance on the subject in the future.

These considerations apply to the 360Lib video conversion software and HDRTools as well.

Software packages that had been developed in prior work of the JCT-VC have similar considerations and are maintained according to the past practice in that work.

## Communication practices

The documents for the meeting can be found at <https://jvet-experts.org/>. It was noted that the previous site <http://phenix.int-evry.fr/jvet/> is still accessible, but was converted to read-only. It was reminded to send a notice to the chairs in cases of changes to document titles, authors etc.

JVET email lists are managed through the site <https://lists.rwth-aachen.de/postorius/lists/jvet.lists.rwth-aachen.de/>, and to send email to the reflector, the email address is [jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de). Only members of the reflector can send email to the list. However, membership of the reflector is not limited to qualified JVET participants.

It was emphasized that reflector subscriptions and email sent to the reflector must use real names when subscribing and sending messages and subscribers must respond to inquiries regarding the nature of their interest in the work. The current number of subscribers on the JVET email list was 1318. Furthermore, the JCT-VC email list currently had 1293 subscribers (as of 6 July 2021). Future discussions should be conducted on the JVET reflector rather than the JCT-VC reflector (or JVT reflector), while the old reflectors should be retained for archiving purposes.

For distribution of test sequences, a password-protected ftp site had been set up at RWTH Aachen University, with a mirror site at FhG-HHI. Accredited members of JVET may contact the responsible JVET coordinators to obtain the password information (but the site is not open for use by others).

## Terminology

* **ACT**: Adaptive colour transform
* **AFF**: Adaptive frame-field
* **AI**: All-intra
* **AIF**: Adaptive interpolation filtering
* **ALF**: Adaptive loop filter
* **AMP**: Asymmetric motion partitioning – a motion prediction partitioning for which the sub-regions of a region are not equal in size (in HEVC, being N/2x2N and 3N/2x2N or 2NxN/2 and 2Nx3N/2 with 2N equal to 16 or 32 for the luma component)
* **AMVP**: Adaptive motion vector prediction
* **AMT or MTS**: Adaptive multi-core transform, or multiple transform selection
* **AMVR**: (Locally) adaptive motion vector resolution
* **APS**: Adaptation parameter set
* **ARC**: Adaptive resolution conversion (synonymous with DRC, and a form of RPR)
* **ARMC**: Adaptive re-ordering of merge candidates
* **ARSS**: Adaptive reference sample smoothing
* **ATMVP** or “subblock-based temporal merging candidates”: Alternative temporal motion vector prediction
* **AU**: Access unit
* **AUD**: Access unit delimiter.
* **AVC**: Advanced video coding – the video coding standard formally published as ITU-T Recommendation H.264 and ISO/IEC 14496-10.
* **BA**: Block adaptive.
* **BC**: See CPR or IBC.
* **BCW**: Biprediction with CU based weighting
* **BD**: Bjøntegaard-delta – a method for measuring percentage bit rate savings at equal PSNR or decibels of PSNR benefit at equal bit rate (e.g., as described in document VCEG-M33 of April 2001).
* **BDOF**: Bi-directional optical flow (formerly known as **BIO**).
* **BDPCM**: Block-wise DPCM.
* **BL**: Base layer.
* **BMS**: Benchmark set (no longer used), a former preliminary compilation of coding tools on top of VTM, which provide somewhat better compression performance, but are not deemed mature for standardzation.
* **BoG**: Break-out group.
* **BR**: Bit rate.
* **BT**: Binary tree.
* **BV**: Block vector (used for intra BC prediction).
* **CABAC**: Context-adaptive binary arithmetic coding.
* **CBF**: Coded block flag(s).
* **CC**: May refer to context-coded, common (test) conditions, or cross-component.
* **CCALF**: Cross-component ALF.
* **CCLM**: Cross-component linear model.
* **CCP**: Cross-component prediction.
* **CE**: Core Experiment – a coordinated experiment conducted toward assessment of coding technology.
* **CG**: Coefficient group.
* **CGS**: Colour gamut scalability (historically, coarse-grained scalability).
* **CIIP**: Combined inter/intra prediction.
* **CL-RAS**: Cross-layer random-access skip.
* **CPB**: Coded picture buffer.
* **CPMV**: Control-point motion vector.
* **CPMVP**: Control-point motion vector prediction (used in affine motion model).
* **CPR**: Current-picture referencing, also known as IBC – a technique by which sample values are predicted from other samples in the same picture by means of a displacement vector called a block vector, in a manner conceptually similar to motion-compensated prediction.
* **CST**: Chroma separate tree.
* **CTC**: Common test conditions.
* **CVS**: Coded video sequence.
* **DCI**: Decoder capability information.
* **DCT**: Discrete cosine transform (sometimes used loosely to refer to other transforms with conceptually similar characteristics).
* **DCTIF**: DCT-derived interpolation filter.
* **DF**: Deblocking filter.
* **DMVR**: Decoder-side motion vector refinement.
* **DoCR**: Disposition of comments report.
* **DPB**: Decoded picture buffer.
* **DPCM**: Differential pulse-code modulation.
* **DPS**: Decoding parameter sets.
* **DRC**: Dynamic resolution conversion (synonymous with ARC, and a form of RPR).
* **DT**: Decoding time.
* **DQ**: Dependent quantization.
* **ECS**: Entropy coding synchronization (typically synonymous with WPP).
* **EMT**: Explicit multiple-core transform.
* **EOTF**: Electro-optical transfer function – a function that converts a representation value to a quantity of output light (e.g., light emitted by a display.
* **EPB**: Emulation prevention byte (as in the emulation\_prevention\_byte syntax element).
* **ECV**: Extended Colour Volume (up to WCG).
* **EL**: Enhancement layer.
* **EOS**: End of (coded video) sequence.
* **ET**: Encoding time.
* **FRUC**: Frame rate up conversion (pattern matched motion vector derivation).
* **GCI**: General constraints information.
* **GDR**: Gradual decoding refresh.
* **GOP**: Group of pictures (somewhat ambiguous).
* **GPM**: Geometry partitioning mode
* **GRA**: Gradual random access
* **HBD**: High bit depth
* **HDR**: High dynamic range.
* **HEVC**: High Efficiency Video Coding – the video coding standard developed and extended by the JCT-VC, formalized by ITU-T as Rec. ITU-T H.265 and by ISO/IEC as ISO/IEC 23008-2.
* **HLS**: High-level syntax.
* **HM**: HEVC Test Model – a video coding design containing selected coding tools that constitutes our draft standard design – now also used especially in reference to the (non-normative) encoder algorithms (see WD and TM).
* **HMVP**: History based motion vector prediction.
* **HRD**: Hypothetical reference decoder.
* **HyGT**: Hyper-cube Givens transform (a type of NSST).
* **IBC** (also **Intra BC**): Intra block copy, also known as CPR – a technique by which sample values are predicted from other samples in the same picture by means of a displacement vector called a block vector, in a manner conceptually similar to motion-compensated prediction.
* **IBDI**: Internal bit-depth increase – a technique by which lower bit-depth (8 bits per sample) source video is encoded using higher bit-depth signal processing, ordinarily including higher bit-depth reference picture storage (ordinarily 12 bits per sample).
* **IBF**: Intra boundary filtering.
* **ILP**: Inter-layer prediction (in scalable coding).
* **ILRP**: Inter-layer reference picture.
* **IPCM**: Intra pulse-code modulation (similar in spirit to IPCM in AVC and HEVC).
* **IRAP**: Intra random access picture.
* **ISP**: Intra subblock partitioning
* **JCCR**: Joint coding of chroma residuals
* **JEM**: Joint exploration model – the software codebase for future video coding exploration.
* **JM**: Joint model – the primary software codebase that has been developed for the AVC standard.
* **JSVM**: Joint scalable video model – another software codebase that has been developed for the AVC standard, which includes support for scalable video coding extensions.
* **KLT**: Karhunen-Loève transform.
* **LB** or **LDB**: Low-delay B – the variant of the LD conditions that uses B pictures.
* **LD**: Low delay – one of two sets of coding conditions designed to enable interactive real-time communication, with less emphasis on ease of random access (contrast with RA). Typically refers to LB, although also applies to LP.
* **LFNST**: Low-frequency non-separable transform
* **LIC**: Local illumination compensation.
* **LM**: Linear model.
* **LMCS**: Luma mapping with chroma scaling (formerly sometimes called “in-loop reshaping”)
* **LP** or **LDP**: Low-delay P – the variant of the LD conditions that uses P frames.
* **LUT**: Look-up table.
* **LTRP**: Long-term reference picture.
* **MANE**: Media-aware network element.
* **MC**: Motion compensation.
* **MCP**: Motion compensated prediction.
* **MCTF**: Motion compensated temporal pre-filtering.
* **MDNSST**: Mode dependent non-separable secondary transform.
* **MIP**: Matrix-based intra prediction
* **MMLM**: Multi-model (cross component) linear mode.
* **MMVD**: Merge with MVD.
* **MPEG**: Moving picture experts group (an alliance of working groups and advisory groups in ISO/IEC JTC 1/‌SC 29, one of the two parent bodies of the JVET).
* **MPM**: Most probable mode (in intra prediction).
* **MRL**: Multiple reference line intra prediction.
* **MV**: Motion vector.
* **MVD**: Motion vector difference.
* **NAL**: Network abstraction layer.
* **NSQT**: Non-square quadtree.
* **NSST**: Non-separable secondary transform.
* **NUH**: NAL unit header.
* **NUT**: NAL unit type (as in AVC and HEVC).
* **OBMC**: Overlapped block motion compensation (e.g., as in H.263 Annex F).
* **OETF**: Opto-electronic transfer function – a function that converts to input light (e.g., light input to a camera) to a representation value.
* **OLS**: Output layer set.
* **OOTF**: Optical-to-optical transfer function – a function that converts input light (e.g. l,ight input to a camera) to output light (e.g., light emitted by a display).
* **operation point**: A temporal subset of an OLS.
* **PDPC**: Position-dependent (intra) prediction combination.
* **PERP**: Padded equirectangular projection (a 360° projection format).
* **PH**: Picture header.
* **PHEC**: Padded hybrid equiangular cubemap (a 360° projection format).
* **PMMVD**: Pattern-matched motion vector derivation.
* **POC**: Picture order count.
* **PoR**: Plan of record.
* **PROF**: Prediction refinement with optical flow
* **PPS**: Picture parameter set (as in AVC and HEVC).
* **PTL**: Profile/tier/level combination.
* **QM**: Quantization matrix (as in AVC and HEVC).
* **QP**: Quantization parameter (as in AVC and HEVC, sometimes confused with quantization step size).
* **QT**: Quadtree.
* **RA**: Random access – a set of coding conditions designed to enable relatively-frequent random access points in the coded video data, with less emphasis on minimization of delay (contrast with LD).
* **RADL**: Random-access decodable leading (type of picture).
* **RASL**: Random-access skipped leading (type of picture).
* **R-D**: Rate-distortion.
* **RDO**: Rate-distortion optimization.
* **RDOQ**: Rate-distortion optimized quantization.
* **RDPCM**: Residual DPCM
* **ROT**: Rotation operation for low-frequency transform coefficients.
* **RPL**: Reference picture list.
* **RPLM**: Reference picture list modification.
* **RPR**: Reference picture resampling (e.g., as in H.263 Annex P), a special case of which is also known as ARC or DRC.
* **RPS**: Reference picture set.
* **RQT**: Residual quadtree.
* **RRU**: Reduced-resolution update (e.g. as in H.263 Annex Q).
* **RVM**: Rate variation measure.
* **SAO**: Sample-adaptive offset.
* **SBT**: Subblock transform.
* **SbTMVP**: Subblock based temporal motion vector prediction.
* **SCIPU**: Smallest chroma intra prediction unit.
* **SD**: Slice data; alternatively, standard-definition.
* **SDH**: Sign data hiding.
* **SDT**: Signal-dependent transform.
* **SE**: Syntax element.
* **SEI**: Supplemental enhancement information (as in AVC and HEVC).
* **SH**: Slice header.
* **SHM**: Scalable HM.
* **SHVC**: Scalable high efficiency video coding.
* **SIF**: Switchable (motion) interpolation filter.
* **SIMD**: Single instruction, multiple data.
* **SMVD**: Symmetric MVD.
* **SPS**: Sequence parameter set (as in AVC and HEVC).
* **STMVP**: Spatial-temporal motion vector prediction.
* **STRP**: Short-term reference picture.
* **STSA**: Step-wise temporal sublayer access.
* **TBA/TBD/TBP**: To be announced/determined/presented.
* **TGM**: Text and graphics with motion – a category of content that primarily contains rendered text and graphics with motion, mixed with a relatively small amount of camera-captured content.
* **TMVP**: Temporal motion vector prediction.
* **TS**: Transform skip.
* **TSRC**: Transform skip residual coding.
* **TT**: Ternary tree.
* **UCBDS**: Unrestricted center-biased diamond search.
* **UGC**: User-generated content.
* **UWP**: Unequal weight prediction.
* **VCEG**: Visual coding experts group (ITU-T Q.6/16, the relevant rapporteur group in ITU-T WP3/16, which is one of the two parent bodies of the JVET).
* **VPS**: Video parameter set – a parameter set that describes the overall characteristics of a coded video sequence – conceptually sitting above the SPS in the syntax hierarchy.
* **VQA**: Visual quality assessment.
* **VT**: Verification testing.
* **VTM**: VVC Test Model.
* **VUI**: Video usability information.
* **VVC**: Versatile Video Coding, the standardization project developed by JVET.
* **WAIP**: Wide-angle intra prediction
* **WCG**: Wide colour gamut.
* **WG**: Working group, a group of technical experts (usually used to refer to WG 11, a.k.a. MPEG).
* **WPP**: Wavefront parallel processing (usually synonymous with ECS).
* Block and unit names in HEVC:
  + **CTB**: Coding tree block (luma or chroma) – unless the format is monochrome, there are three CTBs per CTU.
  + **CTU**: Coding tree unit (containing both luma and chroma, synonymous with LCU), with a size of 16x16, 32x32, or 64x64 for the luma component.
  + **CB**: Coding block (luma or chroma), a luma or chroma block in a CU.
  + **CU**: Coding unit (containing both luma and chroma), the level at which the prediction mode, such as intra versus inter, is determined in HEVC, with a size of 2Nx2N for 2N equal to 8, 16, 32, or 64 for luma.
  + **PB**: Prediction block (luma or chroma), a luma or chroma block of a PU, the level at which the prediction information is conveyed or the level at which the prediction process is performed in HEVC.
  + **PU**: Prediction unit (containing both luma and chroma), the level of the prediction control syntax within a CU, with eight shape possibilities in HEVC:
    - **2Nx2N**: Having the full width and height of the CU.
    - **2NxN (or Nx2N)**: Having two areas that each have the full width and half the height of the CU (or having two areas that each have half the width and the full height of the CU).
    - **NxN**: Having four areas that each have half the width and half the height of the CU, with N equal to 4, 8, 16, or 32 for intra-predicted luma and N equal to 8, 16, or 32 for inter-predicted luma – a case only used when 2N×2N is the minimum CU size.
    - **N/2x2N** paired with **3N/2x2N** or **2NxN/2** paired with **2Nx3N/2**: Having two areas that are different in size – cases referred to as AMP, with 2N equal to 16 or 32 for the luma component.
  + **TB**: Transform block (luma or chroma), a luma or chroma block of a TU, with a size of 4x4, 8x8, 16x16, or 32x32.
  + **TU**: Transform unit (containing both luma and chroma), the level of the residual transform (or transform skip or palette coding) segmentation within a CU (which, when using inter prediction in HEVC, may sometimes span across multiple PU regions).
* Block and unit names in VVC:
  + **CTB**: Coding tree block (luma or chroma) – there are three CTBs per CTU in a P or B slice or in an I slice that uses a single tree, and one CTB per luma CTU and two CTBs per chroma CTU in an I slice that uses separate trees.
  + **CTU**: Coding tree unit (synonymous with LCU, containing both luma and chroma in a P or B slice or in an I slice that uses a single tree, containing only luma or only chroma in an I slice that uses separate trees), with a size of 16x16, 32x32, 64x64, or 128x128 for the luma component.
  + **CB**: Coding block, a luma or chroma block in a CU.
  + **CU**: Coding unit (containing both luma and chroma in P/B slice, containing only luma or chroma in I slice), a leaf node of a QTBT. It’s the level at which the prediction process and residual transform are performed in JEM. A CU can be square or rectangle shape.
  + **PB**: Prediction block, a luma or chroma block of a PU.
  + **PU**: Prediction unit, has the same size as a CU in the VVC context.
  + **TB**: Transform block, a luma or chroma block of a TU.
  + **TU**: Transform unit, has the same size as a CU in the VVC context.

## Opening remarks

Remarks during the opening session of the meeting Wednesday 7 July at 0500 UTC (chaired by GJS and JRO) were as follows.

* Timing and organization of online meetings, calendar
* Standards approval and publication status
  + Working practices using objective metrics report
    - HSTP-VID-WPOM V1: 2020-07-03, published 2020-11
    - ISO/IEC TR 23002-8 (Ed. 1) published 2021-05-20
  + AVC
    - H.264v14 Consented at previous meeting (with annotated regions, shutter interval, and miscellaneous corrections), Last Call to close 2021-07-13 (during the current meeting)
      * A dependency on a future version of VSEI for the annotated regions SEI message was removed in final editing by directly including the technical content
    - Status in ISO/IEC 14496-10 - corresponding aspects are partly in the in-progress CDAM (see below)
  + HEVC
    - H.265 V7 approved 2019-11-29, published 2020-01-10
    - ISO/IEC 23008-2:2020 (Ed. 4) published 2020-08-27
    - H.265 V8 Consented at the previous meeting (Shutter interval information SEI message and miscellaneous corrections), Last Call to close 2021-07-13 (during the current meeting)
    - ISO/IEC 23008-2:2020 FDAM 1 ballot started 2021-04-07 [update] (Shutter interval information SEI message)
  + Usage of code points report
    - H.Sup19 V3 to be Approved at the current meeting (update)
    - ISO/IEC TR 23091-4 (Ed. 3) published 2021-05-23
  + VVC
    - H.266 V1 approved 2020-08-29, published 2020-11-10
    - ISO/IEC 23090-3:2021 (Ed. 1) published 2021-02-16
  + VSEI
    - H.274 V1 approved 2020-08-29, published 2020-11-10
    - ISO/IEC 23002-7:2021 (Ed. 1) published 2021-01-28
  + CICP v2 (incudes errata items)
    - Was FDIS 23091-2 in ISO/IEC [add detail] and ITU-T Consent last meeting
    - H.273v2 (with 4:2:0 sampling alignment and corrections for range of values for sample aspect ratio, ICTCP equations for HLG, and transfer characteristics function for sYCC of IEC 61966-2-1) Last Call to close 2021-07-13 (during the current meeting)
  + The following freely available standards are published here in ISO/IEC:  
    <https://standards.iso.org/ittf/PubliclyAvailableStandards/index.html>
    - ISO/IEC 23002-7:2021 (Ed. 1) VSEI
    - ISO/IEC 23008-2:2020 (Ed. 4) HEVC
    - ISO/IEC 23090-3:2021 (Ed. 1) VVC
    - ISO/IEC 23091-2:2019 (Ed. 1) Video CICP
* Draft standards progression status
  + VVC conformance – under DIS ballot closing 2021-09-29, FDIS in October
  + VVC reference SW – under DIS ballot closing 2021-09-30, FDIS in October
  + AVC additional SEI – DAM this meeting, new draft
    - [m57147](https://dms.mpeg.expert/doc_end_user/current_document.php?id=79339&id_meeting=187) Summary of voting on ISO/IEC 14496-10:2020/CD Amd 1
  + VSEI extensions – DAM this meeting, new draft
    - [m57148](https://dms.mpeg.expert/doc_end_user/current_document.php?id=79340&id_meeting=187) Summary of voting on ISO/IEC 23002-7:2021/CD Amd 1
  + VVC operation range extensions – DAM this meeting, new draft
    - [m57149](https://dms.mpeg.expert/doc_end_user/current_document.php?id=79341&id_meeting=187) Summary of voting on ISO/IEC 23090-3:2021/CD Amd 1
  + The request for free availability in ISO/IEC has to be made for each edition, amendment and corrigendum, and these will also need a request form to be filled out and be approved in the Recommendations. Freely available URL on ITU part should be provided for the following parts:
    - ISO/IEC 23008-2:2020/Amd 1 – done last meeting
    - ISO/IEC DIS 23091-2, 2nd edition – done last meeting
    - ISO/IEC 23002-7:2021/Amd 1 – to be done when finishing
    - ISO/IEC 23090-3:2021/Amd 1 – to be done when finishing
* The meeting logistics, agenda, working practices, policies, and document allocation were reviewed.
  + The meeting was conducted using Zoom
  + Having text and software available is crucial (and not just arriving at the end of the meeting).
  + There were no objections voiced in the opening plenary to the consideration of late contributions.
* The results of the previous meeting and the meeting report JVET-V1000 were reviewed. The following small copy-paste issues in the meeting report were noted and were not considered sufficient to warrant issuing a revision:
  + In section 8.2 on core experiment and exploration experiment planning, the three referenced document numbers had not been updated, so they used “JVET-Uxxxx” rather than “JVET-Vxxxx” numbers.
  + In Annex B, which lists the meeting attendees, the description of how the attendance record was constructed had not been updated, so it said the records were “according to an attendance sheet circulated during the meeting sessions”. Of course, since it was a teleconference meeting, there was no circulation of an attendance sheet and the records had instead been extracted from the Zoom teleconferencing tool that was used to operate the meeting sessions. The meeting idenfication in Annex B also had not been updated, so it said the list was for the twenty-first meeting rather than the twenty-second meeting.
* There was somewhat less of a problem of late non-cross-check documents and no “placeholders” (see section 2.4.2). One late input was noted to be a proposed draft white paper on the VVC standard.
* The primary goals of the meeting were
  + Errata
  + Conformance and software for version2 of VVC & VSEI
  + Verification test results and further planning
  + White paper on VVC/VSEI
  + Extensions of VVC
    - High bit rate / high bit depth
  + Additional SEI messages for VSEI
  + Explorations
    - Neural network-based video coding
    - Enhanced compression beyond VVC
* Funding of verification testing activities: recommendation of thanks, recommendation calling for funding wrt upcoming tests.
* Liaisons?
* Number of documents similar to last meeting
* Scheduling was discussed, and it was agreed to avoid conducting “track” sessions in parallel (some BoG parallelism did occur)
* Principles of standards development were discussed.

## Scheduling of discussions

The plans for the times of meeting sessions were established as follows, in UTC (2 hours behind the time in Geneva, Paris; 7 hours ahead of the time in Los Angeles, etc.). No session should last longer than 2 hrs.

* 0500–0700 1st “morning” session [break after 2 hours]
* 0720–0920 2nd “morning” session
* [“overday” break – nearly 12 hours]
* 2100–2300 1st “night” session [break after 2 hours]
* 2320–0120+1 2nd “night” session

Sessions were announced via the calendar in the JVET document site at least 22 hrs. in advance. Particular scheduling notes are shown below, although not necessarily 100% accurate or complete:

* Wed. 07 July, 1st day
  + Session 1:
    - 0500–0540 Opening remarks, review of practices, agenda, IPR reminder
    - 0540–0700 Reports of AHGs 1–9
  + Session 2:
    - 0725–0750 Reports of AHGs 10–12
    - 0750-0920 Review of CE
  + Session 3:
    - 2100–2300 Review of CE related
  + Session 4:
    - 2320–0120 Review of CE related, Review of EEs and related
* Thu. 8 July, 2nd day
  + Session 5:
    - 0500–0700 Review of EEs and related
  + Session 6:
    - 0720–0920 Review of EEs and related
  + Session 7:
    - 2100–2300 Review of HLS
  + Session 8:
    - 2320–0120 Review of HLS
* Fri. 9 July, 3rd day
  + Session 9:
    - 0500–0700 Review of 5.1.4
  + Session 10:
    - 0720–0920 Review of EE2 related
  + Session 11:
    - 2100–2300 Review of remaining 5.1/5.3/5.2
  + Session 12:
    - 2320–0120 Review of HLS
    - BoGs on NN (A. Segall) and High Throughput (M. Sarwer)
* Mon. 12 July, 4th day
  + 0500–0700 MPEG information sharing session
  + Session 13:
    - 0720–0920 Further planning, BoG reviews, documents section 4
  + Session 14:
    - 2100–2300 Verification test (4.4), remaining documents sections 5.2, 5.3, 4, 6.2
  + Session 15:
    - 2320–0120 Remaining documents sections 5.3, 4, 6.2
    - BoG on NN in parallel
* Tue. 13 July, 5thday
  + Session16:
    - 0500–0700 BoG W0180/W0182 reporting
  + 0730–0830 Joint meeting with WG2, VCEG on VVC V2 profiles, film grain
  + Session 17:
    - 0845–0920 remaining docs review 5.3, further planning
  + Session 18:
    - 2100–2300 remaining docs review 4.8ff
    - BoGs on V2 conformance and NN in parallel
  + Session 19:
    - 2320–0120 remaining docs review 6.2
* Wed. 14 July, 6th day
  + 0500–0630 MPEG information sharing session
  + 0630–0700 Joint meeting with WG3 on VDI and green metadata SEI messages
  + 0720–0820 Joint meeting with AG5, WG4, WG7, VCEG on subjective evaluation guidelines
  + Session 20:
    - 0820–0920 BoG reports, EE/CE planning, revisits
  + Session 21:
    - 2100–2300 Review remaining docs sections 4.x, 5.2
* Thu. 15 July, 7th day
  + Session 22:
    - 0500–0700 Review DoCs, Remaining docs 5.3
  + Session 23:
    - 0720–0945 Remaining docs 6.2, revisits, further planning
* Fri. 16 July, 8th day
  + Plenary:
    - 0500–0700 Revisits JVET-W0133, JVET-W0136, approval of output docs
  + Plenary:
    - 0720–0935 approval of output docs, AHGs, recommendations
  + 2100–XXXX MPEG information sharing session
  + XXXX–XXXX WG5 Closing plenary: Approval of recommendations

## Contribution topic overview

The approximate subject categories and quantity of contributions per category for the meeting were summarized as follows (note that the noted document counts do not include crosschecks, and may not be completely accurate):

* AHG reports (section 3)
* Project development (section 4)
  + Deployment and advertisement of standards (2)
  + Text development and errata reporting (1) (1 TBP)
  + Test conditions (1)
  + Verification testing (4)
  + Test Material (0)
  + Quality assessment (0)
  + Conformance test development (0)
  + Software development (1)
  + Implementation studies (2)
  + Complexity analysis (0)
  + Encoder optimization (4)
  + Profile/tier/level specification (2) (1 TBP)
  + Proposed modification of system interface (0)
* Low-level tool technology proposals (section 5) with subtopics
  + AHG8 and CE: High bit depth and high bit rate coding (22) (section 5.1)
  + AHG11 and EE1: Neural network-based technology (17) (section 5.2) (2 TBP)
  + AHG12 and EE2: Enhanced compression beyond VVC capability (30) (section 5.3) (2 TBP)
* High-level syntax (HLS) proposals (section 6) with subtopics
  + AHG9: SEI message studies and proposals (11) (section 6.1)
  + Non-SEI HLS aspects (5) (section 6.2) (4 TBP)
* Joint meetings, plenary discussions, BoG reports (2), summary of actions (section 7)
* Project planning (section 8)
* Establishment of AHGs (section 9)
* Output documents (section 10)
* Future meeting plans and concluding remarks (section 11)

The document counts above do not include cross-checks and summary reports.

# AHG reports (12)

These reports were discussed Wednesday 7 July 2021 during 0540–0700 and 0725–0750 UTC (chaired by GJS & JRO).

[JVET-W0001](https://jvet-experts.org/doc_end_user/current_document.php?id=10948) JVET AHG report: Project management (AHG1) [J.-R. Ohm, G. J. Sullivan]

The work of the JVET overall had proceeded well in the interim period with approximately same number of input documents (as compared to the previous meeting) submitted to the current meeting. Intense discussion had been carried out on the group email reflector, and all output documents from the preceding meeting had been produced.

Output documents from the preceding meeting had been made initially available at the JVET web site (<https://jvet-experts.org/>) or the ITU-based JVET site (<http://wftp3.itu.int/av-arch/jvet-site/2021_04_V_Virtual/>). It is noted that the previous document site <http://phenix.int-evry.fr/jvet/> is still accessible, but was converted to read-only.

The list of documents produced included the following, particularly:

* The meeting report (JVET-V1000) [Posted 2021-05-27]
* High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Encoder Description Update 15 (JVET-V1002) [Posted 2021-07-06]
* Errata report items for VVC, HEVC, AVC, Video CICP, and CP usage TR (JVET-V1004) [Posted 2021-06-24]
* Algorithm description for Versatile Video Coding and Test Model 13 (VTM 12) (JVET-V2002) [Posted 2021-07-06]
* VVC operation range extensions (Draft 3) (JVET-V2005) [Posted 2021-05-07]
* Additional SEI messages for VSEI (Draft 3) (JVET-V2006) [Posted 2021-05-07]
* VTM common test conditions and evaluation procedures for HDR/WCG video (JVET-V2011) [Posted 2021-05-14]
* Common Test Conditions and evaluation procedures for neural network-based video coding technology (JVET-V2016) [Posted 2021-05-26, last update 2021-06-05]
* Common Test Conditions and evaluation procedures for enhanced compression tool testing (JVET-V2017) [Posted 2021-05-15]
* VVC verification test report for HD SDR and 360° video content (JVET-V2020) [Posted 2021-06-22]
* VVC verification test plan (Draft 6) (JVET-V2021) [Posted 2021-06-24]
* CE on Entropy Coding for High Bit Depth and High Bit Rate Coding (JVET-V2022) [Posted 2021-04-28, last update 2021-05-15]
* EE on Neural Network-based Video Coding (JVET-V2023) [Posted 2021-04-28, last update 2021-05-14]
* EE on Enhanced Compression beyond VVC capability (JVET-V2024) [Posted 2021-04-27, last update 2021-05-22]

The twelve *ad hoc* groups had made progress, and reports from those activities had been submitted. Furthermore, one core experiment (CE) on entropy coding for high bit depth and high bit rate coding, and two exploration experiments (EE) on neural network-based video coding and on enhanced compression beyond VVC capability, were conducted.

Due to issues associated with the COVID-19 pandemic, a conversion of the meeting to be conducted only online was again necessitated.

During the interim period, one meeting of AHG4 (for preparing the HDR verification tests) was held.

Software integration was finalized approximately according to the plan. Significant activities were also conducted on preparation of verification tests, and on development of VVC conformance testing.

Various problem reports relating to asserted bugs in the software, draft specification text, and reference encoder description had been submitted to an informal "bug tracking" system. That system is not intended as a replacement of our ordinary contribution submission process. However, the bug tracking system was considered to have been helpful to the software coordinators and text editors. The bug tracker reports had been automatically forwarded to the group email reflector, where the issues were discussed – and this is reported to have been helpful.

Roughly 90 input contributions (not counting the AHG, CE and EE summary reports and crosschecks) had been registered for consideration at the current meeting.

It is further noted that, starting from the twentieth JVET meeting, work items which had originally been conducted by the Joint Collaborative Team on Video Coding (JCT-VC) were continued to be conducted in JVET as a single joint team, as negotiated by the parent bodies. This particularly consists of work on

* *High Efficiency Video Coding* (HEVC) and its extensions, the development of associated conformance test sets, reference software, verification testing, and non-normative guidance information,
* Specification of *Coding-independent Code Points (Video)* (CICP), and associated technical report(s),
* Maintenance and minor enhancement work on the *Advanced Video Coding* (AVC) standard, associated conformance test sets and reference software.

To retain a consistent numbering scheme, the number range of output documents starting from 1001 was reserved for the previous JCT-VC topic items listed above, whereas the number range starting from 2001 was retained for VVC, VSEI and future exploration activities. Duplication of AHGs was avoided by merging previous JCT-VC AHGs with the corresponding AHGs of JVET.

A preliminary basis for the document subject allocation and meeting notes for the 23rd meeting had been made publicly available on the ITU-hosted ftp site <http://wftp3.itu.int/av-arch/jvet-site/2021_07_W_Virtual/>.

[JVET-W0002](https://jvet-experts.org/doc_end_user/current_document.php?id=10949) JVET AHG report: Draft text and test model algorithm description editing (AHG2) [B. Bross, J. Chen, C. Rosewarne, F. Bossen, J. Boyce, S. Kim, S. Liu, J.-R. Ohm, G. J. Sullivan, A. Tourapis, Y.-K. Wang, Y. Ye]

2.1 Output documents produced

2.1.1 JVET-V2005 VVC operation range extensions (Draft 3)

This document contains the draft text for changes to the Versatile Video Coding (VVC) standard (ITU T H.266 | ISO/IEC 23090-3), for the support of the operation range extensions, the addition of Level 6.3 and the SEI manifest and SEI prefix indication SEI messages, including SEI payload type values and other interfaces for SEI messages added to the VSEI specification, as well as some technical corrections.

Draft 3 incorporated items:

• Method of entropy coding for high bit depth in TSRC(JVET-V0054)

• Method of entropy coding for high bit depth in RRC (JVET-V0106)

• Method for high precision computation of transform scaling (JVET-V0047)

• Bug fix changes to the semantics of dui\_dpb\_output\_du\_delay\_present\_flag and dui\_dpb\_output\_du\_delay (JVET-V0111)

• Addition of the paylaodType values etc. for the display orientation SEI message and the colour transform information SEI message (JVET-V0061, JVET-V0108)

• Fix for ticket #1479.

2.1.2 JVET-V2006 Additional SEI messages for VSEI (Draft 3)

This document contains the draft text for changes to the versatile supplemental enhancement information messages for coded video bitstreams (VSEI) standard (Rec. ITU-T H.274 | ISO/IEC 23002-7), to specify additional SEI messages, including the annotated regions SEI message, the alpha channel information SEI message, the depth representation information SEI message, the multiview acquisition information SEI message, the scalability dimension information SEI message, the extended dependent random access point (DRAP) indication SEI message, the display orientation SEI message, and the colour transform information SEI message. The draft text also includes text changes for some technical corrections and editorial improvements.

Draft 3 incorporated items:

• Addition of the display orientation SEI message (JVET-V0061)

• Addition of the colour transform information SEI message (JVET-V0108)

• Changes on the scalability dimension information SEI message (JVET-V0063)

• Changes on the MAI, DRI, and ACI SEI messages and their interactions with the SDI SEI message (JVET-V0064)

• Changes on the DRAP and EDRAP indication SEI messages (JVET-V0065)

2.1.3 JVET-V1004 Errata report items for VVC, HEVC, AVC, Video CICP, and CP usage TR

This document contains a list of reported errata items for VVC, VSEI, HEVC, AVC, Video CICP, and the TR on usage of video signal type code points, for tracking purposes. Some of the items have been confirmed by the JVET and have been agreed to require fixing, while some other items have not yet been confirmed. This document also provides publication status backgrounds of these standards.

Incorporated items at the JVET-V meeting:

• For VVC (the changes are included in an attachment to this document):

o Fix for ticket #1486

• For HEVC

o Removed the item on the absence of “persistence flag” for the annotated regions SEI message (this was agreed to be removed at the JVET-T meeting in October 2020, but the decision was overlooked)

2.1.4 JVET-V1002 High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Encoder Description Update 15

This document was prepared, incorporating:

• Addition of [AHG10] GOP-based temporal filter improvements (JVET-V0056)

• Addition of AHG 10: QP control for very smooth blocks (JVET-V0078)

2.1.5 JVET-V2002 Algorithm description for Versatile Video Coding and Test Model 13 (VTM 13)

The VVC Test Model 13 (VTM13) algorithm description and encoding method document was prepared, incorporating:

• Updated the description of motion compensated temporal pre-filtering (MCTF)

• Typo Fixes

3 Related input contributions

The following input contributions were noted as relevant to the work of this ad hoc group:

• None at the time of preparing this AHG report.

4 Recommendations

The AHG recommends to:

• Approve JVET-V1002, JVET-V1004, JVET-V2002, JVET-V2005, and JVET-V2006 documents as JVET outputs,

• Compare the VVC documents with the VVC software and resolve any discrepancies that may exist, in collaboration with the software AHG,

• Encourage the use of the issue tracker to report issues with the text of both the VVC specification text and the algorithm and encoder description,

• Continue to improve the editorial consistency of VVC text specification and Test Model documents,

• Ensure that, when considering changes to VVC, properly drafted text for addition to the VVC Test Model and/or the VVC specification text is made available in a timely manner,

• Review AHG2 related contributions (if any) and act on them if found to be necessary.

[JVET-W0003](https://jvet-experts.org/doc_end_user/current_document.php?id=10950) JVET AHG report: Test model software development (AHG3) [F. Bossen, X. Li, K. Sühring, K. Sharman, V. Seregin, A. Tourapis]

The software model versions prior to the start of the meeting were:

• VTM 13.0 (May 2021)

• HM-16.23 (Mar. 2021)

• HM-16.21+SCM-8.8 (Mar. 2020)

• SHM 12.4 (Jan. 2018)

• HTM 16.3 (Jul. 2018)

• JM 19.0

• JSVM 9.19.15

• JMVC 8.5

• 3DV ATM 15.0 (no version history)

• HDRTools 0.22 (June 2021)

Software for MFC and MFCD is only available published by ITU-T and ISO/IEC. It is planned to create repositories with the latest versions available in ITU-T H.264.2 (02/2016). All development history is lost.

2 Software development

Development was continued on the GitLab server, which allows participants to register accounts and use a distributed development workflow based on git.

The server is located at:

https://vcgit.hhi.fraunhofer.de

The registration and development workflow are documented at:

https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware\_VTM/wikis/VVC-Software-Development-Workflow

Although the development process is described in the context of the VTM software, it can be applied to all other software projects hosted on the GitLab server as well.

3 VTM related activities

The VTM software can be found at

https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware\_VTM/

The software development continued on the GitLab server. VTM versions 12.2 and 12.3 were tagged on Apr. 29 and Apr. 30, and VTM version 13.0 was tagged on May 26. VTM 13.1 is expected during the 23rd JVET meeting.

VTM 12.2 was tagged on Apr. 29, 2021. Changes include:

• JVET-U0097: GDR Software

• Fix #1444: fix picture output order in multilayer streams

• Fix #1449: output only layers in active OLS

• Fix #1474: remove erroneous skipped pictures detection in checkRPL

• Fix #1476: fix picture output before GDR

• Fix #1478: correctly set sps\_field\_seq\_flag

• Fix #1480: indexing of initial\_cpb\_removal\_delay arrays

• Fix #1481: LMCS updating issue

• Disable GDR\_LEAK\_TEST

VTM 12.3 was tagged on Apr. 30, 2021. Changes include:

• Remove macros from previous cycle

VTM 13.0 was tagged May 26, 2021. Changes include:

• JVET-V0056: Changes to MCTF

• JVET-V0066: Encoder improvements to palette coding for high bit depth

• JVET-V0095: Using true original samples for SAO and ALF optimization when MCTF is applied

• JVET-V0078: QP control for very smooth blocks

• JVET-V0047: CE3.1 method for the high precision computation of transform scaling

• JVET-V0131: Fixing the forward transform matrices for high bit depth coding

• JVET-V0106: RRC Rice extension

• JVET-V0054: Slice based Rice parameter selection for transform skip residual coding

• Change cfg/hbd/12-bit.cfg to set extended\_precision\_processing\_flag equal to 1

• fix #1485: add conformance check for LMCS

• Conformance window code cleaunp

• Change default for ConformanceWindowMode to "1" (automatic padding)

• Rename loop filter appropriately to deblocking filter

• Fix build error with GCC 11.1

VTM 13.1 is expected to be tagged during the 23rd JVET meeting. Changes include:

• JVET-V0108: Colour transform information SEI

• JVET-U0082: SDI SEI and three other SEIs: MAI, ACI, and DRI

• JVET-U0084: EDRAP SEI message

• JVET-V0061 display orientation SEI message

• Fix for Ticket#1489 - Cleanup of variable names for m\_iSourceWidth and m\_iSourceHeight

3.1 CTC Performance

The following tables show **VTM 13.0** performance over **HM 16.24rc1**:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | **All Intra Main10** |  |  |
|  |  |  | **Over HM-16.24rc1** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 | -29,03% | -32,17% | -34,07% | 1656% | 170% |
| Class A2 | -29,29% | -23,92% | -21,06% | 2657% | 182% |
| Class B | -21,73% | -26,96% | -30,76% | 2872% | 187% |
| Class C | -22,54% | -18,95% | -22,70% | 4026% | 195% |
| Class E | -25,75% | -25,91% | -24,45% | 2369% | 179% |
| **Overall** | -25,06% | -25,37% | -26,85% | 2700% | 184% |
| Class D | -18,46% | -13,31% | -13,41% | 4749% | 166% |
| Class F | -39,33% | -39,73% | -42,22% | 4883% | 178% |
|  |  |  |  |  |  |
|  |  |  | **Random access Main10** |  |  |
|  |  |  | **Over HM-16.24rc1** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 | -39,74% | -39,41% | -46,15% | 654% | 165% |
| Class A2 | -43,15% | -40,53% | -39,75% | 727% | 184% |
| Class B | -36,20% | -48,61% | -47,19% | 723% | 169% |
| Class C | -32,85% | -34,70% | -36,64% | 997% | 183% |
| Class E |  |  |  |  |  |
| **Overall** | -37,41% | -41,45% | -42,68% | 773% | 175% |
| Class D | -30,96% | -31,11% | -30,96% | 1128% | 173% |
| Class F | -45,76% | -49,18% | -50,10% | 560% | 159% |
|  |  |  |  |  |  |
|  |  |  | **Low delay B Main10** |  |  |
|  |  |  | **Over HM-16.24rc1** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  |  |  |
| Class A2 |  |  |  |  |  |
| Class B | -29,24% | -34,80% | -32,41% | 751% | 162% |
| Class C | -25,89% | -17,42% | -17,95% | 916% | 173% |
| Class E | -28,73% | -33,03% | -26,38% | 366% | 147% |
| **Overall** | -28,00% | -28,56% | -26,08% | 670% | 161% |
| Class D | -25,01% | -12,57% | -11,79% | 953% | 182% |
| Class F | -40,20% | -41,56% | -41,87% | 490% | 143% |
|  |  |  |  |  |  |
|  |  |  | **Low delay P Main10** |  |  |
|  |  |  | **Over HM-16.24rc1** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  |  |  |
| Class A2 |  |  |  |  |  |
| Class B | -33,97% | -37,79% | -34,99% | 701% | 171% |
| Class C | -27,68% | -17,28% | -18,05% | 844% | 191% |
| Class E | -32,32% | -36,86% | -30,30% | 366% | 152% |
| **Overall** | -31,46% | -30,72% | -28,17% | 634% | 172% |
| Class D | -26,32% | -11,99% | -10,87% | 890% | 189% |
| Class F | -39,97% | -41,10% | -41,48% | 527% | 149% |

According to common test conditions in random access configuration HM is using a GOP size of 16 pictures compared to VTM using a GOP of 32 pictures. Random access points are inserted approximately every second aligned with a GOP boundary of GOP 32 in both VTM and HM. VTM uses two more reference pictures in random access than HM (due to more memory being availably in typical level settings).

The following tables show **VTM 13.0** performance compared to **VTM 12.1**:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | **All Intra Main10** |  |  |
|  |  |  | **Over VTM-12.1** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 | 0,00% | 0,00% | 0,00% | 106% | 99% |
| Class A2 | 0,00% | 0,00% | 0,00% | 105% | 100% |
| Class B | 0,00% | 0,00% | 0,00% | 103% | 98% |
| Class C | 0,00% | 0,00% | 0,00% | 102% | 99% |
| Class E | 0,00% | 0,00% | 0,00% | 105% | 101% |
| **Overall** | 0,00% | 0,00% | 0,00% | 104% | 99% |
| Class D | 0,00% | 0,00% | 0,00% | 105% | 101% |
| Class F | 0,00% | 0,00% | 0,00% | 102% | 98% |
|  |  |  |  |  |  |
|  |  |  | **Random access Main10** |  |  |
|  |  |  | **Over VTM-12.1** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 | -1,83% | -1,86% | -2,55% | 101% | 100% |
| Class A2 | -1,11% | -0,84% | -0,85% | 101% | 103% |
| Class B | -1,39% | -0,95% | -0,98% | 100% | 103% |
| Class C | -1,19% | -0,58% | -0,29% | 101% | 97% |
| Class E |  |  |  |  |  |
| **Overall** | -1,37% | -1,01% | -1,09% | 101% | 101% |
| Class D | -1,47% | -0,09% | -0,16% | 102% | 100% |
| Class F | 0,00% | 0,00% | 0,00% | 102% | 100% |
|  |  |  |  |  |  |
|  |  |  | **Low delay B Main10** |  |  |
|  |  |  | **Over VTM-12.1** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  |  |  |
| Class A2 |  |  |  |  |  |
| Class B | 0,00% | 0,00% | 0,00% | 101% | 103% |
| Class C | 0,00% | 0,00% | 0,00% | 100% | 99% |
| Class E | 0,00% | 0,00% | 0,00% | 100% | 100% |
| **Overall** | 0,00% | 0,00% | 0,00% | 100% | 101% |
| Class D | 0,00% | 0,00% | 0,00% | 100% | 99% |
| Class F | 0,00% | 0,00% | 0,00% | 101% | 101% |
|  |  |  |  |  |  |
|  |  |  | **Low delay P Main10** |  |  |
|  |  |  | **Over VTM-12.1** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  |  |  |
| Class A2 |  |  |  |  |  |
| Class B | 0,00% | 0,00% | 0,00% | 101% | 101% |
| Class C | 0,00% | 0,00% | 0,00% | 101% | 99% |
| Class E | 0,00% | 0,00% | 0,00% | 101% | 97% |
| **Overall** | 0,00% | 0,00% | 0,00% | 101% | 100% |
| Class D | 0,00% | 0,00% | 0,00% | 100% | 96% |
| Class F | 0,00% | 0,00% | 0,00% | 101% | 99% |

Full results are attached to this AHG report as Excel files.

3.2 Issues in VTM 12.x affecting conformance

The following issues in VTM master branch (Jul. 6, 2021) affect conformance:

• Handling of NoOutputOfPriorPicFlag is disabled due to crash issues (issue #1415)

• Missing HLS features (see sections below)

• Handling of CRA subpicture with mixed NAL units (issue #1490)

However, there are no known issues (other than #1490) in VTM that affect processing of current JVET conformance bitstreams.

3.3 Status of implementation of proposals of previous JVET meetings

The following list contains all adoptions of the Q and R meetings that were not marked as merged (or submitted) or specification only change in the software coordinator tracking sheet:

• JVET-Q0112

• JVET-Q0154: Disallow mixing of GDR and IRAP (Disallow mixing of GDR with any non-GDR).

• JVET-Q0164

• JVET-Q0402

• JVET-Q0443: Modification of the subpicture level SEI message semantics to impose a constraint on MinCR.

• JVET-R0178: Require that when no\_aps\_constraint\_flag is equal to 1, sps\_lmcs\_enabled\_flag and sps\_scaling\_list\_enabled\_flag shall be equal to 0

• JVET-R0221

• JVET-R0046: Change the description of the bitstream extraction process per the value of max\_tid\_il\_ref\_pics\_plus1[ ][ ] (aspect 1.2 per JVET-R0046-v4).

• JVET-R0065: Specify that GDR AUs shall be complete – i.e., all of the layers in the CVS shall have a picture in the AU (as with IRAP AUs).

• JVET-R0191: Update the range value for num\_ols\_hrd\_params\_minus1.

• JVET-R0222 aspect 1: Infer vps\_max\_sublayers\_minus1 to be equal to 6 when sps\_video\_parameter\_set\_id is equal to 0 (i.e. VPS is not present). The exact editorial expression is at the discretion of the editor.

• JVET-S0196 (JVET-S0144 item 17)

• JVET-S0227 (JVET-S0144 item 22)

• JVET-S0077 (JVET-S0139 item 5)

• JVET-S0174 aspect 2 (JVET-S0139 item 18.b)

• JVET-S0156 aspect 3 (JVET-S0139 item 21)

• JVET-S0139 item 26 (no source listed)

• JVET-S0188 aspect 1 (JVET-S0139 item 28)

• JVET-S0139 item 40 (item does not exist)

• JVET-S0042 (JVET-S0142 item 1.b)

• JVET-S0174 aspect 1 (JVET S0143 item 19)

• JVET-S0096 aspect 3 (JVET-S0140 item 10)

• JVET-S0096 aspect 4 (JVET-S0140 item 13)

• JVET-S0159 aspect 3 (JVET-S0140 item 16)

• JVET-S0171 (JVET-S0256)

• JVET-S0118 (JVET-S0141 item 7)

• JVET-S0102 (JVET-S0141 item 9.a)

• JVET-S0117 (JVET-S0141 item 11)

• JVET-S0157 item 2 (JVET-S0141 item 13)

• JVET-S0157 item 4 (JVET-S0141 item 14)

• JVET-S0175 aspect 3 (JVET-S0141 item 16)

• JVET-S0175 aspect 1, 2 (JVET-S0141 item 17)

• JVET-S0175 aspects 4 and 5 (JVET-S0141 item 18)

• JVET-S0175 aspect 6 (JVET-S0141 item 19)

• JVET-S0198/ JVET-S0223 (JVET-S0141 item 24)

• JVET-S0173 aspect 2 (JVET-S0141 item 40.b)

• JVET-S0173 item 1 (JVET-S0141 item 51)

• JVET-S0173 item 3 (JVET-S0141 item 52)

• JVET-S0173 item 5 (JVET-S0141 item 53)

• JVET-S0173 item 6 (JVET-S0141 item 54)

• JVET-S0173 item 4 (JVET-S0141 item 56)

• JVET-S0176 item 4 (JVET-S0141 item 60)

• JVET-S0154 aspect 5 (JVET-S0141 item 68)

• JVET-S0154 aspect 6 (JVET-S0141 item 69)

• JVET-S0154 aspect 8 (JVET-S0141 item 71)

• JVET-S0095 aspect 5 (JVET-S0145 item 5)

• JVET-S0095 aspect 6 (JVET-S0145 item 6)

• JVET-S0100 aspect 1, depends on JVET-R0193 (JVET-S0147 item 2)

• FINB ballot comments Make high tier support up to 960.

3.4 Status of proposals of the 22nd JVET meeting (Online)

The following list contains all adoptions of the V meeting that were not marked as merged or specification only change in the software coordinator tracking sheet:

• JVET-V0063

• JVET-V0064

• JVET-V0065

• JVET-V0111 (MR is available)

4 HM related activities

HM 16.24 is expected to be tagged during the 23rd JVET meeting. Changes include so far:

• JVET-V0056: Changes to MCTF

• JVET-V0078: QP control for very smooth blocks

• JCTVC-AM0023: Illustration of the film grain characteristics SEI message in HEVC

• JCTVC-AM0024: Illustration of the shutter interval info SEI message in HEVC Draft

• Update Conformance Window code (backport from VTM)

The following actions have yet to be included:

• JVET-T0050: Add ability to detect static objects to encoder

Merge requests are available, but have pending discussions.

As reported in the previous report, further information on lambda optimisation in HM would be appreciated, including comparison of allocation of bits within the GOP structures between HM and VTM.

The HEVC bug tracker lists:

• 38 tickets for “HM”, most of which are more than 5 years,

• 1 ticket for “HM RExt”, which was created during this reporting period,

• 7 tickets for “HM SCC”, all of which are at least 3 years old,

Help to address these tickets would be appreciated.

One merge request is available related to HM SCC for ticket #1511 on SCC reference picture marking. We would appreciate help to confirm that the proposed change matches the SCC text.

5 SCM related activities

There had not been any further developments to SCC’s SCM during this meeting cycle.

6 SHM related activities

There had not been any further developments to SHVC’s SHM during this meeting cycle.

7 HTM related activities

There had not been any updates to the HTM of MV-HEVC and 3D-HEVC.

8 HDRTools related activities

HDRTools version 0.22 was tagged on June 22nd, 2021. Changes include:

• Add PNG reading and writing (using libpng)

• Add libpng submodule and configure cmake to update it if the LIBPNG option is enabled

• Setup CI running build on linux and windows instances

• Cleanup README

• Bugfix for parsing the last line of the config file

• Allow ~ (HOME) to be present in config file parameters

• Add JVET based MS-SSIM

• Fix copyright for new JVET based MS-SSIM

• Fix issue with the ouput frame rate not beeing copied from the input

• Add special handling of the "unknown" frame rate (F0:0)

• Add option OverrideContentValues to be able to force values from config or CLI for .y4m files

• Add JVET based PSNR computation

New development is now being added under the branch named 0.23-dev. It is expected that this branch will include the ability of computing distortion metrics for material that may differ in bit-depth, among other features.

9 JM, JSVM, JMVM related activities

There had not been any updates to the JM, JSVM and JMVM software.

10 Bug tracking

The bug tracker for VTM and specification text is located at:

https://jvet.hhi.fraunhofer.de/trac/vvc

The bug tracker uses the same accounts as the HM software bug tracker. Users may need to log in again due to the different sub-domain. For spam fighting reasons account registration is only possible at the HM software bug tracker at

https://hevc.hhi.fraunhofer.de/trac/hevc

Bug tracking for HDRTools is located at:

https://gitlab.com/standards/HDRTools/-/issues

Please file all issues related to the VVC reference software and HDRTools into the appropriate bug tracker. Try to provide all the details, which are necessary to reproduce the issue. Patches for solving issues and improving the software are always appreciated.

11 Software repositories

Git repositories that were previously assigned to the JCT-VC group on the GitLab server were re-assigned to the JVET group. The old URLs are still working and will forward the user to the new location, with the display of a warning suggesting to update bookmarks to the new location.

The subversion repositories for SHM and HTM were converted to git and are now stored on the GitLab server to unify access and development process.

12 CTC alignment

The following differences were found in CTC alignment between HEVC and VVC:

- For HM two test configurations are described: one for 8-bit coding bit depth for Main profile and a second one for 10-bit coding bit depth for Main 10. VTM only specifies a 10-bit test case. These should be aligned, so that the same templates can be used.

Merging of CTC documents related to HEVC and VVC in each area of interest should be considered. JVET-W0152 was submitted to the 23rd JVET providing an overview of the existing CTC documents and suggesting merge options.

13 Recommendations

The AHG recommends to:

- Continue to develop reference software

- Improve documentation, especially the software manual

- Encourage people to test VTM and other reference software more extensively outside of common test conditions.

- Encourage people to report all (potential) bugs that they are finding.

- Encourage people to submit bit-streams/test cases that trigger bugs in VTM and other reference software.

- Encourage people to submit non-normative changes that either reduce encoder run time without significantly sacrificing compression performance or improve compression performance without significantly increasing encoder run time

- Design and add configuration files to the VTM software for testing of HLS features

- Review VTM-related contributions and determine whether features should be added (or removed) from the software

- Continue to investigate the merging of branches.

- Keep common test conditions aligned for the different standards.

- Consider documents (including late documents) related to AHG3 activities

One expert mentions it is believed that docs V0063 to V0065 (bug fixes to SEI messages) had been merged, possibly with reference to the original proposals from the 21st meeting.

(Original proposals for V0063 to V0065 are JVET-U0082 to JVET-U0084)

[JVET-W0004](https://jvet-experts.org/doc_end_user/current_document.php?id=10951) JVET AHG report: Test material and visual assessment (AHG4) [V. Baroncini, T. Suzuki, M. Wien, E. François, S. Liu, A. Norkin, A. Segall, P. Topiwala, S. Wenger, Y. Ye]

2.1 Verification test

Input document JVET-W0041 as discussed in the AHG meeting on 2021-06-10. It summarizes the status of test sequence selection and rate point determination for the HDR VVC verification test category since the last JVET meeting. Areas of progress include:

- Recommendation of QP points and sequences for the HLG category for use in a dry-run.

- Recommendation of sequences for the PQ sequences for use in a dry-run

- Re-encoding of all sequences using HM-16.23 and VTM-12.0 and using 3840x2160 versions of all sequences is complete.

At the AHG meeting, it was recommended to modify the sequence names to make them unique if used in multiple test categories. This specifically applies to DrivingPOV which was recommended to be renamed to DrivingPOV-HDR. It was noted that the document reports Y-PSNR curves. It was commented that it might be appropriate to consider WPSNR for initial rate point selection. It was mentioned that the HLG set is considered to be quite stable in terms of sequences and QP selection. It was recommended to consider propagating the dry run results to verification test results if they are confirmed to be matching the intended quality range. This could also apply to the PQ category.

Based on the outcome of that AHG meeting, dry-run tests were performed using formal visual assessment in order to validate the chosen test sequences and the rate point selection in both, the HDR HLG and HDR PQ categories. The results are reported in the late input document JVET-W0145.

2.2 Test sequences

The test sequences used for CfP/CTC are available on ftp://jvet@ftp.ient.rwth-aachen.de in directory “/jvet-cfp” (accredited members of JVET may contact the JVET chairs for login information).

Due to copyright restrictions, the JVET database of test sequences is only available to accredited members of JVET (i.e. members of ISO/IEC MPEG and ITU-T VCEG).

3 Related contributions

The following related contribution is submitted.

JVET-W0041 “AHG4: Status Report on HDR Video Verification Test Preparation” [A. Segall, M. Wien, V. Baroncini, K. Andersson]

JVET-W0042 “AHG4: Agenda and report of the AHG meeting on the HDR video verification test preparation on 2021-06-10” [A. Segall, M. Wien, V. Baroncini (AHG coordinators)]

JVET-W0145 “Report on dry-run results for VVC compression performance verification testing in the HDR PQ and HLG categories” [A. Segall, M. Wien, V. Baroncini, G. Baroncini]

4 Recommendations

The AHG recommends:

• To review the input contributions related to the verification test preparation.

• To continue to discuss and to update the non-finalized categories of the verification test plan, including those which have not been addressed yet.

• To collect volunteers to conduct the verification test, including volunteers to encode.

• To review the set of available test sequences for the verification tests and potentially collect more test sequences with a variety of content.

• To continue to collect new test sequences available for JVET with licensing statement.

It is noted that potentially the results from the dry run on HDR category could be used for a test report.

[JVET-W0005](https://jvet-experts.org/doc_end_user/current_document.php?id=10952) JVET AHG report: Conformance testing (AHG5) [J. Boyce, W. Wan, E. Alshina, F. Bossen, I. Moccagatta, K. Kawamura, K. Sühring, X. Xu]

The progress on the Conformance testing specification is consistent with the preliminary timeline agreed at the 16th JVET meeting, as follows:

• 17th meeting Jan. 2020: Preliminary guidelines for bitstream preparation (e.g., naming conventions),

improved list of conformance bitstreams

• 18th meeting Apr. 2020: Final guidelines for bitstream preparation and improved list of conformance

bitstreams with identified responsible experts, initial bitstreams provided

• 19th meeting July 2020: Confirmed list of bitstreams to be included in v1, collection of bitstream

candidates for CD ballot at next meeting

• 20th meeting Oct. 2020: CD of conformance specification

• 21st meeting Jan. 2021: Final bitstreams provided, DIS ballot in ISO/IEC

• 22nd meeting April 2021: No action pending DIS ballot

• 23rd meeting July 2021: Final conformance specification

3 Status on bitstream submission

The status at the time of preparation of this report is as follows:

• 104 bitstream categories have been identified

• At least one bitstream has been submitted in each identified category

• 281 total bitstreams have been provided, checked, and made available

• 1 bitstream has an identified problem and is expected to be replaced

4 Activities and Discussion

The AHG activities are on schedule with the preliminary timeline shown in section 2. The final conformance specification is expected to be output from this meeting.

The following bitstreams were replaced to correct identified problems:

• TEMPSCAL\_B\_Panasonic\_6: was failing condition 1 in clause C.4 comparing deltaTime90k and InitCpbRemovalDelay

• SLICES\_A\_Huawei\_3: ticket #1446 “Picture header should be in its own NALU”

A problem was identified with the MNUT\_A\_Nokia\_3 conformance bitstream which requires a change to the VTM13.0, which is addressed in the merge request in ticket #1490.

There are not currently any known issues with the other provided conformance bitstream packages. A VTM13 directory of bitstreams was made available. All provided bitstreams can be decoded using VTM12.2 and VTM13.

Information was provided by the authors of JVT-T0067 “Bit Stream Feature Analyzer (BSFA) for Coding Tool Statistics based on VTM-10.0” that not all IBC block sizes are present in conformance bitstreams using the 4:2:0 chroma format. However, all IBC block sizes are present in conformance bitstreams using the 4:2:2/4:4:4 chroma formats. An additional conformance bitstream could be added to the conformance suite to cover all IBC block sizes for the 4:2:0 chroma format. Such a bitstream can be found in the set commonly generated using CTC.

The regular JVET e-mail reflector was used for discussions (jvet@lists.rwth-aachen.de).

The AHG5 chairs and JVET chairs can be reached at jvet-conformance@lists.rwth-aachen.de. Participants should not subscribe to this list but may send emails to it.

5 Contributions

There are no related input contributions.

6 Ftp site information

The procedure to exchange the bitstream (ftp cite, bitstream files, etc.) is specified in Sec 2 “Procedure” of JVET-R2008. The ftp and http sites for downloading bitstreams are

ftp://ftp3.itu.int/jvet-site/bitstream\_exchange/VVC

https://www.itu.int/wftp3/av-arch/jvet-site/bitstream\_exchange/VVC/

The ftp site for uploading bitstream file is as follows.

ftp://ftp3.itu.int/jvet-site/dropbox/

(user id: avguest, passwd: Avguest201007)

If using FileZilla, the following configuration is suggested:



The AHG recommends the following:

* Solicit a volunteer to provide a bitstream to exercise all IBC block sizes for 4:2:0 in Main 10 profile
* Finalize the conformance specification for VVC v1 as an output document from this meeting
* Begin to prepare a list of conformance bitstreams for VVC v2 and identify volunteers to provide bitstreams and coordinate the activity

For version 1, issues identified in the bitstreams can be fixed along with the finalization (FDIS in ISO in October, ITU consent in January)

Conformance for version 2 was later discussed in a BoG (JVET-W0188), and a corresponding draft was produced as output.

[JVET-W0006](https://jvet-experts.org/doc_end_user/current_document.php?id=10953) JVET AHG report: 360° video coding, software and test conditions (AHG6) [J. Boyce, Y. He, K. Choi, Y. Ye]

2 Activities

(1) There are some software updates for 360Lib for the compilation with VTM-13.0. A patch for VTM-13.0 is provided for compilation with VTM-13.0. The patch has been integrated in the development branch of VTM after VTM-13.0 release.

(2) A new version 360Lib-5.2 was generated in order to support padded CMP coding with HM (https://jvet.hhi.fraunhofer.de/svn/svn\_360Lib/tags/360Lib-5.2/).

3 Software repository and versions

The 360Lib software is developed using a Subversion repository located at:

https://jvet.hhi.fraunhofer.de/svn/svn\_360Lib/

The released version of 360Lib-13.0 can be found at:

https://jvet.hhi.fraunhofer.de/svn/svn\_360Lib/tags/360Lib-13.0/

360Lib-13.0 testing results with VTM-13.0 can be found at:

ftp.ient.rwth-aachen.de/ahg/testresults/360Lib-13.0/VTM-13.0-360Lib-13.0\_CTC.xlsm

360Lib bug tracker

https://hevc.hhi.fraunhofer.de/trac/jem/newticket?component=360Lib

4 360Lib-13.0 results

Table 1 is for the projection formats comparison using VTM-13.0 according to 360-degree video CTC (JVET-U2012) compared to that using VTM-12.0. Table 2 compares generalized cubemap (GCMP) coding and padded equi-rectangular projection (PERP) coding using VTM-13.0.

Table 3 is for PERP coding comparison between VTM-13.0 and HM-16.22. Table 4 is to compare GCMP coding with VTM-13.0 with and PCMP coding with HM-16.22.

**Table 1. VTM-13.0 vs VTM-12.0 (VTM-12.0 as anchor)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **PERP: VTM-13.0 over VTM-12.0** | | | | | |
|  | **End-to-end**  **WS-PSNR** | | | **End-to-end**  **S-PSNR-NN** | | |
|  | Y | U | V | Y | U | V |
| Class S1 | 0.00% | 0.01% | 0.00% | 0.00% | 0.01% | 0.00% |
| Class S2 | 0.00% | 0.01% | 0.00% | 0.00% | 0.01% | 0.00% |
| **Overall** | 0.00% | 0.01% | 0.00% | 0.00% | 0.01% | 0.00% |

**Table 2. VTM-13.0 GCMP vs PERP (PERP as anchor)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **GCMP Over PERP** | | | | | |
|  | **End-to-end WS-PSNR** | | | **End-to-end S-PSNR-NN** | | |
|  | Y | U | V | Y | U | V |
| Class S1 | -11.42% | -5.70% | -6.33% | -11.39% | -5.64% | -6.28% |
| Class S2 | -3.67% | 0.66% | 0.84% | -3.67% | 0.76% | 0.90% |
| **Overall** | -8.32% | -3.15% | -3.47% | -8.30% | -3.08% | -3.41% |

**Table 3. VTM-13.0 PERP vs HM-16.22 PERP (HM-16.22 PERP as anchor)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **VTM-13.0 PERP - Over HM-16.22 PERP** | | | | | |
|  | **End-to-end WS-PSNR** | | | **End-to-end S-PSNR-NN** | | |
|  | Y | U | V | Y | U | V |
| Class S1 | -30.17% | -37.21% | -39.72% | -30.17% | -37.26% | -39.72% |
| Class S2 | -36.22% | -35.94% | -38.27% | -36.20% | -35.98% | -38.32% |
| **Overall** | -32.59% | -36.70% | -39.14% | -32.58% | -36.75% | -39.16% |

**Table 4. VTM-13.0 GCMP vs HM-16.22 PCMP (HM-16.22 PCMP as anchor)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **VTM-13.0 GCMP - Over HM-16.22 PCMP** | | | | | |
|  | **End-to-end WS-PSNR** | | | **End-to-end S-PSNR-NN** | | |
|  | Y | U | V | Y | U | V |
| Class S1 | -34.88% | -39.43% | -41.50% | -34.83% | -39.36% | -41.45% |
| Class S2 | -39.20% | -38.33% | -40.41% | -39.21% | -38.34% | -40.45% |
| **Overall** | -36.61% | -38.99% | -41.06% | -36.58% | -38.95% | -41.05% |

5 Recommendations

The AHG recommends:

• To continue software development of the 360Lib software package.

[JVET-W0007](https://jvet-experts.org/doc_end_user/current_document.php?id=10954) JVET AHG report: Coding of HDR/WCG material (AHG7) [A. Segall, E. François, W. Husak, S. Iwamura, D. Rusanovskyy]

1 Mandates

The AHG was established with the following mandates:

• Study and evaluate available HDR/WCG test content.

• Study objective metrics for quality assessment of HDR/WCG material, including investigation of the correlation between subjective and objective results.

• Compare the performance of the VTM and HM for HDR/WCG content.

• Generate CTC anchors for the VTM according to JVET-V2011.

• Study the luma/chroma bit allocation in the HDR CTC, especially for HLG content.

• Coordinate with AHG4 in preparation for verification testing for HDR video content.

• Study additional aspects of coding HDR/WCG content.

2 Activities

The primary activity of the AhG was related to the mandates of (i) comparing the performance of the VTM for HDR/WCG content and (ii) coordinating with AHG4 in preparation for verification testing for HDR video content. This work is described in the following subsection.

2.1 Anchor Generation

The AhG generated CTC anchors for the VTM according to JVET-V2011. A summary of the performance is provided below, and more detailed information may be found in the included XLS data.

The reported gain in RA configuration is attributed to changes in the MCTF filtering from the adoption of JVET-V0056.

2.1.1 VTM 13.0 versus VTM 12.0

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **All Intra** | | | | | | | | | |
|  | **Over VTM-12.0** | | | | | | | | | |
|  |  |  | **wPSNR** | | | **PSNR** | | |  |  |
|  | DE100 | PSNR-L100 | Y | U | V | Y | U | V | EncT | DecT |
| Class H1 | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 105% | 100% |
| Class H2 |  |  |  |  |  | 0.00% | 0.00% | 0.00% | 104% | 101% |
| **Overall** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 105% | 101% |
|  |  |  |  |  |  |  |  |  |  |  |
|  | **Random Access** | | | | | | | | | |
|  | **Over VTM-12.0** | | | | | | | | | |
|  |  |  | **wPSNR** | | | **PSNR** | | |  |  |
|  | DE100 | PSNR-L100 | Y | U | V | Y | U | V | EncT | DecT |
| Class H1 | -0.40% | -0.66% | -0.68% | -0.49% | -0.69% | -0.61% | -0.38% | -0.73% | 101% | 101% |
| Class H2 |  |  |  |  |  | -0.60% | -0.34% | -0.37% | 101% | 102% |
| **Overall** | -0.40% | -0.66% | -0.68% | -0.49% | -0.69% | -0.61% | -0.37% | -0.60% | 101% | 102% |

2.2 Coordinating with AHG4

The AHG coordinated with AHG4 to prepare for the HDR verification test dry run. This included generating cropped versions of the 4096x2160 content and re-coding all content using VTM-12.0 and HM-16.23.

3 Contributions

There are two contributions related to HDR video coding.

JVET-W0041

AHG4: Status Report on HDR Video Verification Test Preparation A. Segall, M. Wien, V. Baroncini, K. Andersson

JVET-W0042

AHG4: Agenda and report of the AHG meeting on the HDR verification test preparation on 2021-06-10 A. Segall, M. Wien, V. Baroncini (AHG coordinators)

4 Recommendations

The AHG recommends the following:

• Review all input contributions

[JVET-W0008](https://jvet-experts.org/doc_end_user/current_document.php?id=10955) JVET AHG report: High bit depth, high bit rate, and high frame rate coding (AHG8) [A. Browne, T. Ikai, D. Rusanovskyy, M. Sarwer, X. Xiu, Y. Yu]

2 Activities

The AHG used the main JVET reflector, jvet@lists.rwth-aachen.de, with [AHG8] in message headers. A CE upload notification was sent on 25th March, and a formal kick-off message was sent on 31st May. In addition there was frequent communication between the CE coordinators and participants with respect to both the adoptions from the previous CE and the current CE.

In total there are 31 high bit depth related contributions, but none related to high frame rate. The following section lists these contributions

3 Contributions

The contributions can be split into CE and CE-related (19 contributions), and others (12 contributions).

3.1 CE and CE-related

JVET-W0022, “CE: Summary Report on Entropy Coding for High Bit Depth and High Bit Rate Coding”, D. Rusanovskyy, K. Naser, M. G. Sarwer, F. Wang

JVET-W0044, “CE3.1: CABAC-bypass alignment for high bit-depth coding”, M. G. Sarwer, J. Chen, Y. Ye, R. -L. Liao (Alibaba)

JVET-W0045, “CE-3.2: a high throughput mode for high bit depth and high bit rate extensions”, F. Wang, Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO), M. Sarwer, J. Chen, Y. Ye, R. Liao (Alibaba)

JVET-W0046, “CE-1.1: coding of last significant coefficient position for high bit depth and high bit rate extensions”, F. Wang, L. Xu, Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO)

JVET-W0047, “Crosscheck of JVET-W0044: CE3.1: CABAC-bypass alignment for high bit-depth”, A. Browne (Sony)

JVET-W0048, “Cross-check on JVET-W0046: Coding of last significant coefficient position for high bit depth and high bit rate”, D.Rusanovskyy (Qualcomm)

JVET-W0050, “[CE2.1][CE2.2] Content Adaptive Transform Precision”, K. Naser, F. Galpin, T. Poirier, F. Le Leannec (InterDigital)

JVET-W0051, “CE-related: Additional bypass coding for high throughput CABAC”, M. G. Sarwer, J. Chen, Y. Ye, R. -L. Liao ( Alibaba)

JVET-W0052, “CE-related: CABAC skip mode”, K. Abe, T. Toma, V. Drugeon (Panasonic)

JVET-W0058, “crosscheck of CE-3.2: a high throughput mode for high bit depth and high bit rate extensions”, K. Naser (InterDigital)

JVET-W0092, “CE-related: On constraining inverse transform precision for high bit depth and high bitrate video coding”, L. Kerofsky, D. Rusanovskyy, M. Karczewicz (Qualcomm)

JVET-W0094, “Cross-check on JVET-W0050: [CE2.1][CE2.2] Content Adaptive Transform Precision (CE2.2)”, D. Rusanovskyy (Qualcomm)

JVET-W0114, “CE-related: High throughput mode for high bit-depth coding – harmonization of CE-3.2 and CE-1.1”, T. Tsukuba, M. Ikeda, T. Suzuki (Sony)

JVET-W0116, “CE-3 related: a different alignment position for high throughput mode”, F. Wang, Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO)

JVET-W0117, “CE-3 related: a slice level high throughput mode”, F. Wang, Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO)

JVET-W0118, “CE-3 related: High throughput 4:4:4 16 intra profile for VVC”, F. Wang, Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO)

JVET-W0135, “Cross-check on JVET-W0118: CE related: High throughput 4:4:4 16 intra profile for VVC”, T. Tsukuba (Sony)

JVET-W0138, “Crosscheck of JVET-W0114(CE-related: High throughput mode for high bit-depth coding – harmonization of CE-3.2 and CE-1.1)”, F. Wang(OPPO)

JVET-W0139, “Crosscheck of JVET-W0052: CE-related: CABAC skip mode”, T. Zhou, T. Ikai (Sharp)

3.2 Other contributions

JVET-W0055, “AHG8: Signaling TSRC rice parameter in slice header extension”, T. Tsukuba, M. Ikeda, T. Suzuki (Sony)

JVET-W0060, “AHG8: A constraint of max CTU size and tile on WPP for high bit rate coding”, K. Kondo, M. Ikeda (Sony)

JVET-W0064, “AHG8: Constraints on transforms and high precision operation”, T. Ikai, T. Zhou, T. Hashimoto (Sharp)

JVET-W0070, “[AHG8] SPS Cleanup for VVC operation range extension”, K. Naser, F. Galpin, T. Poirier, F. Le Leannec (InterDigital)

JVET-W0091, “AHG8: On constraining of bit depth of ALF classifier and CCLM derivation for coding of high bit-depth video data”, D.Rusanovskyy, M. Karczewicz (Qualcomm)

JVET-W0093, “AHG8: On significance, GT1, and GT2 flag coding for high bit depths”, A. Browne, S. Keating, K. Sharman (Sony)

JVET-W0109, “AHG8: Removal of a prevention mechanism of extended precision for low bit-depth”, T. Tsukuba, M. Ikeda, T. Suzuki (Sony)

JVET-W0115, “AHG8: A study on Bin-to-Bit ratio of VTM-13.0 for high bit depth coding”, T. Tsukuba, M. Ikeda, T. Suzuki (Sony)

JVET-W0121, “AHG8: On signalling of sps\_ts\_residual\_coding\_rice\_present\_in\_sh\_flag”, H.-J. Jhu, X. Xiu, Y.-W. Chen, W. Chen, C.-W. Kuo, X. Wang (Kwai)

JVET-W0136, “Suggested initial profile text for VVC operation range extension”, T. Ikai (Sharp)

JVET-W0140, “Crosscheck of JVET-W0060: AHG8: A constraint of max CTU size and tile on WPP for high bit rate coding”, T. Zhou, T. Ikai (Sharp)

JVET-W0141, “Crosscheck of JVET-W0064 (AHG8: Constraints on transforms and high precision operation)”, K. Kondo (Sony)

4 Benchmarks

The benchmarks for this section were generated using the high bit depth, high bit rate CTC (JVET-U2018). The HM anchor was generated using the RExt lossy and lossless CTC (JVET-U1100) with content and QP settings taken from JVET-U2018.

4.1 VTM13.0 versus VTM12.0

Following the adoption of JVET-V0047, JVET-V0054, and JVET-V0106 from the CE at the V meeting, gains are observed over the low QP range and for lossless coding.

4.1.1 Low QP range

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **HDR PQ** |  |  | **AI** |  |  |  |  |  |
|  |  |  | **Over VTM12.0** |  |  |  |  |  |
|  | wPsnrY | wPsnrU | wPsnrV | psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -2.03% | -2.53% | -2.71% | -2.04% | -2.56% | -2.73% | 108% | 97% |
| PQ422 | -1.93% | -2.15% | -2.29% | -1.93% | -2.15% | -2.27% | 108% | 98% |
| **Overall** | -1.98% | -2.34% | -2.50% | -1.98% | -2.35% | -2.50% | 108% | 97% |
|  |  |  |  |  |  |  |  |  |
|  |  |  | **LDB** |  |  |  |  |  |
|  |  |  | **Over VTM12.0** |  |  |  |  |  |
|  | wPsnrY | wPsnrU | wPsnrV | psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -0.82% | -0.88% | -0.95% | -0.83% | -0.88% | -0.95% | 105% | 102% |
| PQ422 | -0.80% | -0.69% | -0.77% | -0.81% | -0.70% | -0.77% | 105% | 101% |
| **Overall** | -0.81% | -0.79% | -0.86% | -0.82% | -0.79% | -0.86% | 105% | 101% |
|  |  |  |  |  |  |  |  |  |
|  |  |  | **RA** |  |  |  |  |  |
|  |  |  | **Over VTM12.0** |  |  |  |  |  |
|  | wPsnrY | wPsnrU | wPsnrV | psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -0.83% | -0.90% | -0.98% | -0.84% | -0.89% | -0.98% | 105% | 100% |
| PQ422 | -0.82% | -0.78% | -0.87% | -0.82% | -0.78% | -0.86% | 106% | 100% |
| **Overall** | -0.83% | -0.84% | -0.92% | -0.83% | -0.84% | -0.92% | 106% | 100% |
|  |  |  |  |  |  |  |  |  |
| **Overall PQ** | -1.20% | -1.32% | -1.43% | -1.21% | -1.33% | -1.43% | 106% | 100% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HDR HLG** |  |  | **AI** |  |  |
|  |  |  | **Over VTM12.0** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| HLG444 | -1.29% | -1.33% | -1.36% | 116% | 97% |
| HLG422 | -1.36% | -1.38% | -1.43% | 110% | 98% |
| **Overall** | -1.32% | -1.36% | -1.40% | 113% | 98% |
|  |  |  |  |  |  |
|  |  |  | **LDB** |  |  |
|  |  |  | **Over VTM12.0** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| HLG444 | -0.70% | -0.98% | -1.04% | 107% | 97% |
| HLG422 | -0.61% | -0.73% | -0.82% | 107% | 98% |
| **Overall** | -0.66% | -0.86% | -0.93% | 107% | 98% |
|  |  |  |  |  |  |
|  |  |  | **RA** |  |  |
|  |  |  | **Over VTM12.0** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| HLG444 | -0.60% | -0.89% | -0.95% | 107% | 98% |
| HLG422 | -0.53% | -0.72% | -0.81% | 105% | 98% |
| **Overall** | -0.57% | -0.80% | -0.88% | 106% | 98% |
|  |  |  |  |  |  |
| **Overall HLG** | -0.85% | -1.00% | -1.07% | 109% | 98% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SVT RGB** |  |  | **AI** |  |  |
|  |  |  | **Over VTM12.0** |  |  |
|  | psnrG | psnrB | psnrR | EncT | DecT |
| SVT16 | -28.04% | -26.00% | -25.98% | 118% | 105% |
| SVT12 | -4.42% | -4.33% | -4.33% | 106% | 103% |
| **Overall** | -16.23% | -15.17% | -15.16% | 112% | 104% |
|  |  |  |  |  |  |
|  |  |  | **LDB** |  |  |
|  |  |  | **Over VTM12.0** |  |  |
|  | psnrG | psnrB | psnrR | EncT | DecT |
| SVT16 | -18.45% | -17.60% | -17.63% | 122% | 102% |
| SVT12 | -1.64% | -1.56% | -1.56% | 108% | 101% |
| **Overall** | -10.04% | -9.58% | -9.59% | 115% | 101% |
|  |  |  |  |  |  |
|  |  |  | **RA** |  |  |
|  |  |  | **Over VTM12.0** |  |  |
|  | psnrG | psnrB | psnrR | EncT | DecT |
| SVT16 | -17.05% | -17.03% | -17.01% | 120% | 102% |
| SVT12 | -1.53% | -1.50% | -1.50% | 108% | 102% |
| **Overall** | -9.29% | -9.27% | -9.26% | 114% | 102% |
|  |  |  |  |  |  |
| **Overall RGB** | -11.85% | -11.34% | -11.34% | 113% | 102% |

4.1.2 Standard QP range

The gains for random access can be attributed to the adoption of MCTF changes in JVET-V0056.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Random Access** | | | | | | | | | |
|  | **Over VTM12.0** | | | | | | | | | |
|  |  |  | **wPSNR** |  |  | **PSNR** |  |  |  |  |
|  | DE100 | PSNR-L100 | Y | U | V | Y | U | V | EncT | DecT |
| Class H1 | -0.21% | -0.46% | -0.58% | -0.46% | -0.68% | -0.49% | -0.47% | -0.68% | 102% | 102% |
| Class H2 |  |  |  |  |  | -0.44% | -0.28% | -0.23% | 101% | 102% |
| **Overall** | -0.21% | -0.46% | -0.58% | -0.46% | -0.68% | -0.46% | -0.38% | -0.46% | 102% | 102% |
|  |  |  |  |  |  |  |  |  |  |  |
|  | **All Intra** | | | | | | | | | |
|  | **Over VTM12.0** | | | | | | | | | |
|  |  |  | **wPSNR** | | | **PSNR** | | |  |  |
|  | DE100 | PSNR-L100 | Y | U | V | Y | U | V | EncT | DecT |
| Class H1 | -0.06% | -0.01% | 0.00% | 0.02% | -0.09% | 0.00% | -0.01% | -0.10% | 101% | 98% |
| Class H2 |  |  |  |  |  | 0.00% | 0.06% | -0.02% | 100% | 98% |
| **Overall** | -0.06% | -0.01% | 0.00% | 0.02% | -0.09% | 0.00% | 0.02% | -0.06% | 100% | 98% |

4.1.3 Lossless

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PQ** | **All Intra** | | | **Low delay B** | | | **Random Access** | | |
| **ratio** | | bit-rate savings | **ratio** | | bit-rate savings | **ratio** | | bit-rate savings |
| VTM12.0 | VTM13.0 | VTM12.0 | VTM13.0 | VTM12.0 | VTM13.0 |
| PQ444 | 2.5 | 2.6 | -4.12% | 3.0 | 3.1 | -2.51% | 3.0 | 3.1 | -2.38% |
| PQ422 | 2.2 | 2.4 | -5.08% | 2.8 | 2.9 | -2.81% | 2.8 | 2.9 | -2.64% |
| **Overall** | **2.4** | **2.5** | **-4.60%** | **2.9** | **3.0** | **-2.66%** | **2.9** | **3.0** | **-2.51%** |
| Enc Time[%] | 102% | | | 123% | | | 126% | | |
| Dec Time[%] | 113% | | | 118% | | | 114% | | |
|  |  |  |  |  |  |  |  |  |  |
| **HLG** | **All Intra** | | | **Low delay B** | | | **Random Access** | | |
| **ratio** | | bit-rate savings | **ratio** | | bit-rate savings | **ratio** | | bit-rate savings |
| VTM12.0 | VTM13.0 | VTM12.0 | VTM13.0 | VTM12.0 | VTM13.0 |
| HLG444 | 1.7 | 1.8 | -6.74% | 1.9 | 2.0 | -2.38% | 1.9 | 2.0 | -2.23% |
| HLG422 | 1.6 | 1.7 | -8.05% | 1.9 | 1.9 | -2.67% | 1.9 | 1.9 | -2.51% |
| **Overall** | **1.6** | **1.7** | **-7.40%** | **1.9** | **1.9** | **-2.52%** | **1.9** | **1.9** | **-2.37%** |
| Enc Time[%] | 95% | | | 144% | | | 152% | | |
| Dec Time[%] | 125% | | | 130% | | | 127% | | |
|  |  |  |  |  |  |  |  |  |  |
| **SVT** | **All Intra** | | | **Low delay B** | | | **Random Access** | | |
| **ratio** | | bit-rate savings | **ratio** | | bit-rate savings | **ratio** | | bit-rate savings |
| VTM12.0 | VTM13.0 | VTM12.0 | VTM13.0 | VTM12.0 | VTM13.0 |
| SVT16 | 0.7 | 1.2 | -42.41% | 0.7 | 1.2 | -44.39% | 0.7 | 1.2 | -44.44% |
| SVT12 | 1.0 | 1.3 | -24.05% | 1.0 | 1.3 | -22.94% | 1.0 | 1.3 | -22.65% |
| **Overall** | **0.8** | **1.2** | **-33.23%** | **0.8** | **1.3** | **-33.67%** | **0.8** | **1.3** | **-33.54%** |
| Enc Time[%] | 117% | | | 135% | | | 137% | | |
| Dec Time[%] | 109% | | | 94% | | | 93% | | |

4.2 VTM13.0 versus HM16.23

4.2.1 Low QP range

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HDR PQ** | **AI** | | |  |  |
|  | **Over HM 16.23** | | |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -7.28% | -8.43% | -9.16% | 3221% | 169% |
| PQ422 | -9.94% | -14.14% | -14.75% | 2743% | 165% |
| **Overall** | -8.61% | -11.28% | -11.96% | 2982% | 167% |
|  |  |  |  |  |  |
|  | **LDB** | | |  |  |
|  | **Over HM 16.23** | | |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -6.73% | -5.00% | -6.63% | 314% | 167% |
| PQ422 | -7.95% | -11.75% | -12.42% | 413% | 165% |
| **Overall** | -7.34% | -8.38% | -9.52% | 364% | 166% |
|  |  |  |  |  |  |
|  | **RA** | | |  |  |
|  | **Over HM 16.23** | | |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -6.88% | -5.15% | -6.48% | 362% | 164% |
| PQ422 | -8.21% | -12.16% | -12.44% | 501% | 163% |
| **Overall** | -7.54% | -8.66% | -9.46% | 431% | 164% |
|  |  |  |  |  |  |
| **Overall PQ** | -7.83% | -9.44% | -10.31% | 1259% | 166% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HDR HLG** |  |  | **AI** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| HLG444 | -4.63% | -6.64% | -6.95% | 3706% | 166% |
| HLG422 | -5.27% | -8.00% | -8.02% | 3241% | 164% |
| **Overall** | -4.95% | -7.32% | -7.49% | 3473% | 165% |
|  |  |  |  |  |  |
|  |  |  | **LDB** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| HLG444 | -6.01% | -8.34% | -7.72% | 430% | 154% |
| HLG422 | -5.48% | -9.48% | -7.96% | 527% | 160% |
| **Overall** | -5.74% | -8.91% | -7.84% | 479% | 157% |
|  |  |  |  |  |  |
|  |  |  | **RA** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| HLG444 | -6.48% | -8.85% | -8.04% | 479% | 154% |
| HLG422 | -6.00% | -9.80% | -8.15% | 645% | 159% |
| **Overall** | -6.24% | -9.32% | -8.09% | 562% | 156% |
|  |  |  |  |  |  |
| **Overall HLG** | -5.64% | -8.52% | -7.81% | 1505% | 159% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SVT RGB** |  |  | **AI** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrG | psnrB | psnrR | EncT | DecT |
| SVT16 | -2.36% | -1.56% | -1.58% | 5808% | 151% |
| SVT12 | -4.26% | -3.69% | -3.79% | 5524% | 142% |
| **Overall** | -3.31% | -2.62% | -2.68% | 5666% | 147% |
|  |  |  |  |  |  |
|  |  |  | **LDB** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrG | psnrB | psnrR | EncT | DecT |
| SVT16 | -2.29% | -0.78% | -0.77% | 715% | 147% |
| SVT12 | -4.46% | -2.12% | -2.10% | 676% | 142% |
| **Overall** | -3.37% | -1.45% | -1.44% | 696% | 145% |
|  |  |  |  |  |  |
|  |  |  | **RA** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrG | psnrB | psnrR | EncT | DecT |
| SVT16 | -2.40% | -0.95% | -0.94% | 867% | 147% |
| SVT12 | -4.72% | -2.32% | -2.30% | 843% | 143% |
| **Overall** | -3.56% | -1.63% | -1.62% | 855% | 145% |
|  |  |  |  |  |  |
| **Overall RGB** | -3.41% | -1.90% | -1.91% | 2406% | 145% |

4.2.2 Lossless

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PQ** | **All Intra** | | | **Low delay B** | | | **Random Access** | | |
| **ratio** | | bit-rate savings | **ratio** | | bit-rate savings | **ratio** | | bit-rate savings |
| HM16.23 | VTM13.0 | HM16.23 | VTM13.0 | HM16.23 | VTM13.0 |
| PQ444 | 2.5 | 2.6 | -1.27% | 3.1 | 3.1 | 1.59% | 3.1 | 3.1 | 1.36% |
| PQ422 | 2.3 | 2.4 | -2.43% | 2.9 | 2.9 | 0.23% | 2.9 | 2.9 | 0.06% |
| **Overall** | **2.4** | **2.5** | **-1.85%** | **3.0** | **3.0** | **0.91%** | **3.0** | **3.0** | **0.71%** |
| Enc Time[%] | 7538% | | | 1077% | | | 1093% | | |
| Dec Time[%] | 178% | | | 163% | | | 159% | | |
|  |  |  |  |  |  |  |  |  |  |
| **HLG** | **All Intra** | | | **Low delay B** | | | **Random Access** | | |
| **ratio** | | bit-rate savings | **ratio** | | bit-rate savings | **ratio** | | bit-rate savings |
| HM16.23 | VTM13.0 | HM16.23 | VTM13.0 | HM16.23 | VTM13.0 |
| HLG444 | 1.8 | 1.8 | -0.63% | 1.9 | 2.0 | -0.33% | 2.0 | 2.0 | -0.41% |
| HLG422 | 1.7 | 1.7 | -1.12% | 1.9 | 1.9 | -0.65% | 1.9 | 1.9 | -0.73% |
| **Overall** | **1.7** | **1.7** | **-0.88%** | **1.9** | **1.9** | **-0.49%** | **1.9** | **1.9** | **-0.57%** |
| Enc Time[%] | 7894% | | | 1098% | | | 1099% | | |
| Dec Time[%] | 172% | | | 151% | | | 148% | | |
|  |  |  |  |  |  |  |  |  |  |
| **SVT** | **All Intra** | | | **Low delay B** | | | **Random Access** | | |
| **ratio** | | bit-rate savings | **ratio** | | bit-rate savings | **ratio** | | bit-rate savings |
| HM16.23 | VTM13.0 | HM16.23 | VTM13.0 | HM16.23 | VTM13.0 |
| SVT16 | 1.2 | 1.2 | -0.28% | 1.2 | 1.2 | 0.19% | 1.2 | 1.2 | 0.14% |
| SVT12 | 1.3 | 1.3 | -0.51% | 1.3 | 1.3 | 0.05% | 1.3 | 1.3 | -0.01% |
| **Overall** | **1.2** | **1.2** | **-0.39%** | **1.3** | **1.3** | **0.12%** | **1.3** | **1.3** | **0.06%** |
| Enc Time[%] | 9042% | | | 1120% | | | 1097% | | |
| Dec Time[%] | 145% | | | 142% | | | 142% | | |

5 Recommendations

The AHG recommends the following:

• To review all related contributions;

• To continue high bit depth, high bit rate, and high frame rate studies.

• To consider the timeline for high bit depth profile development in VVC version 2.

It is pointed out that in the standard QP range the performance compared to HEVC is expected to be similar to 10 bit case.

It is asked what a useful QP range for 12 bit HDR consumer quality might be? Could this be somewhat between low and standard QP range. It is mentioned that JVET-W0115 refers to that.

[JVET-W0009](https://jvet-experts.org/doc_end_user/current_document.php?id=10956) JVET AHG report: SEI message studies (AHG9) [J. Boyce, S. McCarthy, C. Fogg, P. de Lagrange, A. Luthra, G. J. Sullivan, A. Tourapis, Y.-K. Wang, S. Wenger]

2 Related contributions

A total of 12 contributions, not including cross-checks, are identified relating to AHG9, of which:

• 4 contributions relate to the mandate to study the SEI messages in VSEI, VVC, HEVC, and AVC;

• 2 contributions relate to the mandate to collect software and showcase information of SEI messages;

• 4 contributions relate to the mandate to identify potential needs for additional SEI messages;

• no contributions relate to the mandate to investigate the possible need of mandatory post processing in the context of SEI messages; and

• 2 contributions relate to the mandate to study SEI messages defined in HEVC and AVC for potential use in the VVC context.

The following is a list of contributions related to AHG9.

2.1 Study the SEI messages in VSEI, VVC, HEVC and AVC

2.1.1 Errata, bug fixes, and clarifications

JVET-W0077 AHG9: Comments on multiview-related SEI messages in VSEI [B. Choi, S. Wenger, S. Liu (Tencent)]

JVET-W0080 AHG9: some errata and clarification items for Additional SEI messages for VSEI [E. François, M. Radosavljevic (InterDigital)]

JVET-W0083 AHG9: Bug fixes for some SEI messages in the VSEI amendment [Y.-K. Wang, Y. Wang, L. Zhang (Bytedance)]

2.1.2 Grain blending process for film grain characteristics SEI message

JVET-W0095 AHG9: Bit-accurate grain blending process for film grain characteristics SEI message [S. McCarthy, P. Yin, W. Husak, F. Pu, T. Lu, T. Chen (Dolby), E. François, M. Radosavljević (InterDigital), V. G R, K. Patankar, S. Kadaramandalgi, Ajayshyam (Ittiam)]

2.2 Collect software and showcase information

JVET-W0072 AHG9: Enhancement of film grain parameter estimation for different intensity intervals [M. Radosavljević, E. François (Interdigital), W. Hamidouche, T. Amestoy, G. Gautier (INSA)]

JVET-W0096 AHG9: Demonstration of AVC FGC SEI in real-time bit-accurate grain blending process on smartphone [V. G R, J. Shingala, S. Kadaramandalgi, Ajayshyam (Ittiam), S. McCarthy, P. Yin, W. Husak, F. Pu, T. Lu, T. Chen (Dolby)]

2.3 Identify potential needs for additional SEI messages

JVET-W0074 HLS for ALF parameters for RPR [P. Bordes, F. Galpin, K. Naser, F. Leleannec (InterDigital)]

(Method 1 and an aspect of Method 2 relate to this mandate)

JVET-W0076 AHG9: Independently coded region output SEI message [B. Choi, S. Wenger, X. Li, S. Liu (Tencent)]

JVET-W0085 AHG9: Picture quality metrics SEI message [Y. He, M. Coban, D. Rusanovskyy, M. Karczewicz (Qualcomm)]

JVET-W0104 AHG9: Resampling SEI message [T. Poirier, G. Martin-Cocher, F. Le Léannec, K. Naser (InterDigital)]

2.4 Study SEI messages defined in HEVC and AVC

JVET-W0071 AHG9: Green Metadata SEI message for VVC [C. Herglotz, M. Kränzler, A. Kaup (FAU), E. Francois, M. Radosavljevic, E. Reinhard (InterDigital), X. Ducloux (Harmonic), D. Menard (INSA)]

JVET-W0078 AHG9: Multiview view position SEI message [B. Choi, S. Wenger, S. Liu (Tencent)]

3 Activities

The regular JVET e-mail reflector was used for discussions (jvet@lists.rwth-aachen.de) with [AHG9] in message headers. There were no emails sent to the JVET reflector during the AHG period.

4 Recommendations

The AHG recommends to:

• Review all related contributions;

• Continue SEI messages studies.

[JVET-W0010](https://jvet-experts.org/doc_end_user/current_document.php?id=10957) JVET AHG report: Encoding algorithm optimization (AHG10) [A. Duenas, R. Sjöberg, A. Tourapis]

2 Overview of input documents related to the AHG

The following input documents were identified to be related to this AHG:

2.1 JVET-W0043 AHG 10: Alignment of smooth QP control with adaptive QP in VTM

This contribution asserts that the smooth QP control method in VTM does not work properly with adaptive QP because the correct lambda is not used. It then suggests fixing the lambda but also changing the default settings for smooth QP control to only be applied on intra pictures since, it is claimed, that such modification will result in the most impact on visual quality. Although not presented, it was also suggested that additional related modifications and the creation of appropriate configuration files for inter pictures could also be made so as to provide additional flexibility. It is claimed that some subjective gains can be seen especially for some of the HDR material.

2.2 JVET-W0061 AHG10: Encoder MV selections and DMVR

This contribution presents a non-normative method that can improve the performance of the decoder side motion vector refinement method that exists in VVC.

In particular, DMVR is a method that permits the refinement of motion vectors for inter blocks that are coded using merge mode. The refinement is performed on a subblock basis. It is asserted that the refined MVs for the subblocks may sometimes have poor correlation since the refinement is carried out for each subblock independently. It is further asserted that the poorly correlated subblock MVs are likely to introduce noticeable subblock boundary artifacts, especially for low spatial activity regions.

The contribution proposes an encoder-only method to reduce the risk of subblock boundary artifacts being introduced by DMVR. The method restricts the encoder from choosing a pair of DMVR-applicable bi-MVs when a current block together with the corresponding DMVR-refined MVs are determined to be prone to give the subblock boundary artifacts.

The method has been implemented in VTM-13.0. The BD-rate PSNR impact is reported to be 0.19% for SDR RA CTC and 0.20% for HDR RA CTC. It is asserted that the proposed method can improve subjective quality when refinement is based on low spatial activity regions, especially when the source block to be predicted happens to contain an edge and the residual coding is too coarse to take care of the artifact. It is proposed to include the method in VTM to enable improved subjective robustness of DMVR.

2.3 JVET-W0082 AHG10: GOP-based RPR encoder control

This contribution proposes a GOP-based selection method for reference picture resampling (RPR) in the VTM that decides when to encode pictures in reduced resolution with RPR or if its better to encode them in the source resolution. The selection is based on QP and picture self-similarity after re-scaling. The self-similarity test is only conducted on the first source picture in display order within each GOP. If the re-scaled picture is determined to have sufficient similarity with the source picture, all pictures in the GOP are encoded at reduced resolution, otherwise they are encoded in the source resolution.

The BD-rate performance (Y/Cb/Cr) compared to CTC with VTM-13.0 as anchor is as follows:

RA (luma/Cb/Cr): -0.03%/0.61%/0.58%

LDB (luma/Cb/Cr): 0.27%/0.35%/0.35%

AI (luma/Cb/Cr): 0.11% /0.63%/0.67%

The BD-rate performance for QP 27 to 47 with VTM-13.0 as anchor:

RA (luma/Cb/Cr): -1.57%/6.94%/5.94%

LDB (luma/Cb/Cr): 1.42%/4.06%/3.41%

AI (luma/Cb/Cr): -0.41%/6.39%/4.76%

It is asserted that the method can, in some cases, improve subjective quality compared to encoding in the source resolution and at low bitrates.

2.4 JVET-W0129 AHG10: Using original samples for ALF optimization

In VTM-13.0, motion compensated temporal pre-filtering (MCTF) is used for the random access (RA) configuration. Samples filtered by MCTF are used for sample adaptive offset (SAO) and adaptive loop filter (ALF) optimization in the common test conditions (CTC). It is configurable to use original samples for SAO and ALF optimization as JVET-V0095. However, in HM and ECM, filtered samples are used for SAO and bilateral filter optimization and original ones are used for ALF in CTC. To align the different reference software implementations, this contribution proposes to use original samples only for ALF optimization and enable it in the CTC for VTM. The proposed method was tested under SDR, HDR, and high bit depth/high bit rate test conditions.

Recommended to review these input docs.

[JVET-W0011](https://jvet-experts.org/doc_end_user/current_document.php?id=10958) JVET AHG report: Neural network-based video coding (AHG11) [E. Alshina, S. Liu, A. Segall, J. Chen, F. Galpin, J. Pfaff, S. S. Wang, Z. Wang, M. Wien, P. Wu, J. Xu]

2 Activities

The AHG used the main JVET reflector, jvet@lists.rwth-aachen.de, for email exchange with AHG11 included in the subject lines. Two emails were exchanged on the reflector.

2.1 EE Coordination

The AHG finalized, conducted and discussed the EE on NN based video coding. This was accomplished via the reflector. And, the final version of the EE description was announced on May 14, 2021.

A summary report for the EE is available at this meeting as:

JVET-W0023 EE1: Summary of Exploration Experiments on Neural Network-based Video Coding E. Alshina, S. Liu, W. Chen, F. Galpin, Y. Li, Z. Ma, H. Wang

2.2 CTC Refinement and Support

The AHG refined and released the CTC test conditions on May 27, 2021. (The CTC was updated on June 5, 2021, but the only change was in the document header.)

Additionally, to better support and automate the calculation of complexity metrics, the AHG created a script that computes the number of additions and multiplications from either a pytorch or tensorflow model. The script was made available on the AHG git repository and can be found at https://vcgit.hhi.fraunhofer.de/jvet-ahg-nnvc/nnvc-ctc/-/tree/master/scripts

Furthermore, to support the decision at the previous meeting to update the NNVC anchor to include JVET-V0056, two steps were taken:

1. A repository was created that included VTM-11.0 with the JVET-V0056 backport. This repository was then referenced as the anchor within the CTC document. The tagged repository can be found at: https://vcgit.hhi.fraunhofer.de/jvet-ahg-nnvc/VVCSoftware\_VTM/-/tree/VTM-11.0\_nnvc

2. A patch to apply the JVET-V0056 backport to VTM-11.0 was created and also distributed via the AhG git repository. This patch should be useful for participants who have existing software that is based on VTM-11.0 and would like to incorporate the JVET-V0056 technoloygtechnology. The patch can be found at: https://vcgit.hhi.fraunhofer.de/jvet-ahg-nnvc/nnvc-ctc/-/tree/master/Software%20Patches

Finally, several bugs in the Excel reporting template were addressed. For example, the rate-distortion curve functionality has been broken in earlier versions but was fixed during the AhG period.

The AHG11 coordinators would like to thank Franck Galpin, Christopher Hollmann, Jacob Ström for their generous contributions.

2.3 Anchor Encoding

Anchors for the NN-based video coding activity were created and released on June 4, 2021, within the EE activity. The anchors were also made available on the Git repository used for the AHG activity: https://vcgit.hhi.fraunhofer.de/jvet-ahg-nnvc/nnvc-ctc/-/tree/master

It is noted that the anchor data includes performance data for QP points 22, 27, 32, 37, 42 and 47. The largest QP point was generated to support the EE test on super-resolution methods.

2.4 Coordination with SC29/AG5

The AHG coordinated with SC29/AG5 to prepare a viewing procedure for EE contributions. Sequences were selected for three proposals included in the EE, with each proponent selecting three sequences/QP pair that illustrated the visual benefit of the proposal. These selections emphaszied the RA configurations.

With close coordination with SC29/AG5, it is intended to perform remote viewing sessions to understand the visual benefit of the approaches. This will tentatively be done by viewing each sequence multiple times, in order to give participants the opportunity to focus and evaluate different aspects of the sequences.

2.5 Technical Evaluation

The AHG made meaningful progress on the mandate to evaluate and quantify potential NN based video coding technologies. A summary of the non-EE contribution provided as input to the 23rd meeting is provided below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Performance** | **Dec.T.** | **Features** | **#params** | **kMAC /pixel** |
| [JVET-W0057](https://jvet-experts.org/doc_end_user/current_document.php?id=10872) | RA(BC classes): **3.0**%(Y), 6%(Ch.) | ×500 | fine-tuning per content, CTU level Luma/Chroma independent on/off; input: 128x128 YUV444 + QPmap, FineTune per video sequence (only Bias); LZMA2 for updated weights compression | **0.15M** (323 variable) | 151 (SmallNN) 9469(BigNN) |
| [JVET-W0100](https://jvet-experts.org/doc_end_user/documents/23_Teleconference/wg11/JVET-W0100-v1.zip) | RA :**12** % (Y) 27% (Ch.) AI: **9** % (Y) 22% (Ch.) | ×1000 (RA)  ×545(AI) | Picture on/off control + CTU level model selection (not in AI); Filter dictionary dependes on QP, filter switch granu;larity - on resolution and QP. Number resBlocks **32**. New: Attention residual block = Pre & Rec Frames --> CONV; ReLu;CONV | **528**M (24\*22M) | 1430 |
| JVET-W0101 | RA (BC):**10** % (Y) 25% (Ch.) AI: **7** % (Y) 20% (Ch.) | ×225 (RA) ×255 (AI) | JVET-W0100 but on top of ECM-1.0 | **528**M (24\*22M) | 1430 |
| [JVET-W0081](https://jvet-experts.org/doc_end_user/current_document.php?id=10897) | RA : 1.7 % (Y) --> 1.6% AI: **3.3** % (Y)--> 2.9% | ×64--> 61(AI) | Lower complexity version of JVET-T0073 and JVET-V0105; 16 bits version provided | ~**10**M | 11~100-->6~11 |
| [JVET-W0059](https://jvet-experts.org/doc_end_user/current_document.php?id=10874) | AI (CD): **5** % (Y) |  | After SAO, before ALF; 5 models, switched based on QP, 16 residual blocks, Luma: 160´160 support (CTU + 16 samples on a border); Chroma up-sampled and used as additional input for Luma filtering Chroma: 80´80 support (CTU + 8 samples on a border); Luma down-sampled and used as additional input for Chroma filtering | 5\*(1.4+0.7)M | 380000(?) |
| [JVET-W0099](https://jvet-experts.org/doc_end_user/current_document.php?id=10915) | AI(4K):**12**%(Y); drop 5% (Ch) |  | Input: Predicted Frame, reconstructed Frame, QP map. 16 ResBlocks; L1 loss; Luma as extra input for Chroma | 4.8(Y)+1.1(Ch) M | 5770 (4400 Y + 1370 Ch) |
| [JVET-W0111](https://jvet-experts.org/doc_end_user/current_document.php?id=10927) | AI: **0.3** % (Y) 0.5% (Ch.) | ×40 (AI) | Luma ¯& 2D Transformation --> 3 channes; Ch 2\*1D Tr -->4 channels; 7 channels input --> NN with 8 ResBlocks (3 CONV each); "skip connection " through 5x5 CONV. | 0.15M | 150 |
| [JVET-W0113](https://jvet-experts.org/doc_end_user/current_document.php?id=10929) | RA : **3.5** % (Y) 13% (Ch.)%  Al: **4.8**%(Y) 11% (Ch.) | ×800 (RA) ×439 (AI) | Replaces Deblock and SAO; 8 ResBlocks; (3 CONV each); "skip connection " through 5x5 CONV. | 8\*2.3 M | 37 |
| [JVET-W0131](file:///D:\sciebo\Meetings\20210707_MPEG135_JVET_online\JVET\JVET-W0011_AHG11\current_document.php%3fid=10960) | RA : **5.8** % (Y) 5.9% (Ch.)%  Al: **4.8**%(Y) 5.7% (Ch.) | x41 (RA) x23 (AI) | Replaces Deblock, 3x3 and 1x1 conv layer and leaky ReLU | 4 \* 0.107M | 33.6 |
| [JVET-W0132](https://jvet-experts.org/doc_end_user/current_document.php?id=10961) | RA(4K)´2¯:**4**%(Y); 4% (Ch) RA(4K)´1.5¯:**2.5**%(Y); (drop)1% (Ch) |  | Different variats of up-sampling filters of "simlified ESRGAN " architecture: CONV\*PixelShufflePReLu and just biCubic --> no big difference in performance | 0.5M (PixelShuffle) 0.3M(BiCubic) |  |

3 Input contributions

There are 17 input contriubtions related to the AHG mandates. Seven of the contributions are directly relatd to the EE activity, while the remaining 10 contributions are related to AHG11 but not directly part of the EE. The list of input contributions is provided below.

3.1 EE Related Input Contributions

JVET-W0062

EE1: SSIM based CNN model for in-loop filtering T. Ouyang, Y. Guo, Z. Chen (WHU), X. Xu, S. Liu (Tencent)

JVET-W0063

EE1.2.2: NN-based super resolution (JVET-V0073) T. Chujoh, Y. Yasugi, T. Ikai (Sharp)

JVET-W0105

EE1-2.3: Neural Network-based Super Resolution A. M. Kotra, K. Reuzé, J. Chen, H. Wang, M. Karczewicz, J. Li (Qualcomm)

JVET-W0125

EE1-related: Tests on Decomposition, Compression and Synthesis (DCS)-based Technology (JVET-V0149) Ming Lu, Zhan Ma (NJU), Zhenyu Dai, Dong Wang (OPPO)

JVET-W0130

EE1-1.4: Test on Neural Network-based In-Loop Filter with Large Activation Layer H. Wang, J. Chen, K. Reuze, A.M. Kotra, M. Karczewicz (Qualcomm)

JVET-W0151

EE1: neural network based in-loop filter using depthwise separable convolution and regular convolution L. Wang, S. Lin, X. Xu, S. Liu, C. Auyeung, X. Li (Tencent)

3.2 Non-EE Input Contributions

Reporting

JVET-W0011

JVET AHG report: Neural network-based video coding (AHG11) E. Alshina, S. Liu, A. Segall, J. Chen, F. Galpin, J. Pfaff, S. S. Wang, Z. Wang, M. Wien, P. Wu, J. Xu

Loop Filtering

JVET-W0057

AHG11: Content-adaptive neural network post-processing filter M. Santamaria, Y.-H. Lam, J. Lainema, F. Cricri, R. Ghaznavi-Youvalari, H. Zhang, A. Zare, H. R. Tavakoli, M. Hannuksela (Nokia)

JVET-W0059

AHG11: A Deep In-Loop Filter Method X. Zhang, C. Fang, D. Jiang, J. Lin (Dahua)

JVET-W0100

AHG11: Deep In-Loop Filter with Adaptive Model Selection and External Attention Y. Li, K. Zhang, L. Zhang (Bytedance)

JVET-W0101

AHG11 & AHG12: Deep In-Loop Filter with Adaptive Model Selection and External Attention for Enhanced Compression Beyond VVC Capability Y. Li, K. Zhang, L. Zhang (Bytedance)

JVET-W0113

AHG11: neural network based in-loop filter L. Wang, W. Jiang, X. Xu, S. Liu (Tencent)

Super Resolution

JVET-W0099

AHG11: CNN-based Super Resolution for Video Coding Using Decoded Information C. Lin, L. Zhang, K. Zhang, Y. Li (Bytedance)

JVET-W0131

EE1-related: Neural Network-based in-loop filter with constrained computational complexity H. Wang, J. Chen, K. Reuze, A.M. Kotra, M. Karczewicz (Qualcomm)

JVET-W0132

AHG11: 1.5x/2.0x Upsample method for NN-Based Super-Resolution Post-Filters Y. Yasugi, T. Chujoh, T. Ikai (Sharp)

Intra Prediction

JVET-W0081

AHG11: BD-rate gains vs complexity of NN-based intra prediction T. Dumas, F. Galpin, P. Bordes, F. Leleannec (InterDigital)

JVET-W0111

AHG11: neural network based cross-component prediction model L. Wang, S. Lin, R. Chang, X. Xu, S. Liu (Tencent)

4 Recommendations

The AHG recommends:

• Review all input contributions.

• Perform viewing of the neural network-based video coding tools during the meeting, in coordination with SC29/AG5.

• Continue investigating neural network-based video coding tools, including coding performance and complexity.

It is pointed out that some scripts used in the NN exploration (invoking external libraries) may not be consistent with usual software headers.

[JVET-W0012](https://jvet-experts.org/doc_end_user/current_document.php?id=10959) JVET AHG report: Enhanced compression beyond VVC capability (AHG12) [M. Karczewicz, Y. Ye, L. Zhang, B. Bross, X. Li, K. Naser, H. Yang]

The primary activity of the AHG was the “Exploration experiment on enhanced compression beyond VVC capability” (JVET-V2024). The first indication of the achievable improvement over VVC are the results of the EE2 software base (ECM-1.0) as agreed at JVET-V meeting. The combined improvement of the ECM-1.0 over VTM-11.0 with the improved MCTF from JVET-V0056 for AI, RA and LB configurations are:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | All Intra Main10 | | | | |
|  | Y | U | V | EncT | DecT |
| Class A1 | -5.78% | -9.67% | -10.61% | 219% | 216% |
| Class A2 | -5.47% | -7.85% | -5.90% | 212% | 200% |
| Class B | -5.00% | -7.15% | -7.52% | 246% | 213% |
| Class C | -6.04% | -7.43% | -7.71% | 247% | 198% |
| Class E | -6.20% | -7.84% | -6.72% | 251% | 241% |
| Overall | **-5.64%** | **-7.86%** | **-7.67%** | **236%** | **212%** |
| Class D | -4.86% | -5.53% | -5.56% | 251% | 200% |
| Class F | -10.03% | -11.63% | -12.07% | 214% | 258% |
| Class TGM | -14.61% | -16.09% | -15.97% | 213% | 282% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Random Access Main 10 | | | | |
|  | Y | U | V | EncT | DecT |
| Class A1 | -12.21% | -13.33% | -15.84% | 279% | 453% |
| Class A2 | -13.45% | -14.39% | -14.09% | 268% | 546% |
| Class B | -11.52% | -12.24% | -11.79% | 271% | 481% |
| Class C | -13.26% | -13.23% | -13.39% | 266% | 415% |
| Class E |  |  |  |  |  |
| Overall | **-12.51%** | **-13.15%** | **-13.49%** | **271%** | **469%** |
| Class D | -14.39% | -13.40% | -13.32% | 280% | 449% |
| Class F | -12.13% | -13.26% | -13.60% | 253% | 348% |
| Class TGM | -12.67% | -14.71% | -14.89% | 245% | 255% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Low Delay B Main 10 | | | | |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  |  |  |
| Class A2 |  |  |  |  |  |
| Class B | -9.55% | -9.45% | -9.76% | 237% | 287% |
| Class C | -10.99% | -9.24% | -10.94% | 251% | 254% |
| Class E | -9.22% | -10.12% | -8.20% | 200% | 241% |
| Overall | **-9.95%** | **-9.55%** | **-9.76%** | **231%** | **264%** |
| Class D | -13.37% | -11.54% | -12.00% | 257% | 264% |
| Class F | -10.91% | -10.19% | -10.47% | 233% | 232% |
| Class TGM | -10.70% | -12.82% | -13.62% | 191% | 186% |

3 Contributions

In addition to twelve EE2 contributions, 19 related technical contributions and 2 contributions on the ECM software were received, which can be subdivided as follows:

3.1 Partitioning

JVET-W0110, “AHG12: GPM with inter and intra prediction”, Y. Kidani, H. Kato, K. Kawamura (KDDI)

3.2 In Loop Filters

JVET-W0073, “ALF parameters for RPR”, P. Bordes, F. Galpin, F. Léannec (InterDigital)

JVET-W0074, “HLS for ALF parameters for RPR”, P. Bordes, F. Galpin, K. Naser, F. Léannec (InterDigital)

JVET-W0079, “AHG12: CTB level filter shape selection of CCALF”, M. G. Sarwer, R. -L. Liao, J. Chen, Y. Ye, X. Li (Alibaba)

JVET-W0098, “Non-EE2: Bilateral Inloop Filter on Chroma”, W. Yin, K. Zhang, L. Zhang (Bytedance)

JVET-W0101, “AHG11 & AHG12: Deep In-Loop Filter with Adaptive Model Selection and External Attention for Enhanced Compression Beyond VVC Capability”, Y. Li, K. Zhang, L. Zhang (Bytedance)

JVET-W0128, “AHG12: Alternative classifiers for ALF”, N. Hu, V. Seregin, M. Karczewicz (Qualcomm)

3.3 Intra

JVET-W0067, “EE2-related: Implicit derivation of DIMD blend modes”, X. Li, R.-L. Liao, J. Chen, Y. Ye (Alibaba)

JVET-W0069, “[AHG12] On Intra TMP Boundary Conditions”, K. Naser, T. Poirier, F. Le Léannec, G. Martin-Cocher (InterDigital)

JVET-W0123, “EE2-related: Fusion for template-based intra mode derivation”, K. Cao, N. Hu, V. Seregin, M. Karczewicz (Qualcomm), Y. Wang, K. Zhang, L. Zhang (Bytedance)

JVET-W0124, “EE2-related: Template based intra most probable modes sorting”, K. Cao, N. Hu, V. Seregin, M. Karczewicz (Qualcomm)

JVET-W0126, “EE2 Related - DIMD with implicitly derived multiple blending modes” J. Zhao, S. Paluri, S. Kim (LGE)

3.4 Inter

JVET-W0068, “EE2-related: A combination of CIIP and DIMD/TIMD”, X. Li, R.-L. Liao, J. Chen, Y. Ye (Alibaba)

JVET-W0097, “EE2-related: Combination of EE2-3.3, EE2-3.4 and EE2-3.5”, X. Xiu, C.-W. Kuo, X. Wang (Kwai), R.-L. Liao, Y. Ye, X. Li, J. Chen (Alibaba), Z. Deng, K. Zhang, L. Zhang, N. Zhang, Y. Wang (Bytedance), Y.-J. Chang, H. Huang, V. Seregin, C.-C. Chen, M. Karczewicz (Qualcomm)

JVET-W0106, “EE2-related: Bilateral matching AMVP-merge mode” Z. Zhang, H. Huang, C.-C. Chen, V. Seregin, M. Karczewicz (Qualcomm)

JVET-W0107, “EE2-related: Adaptive decoder side motion vector refinement”, H. Huang, Z. Zhang, V. Seregin, W.-J. Chien, C.-C. Chen, M. Karczewicz (Qualcomm)

JVET-W0112, “AHG12: Diagonal MMVD with ARMC”, Y. Kidani, K. Kawamura (KDDI)

JVET-W0122, “EE2-related: On spatial MV propagation and neighboring template block access for template matching and multi-pass DMVR”, C.-C. Chen, C.-T. Hsieh, H. Huang, V. Seregin, W.-J. Chien, Y.-J. Chang, Z. Zhang, Y. Zhang, M. Karczewicz (Qualcomm)

3.5 Coefficient Coding

JVET-W0108, “EE2-related: Low complexity sign prediction” C. Auyeung, X. Li, S. Liu (Tencent)

3.6 ECM software related

JVET-W0049, “HG12: on the status of the ECM software”, Y. Ye (Alibaba), M. Karczewicz (Qualcomm), Y.-W. Huang (MediaTek), P. Yin (Dolby), D. Wang (OPPO), X. Wang (Kwai), J. Ström (Ericsson), F. Le Leannec (InterDigital), L. Zhang (Bytedance), S.-H. Kim (LGE), M. Hannuksela (Nokia), P. Wu (ZTE)

JVET-W0102, “Preliminary draft of algorithm description for Enhanced Compression Model 1 Software (ECM 1)”. M. Coban (Qualcomm), F. Le Léannec (InterDigital), J. Ström (Ericsson), Y. Ye (Alibaba)

4 Recommendations

The AHG recommends to:

• To review all the related contributions.

• To encourage contribution of new test sequences, especially 8k resolution, which might be included in future test conditions.

# Project development (14)

## Deployment and advertisement of standards (2)

Contributions in this area were discussed in session 13 at 0840–0900 UTC on Monday 12 July 2021 (chaired by JRO).

[JVET-W0021](https://jvet-experts.org/doc_end_user/current_document.php?id=10884) Deployment status of the VVC standard [G. J. Sullivan]

This contribution contains a survey of deployed products and services, publicly available software source code, and related tools supporting the VVC standard (Rec. ITU-T H.266 | ISO/IEC 23090-3).

Except for a restructuring of the document to merge two sections without changing content, revision marking is included to show changes relative to JVET-W0021-v2 of April 2021.

Updates are in 7), 10) and 15) below.

An additional version -v4 was later uploaded. Item 7) below was split into two items since the new information in JVET-W0127 is about an encoder rather than a decoder. Item 3) was also updated with information from JVET-W0184. [Update notes to reflect -v4.]

**Publicly available software source code**

1. **JVET** has developed the **VVC Test Model (VTM)** as its reference software encoder and decoder codebase [6]. It is intended primarily to demonstrate coding efficiency capability and proper interpretation of the syntax and decoding process specified in the standard (but not as a speed-optimized implementation), and is intended to be usable as a starting basis for product implementations. The software is available under a BSD copyright licence. The reference software is under development by JVET as H.266.2 (ex H.VVC.2) and ISO/IEC CD 23090-16. It reached the CD stage of the ISO/IEC approval process in October 2020 and the DIS stage in January 2021 (with a ballot issued to close on 2021-09-30). A description of the encoding algorithms in the test model is also maintained and made available by JVET [13].
2. **InterDigital** developed a multi-threaded VTM decoder, and reported 6–10× speed-up relative to the single-threaded reference software in October 2020 [14]. It is intended to support all features of the VTM. The software was later placed in an accessible repository, and it is available under the same BSD copyright licence as the VTM software [15]. A further update to correspond with the newer VTM and improve speed by about 10% using various scheduling and low-level changes was reported in April 2021 [16].
3. **Fraunhofer HHI** announced the **VVenC** encoder and **VVdeC** decoder open-source software (release 0.1) in September 2020 [17][18][19][20][21]. It includes support for multithreading operation, single-pass rate control, perceptual QP adaptation, and motion-compensated temporal filtering (MCTF). The software has four defined presets for quality/speed tradeoff (called “slow”, “medium”, “fast”, and “faster”). Subjective testing reported in October 2020 indicated that the VVenC encoder had about the same or better subjective compression performance as the VTM encoder when operating in its “medium” speed configuration (operating with MCTF and QP adaptation disabled in the VTM and enabled in VVenC and with rate control disabled in both) with encoding speed more than 100× that of the VTM, for 4K UHD SDR video content [21][22]. As of December 2020, a “slower” preset was added, along with an improved single-pass rate control and a new two-pass rate control [23]. The “slower” preset mode reportedly achieves approximately the same BD-rate coding efficiency as the VTM while providing a speedup of more than 8.6× for UHD and 5.2× for HD sequences relative to the VTM. As of December 2020, with release 0.2, the software is available under a BSD copyright licence. Release 0.3 of March 2021 includes substantial further speed and multithreading improvements [24].
4. **Friedrich–Alexander University** Erlangen–Nürnberg released an open-source bitstream analyser as an add-on for the VTM decoder [25][26]. The analyzer counts the occurrence of coding tools and coding modes used in a decoded bitstream and can be used for evaluating the decoding energy and time demands of VVC features. The software is available under a BSD copyright licence.

**Software decoders**

1. **Sharp** announced a real-time software decoder in June 2020, and issued a corresponding press release in December 2020 [27][28]. As of June 2020, it was reportedly capable of decoding 4K common test conditions (CTC) UHD bitstreams at up to 40 Mbps at more than 60 fps.
2. **Tencent** announced its **O266dec** software decoder with SIMD and multithreading support and an associated FFmpeg/VLC-based video player in October 2020 [21][29][30]. As of December 2020, it is reportedly more than 3× the speed of the VTM reference software decoder when tested under VVC CTC in single-threaded operation and about 20× the VTM decoder speed in 8-thread operation. It could reportedly decode UHD video at more than 60 fps at up to 40 Mbps and decode full HD video at more than 200 fps. In December 2020, a version with mobile platform support based on Arm Neon technology was reported. On an Apple A14 processor (iPhone 12pro) in single-threaded operation, it could reportedly decode 8-bit 1080p CTC bitstreams at more than 50 fps, and in multi-threaded operation it could decode such bitstreams at more than 100 fps and could decode 8-bit 4K UHD bitstreams at more than 30 fps in the random access (RA) example configuration [31]. Although 8-bit operation was more optimized, the decoder also supports 10-bit operation.
3. **Alibaba** announced its **Ali266** decoder for mobile devices (e.g., Android and iPhone) in December 2020 [32]. It includes optimizations for multi-threading, ARM assembly, cache efficiency, and memory usage, particularly for 8 bit video content. Real-time speed was reported for 8 bit 720p, 1080p (using 2–4 threads for up to 60 fps) and 4K (up to 7 Mbps) video content with the ALF feature disabled. In April 2021, updated information was provided with both ALF and CCALF enabled [33]. This version of Ali266 reached 15× speed-up compared to VTM-11.0 on Android P40 with 8 threads on the CTC bitstreams of class A (4K) and class B (1080p), and performed real-time decoding of tested 720p e-commerce content on most phones with 3 threads. In July 2021, updated information was provided about another new version [34]. For CTC classs A1 (4K), A2 (4K), B (1080p) and C (720p), its “slow” preset reportedly achieves BD rates for {Y, U, V, YUV} of {10.95%, −6.11%, −4.50%, 7.11%} under the RA configuration with an average speedup factor of 227 times over VTM 13.0. For internal VOD content, Ali266 “fast” preset was compared to the x265 open source HEVC software encoder “medium” preset, and reportedly achieved BD rates for {Y, U, V, YUV} of {−32.87%, −46.07%, −42.70%, −35.35%} under the RA configuration. Ali266 “fast” preset is aimed at real-time VVC encoding, and currently achieved average encoding speeds of 19.3 fps for classes 1920×1080/1080×1920 and 37.7 fps for classes 1280×720/720×1080.
4. **Kwai** announced **K266Dec** in April 2020 [35]. The decoder is designed from scratch following the VVC specification and can support all coding tools specified in the VVC Main profile. For the Android platform, K266Dec can decode 2K 8-bit CTC bitstreams at 33 fps with single thread on Huawei P40, which is 4.11× of VTM-11.0 decoder. In addition, CTC bitstreams and UGC videos were tested on various Android phones based on the 50B+ daily playback data from Kwai App. For iOS platform, K266Dec can decode 2K 8-bit CTC bitstreams at 94 fps with single thread on iPhone 12 Pro Max, which is 4.78× compared to VTM-11.0 reference decoder.
5. **Bytedance** announced its **BVC** decoder in April 2021 [36]. The BVC decoder supports various platforms such as Android, iOS, Linux/MacOS and Windows. It was optimized especially for mobile platforms with different decoding capabilities. For decoding of 4K 8-bit CTC bitstreams on iPhone 12 (core A14), it is reported that the BVC decoder achieves 22 fps on average with single thread, which is 10× faster than the VTM-11.0 reference decoder, and 55 fps can be achieved with 6 threads. For decoding of 2K 8-bit CTC bitstreams, it is reported that the BVC decoder achieves 86 fps on average with single thread, which is 8.8× faster than the VTM-11.0 reference decoder.
6. **Spin Digital** announced a real-time software decoder and media player supporting VVC in June 2021 [37]. It includes support of the Scalable Main 10 profile and 8K capability at 60 and 120 fps on recent-generation CPUs.

**Services and encoding products**

1. **Bytedance** announced its **BVC** encoder in April 2021 [38]. The majority of the coding tools and features introduced in VVC have been integrated and optimized in the BVC encoder. In addition, it also includes real-world encoder features like scene change detection, rate control and multi-threading. When compared to the HM and VTM, the corresponding global BD-rate savings and speedup factors under the RA configuration are summarized as follows: “fast” preset (6 threads): 34.13% (vs. HM), 17× (HM), 131× (VTM); “slow” preset (6 threads): 41.63% (vs. HM), 1× (HM), 8× (VTM).
2. **KDDI Research** announced a real-time VVC encoder with 4K @60 fps capability in September 2020 [39].
3. **Ateme** launched support for VVC in its Titan family of products, and demonstrated the technology in an OTT channel launched in November 2020 [40].
4. **Bitmovin**, in partnership with Fraunhofer HHI based on VVencC as described in item 3), announced support of VVC in its video encoding platform in November 2020 [41].
5. **MX Player** and **MX TakaTak**, a streaming service in India, began OTT services using VVC in June 2021 [42].

**Bitstream analyser products**

1. **Elecard** announced support for VVC in its StreamEye and StreamAnalyzer products in April 2020 [43].
2. **ViCueSoft** supports VVC in its VQ Analyzer bitstream analysis product, as of late 2020 [44].

**Conformance test sets**

1. A conformance test set is under development by JVET as H.266.1 (ex H.VVC.1) and ISO/IEC CD 23090-15 [45]. It reached the CD stage of the ISO/IEC approval process in October 2020 and the DIS stage in January 2021 (with a ballot issued to close on 2021-09-29).
2. **Allegro DVT** began offering a conformance test set for VVC as early as January 2020 (i.e., initially in preliminary form before the finalization of the standard) [46][47][48].

[JVET-W0165](https://jvet-experts.org/doc_end_user/current_document.php?id=10996) Proposed draft white paper on the VVC standard [G. J. Sullivan, B. Bross]

This contribution contains a proposed draft white paper on the VVC standard (Rec. ITU-T H.266 | ISO/IEC 23090-3) for potential issuance as a JVET output document of the current meeting. The contribution was prepared in response to previous informal requests for such a white paper, especially by the convenors of ISO/IEC JTC 1/SC 29/AG 2 (MPEG Technical Coordination) and ISO/IEC JTC 1/SC 29/AG 3 (MPEG Liaison and Communication).

To be converted into an output (with editing period).

## Text development and errata reporting (1)

Contributions in this area were discussed in session 21 at 2100–2115 UTC on Wednesday 14 July 2021 (chaired by JRO).

[JVET-W0187](https://jvet-experts.org/doc_end_user/current_document.php?id=11018) AVC and HEVC draft text in response to JVET-W0080 and JVET-W0083 [G. J. Sullivan (Microsoft), Y.-K. Wang (Bytedance)]

This contribution proposes draft text changes for the AVC and HEVC standards that were prepared in response to JVET-W0080 and JVET-W0083, relative to the text circulated by ITU-T for AAP Last Call.

JVET contributions JVET-W0080 and JVET-W0083 have reported some problems in the text drafted for the VSEI standard (Rec. ITU-T H.274 | ISO/IEC 23002-7).

Most of the reported problems are also evident in the text of AVC (Rec. ITU-T H.264 | ISO/IEC 14496-10) and HEVC (Rec. ITU-T H.265 | ISO/IEC 23008-2) as well as in VSEI. In the ITU-T, these other standards are currently in a “Last Call” stage of the ITU-T Alternative Approval Process for edition 14 of H.264 and edition 8 of H.265, and the Last Call period is closing on the day of submission of this document.

The attached drafts contain suggested edits to address the identified issues, and have been submitted as Microsoft input to the Last Call.

We suggest that these proposed changes are clearly merely editorial in spirit as they are only made to provide obviously missing material, correct small errors, and improve clarity and consistency.

The changes are summarized as follows. Further commentary is available in JVET-W0080 and JVET-W0083.

• Correcting an “off by one” error in the loop counter for ar\_num\_object\_updates, which indicates the number of updated annotated objects for the annotated regions SEI message.

• Adding a description of the purpose of some variables that are derived in the semantics of the annotated regions SEI message

• Adding a missing descriptor to specify the type of coding for the depth\_nonlinear\_representation\_model[ i ] syntax element (the same as used in another specification, so clearly what was intended) for the depth representation information SEI message

• Adding missing range specifications for a few syntax elements of the depth representation information SEI message

• Miscellaneous very small cleanups of consistency, grammar and clarity, and correction of typographical errors.

The edits are marked using edit tracking with a special username identifier. For H.264v14, the edit identifier is “V14+”, and for H.265v8, the edit identifier is “V8+”.

It is reported that, beyond text that would resolve the issues pointed out in W0080 and W0083, some other editorial improvements are included that were found during preparation of the texts in response to the ITU Call for Comments (had already been submitted by the time this was presented in JVET).

Decision(BF/ED): No concerns were expressed by the time of presentation. It was agreed that, unless concerns would still be raised before the end of the meeting, the suggested modifications are agreed by JVET.

## Test conditions (1)

Contributions in this area were discussed in session 13 at 0900–0920 UTC on Monday 12 July 2021 (chaired by JRO).

[JVET-W0152](https://jvet-experts.org/doc_end_user/current_document.php?id=10982) AHG3: Overview of CTC documents and possible merging [K. Sühring, F. Bossen, X. Li, V. Seregin, K. Sharman, A. Tourapis]

The following CTC documents exist for VVC and HEVC:

|  |  |  |  |
| --- | --- | --- | --- |
| Area of interest | HEVC | VVC | Alignment |
| SDR video | JVET-U1100 | JVET-T2010 | Yes (except for RA GOP, additional 8-bit profile in HEVC) |
| High bit depth / rate | JVET-U1100 (RExt-HBD) | JVET-U2018 | No |
| Lossless, near-lossless, mixed | JVET-U1100 (RExt-HBD lossless) | JVET-Q2014 | No |
| Non-4:2:0 colour formats | JVET-U1100 (RExt) | JVET-T2013 | No |
| HDR/WCG video | JCTVC-Z1020 | JVET-V2011 | No, but some similarities |
| Screen content | JCTVC-Z1015 | JVET-T2013 | No, but some similarities |
| 360° video |  | JVET-U2012 | - |
| Scalable video coding | JCTVC-X1009 |  | - |
| Reference picture resampling | (not supported) | JVET-Q2015 | - |
|  |  |  |  |
| Neural network | - | JVET-Q2016 (new JVET work topic) | - |
| Enhanced compression tool testing | - | JVET-Q2017 (new JVET work topic) | - |

For alignment of experiments and reduction of number of documents to maintain, it is suggested to keep only one CTC document per area of interest. Settings specific to each standard should be described in different sections of that document.

If CTC documents exist for VVC and HEVC in the same area, and the conditions are not aligned, it is suggested to keep the VVC related coding conditions and to produce comparable settings for HEVC, so that comparing these settings is possible.

If a CTC document exists in an area only for one of the standards, it is suggested to keep this document, unless there is a need to create a comparison between the standards, or exploration in the context of a newer standard. If it becomes necessary, the CTC should be updated according to the rules above.

|  |  |  |  |
| --- | --- | --- | --- |
| Area of interest | HEVC | VVC | Suggested action |
| SDR video | JVET-U1100 | JVET-T2010 | Use VVC CTC, create HEVC specific section |
| High bit depth / rate | JVET-U1100 (RExt-HBD) | JVET-U2018 | Use VVC CTC, create HEVC specific section |
| Lossless, near-lossless, mixed | JVET-U1100 (RExt-HBD lossless) | JVET-Q2014 | Use VVC CTC, create HEVC specific section |
| Non-4:20 colour formats | JVET-U1100 (RExt) | JVET-T2013 | Use VVC CTC, create HEVC specific section |
| HDR/WCG video | JCTVC-Z1020 | JVET-V2011 | Use VVC CTC, create HEVC specific section |
| Screen content | JCTVC-Z1015 | JVET-T2013 | Use VVC CTC, create HEVC specific section |
| 360° video |  | JVET-U2012 | Keep |
| Scalable video coding | JCTVC-X1009 |  | Update for VVC when needed |
| Reference picture resampling | (not supported) | JVET-Q2015 | Keep |

One expert points out that there is some problem when computing the HDR-specific metrics internally in VTM, or externally using HDRtools based on decoded vs. original yuv.

Another expert explains that there was some issue in VTM11, which should be resolved now (only minor difference in third decimal point).

For HM (also in CE/EE) it is run externally.

It was agreed that there is no urgency of doing this, the AHG should plan to propose integrated version(s) of CTC document(s) as input to the next meeting.

## Verification testing (4)

Contributions in this area were discussed in session 14 at 2110–2150 UTC on Monday 12 July 2021 (chaired by JRO).

[JVET-W0041](https://jvet-experts.org/doc_end_user/current_document.php?id=10856) AHG4: Status Report on HDR Video Verification Test Preparation [[A. Segall](mailto:asegall@sharplabs.com), M. Wien, V. Baroncini, K. Andersson]

Had been included in AHG report – no need for presentation

[JVET-W0042](https://jvet-experts.org/doc_end_user/current_document.php?id=10857) AHG4: Agenda and report of the AHG meeting on the HDR video verification test preparation on 2021-06-10 [A. Segall, M. Wien, V. Baroncini]

Had been included in AHG report – no need for presentation

[JVET-W0145](https://jvet-experts.org/doc_end_user/current_document.php?id=10974) Report on dry-run results for VVC compression performance verification testing in the HDR PQ and HLG categories [M. Wien, A. Segall, V. Baroncini, G. Baroncini]

This document reports the dry-run verification test results comparing VVC to its predecessor, HEVC, according to the verification test plan JVET-V2021 that was issued at the previous meeting. The compression performance capabilities of the two specifications are compared based on the HEVC reference software HM-16.23 and the VVC reference software VTM-12.0. A visual assessment of VVC compared to HEVC was performed, and the result of a formal subjective assessment evaluation (with the participation of naïve test subjects) is reported. The assessment includes testing with five sequences with HDR PQ content and five sequences with HDR HLG content. MOS-over-bitrate plots indicate the selected rate points are suitable for the purpose of verification testing of VVC compared to HEVC. The curves show good coverage of the MOS range from slightly perceptible impairments down to annoying impairments for all sequences. The goal of approximately matching MOS values for the HM and VTM rate points has been achieved within the confidence intervals for the large majority of the test sequences and rate points. The Bjøntegaard delta rate figures show an overall gain of about 49% for the HLG test sequences and an overall gain of about 52% the PQ test sequences in this dry-run test. The curves for both the HLG and the PQ test sequences show a substantial and consistent benefit of VTM-12.0 over HM-16.23 in terms of MOS over bitrate. Thereby, the dry run confirms the suitability of the rate points proposed in the verification test plan document.

It is concluded that the dry run was conducted under conditions that were fully matching a formal subjective test, and the results are fully conclusive, such that no further test is necessary. It was decided that the final report on HDR test cases shall be issued from these results.

It was asked if the results would be reproducible. It was answered that the config files were included in JVET-W0041. These should also be attached to the report.

[JVET-W0150](https://jvet-experts.org/doc_end_user/current_document.php?id=10980) Information on and analysis of the VVC encoders in the SDR HD RA verification test [C. Helmrich, B. Bross, J. Pfaff, H. Schwarz, D. Marpe, T. Wiegand (HHI)] [late]

This contribution provides further information on the configuration of the VVC reference encoder (VTM) and the alternative encoder (VVenC) evaluated in the VVC verification test for SDR high definition content in random access (RA) configuration, specified in JVET-V2021 (test plan, draft 6).The contribution follows contribution JVET-T0103 (UHD).Based on an analysis of the bitstreams and Bjøntegaard delta-rate values obtained using numerous video quality assessment (VQA) methods, the following conclusion can be drawn:

The results of the SDR HD RA verification test, reported in JVET-V2020, indicate that VVenC 0.3 matches or, in some cases, tends to outperform the VTM 11.0 encoder in terms of visual quality.It was observed,similarly to JVET-T0103,that only some of the VQA methods used for BD-rate eva­luation, namely, the SSIM and MS-SSIM methods, allow to predict this tendency with satisfactory accuracy.The PSNR and VMAF metrics, in particular, fail to model the subjective judgments well.

Actual BD rate savings based on MOS are not reflected by BD rate based on SSIM as well, but at least the MOS tendency between the two codecs is estimated more correct than with PSNR, XPSNR, and VMAF.

[JVET-W0185](https://jvet-experts.org/doc_end_user/current_document.php?id=11016) Draft SC29/AG5 Guidelines for remote experts viewing sessions [J. Jung (Tencent), M. Wien (RWTH Aachen University), V. Baroncini (Vabtech)]

(for joint meeting Wednesday 14 July 0720)

## Test material (0)

Contributions in this area were discussed in session X at XXXX–XXXX UTC on XXday X July 2021 (chaired by XXX).

## Quality assessment (0)

Contributions in this area were discussed in session X at XXXX–XXXX UTC on XXday X July 2021 (chaired by XXX).

## Conformance test development (0)

See the AHG5 report in section 3.

## Software development (1)

Contributions in this area were discussed in session 18 at 2130–2145 UTC on Tuesday 13 July 2021 (chaired by JRO).

[JVET-W0134](https://jvet-experts.org/doc_end_user/current_document.php?id=10963) AHG3: Uniform logs output [F. Galpin (InterDigital)] [late]

This contribution proposes to modify the output format of the log of VTM software in order to make it easier to process.

VTM software output logs are used to extract important metrics used to assess the codec performance.

Historically, the log format output is the same since the HM software and allows a “line” parsing of the logs to extract the information. However, in VTM, several new modes are available and output information in a way not completely compatible with the current format:

* RPR mode where 2 metrics are output: one for the coded resolution and one for the original resolution
* Scalable mode: a set of metrics per layers
* New metrics from HDR, w-PSNR, 360 metrics etc.

When summarizing the results, an automatic process is usually used to extract relevant information from the log. Moreover, in chunk mode, the whole log should be processed in order to create a concatenated log. We propose to correct some aspects of the current log to simplify the logs parsing by automatic scripts.

It is noted that the purpose of this contribution is for rising awareness about the proposed change, as it may have impact on script used for simulations in experiments.

Decision(SW/cleanup): Adopt JVET-W0134.

## Implementation studies (2)

Contributions in this area were discussed in session 18 at 2100–2130 UTC on Tuesday 13 July 2021 (chaired by JRO).

[JVET-W0127](https://jvet-experts.org/doc_end_user/current_document.php?id=10944) Ali266: an optimized VVC software encoder implementation [S. Fang, J. Guo, Z. Huang, J. Liu, S. Xu, L. Yu, J. Chen, R.-L. Liao, Y. Ye (Alibaba)] [late]

This contribution reports the performance of Ali266, an optimized software VVC encoder developed by Alibaba. Two sets of content are used to test the Ali266 encoding performance and speed: the JVET CTC content, including class A1, class A2, class B and class C; and internal video-on-demand (VOD) content, including classes 1920x1080, 1080x1920, 1280x720, 720x1280. For the JVET CTC content, Ali266 slow preset is compared to VTM-13.0, and achieves BD rates for {Y, U, V, YUV} of {10.95%, -6.11%, -4.50%, 7.11%} under the Random Access configuration with an average speedup factor of 227 times over VTM. For the internal VOD content, Ali266 fast preset is compared to the open source HEVC software encoder x265 medium preset, and achieves BD rates for {Y, U, V, YUV} of {-32.87%, -46.07%, -42.70%, -35.35%} under the Random Access configuration. Ali266 fast preset is aimed at real-time VVC encoding, and currently achieves average encoding speeds of 19.3 fps for classes 1920x1080/1080x1920 and 37.7 fps for classes 1280x720/720x1080. Based on such data, it is our conclusion that Ali266 is a VVC encoder capable of real-time encoding 720p sequences, and it is suitable for deployment in the real world in the near future.

Implementation with 10 bit.

Fixed QP, rate control disabled.

Multicore implementation, 12 threads, hyperthreading disabled.

[JVET-W0184](https://jvet-experts.org/doc_end_user/current_document.php?id=11015) Update on open, optimized VVC implementations VVenC and VVdeC [A. Wieckowski, J. Brandenburg, C. Bartnik, V. George, J. Güther, G. Hege, C. Helmrich, A. Henkel, T. Hinz, C. Lehmann, C. Stoffers, I. Zupancic, B. Bross, H. Schwarz, D. Marpe, T. Schierl (HHI)] [late]

This document provides updated information on features, coding efficiency and runtime for version 1.0.0 of the open VVC software encoder VVenC released in May 2021 and version 1.1.2 of the open VVC software decoder VVdeC released in July 2021.

Main changes for VVenC v1.0.0 since version 0.2.1 include:

* **Presets**: Increased coding efficiency and speed for existing presets (faster, fast, medium, slow, slower).
* **New coding tools**: added support and parameters for high dynamic range (HDR) and wide color gamut (WCG) video, Intra Block Copy (IBC) for improved coding of screen content coding (SCC), constrained reference picture resampling (RPR) encoding mode to enable low-drift adaptive resolution switching with open GOPs.
* **Multithreading**: Added frame-level parallelism to improve multithreading with increasing number of threads.
* **Rate Control (RC)**: improved 1- and 2-pass RC, especially in combination with subjective QP adaptation.
* **Application Programming Interface (API)**: change libvvenc interface from C++ to C.

Without QP adaptation for subjective optimization and 8 threads the following PSNR-based YUV BD-rates and speedup factors compared to HM and VTM-12.0 (GOP32+MCTF) are reported for different presets:

Faster HD: −7.9%, 119x (HM), 915x (VTM) UHD: −13.0%, 142x (HM), 1042x (VTM)

Fast HD: −25.9%, 68x (HM), 519x (VTM) UHD: −28.2%, 81x (HM), 591x (VTM)

Medium HD: −36.4%, 21x (HM), 162x (VTM) UHD: −38.3%, 29x (HM), 211x (VTM)

Slow HD: −39.8%, 7.6x (HM), 58x (VTM) UHD: −41.7%, 11x (HM), 77x (VTM)

Slower HD: −43.2%, 1.6x (HM), 12x (VTM) UHD: −45.1%, 2.2x (HM), 16x (VTM)

With QP adaptation for subjective optimization and 8 threads, the following MS-SSIM-based YUV BD-rates and speedup factors compared to HM and VTM-12.0 (GOP32+MCTF) are reported:

Faster HD: −17.3%, 111x (HM), 885x (VTM) UHD: −12.4% 134x (HM), 1097x (VTM)

Fast HD: −31.2%, 65x (HM), 511x (VTM) UHD: −29.6% 78x (HM), 637x (VTM)

Medium HD: −39.9%, 20x (HM), 162x (VTM) UHD: −40.3%, 28x (HM), 225x (VTM)

Slow HD: −43.3%, 7.5x (HM), 60x (VTM) UHD: −43.8%, 10x (HM), 82x (VTM)

Slower HD: −45.7%, 1.6x (HM), 13x (VTM) UHD: −46.9%, 2.2x (HM), 18x (VTM)

Main changes for VVdeC v1.1.2 since version 0.2.0.1 include:

* **Subpictures** support including nested SEIs.
* **Many fixes and improvements** including code cleanup, fixed memory leaks, memory optimizations, stability improvements and conformance fixes.
* **Application Programming Interface (API)**: change libvvdec interface from C++ to C.

Multicore configured for 8 threads.Medium and fast presets became still faster.

Improved two-pass rate control achieves same performance as fixed QP.

## Complexity analysis (0)

No contributions in this area were noted.

## AHG10: Encoding algorithm optimization (4)

Contributions in this area were discussed in session 18 at 2145–2300 UTC on Tuesday 13 July 2021 (chaired by JRO).

[JVET-W0043](https://jvet-experts.org/doc_end_user/current_document.php?id=10858) AHG10: Alignment of smooth QP control with adaptive QP in VTM [K. Andersson, J. Enhorn, J. Ström (Ericsson)]

The smooth QP control method in VTM do not work properly with adaptive QP because the correct lambda is not used. This contribution suggests fixing the lambda but also change the default settings for smooth QP control to only be applied on intra pictures since that gives most of the visual effect. It could be considered to also add some separate config parameters for inter pictures for additional flexibility.

The BD-rate performance (Y/Cb/Cr) compared to CTC with VTM-13.0 as anchor is as follows:

RA SDR (luma, cb, cr): 0.05%, -0.01%, 0.00%

RA HDR (DE100 PSNR-L100, wPSNR, Y-PSNR): -0.78%, -0.14%, 0.07%, 0.05%

With PerceptQPA enabled in both anchor and proposal:

RA SDR (luma, cb, cr): 0.00%, 0.01%, -0.02%

RA HDR (DE100 PSNR-L100, wPSNR, Y-PSNR): -0.05%, -0.00%, 0.00%, 0.00%

It is asserted that the method improves visual quality over the percept QPA method for some HDR content.

Crosscheckers report that subjective benefit was observed in particular at low QP ranges.

Same method of smooth region QP is in HM, but not the adaptive QP method. Would in principle be possible to implement. However currentlc no change tom HM necessary as the ad. QP is not included there.

Decision(SW): Adopt JVET-W0043.

[JVET-W0147](https://jvet-experts.org/doc_end_user/current_document.php?id=10977) Crosscheck of JVET-W0043: AHG 10: Alignment of smooth QP control with adaptive QP in VTM [I. Zupancic, C. Helmrich (HHI)] [late]

[JVET-W0061](https://jvet-experts.org/doc_end_user/current_document.php?id=10876) AHG10: Encoder MV selections and DMVR [R. Yu, K. Andersson, J. Ström (Ericsson)]

VVC includes a tool called decoder side motion vector refinement (DMVR) to refine motion vectors for inter blocks that are coded using merge mode. The refinement is performed on a subblock basis. It is asserted that the refined MVs for the subblocks may sometimes have poor correlation since the refinement is carried out for each subblock independently. It is further asserted that the poorly correlated subblock MVs are likely to introduce noticeable subblock boundary artifacts, especially for low spatial activity regions.

The contribution proposes an encoder-only method to reduce the risk of subblock boundary artifacts being introduced by DMVR. The method restricts the encoder from choosing a pair of DMVR-applicable bi-MVs when a current block together with the corresponding DMVR-refined MVs are determined to be prone to give the subblock boundary artifacts.

The method has been implemented in the VTM-13.0. The BD-rate PSNR impact is reported to be 0.19% for SDR RA CTC and 0.20% for HDR RA CTC. It is asserted that the proposed method can improve subjective quality when refinement is based on low spatial activity regions, especially when the source block to be predicted happens to contain an edge and the residual coding is too coarse to take care of the artifact. It is proposed to include the method in VTM to enable improved subjective robustness of DMVR.

Question: Bitrate increase is relatively large, though the change is mostly relevant only for QP 37. Would it not be better disabling DMVR completely. The proponent answers that the gain would be clearly higher.

BD rate losses are likely due to discarding some merge candidates from the selection.

Were the artifacts visible in video? According to proponents, yes.

Crosschecker also reports that subjective quality was improved (only in BQ terrace)

It was commented by crosschecker that BD rate loss appears high.

Would it not even be possible to improve th BD results? Yes, but might be complicated.

Considering that the problem is occurring in specific sequences, and considering that the loss in terms of objective quality is comparably high, further study is recommended. It is howeer highly welcome that this problem was identified.

[JVET-W0167](https://jvet-experts.org/doc_end_user/current_document.php?id=10998) Crosscheck of JVET-W0061 (AHG10: Encoder MV selections and DMVR) [H. Chen, H. Yang (Huawei)] [late]

[JVET-W0082](https://jvet-experts.org/doc_end_user/current_document.php?id=10898) AHG10: GOP-based RPR encoder control [K. Andersson, J. Ström, R. Yu (Ericsson)]

VTM currently lacks encoder control for selecting when to use Reference Picture Resampling (RPR). This contribution proposes a GOP-based selection method that decides when to encode pictures in reduced resolution with RPR or if its better to encode them in the source resolution. The selection is based on QP and picture self-similarity after re-scaling. The self-similarity test is only conducted on the first source picture in display order within each GOP. If the re-scaled picture is determined to have sufficient similarity with the source picture, all pictures in the GOP are encoded at reduced resolution, otherwise they are encoded in the source resolution.

The BD-rate performance (Y/Cb/Cr) compared to CTC with VTM-13.0 as anchor is as follows:

RA (luma/Cb/Cr): -0.03%/0.61%/0.58%

LDB (luma/Cb/Cr): 0.27%/0.35%/0.35%

AI (luma/Cb/Cr): 0.11% /0.63%/0.67%

The BD-rate performance for QP 27 to 47 with VTM-13.0 as anchor:

RA (luma/Cb/Cr): -1.57%/6.94%/5.94%

LDB (luma/Cb/Cr): 1.42%/4.06%/3.41%

AI (luma/Cb/Cr): -0.41%/6.39%/4.76%

It is asserted that the method, when it is selected, in some cases can improve subjective quality compared to encoding in source resolution for low bitrates.

Method only applied for QP37 and beyond

It is pointed out that this might well be combined with segment-wise encoding and open GOP.

It is also pointed out that it mainly seems to be efficient for class A.

It is also mentioned that the effectiveness highly depends on sequence characteristics. For some sequences, gain is achieved only at QP50. In this conext, it is recommended to seek for individual “transition points”.

Has it been investigated why there is loss in chroma? See discussion on this elsewhere.

Very interesting approach to save bit rate by using coding at lower resolution whenever appropriate. Further study is recommended for further improvements as per the hints above.

It is suggested to also look into subjective quality.

[JVET-W0173](https://jvet-experts.org/doc_end_user/current_document.php?id=11004) Crosscheck of JVET-W0082: AHG10: GOP-based RPR encoder control [T. Chujoh (Sharp)] [late]

[JVET-W0129](https://jvet-experts.org/doc_end_user/current_document.php?id=10946) AHG10: Using original samples for ALF optimization [N. Hu, D. Rusanovskyy, V. Seregin, M. Karczewicz (Qualcomm)]

In VTM-13.0, motion compensated temporal pre-filtering (MCTF) is used for random access (RA) configuration and samples filtered by MCTF are used for sample adaptive offset (SAO) and adaptive loop filter (ALF) optimizations in common test condition (CTC). It is configurable to use original samples for SAO and ALF optimization as JVET-V0095. However, in HM and ECM, filtered samples are used for SAO and bilateral filter optimization and original ones are used for ALF in CTC. To align different reference softwares, this contribution proposes to use original samples only for ALF optimization and enable it in CTC for VTM. The proposed method was tested under SDR, HDR, and high bit depth/high bit rate test conditions, the results of the informal subjective evaluation are reported as well.

In the last meeting, a method was proposed using unfiltered samples for both SAO and ALF. It is now suggested to do that only for ALF, and use samples after MCTF for SAO. The difference in terms of compression is very small.

It is suggested to use identical data for deriving in both cases. It is also pointed out that this requires an additional picture buffer.

However, since the method iof using unfiltered samples for ALF and flitered samples for SAO is also used in ECM, and it is a pure encoder choice, it makes sense to do the same in VTM. The compression benefit on whether doing SAO with filtered or unfiltered samples is marginal.

Decision (SW/CTC): adopt JVET-W0129

## Profile/tier/level specification (2)

Contributions in this area were discussed in session 21 at 2115–2155 UTC on Wednesday 14 July 2021 (chaired by JRO).

[JVET-W0136](https://jvet-experts.org/doc_end_user/current_document.php?id=10965) Suggested initial profile text for VVC operation range extension [T. Ikai, T. Chujoh, T. Aono, S. Deshpande (Sharp)] [late]

Was presented in joint meeting Tuesday 13 July, 0730-0845

A new version of this doc wss presented in session 21, which is updated with the set of profiles decided in the joint session. It furthermore suggests allocations of profile idc values and constraint flags, which follow the general principles used traditionally.

Furthermore, a general\_lower\_bit\_rate\_constraint\_flag is proposed which allows specifying two different ranges of bit rate in the high tier. This can be set to zero or one in still picture and Intra profiles (except main 12). It is noted that a similar mechanism is used in HEVC intra profiles to accommodate the case that the bit rate in intra could get significantly higher.

It is pointed out by one expert that the document defines activation/deactivation of palette mode and other tools in main12 via GCI flags. This is different from the approach used in V1, where this is typically defined by prescribing the value of the corresponding SPS flag.

Decision: Adopt the profile specification of JVET-W0136v5 (including the modification of the tool enabling/disabling via SPS flags rather than GCI, and aligns the level definition with that of HEVC).

Was further discussed in the plenary on Friday 16 July at 0510 UTC.

Due to the fact that VVC has a larger number range of profile idc values, the numbering scheme from HEVC appears not desirable. Furthermore, GCI is optional, therefore it does not fulfill exactly the same purpose as constraint flags in HEVC.

Also in version 1, individual profile idc values had been used.

It is left to the discretion of the editors to select appropriate profile idc values for the new profiles, and also to take care that the nesting relationship is appropriately reflected. The same GCI flag for bit rate constraint will be defined for the different profiles but not required to have particular values associated with profile idc. It is further agreed that the new flag would not be conditioned on profile idc.

It is further pointed out that currently no GCI flags have been defined for the other new tools. This is for further study.

[JVET-W0183](https://jvet-experts.org/doc_end_user/current_document.php?id=11014) Maximum picture rates for level 6.3 [F. Bossen] [late]

Section A.4.3 of the VVC specification is informative and provides examples of the relationship between level, picture size, and maximum picture rate. Values for level 6.3 are provided in this document to support the disposition of a related ballot comment.

Decision(ED): Adopt JVET-W0183. Some more editorial alignment could be desirable to bring all 6.x levels together.

## Proposed modification of system interface (0)

Contributions in this area were discussed in session X at XXXX–XXXX UTC on XXday X July 2021 (chaired by XXX).

# Low-level tool technology proposals (65)

## AHG8: High bit rate and high bit depth coding for VVC (20)

### General (0)

Contributions in this area were discussed in session X at XXXX–XXXX UTC on XXday X July 2021 (chaired by XXX).

### CE contributions: Entropy Coding for High Bit Depth and High Bit Rate Coding (4)

Contributions in this area were discussed in session 2 at 0750–0920 UTC on Wednesday 7 July 2021 (chaired by JRO and GJS).

[JVET-W0022](https://jvet-experts.org/doc_end_user/current_document.php?id=10943) CE: Summary Report on Entropy Coding for High Bit Depth and High Bit Rate Coding [D. Rusanovskyy, K. Naser, M. G. Sarwer, F. Wang]

This contribution provides a summary report for the Core Experiment on entropy coding for high bit depth and high bit rate proposed at V meeting of JVET. A total of 5 tests have been completed within the CE between the JVET V and W meetings to study and evaluate technologies related to entropy coding for High Bitdepth (HBD) and High Bitrate (HBR) coding. The scope of this experiment covered tests on changes to coding of the last significant coeffitient (1 test), tests on content adaptive transform precision (2 tests) and tests on CABAC bypass coding methods (2 tests). The software basis for this CE is VTM-13.0. For the test sequences, configurations and test conditions, the High Bit-depth CTC described in JVET-U2018 is used, unless otherwise specified in the CE description. Test results against anchors are provided to show the coding efficiency and complexity trade-off of each tool. Where appropriate results are also provided for lossless conditions. For tests CE-3.1 and CE-3.2 targeting high throughput configuration, comparative results against the HM in the corresponding configuration. Cross-check results for the performed tests are submitted as separate documents and their summaries are integrated in this contribution.

1 Test conditions and evaluation criteria

In LowQP configuration, extended precision processing is enabled (parameter ExtendedPrecision is set equal to 1), and all results are compared with an anchor with the same setting.

For Normal QP configuration (12 bits 4:2:0 PQ/HLG test sequences), extended precision is disabled (parameter ExtendedPrecision is set equal to 0) and the results are be compared with an anchor with extended precision disabled.

Adaptive transform precision tests (i.e. CE-2) are also report results for extended precision processing being disabled for 12 bits data (HDR PQ/HLG and SVT12) in LowQP configuration for both, the anchor and the test.

For tests targeting high throughput profile ( i.e., CE-3), HM configured in HEVC high throughput profile as an additional reference point to be used. The latest version of HM (HM-16.23) will be used for high throughout profile. In HM-16.23, the High Throughput 4:4:4 16 Intra profile is enabled using encoder\_intra\_high\_throughput\_rext.cfg.

2 Tests on coding of the last significant coeffitient

In this category, one test on proposal JVET-V0121 has been conducted. Proposed method features encoding method to determine at encoder side the value of sh\_reverse\_last\_sig\_coeff\_flag for each slice that determines the derivation of LastSignificantCoeffX and LastSignificantCoeffY variables. When sh\_reverse\_last\_sig\_coeff\_flag is equal to 1, the last coefficient position is coded by using the proposed alternative coordinates for a whole slice. Otherwise, the current coordinates for the last coefficient position are used.

Simulation results for the tests in this test category, have been provided for HBD/HBR/HFR CTC in the lowQP, normal QP range and lossless configurations and reported in corresponding JVET input documents. Summary of the results form reporting template is provided below.

Table 2.2. Simulation results for CE1.1 test, 12 bits data, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
|  | wY | wU | wV | Y | U | V | Aver.GBR | G | B | R |
| **AI** | **CE1.1** | -0.02% | -0.04% | -0.05% | -0.22% | -0.26% | -0.31% | -0.30% | -0.33% | -0.29% | -0.29% |
| **LDB** | **CE1.1** | -0.01% | 0.02% | 0.01% | -0.18% | -0.18% | -0.28% | -0.27% | -0.29% | -0.26% | -0.26% |
| **RA** | **CE1.1** | 0.01% | 0.00% | 0.02% | -0.19% | -0.19% | -0.28% | -0.24% | -0.26% | -0.23% | -0.23% |

Table 2.3. Simulation results for CE1.1 test, 16 bits data, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **SVT16 RGB** | | | |
|  | Aver.GBR | G | B | R |
| **AI** | **CE1.1** | -0.28% | -0.29% | -0.28% | -0.28% |
| **LDB** | **CE1.1** | -0.27% | -0.28% | -0.26% | -0.26% |
| **RA** | **CE1.1** | -0.26% | -0.27% | -0.25% | -0.25% |

Table 2.4. Reported run-time estimates for CE1.1 tests, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | **HDR HLG** | | **SVT12 RGB** | | **SVT16 RGB** | |
|  | Enc | Dec | Enc | Dec | Enc | Dec | Enc | Dec |
| **AI** | **CE1.1** | 101% | 99% | 100% | 96% | 100% | 97% | 98% | 95% |
| **LDB** | **CE1.1** | 100% | 100% | 101% | 100% | 103% | 99% | 102% | 99% |
| **RA** | **CE1.1** | 99% | 100% | 101% | 98% | 102% | 99% | 103% | 99% |

Table 2.5. Simulation results for CE1.1 tests, HBD/HBR CTC, NormQP test configuration.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | | | **HDR HLG** | | |
|  | DE100 | PSNR-L100 | wY | wU | wV | Y | U | V |
| **AI** | **CE1.1** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **RA** | **CE1.1** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |  |  |  |

Table 2.6. Run-time estimates for CE1.1 tests, HBD/HBR CTC, NormQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | **HDR HLG** | |
|  | Enc | Dec | Enc | Dec |
| **AI** | **CE1.1** | 99% | 99% | 102% | 100% |
| **RA** | **CE1.1** | 100% | 101% | 100% | 100% |

Table 2.7. Simulation results for CE1.1 tests, HBD/HBR CTC, lossless test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | bit-rate saving | | | |
|  |  | HDR PQ | HDR HLG | SVT12 | SVT16 |
| **AI** | **CE1.1** | -0.31% | -0.30% | -0.25% | -0.15% |
| **LDB** | **CE1.1** | -0.42% | -0.45% | -0.26% | -0.20% |
| **RA** | **CE1.1** | -0.41% | -0.44% | -0.26% | -0.19% |

It is asserted that the proposal is straightforward, simple to implement, and provides reasonable tradeoff in terms of improved compression. Cross-checkers reported that the specification text is appropriate.

Decision: Adopt JVET-W0046 (for VVC version 2 draft)

3 Tests on Content Adaptive Transform Precision

In this category, one test on proposal JVET-V0133 has been conducted. This contribution proposes to adaptively assign the downshifts in the scaling process for transform coefficients (clause 8.7.3 of VVC) and in the transformation process for scaled transform coefficients (clause 8.7.4 of VVC) by examining the coefficients values of upon dequantization and before inverse transform.

For tests in this category, additional results are provided for 12 bits data (HDR PQ/HLG and SVT12) in LowQP configuration with extended precission flag being disabled for both, the test and the VTM anchor.

Additional results will be provided for 16 bit data (SVT16) being coded in LowQP configuration with extended precission flag being enabled, and Log2TransformRange in the test being set equal to 16. Results will be compared against the VTM anchor coded in LowQP configuration with extended precission flag being enabled, and also against the same anchor with Log2TransformRange being set equal to 16.

Test Description document Cross-checker document

CE2.1 JVET-W0050 JVET-W0056

CE2.2 JVET-W0050 JVET-W0094

In test CE-2.1, the maximum dequantized coefficient value before applying the downshift is considered for computing the adaptive dequantization downshift. The susbsequent inverse transform shifts (horizontal and vertical) are modified accordingly.

In test CE-2.2, in addition to CE2.1, the maximum coefficient value of the horizontal transform is determined for computing the inverse horizontal transform shift. The subsequent vertical transform shift is modified accordingly.

***Common Test Condition, Anchor is VTM13.0, Extended Precision is ON for Test and Anchor.***

Table 3.2. Simulation results for CE2.x tests, 12 bits data, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
|  | wY | wU | wV | Y | U | V | Aver.GBR | G | B | R |
| **AI** | **CE2.1** | -0.01% | -0.01% | -0.01% | 0.00% | -0.01% | -0.01% | 0.00% | 0.00% | -0.01% | 0.00% |
| **CE2.2** | -0.01% | -0.02% | -0.02% | -0.01% | -0.01% | -0.01% | -0.01% | 0.00% | -0.01% | -0.01% |
| **LDB** | **CE2.1** | -0.01% | 0.00% | 0.02% | 0.01% | 0.00% | 0.00% | 0.00% | -0.01% | 0.00% | 0.00% |
| **CE2.2** | -0.01% | -0.01% | 0.01% | 0.00% | -0.02% | -0.01% | 0.00% | -0.01% | 0.00% | 0.00% |
| **RA** | **CE2.1** | 0.00% | 0.00% | 0.00% | 0.00% | -0.01% | 0.01% | 0.00% | -0.01% | 0.00% | 0.00% |
| **CE2.2** | -0.01% | 0.00% | 0.00% | -0.01% | -0.01% | -0.01% | 0.00% | -0.01% | 0.00% | 0.00% |

Table 3.3. Simulation results for CE2.x tests, 16 bits data, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **SVT16 RGB** | | | |
|  | Aver.GBR | G | B | R |
| **AI** | **CE2.1** | -0.01% | -0.01% | -0.01% | -0.01% |
| **CE2.2** | -0.02% | -0.01% | -0.02% | -0.02% |
| **LDB** | **CE2.1** | -0.02% | 0.00% | -0.03% | -0.03% |
| **CE2.2** | -0.03% | -0.01% | -0.04% | -0.04% |
| **RA** | **CE2.1** | -0.01% | 0.00% | -0.02% | -0.02% |
| **CE2.2** | -0.02% | -0.01% | -0.03% | -0.03% |

Table 3.4. Reported run-time estimates for CE2.x tests, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | **HDR HLG** | | **SVT12 RGB** | | **SVT16 RGB** | |
|  | Enc | Dec | Enc | Dec | Enc | Dec | Enc | Dec |
| **AI** | **CE2.1** | 109% | 112% | 108% | 115% | 109% | 116% | 111% | 122% |
| **CE2.2** | 110% | 114% | 109% | 117% | 111% | 123% | 108% | 118% |
| **LDB** | **CE2.1** | 106% | 109% | 100% | 101% | 101% | 104% | 103% | 105% |
| **CE2.2** | 107% | 110% | 103% | 108% | 102% | 106% | 102% | 107% |
| **RA** | **CE2.1** | 103% | 106% | 104% | 109% | 104% | 111% | 104% | 114% |
| **CE2.2** | 107% | 110% | 105% | 107% | 104% | 113% | 104% | 110% |

***Additional tests, Anchor is VTM13.0, Extended Precision is OFF for test and Anchor.***

Table 3.5. Additional results for CE2.x tests, 12 bits data, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
|  | wY | wU | wV | Y | U | V | Aver.GBR | G | B | R |
| **AI** | **CE2.1** | -0.28% | -0.47% | -0.48% | -0.11% | -0.22% | -0.22% | -0.12% | -0.05% | -0.15% | -0.15% |
| **CE2.2** | -0.49% | -0.76% | -0.82% | -0.18% | -0.36% | -0.38% | -0.20% | -0.11% | -0.24% | -0.24% |
| **LDB** | **CE2.1** | -0.21% | -0.18% | -0.20% | -0.19% | -0.24% | -0.27% | -0.15% | -0.10% | -0.17% | -0.17% |
| **CE2.2** | -0.38% | -0.35% | -0.38% | -0.29% | -0.38% | -0.42% | -0.21% | -0.17% | -0.23% | -0.23% |
| **RA** | **CE2.1** | -0.12% | -0.13% | -0.14% | -0.12% | -0.18% | -0.18% | -0.14% | -0.05% | -0.18% | -0.18% |
| **CE2.2** | -0.31% | -0.33% | -0.34% | -0.20% | -0.31% | -0.33% | -0.19% | -0.11% | -0.23% | -0.24% |

***Additional tests, Anchor is VTM13.0, Log2Transform is set equal to 16 for Test and Anchor.***

Table 3.6. Simulation results for additional CE2.x tests, 16 bits data, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **SVT16 RGB** | | | |
|  | Aver.GBR | G | B | R |
| **AI** | **CE2.1** | -0.22% | -0.19% | -0.23% | -0.23% |
| **CE2.2** | -0.62% | -0.63% | -0.61% | -0.62% |
| **LDB** | **CE2.1** | -0.37% | -0.15% | -0.48% | -0.49% |
| **CE2.2** | -0.68% | -0.46% | -0.78% | -0.79% |
| **RA** | **CE2.1** | -0.32% | -0.11% | -0.43% | -0.43% |
| **CE2.2** | -0.60% | -0.39% | -0.70% | -0.71% |

Table 3.7. Reported run-time estimates for additional CE2.x tests, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | **HDR HLG** | | **SVT12 RGB** | | **SVT16 RGB** | |
|  | Enc | Dec | Enc | Dec | Enc | Dec | Enc | Dec |
| **AI** | **CE2.1** | 107% | 110% | 108% | 113% | 107% | 111% | 108% | 110% |
| **CE2.2** | 107% | 110% | 110% | 113% | 107% | 113% | 107% | 112% |
| **LDB** | **CE2.1** | 101% | 104% | 101% | 110% | 105% | 108% | 108% | 111% |
| **CE2.2** | 102% | 109% | 102% | 107% | 107% | 110% | 107% | 108% |
| **RA** | **CE2.1** | 108% | 112% | 106% | 111% | 107% | 106% | 107% | 107% |
| **CE2.2** | 108% | 113% | 106% | 109% | 107% | 106% | 107% | 110% |

***Additional tests: Log2Transform is set equal to 16 for the Test, VTM13.0 Anchor is not modified.***

Additional test results of CE2.1 and CE2.2 with Log2Transform set equal to 16, compared against the VTM13.0 anchor (not modified Log2Transform initialization) are reported in Table below.

Table 3.8. Additional results for CE2.x tests, 12 bits data, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
|  | wY | wU | wV | Y | U | V | Aver.GBR | G | B | R |
| **AI** | **CE2.1** | 0.93% | 1.22% | 1.33% | 0.35% | 0.55% | 0.57% | 0.35% | 0.19% | 0.43% | 0.44% |
| **CE2.2** | 0.72% | 0.93% | 1.00% | 0.27% | 0.41% | 0.41% | 0.27% | 0.13% | 0.34% | 0.34% |
| **LDB** | **CE2.1** | 0.46% | 0.45% | 0.50% | 0.43% | 0.56% | 0.59% | 0.33% | 0.42% | 0.28% | 0.28% |
| **CE2.2** | 0.30% | 0.29% | 0.32% | 0.32% | 0.42% | 0.44% | 0.26% | 0.35% | 0.21% | 0.22% |
| **RA** | **CE2.1** | 0.44% | 0.44% | 0.50% | 0.37% | 0.55% | 0.60% | 0.30% | 0.33% | 0.29% | 0.29% |
| **CE2.2** | 0.30% | 0.29% | 0.34% | 0.29% | 0.42% | 0.45% | 0.24% | 0.27% | 0.23% | 0.23% |

Table 3.9. Simulation results for additional CE2.x tests, 16 bits data, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **SVT16 RGB** | | | |
|  | Aver.GBR | G | B | R |
| **AI** | **CE2.1** | 1.79% | 1.63% | 1.86% | 1.87% |
| **CE2.2** | 1.38% | 1.19% | 1.47% | 1.48% |
| **LDB** | **CE2.1** | 1.68% | 2.11% | 1.45% | 1.48% |
| **CE2.2** | 1.37% | 1.79% | 1.15% | 1.17% |
| **RA** | **CE2.1** |  | n/a | n/a | n/a |
| **CE2.2** |  | n/a | n/a | n/a |

Table 3.10. Reported run-time estimates for additional CE2.x tests, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | **HDR HLG** | | **SVT12 RGB** | | **SVT16 RGB** | |
|  | Enc | Dec | Enc | Dec | Enc | Dec | Enc | Dec |
| **AI** | **CE2.1** | 109% | 120% | 103% | 123% | 113% | 128% | 113% | 119% |
| **CE2.2** | 109% | 119% | 104% | 124% | 113% | 131% | 112% | 121% |
| **LDB** | **CE2.1** | 106% | 109% | 102% | 117% | 106% | 114% | 101% | 125% |
| **CE2.2** | 107% | 112% | 103% | 114% | 107% | 115% | 101% | 123% |
| **RA** | **CE2.1** | 104% | 109% | 104% | 109% | 107% | 119% |  |  |
| **CE2.2** | 102% | 106% | 104% | 107% | 107% | 120% |  |  |

No results have been provided for normal QP configuration. Tests on of CE2.x are not applicabe for lossless coding configuration.

From JVET discussion:

In summary, the method does not show benefit when the extended precision flag is enabled. When the extended precision flag is disabled, there is some gain (however less than with the extended precision flag). The benefit would be that it requires less internal bit depth than the extended precision flag.

It is further verbally reported that this applies only to the low QP case. For normal QP, the extended precision flag is disabled.

In fact, it can be concluded that the proposal would be competing as a replacement of the extended precision flag (does not provide additional gain). Considering the loss compared to extended precision flag usage (and considering that this loss would rather occur in the operation range of professional application), the benefit of internal bit depth reduction would not justify its adoption.

It is however noted that a CE related contribution reports better results.

No action on the proposal investigated in the CE.

It is further pointed out that, even though the proposal may reduce the internal bit depth, it requires additional logic and may also reduce the throughput (latency between inverse quantization and inverse transform) and requires additional buffer.

4 Tests on CABAC bypass coding methods

In this category, one test based on proposal JVET-V0059 and another test on JVET-V0178 have been conducted.

4.1 Bypass alignment and bypass coding

In this category, methods targeting high throughput profile and presented in JVET-V0059/JVET-V0178 are being tested.

Test Description document Cross-checker document

CE3.1 JVET-W0044 JVET-W0047

CE3.2 JVET-W0045 JVET-W0058

In test CE3.1, option 2 presented in JVET-V0059 was tested. The CABAC bypass alignment will be done before bypass coding of each coefficient group. After reaching to the maximum number of context coded bins ( defined in existing VVC standard), the bypass alignment will be done only once for that transform block before CABAC engine switched to fully bypass coding of transform coefficients. If the remaining number of context coded bins < 4, sb\_coded\_flag (of both RRC and TSRC) will be context coded, Otherwise, sb\_coded\_flag (of both RRC and TSRC) will be bypasss coded.

In test CE3.2, the throughput benefits of the method proposed in JVET-V0178 was tested. When sps\_cabac\_bypass\_alignment\_enabled\_flag is equal to 1:

• the corresponding remBinsPass1 and RemCcbs are set to be 0;

• the coding of last\_sig\_coeff\_x\_prefix, last\_sig\_coeff\_y\_prefix, last\_sig\_coeff\_x\_suffix and last\_sig\_coeff\_y\_suffix remain the same (context coded);

• bypass alignment happens before the first bypass coded bin in both RRC and TSRC;

• all coefficient or residual levels are bypass coded with corresponding syntax elements for both RRC and TSRC;

• all sb\_coded\_flag are bypass coded as well.

In this test, the VTM throughput will be studied and tested in reference to HEVC. The measurements of bitrate, binrate, bin2bit ratio (weighted, unweighted, peak, average) and other relevant test results will be reported by using the templates in JVET-N0049 and JVET-V0150.

Table 4.2. Simulation results for CE3.x tests, 12 bits data, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
|  | wY | wU | wV | Y | U | V | Aver.GBR | G | B | R |
| **AI** | **CE3.1** | 1.13% | 1.18% | 1.14% | 0.87% | 0.93% | 0.94% | 0.56% | 0.48% | 0.59% | 0.60% |
| **CE3.2** | 10.33% | 12.57% | 12.70% | 6.34% | 8.09% | 8.59% | 1.63% | 1.51% | 1.70% | 1.68% |
| **LDB** | **CE3.1** | 1.49% | 1.60% | 1.62% | 1.43% | 1.47% | 1.51% | 0.75% | 0.72% | 0.76% | 0.76% |
| **CE3.2** | 17.40% | 20.27% | 20.36% | 16.89% | 18.27% | 18.82% | 3.19% | 3.00% | 3.29% | 3.27% |
| **RA** | **CE3.1** | 1.49% | 1.57% | 1.58% | 1.43% | 1.47% | 1.51% | 0.78% | 0.74% | 0.80% | 0.80% |
| **CE3.2** | 17.11% | 19.76% | 20.20% | 17.19% | 18.16% | 18.61% | 3.59% | 3.40% | 3.69% | 3.67% |

Table 4.3. Simulation results for CE3.x tests, 16 bits data, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **SVT16 RGB** | | | |
|  | Aver.GBR | G | B | R |
| **AI** | **CE3.1** | 0.31% | 0.28% | 0.33% | 0.33% |
| **CE3.2** | 0.07% | 0.06% | 0.09% | 0.07% |
| **LDB** | **CE3.1** | 0.47% | 0.46% | 0.47% | 0.47% |
| **CE3.2** | 0.34% | 0.32% | 0.36% | 0.35% |
| **RA** | **CE3.1** | 0.48% | 0.46% | 0.49% | 0.49% |
| **CE3.2** | 0.39% | 0.35% | 0.41% | 0.41% |

Table 4.4. Reported run-time estimates for CE3.x tests, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | **HDR HLG** | | **SVT12 RGB** | | **SVT16 RGB** | |
|  | Enc | Dec | Enc | Dec | Enc | Dec | Enc | Dec |
| **AI** | **CE3.1** | 100% | 98% | 100% | 99% | 101% | 99% | 101% | 93% |
| **CE3.2** | 91% | 88% | 84% | 83% | 86% | 81% | 85% | 76% |
| **LDB** | **CE3.1** | 100% | 98% | 99% | 99% | 102% | 101% | 101% | 96% |
| **CE3.2** | 86% | 92% | 73% | 81% | 81% | 83% | 86% | 82% |
| **RA** | **CE3.1** | 100% | 98% | 99% | 99% | 102% | 101% | 101% | 96% |
| **CE3.2** | 86% | 90% | 72% | 84% | 82% | 86% | 87% | 86% |

Table 4.5. Simulation results for CE3.x tests, HBD/HBR CTC, NormQP test configuration.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | | | **HDR HLG** | | |
|  | DE100 | PSNR-L100 | wY | wU | wV | Y | U | V |
| **AI** | **CE3.1** | 3.06% | 1.97% | 1.91% | 3.26% | 3.29% | 1.90% | 3.31% | 3.42% |
| **CE3.2** | 11.73% | 8.87% | 9.07% | 12.86% | 13.64% | 15.56% | 8.76% | 9.14% |
| **RA** | **CE3.1** | 2.68% | 1.52% | 1.48% | 2.64% | 2.96% | 1.50% | 2.73% | 3.04% |
| **CE3.2** | 10.79% | 7.92% | 7.75% | 11.38% | 11.83% | 10.42% | 6.67% | 7.51% |

Table 4.6. Run-time estimates for CE2.x tests, HBD/HBR CTC, NormQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | **HDR HLG** | |
|  | Enc | Dec | Enc | Dec |
| **AI** | **CE3.1** | 100% | 99% | 99% | 98% |
| **CE3.2** | 91% | 97% | 91% | 100% |
| **RA** | **CE3.1** | 99% | 99% | 99% | 99% |
| **CE3.2** | 93% | 101% | 89% | 100% |

Table 4.7. Simulation results for CE3.x tests, HBD/HBR CTC, lossless test configuration.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | bit-rate saving | | | | |
|  |  | HDR PQ | HDR HLG | SVT12 | SVT16 |
| **AI** | **CE3.1** | 0.55% | 0.38% | 0.29% | 0.20% |
| **CE3.2** | 2.91% | 0.93% | 0.42% | -0.29% |
| **LDB** | **CE3.1** | 0.67% | 0.47% | 0.39% | 0.26% |
| **CE3.2** | 6.11% | 1.29% | 0.39% | -0.28% |
| **RA** | **CE3.1** | 0.69% | 0.48% | 0.40% | 0.27% |
| **CE3.2** | 6.10% | 1.30% | 0.41% | -0.25% |

4.2.2 Throughput related simulation results vs. VTM13

Table 4.8. BD-binrate (average, weighted) simulation results for CE3.2 tests, 12 bits data, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
|  | Y | U | V | Y | U | V | Aver.GBR | G | B | R |
| **AI** | **CE3.2** | -53.71% | -52.96% | -52.64% | -53.77% | -53.12% | -52.71% | -45.36% | -45.42% | -45.33% | -45.32% |
| **LDB** | **CE3.2** | -54.92% | -53.96% | -53.68% | -61.88% | -61.57% | -61.25% | -50.72% | -50.77% | -50.71% | -50.70% |
| **RA** | **CE3.2** | -55.30% | -54.36% | -53.93% | -61.98% | -61.86% | -61.63% | -51.64% | -51.70% | -51.61% | -51.61% |

Table 4.9. BD-binrate (average, weighted) simulation results for CE3.2 tests, 16 bits data, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **SVT16 RGB** | | | |
|  | Aver.GBR | G | B | R |
| **AI** | **CE3.2** | -37.57% | -37.59% | -37.56% | -37.57% |
| **LDB** | **CE3.2** | -39.98% | -40.01% | -39.96% | -39.96% |
| **RA** | **CE3.2** | -40.33% | -40.39% | -40.30% | -40.30% |

Table 4.10. bin to bit ratio (peak, weighted and unweighted) simulation results for CE3.2 tests, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | **HDR HLG** | | **SVT12 RGB** | | **SVT16 RGB** | |
|  | unweighted | weighted | unweighted | weighted | unweighted | weighted | unweighted | weighted |
| **AI** | **CE3.2** | -15.85% | -57.49% | -14.55% | -56.41% | -10.40% | -46.46% | -10.57% | -37.45% |
| **LDB** | **CE3.2** | -14.11% | -55.32% | -14.40% | -55.85% | -10.39% | -45.72% | -10.56% | -37.16% |
| **RA** | **CE3.2** | -12.80% | -53.70% | -13.62% | -55.00% | -10.26% | -44.23% | -10.34% | -36.32% |

In CE3.2, transform coefficients are bypass coded. The losses are particularly large for HDR cases (10% and more; it is reported that this may be due to the fact that a relatively large number of coefficients still has a low value even in the low QP range). For CE3.1, the losses are more homogeneous over the different test data (HDR and SVT), and also lower.

From the results reported in the CE, it is difficult to quantify what the benefit of the two methods in terms of increasing the throughput is. CE3.2 probably increases the throughput more radically, but also has quite significant loss.

More study is necessary to understand the problem and understand if any of the two proposals is an appropriate solution. There are CE related conributions which provide more data and propose additional ways to increase the throughput.

**Simulation results vs. HM16.23**

Table 4.11. Simulation results for CE3.x tests, 16 bits data, HBD/HBR CTC, LowQP test configuration, HM16.23 with High Throughput 4:4:4 Intra 16 profile.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **SVT16 RGB** | | | |
|  | Aver.GBR | G | B | R |
| **AI** | **CE3.1** | -1.82% | -2.40% | -1.53% | -1.55% |
| **CE3.2** | -2.05% | -2.61% | -1.76% | -1.80% |

Table 4.12. Simulation results for CE3.x tests, 12 bits data, HBD/HBR CTC, LowQP test configuration, HM16.23 with High Throughput 4:4:4 Intra 16 profile

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
|  | wY | wU | wV | Y | U | V | Aver.GBR | G | B | R |
| **AI** | **CE3.1** | -7.03% | -9.37% | -10.06% | -4.62% | -6.50% | -6.67% | -2.27% | -4.03% | -2.86% | -2.94% |
| **CE3.2** | 1.15% | 0.51% | -0.21% | 0.54% | 0.16% | 0.42% | -2.24% | -3.04% | -1.79% | -1.89% |

Throughput **related simulation results vs. HM16.23**

Table 4.13. BD-binrate (average, weighted) simulation results for CE3.2 tests, 16 bits data, HBD/HBR CTC, LowQP test configuration, HM16.23 with High Throughput 4:4:4 Intra 16 profile.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **SVT16 RGB** | | | |
|  | Aver.GBR | G | B | R |
| **AI** | **CE3.2** | -37.86% | -38.16% | -37.71% | -37.72% |

Table 4.14. BD-binrate (average, weighted) simulation results for CE3.2 tests, 12 bits data, HBD/HBR CTC, LowQP test configuration, HM16.23 with High Throughput 4:4:4 Intra 16 profile.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
|  | Y | U | V | Y | U | V | Aver.GBR | G | B | R |
| **AI** | **CE3.2** | -50.16% | -50.44% | -50.51% | -52.39% | -52.58% | -52.31% | -46.08% | -46.36% | -45.93% | -45.96% |

Table 4.15. bin to bit ratio (peak, weighted and unweighted) simulation results for CE3.2 tests, HBD/HBR CTC, LowQP test configuration, HM16.23 with High Throughput 4:4:4 Intra 16 profile..

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | **HDR HLG** | | **SVT12 RGB** | | **SVT16 RGB** | |
|  | unweighted | weighted | unweighted | weighted | unweighted | weighted | unweighted | weighted |
| **AI** | **CE3.2** | -10.59% | -50.09% | -12.37% | -52.02% | -13.78% | -45.14% | -13.28% | -36.50% |

The results indicate that the method from CE3.1 has comparable gain to HEVC high throughput profile, as found when comparing the “normal throughput” configurations (see AHG8 report). CE3.2 has losses compared to HEVC high throughput in the HDR test cases (whereas similar to CE3.1 for SVT12).

In general, it is not obvious that the numbers above would give an appropriate quantification of the throughput benefit. More analysis was performed in BoG JVET-W0180.

[JVET-W0044](https://jvet-experts.org/doc_end_user/current_document.php?id=10859) CE3.1: CABAC-bypass alignment for high bit-depth coding [M. G. Sarwer, J. Chen, Y. Ye, R. -L. Liao (Alibaba)]

[JVET-W0047](https://jvet-experts.org/doc_end_user/current_document.php?id=10862) Crosscheck of JVET-W0044: CE3.1: CABAC-bypass alignment for high bit-depth coding [A. Browne (Sony)] [late]

[JVET-W0045](https://jvet-experts.org/doc_end_user/current_document.php?id=10860) CE-3.2: a high throughput mode for high bit depth and high bit rate extensions [F. Wang, Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO), M. Sarwer, J. Chen, Y. Ye, R. Liao (Alibaba)]

[JVET-W0058](https://jvet-experts.org/doc_end_user/current_document.php?id=10873) Crosscheck of CE-3.2: a high throughput mode for high bit depth and high bit rate extensions [K. Naser (InterDigital)] [miss]

[JVET-W0046](https://jvet-experts.org/doc_end_user/current_document.php?id=10861) CE-1.1: coding of last significant coefficient position for high bit depth and high bit rate extensions [F. Wang, L. Xu, Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO)]

[JVET-W0048](https://jvet-experts.org/doc_end_user/current_document.php?id=10863) Cross-check on JVET-W0046: Coding of last significant coefficient position for high bit depth and high bit rate extensions (CE1.1) [D. Rusanovskyy (Qualcomm)] [late]

[JVET-W0050](https://jvet-experts.org/doc_end_user/current_document.php?id=10865) [CE2.1][CE2.2] Content Adaptive Transform Precision [K. Naser, F. Galpin, T. Poirier, F. Le Leannec (InterDigital)]

[JVET-W0056](https://jvet-experts.org/doc_end_user/current_document.php?id=10871) Crosscheck of JVET-W0050: [CE2.1][CE2.2] Content Adaptive Transform Precision [A. Browne (Sony)] [late]

[JVET-W0094](https://jvet-experts.org/doc_end_user/current_document.php?id=10910) Cross-check on JVET-W0050: [CE2.1][CE2.2] Content Adaptive Transform Precision (CE2.2) [D. Rusanovskyy (Qualcomm)] [late]

### CE related contributions: Entropy Coding for High Bit Depth and High Bit Rate Coding (7)

Contributions in this area were discussed in session 3 at 2100–2310 UTC on Wednesday 7 July 2021 (chaired by JRO).

[JVET-W0051](https://jvet-experts.org/doc_end_user/current_document.php?id=10866) CE-related: Additional bypass coding for high throughput CABAC [M. G. Sarwer, J. Chen, Y. Ye, R. -L. Liao (Alibaba)]

CE3.2 proposes a high throughput mode for operation range extension of VVC. In this contribution, in order to further improve the throughput of CE3.2, following modifications are proposed on top of CE-3.2.

* Disable signaling of the position of the last significant coefficient
* Bypass coding of the transform skip flag, and
* TU level bypass alignment.

It is proposed to enable the proposed methods only in High Throughput 4:4:4 16 Intra Profile. Following results are reported.

As compared to CE3.2:

* SVT-16-bit sequences (low QP):
  + All Intra: 0.00%(G), -0.04% (B), -0.04 % (R), 95% (EncT), 95%(DecT)
  + LDB: -0.11%(G), -0.10 % (B), -0.11 % (R), 99% (EncT), 97%(DecT)
  + Random access: -0.11%(G), -0.11 % (B), -0.12 % (R), 98% (EncT), 97%(DecT)
* SVT-16-bit sequences (lossless):
  + All Intra: -0.03%, 98% (EncT), 93%(DecT)
  + LDB: -0.07%, 102% (EncT), 96%(DecT)
  + Random access: -0.08%, 102% (EncT), 98%(DecT)

As compared to VTM-13.0:

* SVT-16-bit sequences (low QP):
  + All Intra: 0.06%(G), 0.05% (B), 0.03 % (R), 83% (EncT), 73%(DecT)
  + LDB: 0.21%(G), 0.25 % (B), 0.24 % (R), 86% (EncT), 79%(DecT)
  + Random access: 0.24%(G), 0.30 % (B), 0.29 % (R), 85% (EncT), 79%(DecT)
* SVT-16-bit sequences (lossless):
  + All Intra: -0.32%, 89% (EncT), 59%(DecT)
  + LDB: -0.35%, 89% (EncT), 62%(DecT)
  + Random access: -0.33%, 88% (EncT), 62%(DecT)

Loss is larger than for CE3.2. The Non-uniformity of the loss (much larger for HDR classes than for SVT) is also found here (and even increased).

In the associated powerpoint deck, the following estimation is given about number of necessary CPU cycles (for Crowd Run AI QP=-33):

|  |  |  |
| --- | --- | --- |
|  | bypassWeight | Estimated Cycle in MHz |
| **HM High throughput profile** | 12 | 853.00 |
| **VTM-13.0** | 4 | 1423.52 |
| **CE-3.1** | 12 | 885.41 |
| **CE-3.2** | 12 | 362.99 |
| **W0051** | 12 | 281.74 |

The compression performance is affected significantly for some cases here (for HDR, worse than HEVC high throughput profile for the given proposal and CE3.2).

Before taking any action (if at all) a better understanding is needed how to measure the throughput, and how much reduction of throughput would be necessary relative to VTM.

Some worst case throughput analysis also appears necessary.

[JVET-W0149](https://jvet-experts.org/doc_end_user/current_document.php?id=10979) Cross-check report on JVET-W0051: CE-related: Additional bypass coding for high throughput CABAC [D. Rusanovskyy (Qualcomm)] [late]

[JVET-W0052](https://jvet-experts.org/doc_end_user/current_document.php?id=10867) CE-related: CABAC skip mode [K. Abe, T. Toma, V. Drugeon (Panasonic)]

This contribution proposes a CABAC skip mode for the high bit rate and super low delay use cases. CABAC skip mode was already proposed in JVET-M0089, JVET-N0207, and JVET-O0308 for VVC version1. The proposed scheme directly outputs binarized bins as a bitstream without any CABAC processing, therefore avoiding CABAC throughput issues without any additional building blocks. There are no technical changes compared to the previous proposals. This document reports the simulation results on high bit depth CTC. Simulation results show that CABAC skip mode can solve the throughput issues at the cost of 12%, 17%, 17% bitrate increase for AI, RA, LDB on PQ for HBD Low QP, 8%, 15%, 15% bitrate increase for AI, RA, LDB on HLG for HBD Low QP, and 3%, 3%, 3% bitrate increase for AI, RA, LDB on SVT for HBD Low QP.

The proposal is more radically disabling CABAC than the CE proposals (also compared to CE 3.2)

Compared to HEVC high throughput profile, this would have worse compression performance (up to 10% bit rate increase for HDR PQ 444)

An analysis shows that the symbol rate (number of syntax elements per second) is approximately half compared to VTM13, almost consistently over all sequences. Reduction in number of context coded bins is also significant, but more sequence dependent (for some sequences more than a factor of 10, for others less). The fact that the symbol rate is not as dramatically decreased as the number the number of context coded bins might be an indicator that the debinarization becomes the main bottleneck at high bit rate.

Question to be answered: Is the throughput of HEVC high throughput profile not good enough? And would a VVC profile with less compression performance and possibly more complex encoder?

[JVET-W0139](https://jvet-experts.org/doc_end_user/current_document.php?id=10968) Crosscheck of JVET-W0052: CE-related: CABAC skip mode [T. Zhou, T. Ikai (Sharp)] [late]

[JVET-W0092](https://jvet-experts.org/doc_end_user/current_document.php?id=10908) CE-related: On constraining inverse transform precision for high bit depth and high bitrate video coding [L. Kerofsky, D. Rusanovskyy, M. Karczewicz (Qualcomm), K. Naser, F. Galpin, F. Le Leannec, P. de Lagrange (Interdigital)]

Was presented July 7 2330-0010

At the 22nd JVET meeting extended precision mode was adopted to VVCv2 draft specification for high bit depth and high bitrate video coding. In this mode, the precision of the inverse transform is extended to 19 bits for coding of 12 bits video and to 21 bits for coding of 16 bits video. This contribution proposes to employ Content Adaptive Transform Precision (CATP) method to constrain the bit depth of inverse transform implementation while preserving coding performance of the Anchor (VTM13.0, extended precision ON). The method proposed here is a modification of the CATP of JVET-W0050 (CE2.2) and targets complexity reduction and constraining precision used in the inverse transform.

Method has been tested under HBD / HBR CTC with 16-bit, 17-bit and 18-bit constraints vs. the Anchor. No run time increase was reported for tested configurations, no noticeable penalty is reported for Normal QP configuration. Summary of the results for Low QP are following:

*Test on 16 bit constraints:*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Low QP | HDR PQ12 | | | HDR HLG12 | | | SVT12 | SVT16 |
| wY | wU | wV | Y | U | V | Aver.GBR | Aver.GBR |
| **AI** | -0.01% | -0.01% | -0.02% | 0.00% | -0.01% | -0.01% | 0.03% | 0.28% |
| **LDB** | -0.02% | 0.01% | 0.00% | 0.00% | -0.02% | -0.02% | 0.01% | 0.14% |
| **RA** | 0.00% | -0.01% | 0.00% | 0.00% | -0.01% | 0.00% | 0.01% | 0.15% |

*Test on 17 bit constraints:*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Low QP | HDR PQ12 | | | HDR HLG12 | | | SVT12 | SVT16 |
| wY | wU | wV | Y | U | V | Aver.GBR | Aver.GBR |
| **AI** | -0.01% | -0.01% | -0.03% | -0.01% | -0.01% | -0.01% | 0.01% | 0.10% |
| **LDB** | 0.00% | -0.01% | 0.01% | 0.00% | -0.01% | -0.01% | 0.00% | 0.03% |
| **RA** | 0.00% | -0.01% | 0.00% | 0.00% | 0.00% | -0.01% | 0.00% | 0.04% |

*Test on 18 bit constraints:*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Low QP | HDR PQ12 | | | HDR HLG12 | | | SVT12 | SVT16 |
| wY | wU | wV | Y | U | V | Aver.GBR | Aver.GBR |
| **AI** | -0.01% | -0.03% | -0.03% | 0.00% | -0.01% | -0.01% | 0.00% | 0.02% |
| **LDB** | -0.01% | -0.01% | 0.01% | -0.01% | -0.02% | 0.00% | 0.00% | -0.01% |
| **RA** | 0.00% | -0.01% | 0.01% | -0.01% | -0.01% | 0.00% | 0.00% | 0.00% |

It is proposed to adopt proposed CATP method and constrained bit depth implementation of the inverse transform to VVCv2.

The proponents report that the extended precision flag of current v2 draft increases buffer size of transform coefficients by 18% (for 12 bit data), whereas the proposal does not have any increase in buffer size.

For 16 bit data, the increase of buffer size by the extended precision flag is larger (about 30%, from 16 to 21 bits). When keeping 16 bit internal bit depth, the proposal has a compression loss of roughly 0.3% for AI, 0.15% for RA/LB. When the proposal is using 18 bit internal bit depth, there is no loss for 16 bit data, but the buffer size is increased by 12.5% relative to VTM; but decreased by 14% compared to extended precision flag.

In order to achieve that, it is necessary to parse the entire TU and determine the maximum transform coefficient in that TU before the final inverse quantization and inverse transform can be done.

The absolute saving in memory (considering the cases where no loss occurs) is 3 bits per coefficient, which gives 768 bytes for the entire 64x32 block (considering the zero-out of secondary transform).

Implementation-wise, it would be either necessary to add another buffer for the inverse quantization (which would end up with more buffer requirement than for the usage of extended precision flag), or perform inverse quantization twice (in which latter case, the latency would be increased).

No action.

[JVET-W0161](https://jvet-experts.org/doc_end_user/current_document.php?id=10992) Cross-check report on JVET-W0092 (CE-related: On constraining inverse transform precision for high bit depth and high bitrate video coding) [M. G. Sarwer (Alibaba)] [late]

[JVET-W0114](https://jvet-experts.org/doc_end_user/current_document.php?id=10930) CE-related: High throughput mode for high bit-depth coding – harmonization of CE-3.2 and CE-1.1 [T. Tsukuba, M. Ikeda, T. Suzuki (Sony)]

In JVET-W0045 (CE-3.2), CABAC high throughput mode is proposed, where all the syntax elements in residual coding, except for last significant coefficient position in RRC, are fully bypass-coded and CABAC bypass alignment is carried out once after the last significant coefficient position in RRC and at the beginning of a TB in TSRC. It is asserted that for SVT 16-bit sequences this technique improves weighted Delta Binrate by -37.53% for AI, -40.3% for RA and -40.40% for LB with minor coding losses of {0.06%, 0.09%, 0.07%} for AI, {0.32%, 0.36%, 0.35%} for RA and {0.35%, 0.41%, 0.41%}, respectively. However, there is room to further improve CABAC throughput by bypass-coding last significant coefficient position in RRC.

In JVET-W0046 (CE-1.1), a modification to last significant coefficient position is proposed, which codes the relative coordinates of last significant coefficient with reference to the bottom-right corner of a zero-out transform block. It is asserted that for SVT 16-bit sequences this technique improves weighted Delta Binrate by -4.02% for AI, -2.17% for RA and -2.15% for LB with coding gain of {-0.29%, -0.28%, -0.28%} for AI, {-0.28%, -0.26%, -0.26%} for RA and {-0.27%, -0.25%, -0.25%} for LB, respectively. It implies that CE-1.1 is also useful to reduce the number of bins of last significant coefficient position as well as improve coding efficiency.

In order to fully bypass-code all the syntax elements in residual coding, this contribution proposes a harmonization of CE-3.2 and CE-1.1 with bypass-coding of last significant coefficient position, where all the syntax elements in residual coding, including last significant coefficient position in RRC, are fully bypass-coded and CABAC bypass alignment is carried out once at the beginning of TB in both RRC and TSRC. It is suggested to apply the proposed method only in 16-bit high throughput profile.

The simulation results for SVT 16-bit sequences are shown below. The full simulation results are shown in section 3.

BD-bitrate compared to VTM-13.0 (low QP):

* AI: 0.26% (G), 0.17% (B), 0.15% (R), 81% (EncT), 71% (DecT)
* LDB: 0.31% (G), 0.31% (B), 0.31% (R), 90% (EncT), 79% (DecT)
* RA: 0.35% (G), 0.35% (B), 0.35% (R), 92% (EncT), 78% (DecT)

Delta Binrate and Bin-to-Bit ratio compared to VTM-13.0 (low QP):

* AI: -43.79% (weighted Delta Binrate), -12.91% (peak unweighted Bin2Bit), -43.76 % (peak weighted Bin2Bit)
* LDB: -44.73% (weighted Delta Binrate), -12.84% (peak unweighted Bin2Bit), -43.37 % (peak weighted Bin2Bit)
* RA: -45.11% (weighted Delta Binrate), -12.55% (peak unweighted Bin2Bit), -42.40% (peak weighted Bin2Bit)

BD-bitrate compared to CE-3.2 (low QP):

* AI: 0.19% (G), 0.08% (B), 0.07% (R), 95% (EncT), 90% (DecT)
* LDB: -0.01% (G), -0.05% (B), -0.04% (R), 99% (EncT), 94% (DecT)
* RA: 0.00% (G), -0.06% (B), -0.06% (R), 100% (EncT), 94% (DecT)

Delta Binrate and Bin-to-Bit ratio compared to CE-3.2 (low QP):

* AI: -10.02% (weighted Delta Binrate), -2.61% (peak unweighted Bin2Bit), -10.08% (peak weighted Bin2Bit)
* LDB: -7.81% (weighted Delta Binrate), -2.55% (peak unweighted Bin2Bit), -9.88% (peak weighted Bin2Bit)
* RA: -7.86% (weighted Delta Binrate), -2.46% (peak unweighted Bin2Bit), -9.54% (peak weighted Bin2Bit)

The proposal is not a plain combination of the two CE proposals, as bypass coding is also applied to the last significant coefficient position syntax for RRC

[JVET-W0138](https://jvet-experts.org/doc_end_user/current_document.php?id=10967) Crosscheck of JVET-W0114 (CE-related: High throughput mode for high bit-depth coding – harmonization of CE-3.2 and CE-1.1) [F. Wang (OPPO)] [late]

[JVET-W0116](https://jvet-experts.org/doc_end_user/current_document.php?id=10932) CE-3 related: a different alignment position for high throughput mode [F. Wang, Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO)]

CE-3.2 proposes a high throughput mode for high bit depth high bit rate coding. In the CE-3.2 method, the bit alignment mechanism is applied after coding the last significant coefficient position in RRC where last\_sig\_coeff\_x\_prefix and last\_sig\_coeff\_y\_prefix are first context coded then last\_sig\_coeff\_x\_suffix and last\_sig\_coeff\_y\_suffix are subsequently bypass coded. This contribution evaluates an alternative solution where the bit alignment process is performed after the context coded bins of last\_sig\_coeff\_x\_prefix and last\_sig\_coeff\_y\_prefix and right before the bypass-coded bins of last\_sig\_coeff\_x\_suffix and last\_sig\_coeff\_y\_suffix.

Under the high bit-depth common test condition (CTC), the simulation results show negligible change in BD-bitrate, BD-binrate and Bin2Bit compared with that from CE-3.2.

The method is not changing the throughput.

Benefit not obvious.

No action.

[JVET-W0162](https://jvet-experts.org/doc_end_user/current_document.php?id=10993) Cross-check report on JVET-W0116 (CE-3 related: a different alignment position for high throughput mode) [M. G. Sarwer (Alibaba)] [late]

[JVET-W0117](https://jvet-experts.org/doc_end_user/current_document.php?id=10933) CE-3 related: a slice level high throughput mode [F. Wang, Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO)]

CE-3.2 proposes a high throughput mode for high bit depth and high bit rate coding. It is reported that the high throughput mode significantly improves the BD-bin rate and bin to bit ratio and it is suggested to apply the proposed method in 16-bit high throughput configurations. This contribution proposes a slice level control mothed over the CE-3.2 high throughput mode. With the proposed mothed, the high throughput mode can be invoked in more throughput-demanding use-cases and applications.

Under the high bit-depth common test condition (CTC), the simulation results compared to VTM-13.0 are summarized as below:

 For PQ and HLG: no noticeable change (low QP)

 For SVT:

BD-bitrate (low QP):

 AI: 0.12 % (G), 0.17 % (B), 0.16 % (R), 89 % (EncT), 87 % (DecT)

 LDB: 0.23 % (G), 0.26 % (B), 0.26 % (R), 92 % (EncT), 92 % (DecT)

 RA: x.xx % (G), x.xx % (B), x.xx % (R), xx % (EncT), xx % (DecT)

BD-binrate and Bin2Bit (low QP):

 AI: -28.19 % (weighted average Delta Binrate), -7.09 % (peak unweighted Bin2Bit), -23.20 % (peak weighted Bin2Bit)

 LDB: -24.73 % (weighted average Delta Binrate), -6.22 % (peak unweighted Bin2Bit), -19.57 % (peak weighted Bin2Bit)

 RA: xx % (weighted average Delta Binrate), xx % (peak unweighted Bin2Bit), xx % (peak weighted Bin2Bit)

It is pointed out that leaving the enabling of the high-throughput flag to the encoder would not solve the throughput problem which is a problem of the decoder complexity. It would be necessary to make the enabling of the high throughput flag mandatory under certain conditions, e.g. if the bin number in the slice becomes very large.

The proponents claim that one of the purposes of their proposal is to point out that the throughput problem occurs only at the very high bit rates (e.g. QP <= -5), and restricting the slice-level switching to such cases and not running a high throughput mode for higher QPs avoids compression losses.

Another example is the appearance of scene cuts, where the number of bits becomes large.

There must be some bitstream constraint to limit throughput, and an encoder could then decide to use a mechanism like this to obey this constraint.

It is pointed out that already version 1 has such constraints, but those are mainly assuming single core implementations. For extreme high bit rate, this is not straightforward for extreme high bit rate

[JVET-W0118](https://jvet-experts.org/doc_end_user/current_document.php?id=10934) CE-3 related: High throughput 4:4:4 16 intra profile for VVC [F. Wang, Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO)]

Only the aspect of this contribution reporting results from combining tools from the CE belongs into this section. Otherwise it should be considered in 4.12.

It is reported that combining two coding tools, namely the alternative derivation method of last significant coefficient position proposed in CE-1.1, and the bit-aligned full-bypass method proposed in CE-3.2 I beneficial.

Under the high bit-depth common test condition (CTC), the simulation results for SVT 16-bit intra are reported together with BD-bitrate, BD-binrate and Bin2Bit. These test results clearly demonstrated the improved overall performance, including both BD-bitrate and throughput, when running both tools together in high bitrate 4:4:4 16-bit Intra coding.

BD-bitrate: -0.23 % (G), -0.19 % (B), -0.20 % (R), 81 % (EncT), 74 % (DecT)

BD-binrate and Bin2Bit: -41.50% (weighted average Delta Binrate), -12.16 % (peak unweighted Bin2Bit), -41.43 % (peak weighted Bin2Bit)

Similar combination with a modified version of CE3.2 is proposed in JVET-W0114.

The gain probably comes due to CE-1.1.

Breakout group (M. Sarwer) to discuss methods and criteria to analyse the entropy coding throughput. Various of such approaches are proposed in different documents of this section, and should be aligned. It should also be discussed how to analyse and limit worst case throughput. See JVET-W0180

[JVET-W0135](https://jvet-experts.org/doc_end_user/current_document.php?id=10964) Cross-check on JVET-W0118: CE related: High throughput 4:4:4 16 intra profile for VVC [T. Tsukuba (Sony)] [late]

### Adaptation of other tools for high bit rate and high bit depth (11)

Contributions in this area were discussed in session 9 at 0510–0725 UTC and session 11 2100-2300 UTC on Friday 9 July 2021 (chaired by JRO and GJS).

[JVET-W0055](https://jvet-experts.org/doc_end_user/current_document.php?id=10870) AHG8: Signaling TSRC Rice parameter in slice header extension [T. Tsukuba, M. Ikeda, T. Suzuki (Sony)]

The slice header of VVC version 1 contains an extension field, called slice header extension, as a hook for the future extension; It makes it possible to signal new syntax elements to the slice header that are ignored by a version 1 decoder, but can be used by a decoder conforming to a future version. In the latest WD for VVC operation range extension, a Rice parameter index for transform-skip residual coding can be signaled in the slice header but not in the slice header extension. This contribution proposes to relocate it to the slice header extension. A patch for the proposed changes relative to VTM-13.0 is attached.

It is mentioned that this aspect had been discussed between editors earlier. It had been concluded that in the case of the flags added in operation range extensions it is not necessary to use the extension mechanism, as a v1 decoder should not be capable to decode bitstreams of a v2 profile. This depends on of whether the slice header extension is intended for non-backward-compatible extensions or only for backward compatibility.

Some participants thought that decoders should check the profile indicator and not attempt to decode the headers of bitstreams of profiles that it does not recognize. If that is the case, it does not necessarily matter whether a new syntax element is placed in the SH extension or directly into a profile-dependent part of the SH syntax structure. It was commented that both ways of extending the syntax would have about the same number of condition checks.

Further, the number of condition checks is exactly the same in the proposal as in the current syntax.

It could be done as proposed, but not necessary to take action.

[JVET-W0060](https://jvet-experts.org/doc_end_user/current_document.php?id=10875) AHG8: A constraint of max CTU size and tile on WPP for high bit rate coding [K. Kondo, M. Ikeda (Sony)]

This contribution proposes a constraint for max CTU size to 64 for high bit rate coding. When high throughput is required, wavefront parallel processing can be used by setting sps\_entropy\_coding\_sync\_enabled\_flag to 1. In VVC version 1, max CTU size was extended from 64 (HEVC) to 128. It might make max delay of CABAC of a CTU row increased. This proposed constraint can decrease the max delay. Also, the other solution, CTU row splitting using tile, is proposed. To investigate the impact of coding efficiency, max CTU size 64 and CTU row split by tiling were tested. In the experiment, an impact of coding efficiency is reported as below. The numbers in the bracket mean BD rates for Y/G, U/B and V/R.

* Test-1: sps\_entropy\_coding\_sync\_enabled\_flag: 1 (CTU size: 128)
  + {0.08%, 0.07%, 0.02%} for wPSNR and PQ contents
  + {0.08%, 0.07% ,0.02%} for PSNR and PQ contents
  + {0.06%, 0.03%, 0.01%} for HLG contents
  + {0.01%, 0.01%,0.01%} for SVT contents.
* Test-2: sps\_entropy\_coding\_sync\_enabled\_flag: 1 and max CTU size: 64
  + {0.15%, 0.01%, -0.17%} for wPSNR and PQ contents
  + {0.15%, 0.02%, -0.14%} for PSNR and PQ contents
  + {0.15%, 0.02%, -0.08%} for HLG contents
  + {0.05%, 0.03%, 0.02%} for SVT contents.
* Test-3: sps\_entropy\_coding\_sync\_enabled\_flag: 1 and Tile 2x1
  + {0.19%, 0.18%, 0.11%} for wPSNR and PQ contents
  + {0.18%, 0.17%, 0.10%} for PSNR and PQ contents
  + {0.10%, 0.08%, 0.03%} for HLG contents
  + {0.04%, 0.03%, 0.02%} for SVT contents.

It is noted that the losses are comparably lower than in some other proposals targeting high throughput

One experts has the opinion that the restrictions might be too strong (in particular, the 64 CTU restriction might not be necessary for 4K content)

It is furthermore noted that a v2 decoder should still be able to decode v1 bitstreams, i.e. restricting to 64 CTU in a potential high throughput mode would not mean that the decoder could save memory for 128 CTU storage.

Include in the analysis of the BoG on high throughput.

[JVET-W0140](https://jvet-experts.org/doc_end_user/current_document.php?id=10969) Crosscheck of JVET-W0060: AHG8: A constraint of max CTU size and tile on WPP for high bit rate coding [T. Zhou, T. Ikai (Sharp)] [late]

[JVET-W0064](https://jvet-experts.org/doc_end_user/current_document.php?id=10879) AHG8: Constraints on transforms and high precision operation [T. Ikai, T. Zhou, T. Hashimoto (Sharp)]

This contribution proposes to reduce complexity by utilizing exclusive usage on 64 point transform and LFNST for high bit depth coding. Specifically, 64 point transform and LFNST are normatively disabled when extended precision flag is equal to 1. In low QP HBD CTC experiment, the proposal shows up to 0.01 % loss as follows.

- 0.04 %, 0.06 %, 0.05 % in PQ

- 0.00%, 0.00% and -0.00% in HLG

- 0.00%, 0.00% and 0.00 % in SVT

The more QPs condition from QP = -13 to 27 was also conducted. It shows proposed constraint (disabling 64 point transform and LFNST) show 0.37 %, 2.12 % and 2.16 % loss in Y, U, and V at QP 12, 17, 22, 27 range in AI PQ 4:2:2. At the same time, this loss reduced to 0.11 %, 0.42 %, 0.27 % at QP 7, 12, 17, 22 by enabling 64 point transform and disabling extended precision (this combination remains valid in this proposal).

Considering performance and complexity balance, this contribution recommends exclusive usage.

v2 fixed PQ simulation results to include FireEater for average bd-rate value in accordance to JVET-U2018.

v3 include alternative proposal where only LFNST is disabled when extended precision flag is equal to 1 as option1b to 3b. Also find the attached slide.

Summary results:

* Proposal
  + Restrict 64 point transform and LFNST when extended precision is on
  + 0.05% to 0.15 % loss in QP=-13 to 27 in PQ AI (0.09 % loss in HBT CTC)
* Alternative proposal
  + Restrict LFNST when extended precision is on
  + 0.06 % possible loss in QP=12 to 27 in PQ AI (No loss in HBT CTC)

The main motivation is reduction of design changes relative to v1. However, design changes are anyway needed (e.g., disabling 64 point transform does not help much, as other transform sizes still need to be modified).

Disabling LFNST might help, but it also has small loss, and not obvious that the benefit is substantial – many other building blocks have to be redesigned, anyway.

No action.

[JVET-W0141](https://jvet-experts.org/doc_end_user/current_document.php?id=10970) Crosscheck of JVET-W0064 (AHG8: Constraints on transforms and high precision operation) [K. Kondo (Sony)] [late]

[JVET-W0070](https://jvet-experts.org/doc_end_user/current_document.php?id=10886) [AHG8] SPS Cleanup for VVC operation range extension [K. Naser, F. Galpin, T. Poirier, F. Le Leannec (InterDigital)]

This contribution proposes a syntax cleanup of SPS flags related to VVC range extension. Specifically, it proposes signaling all/some of the range extension SPS flags only when BitDepth is higher than 10, otherwise it is inferred to 0. In addition, it proposes to signal SPS for TSRC rice parameter only when transform skip is enabled. The goal is to remove the redundant signaling and improve the readability of the specification.

W0109 (related to proposals #1 and #2) and W0121 (identical with proposal #3) are related.

It is asked how common it is that a v2 bitstream carries data with 10 bits or less.

For proposal #2, it is commented that one condition check could be avoided by re-ordering syntax elements

Decision: Adopt JVET-W0070 proposal #3

[JVET-W0091](https://jvet-experts.org/doc_end_user/current_document.php?id=10907) AHG8: On constraining of bit depth of ALF classifier and CCLM derivation for coding of high bit-depth video data [D. Rusanovskyy, M. Karczewicz (Qualcomm)]

This contribution presents results of a study on reusage of 10 bits implementations for ALF classifier and CCLM parameters derivation for higher bit-depth VVC coding. Tools under consideration has been tested under CTC for HBD/HBR video coding, JVET-T2018.

Under LowQP CTC, ALF reportedly provides around 0.4%/1.5%/1.4% of luma gain for AI/LDB/RA configuration of PQ12, respectively, and 0.05%/1.2%/1.2% of luma gain for AI/LDB/RA configurations of HLG12, respectively. It is reported that constraining operations of ALF classifiers to 10 bits input, no penalty on average is observed over PQ12/HLG12 and SVT classes in low QP configuration.

In normal QP configuration, ALF reportedly provides around 2.1% of luma gain and 28.4% of chroma gain on average across PQ12/HLG12 classes and AI/RA configurations. Constraining operations of ALF classifiers to 10 input bits, brings ~0.1% of penalty for luma and ~0.1% of chroma gain on average in PSNR metric.

CCLM gain in AI/LDB/RA configuration of HBD coding are 2.3%/0.2%/0.4% respectively, for PQ12, and 0.7%/0.1%/0.1%, respectively for HLG12. For SVT, reported CCLM gain is 1.2%/0.14%/0.5% on average over RGB, in AI/LDB/RA, respectively. It is reported that constraining operations of CCLM derivation to 10 bits, brings 0.01% penalty on average over PQ12/HLG12 classes in low QP configuration over.

In normal QP, CCLM reportedly provides around 1.3% of luma gain and 24.2% of chroma gain on average across PQ12/HLG12 classes and AI/RA configurations. Constraining operations of CCLM derivation to 10 bits, brings ~0.1% of luma and ~0.5% of chroma penalty on average in PSNR metric.

It is proposed to consider constraining ALF classifier and CCLM derivation process to 10 bits video data to allow reusage of HW implementations of VVC Main10 for high bit-depth, high bitrate VVC coding.

It is commented that it is not fully clear how large the benefit in complexity reduction would be. In terms of high bit depth, the entropy decoding is likely much more serious in terms of complexity. A more thorough analysis would be necessary

It is asked why other common HDR metrics are not reported for HDR – they are not defined in CTC

It is confirmed that the results are not highly varying per sequence.

The benefit is not too obvious without further analysis of the reduction of the complexity.

The proponent comments that for normal QP range where some loss might occur, the proposal should not be applied. Mainly intended for very high rates.

At this moment, no other experts believed that there is enough evidence about benefit.

For further study.

[JVET-W0093](https://jvet-experts.org/doc_end_user/current_document.php?id=10909) AHG8: On significance, GT1, and GT2 flag coding for high bit depths [A. Browne, S. Keating, K. Sharman (Sony)]

This document describes a study of the use of significance, GT1 and GT2 flags in regular residual coding for high bit depths at low QP and lossless operating points. An analysis using VTM 13.0 is provided which lists the number of these flags used and the fraction of them that are set for each QP in the low QP high bit depth CTC. This analysis outlines how as QPs decrease, the fraction of the flags that are set increases to almost 100%. A modification is described which effectively shifts the significance, GT1 and GT2 tests left by a number of bits signalled from the encoder. A non-normative algorithm is also described which computes a suitable value for this shift in a pre-encode step. The analysis, previously done for VTM 13.0, is repeated for the modification and indicates that the fraction of flags that are set does not rise as QPs decrease, unlike VTM 13.0.

The following results are reported for the modification against an anchor of VTM 13.0:

|  |  |  |  |
| --- | --- | --- | --- |
| Lossless | AI | LDB | RA |
| PQ | -0.36% | -0.41% | -0.41% |
| HLG | -0.77% | -1.13% | -1.13% |
| SVT-12 | -1.10% | -1.33% | -1.34% |
| SVT-16 | -1.12% | -1.45% | -1.46% |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Low QP | AI | | | LDB | | | RA | | |
|  | Y/G | U/B | V/R | Y/G | U/B | V/R | Y/G | U/B | V/R |
| PQ | 0.05% | 0.05% | 0.06% | 0.03% | 0.03% | 0.04% | 0.03% | 0.04% | 0.04% |
| HLG | -0.08% | -0.10% | -0.10% | 0.02% | 0.03% | 0.03% | 0.01% | 0.03% | 0.03% |
| SVT-12 | -0.25% | -0.23% | -0.25% | -0.09% | -0.06% | -0.07% | -0.10% | -0.05% | -0.06% |
| SVT-16 | -0.87% | -0.89% | -0.89% | -0.87% | -0.80% | -0.80% | -0.85% | -0.77% | -0.77% |

If GolombRiceParameterAdaptation is disabled for both VTM 13.0 and the modification the following results are reported:

|  |  |  |  |
| --- | --- | --- | --- |
| Lossless | AI | LDB | RA |
| PQ | -1.00% | -1.06% | -1.00% |
| HLG | -2.74% | -2.65% | -2.51% |
| SVT-12 | -3.16% | -2.91% | -2.90% |
| SVT-16 | -2.81% | -2.41% | -2.42% |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Low QP | AI | | | LDB | | | RA | | |
|  | Y/G | U/B | V/R | Y/G | U/B | V/R | Y/G | U/B | V/R |
| PQ | 0.02% | 0.04% | 0.05% | 0.03% | 0.03% | 0.03% | 0.02% | 0.02% | 0.03% |
| HLG | -0.30% | -0.16% | -0.17% | -0.01% | 0.01% | 0.01% | -0.01% | 0.01% | 0.01% |
| SVT-12 | -0.82% | -0.55% | -0.56% | -0.38% | -0.34% | -0.35% | -0.38% | -0.28% | -0.29% |
| SVT-16 | -2.33% | -2.08% | -2.06% | -1.78% | -1.82% | -1.82% | -1.76% | -1.69% | -1.69% |

Amount of shift is signalled in th slice header (separate for luma and chroma)

The shift is likely somewhat sequence dependent – does this cause problems?

The method to determine the shift is a two-pass algorithm (based on estimate of the residual)

Run for every picture, though it is likely that the change of is similar over the sequence

Most gain achieved for SVT (noisy sequences) and lossless. 10 bit performance not studied

Some small losses in normal QP range

LSBs (lower than the position of shift offset) are always coded

It is recommended to measure the benefit in terms of throughput

Further study recommended

[JVET-W0160](https://jvet-experts.org/doc_end_user/current_document.php?id=10991) Cross-check report on JVET-W0093 (AHG8: On significance, GT1, and GT2 flag coding for high bit depths) [M. G. Sarwer (Alibaba)] [late]

[JVET-W0109](https://jvet-experts.org/doc_end_user/current_document.php?id=10925) AHG8: Removal of a prevention mechanism of extended precision for low bit-depth [T. Tsukuba, M. Ikeda, T. Suzuki (Sony)]

In the latest WD for VVC ver.2, there is a prevention mechanism to deactivate extended precision even when sps-level enabled flag (extended\_precision\_processing\_flag) is set to 1. This prevention mechanism loses an encoder flexibility to use extended precision for low bit-depth. In addition, extended precision provides minor gain (roughly -0.1% for AI, -0.1% for RA and -0.1% for LB) for YUV420 10-bit sequences at low QP. This contribution proposes to remove its prevention mechanism, which gives encoders flexibility to use extended precision for low bit-depth (proposal#1).

If necessary to disable extended precision for low bit-depth coding, rather than having a prevention mechanism, this contribution suggests to

- impose a bitstream constraint which indicates such a tool shall be disabled for bit-depth less than or equal to 10 in the profile like HEVC or semantics (proposal#2), or

- omit signaling of extended\_precision\_processing\_flag depending on BitDepth to avoid unnecessary signaling (proposal#3).

The contribution shows that there is a small benefit of extended precision usage and Rice parameter modification even for 10 bit video in low QP range.

Certain application domain standards prescribe using specific profiles. This could be incurring cases where content with lower bit depth is coded in a higher bit depth profile.

The current design prevents usage of the extended precision mechanism for bit depth <= 10 (even if set on by the encoder) by a block-level check. A smart decoder could just overwrite such a flag. Contributions W0070 and W0109 try to resolve this contradiction by either a bitstream restriction, or a high-level mechanism.

After some discussion, it is concluded that imposing a semantics constraint (extended precision flag shall be equal to zero when bit depth is <= 10) is the most appropriate solution. This would be approximately “proposal 2 case 1” from W0109. At the same time, removal of the block-level condition becomes purely editorial.

Decision: Adopt JVET-W0109 “proposal 2 case 1”, up to discretion of editors to align with block-level.

A similar approach should be taken for the Rice parameter extension mechanism (or potentially other mechanisms of v2 that currently have low-level check for <= 10 bits).

Proponents of JVET-W0070 and JVET-W0109 were asked to discuss this offline and propose an appropriate text. See JVET-W0178.

[JVET-W0178](https://jvet-experts.org/doc_end_user/current_document.php?id=11009) [AHG8] Combination of JVET-W0070 and JVET-W0109 on Constraining SPS flags for HBD [K. Naser, F. Galpin, T. Poirier, F. Le Leannec (InterDigital), T. Tsukuba, M. Ikeda, T. Suzuki (Sony)] [late]

This contribution is a combination of JVET-W0070 and W0109 where SPS flags of HBD are constrained for bit-depth less than or equal to 10.

Was presented in session 23 0820 UTC.

It is asserted that “option 3” is reflecting the intended purpose with minimum amount of changes (setting the constraint at highest level of SPS). However, still some editorial changes may be desirable, e.g. not introducing variables where unnecessary.

Decision: Adopt JVET-W0178 option 3.

[JVET-W0176](https://jvet-experts.org/doc_end_user/current_document.php?id=11007) Crosscheck of AHG8: Removal of a prevention mechanism of extended precision for low bit-depth [K. Naser (InterDigital)] [miss]

[JVET-W0115](https://jvet-experts.org/doc_end_user/current_document.php?id=10931) AHG8: A study on Bin-to-Bit ratio of VTM-13.0 for high bit depth coding [T. Tsukuba, S. Keating, M. Ikeda, T. Suzuki (Sony)]

Was discussed in BoG JVET-W0180.

This proposal is reviewed in BoG.

This contribution reports the coding performance (BD-bitrate, BD-binrate) and bin-to-bit ratio of VTM-13.0 relative to HM-16.23 for high bit depth coding under low QP setting (QP=-13,…,12 for 12-bit, QP=-33,…,-8 for 16-bit) and normal QP setting (QP=22, 27, 32, 37). The average results are summarized as follows:

In the normal QPs,

* BD-bitrate changes are -26.21% for AI and -17.17% for RA, respectively.
* BD-binrate changes are -18.01% for AI and -2.43% for RA, respectively.
* The unweighted bin to bit ratios of VTM are roughly 13~15% higher than HM’s. The weighted ones are roughly 20% higher than HM’s.

In the low QPs,

* BD-bitrate changes for AI/LDB/RA is {-8.61%, -7.34%, -7.54%} for PQ, {-4.95%, -5.74%, -6.24%} for HLG, and {-3.31%, -3.37%, -3.56%} for SVT, respectively.
* BD-binrate changes for AI/LDB/RA is {7.71%, 6.41%, 6.51%} for PQ, {3.75%, 3.32%, 3.03%} for HLG, and {-1.77%, -2.39%, -2.07%} for SVT, respectively.
* The unweighted bin to bit ratio changes for AI/LDB/RA are {6.35%, 5.36%, 3.66%} for PQ, {2.72%, 3.60%, 2.42%} for HLG, and {-3.65%, -3.56%, -3.80%} for SVT, respectively. The weighted bin to bit ratio changes for AI/LDB/RA are {17.77%, 18.30%, 17.40%} for PQ, {10.75%, 12.34%, 11.74%} for HLG and {1.68%, 1.88%, 1.37 %} for SVT, respectively.

In the weighted measurements, a bypass bin is counted at 0.25 context coded bins (1:4); in the unweight measurements, a bypass bin is counted at 1 context coded bin (1:1). All the percentage numbers are measured against the HM-16.23.

[JVET-W0121](https://jvet-experts.org/doc_end_user/current_document.php?id=10937) AHG8: On signalling of sps\_ts\_residual\_coding\_rice\_present\_in\_sh\_flag [H.-J. Jhu, X. Xiu, Y.-W. Chen, W. Chen, C.-W. Kuo, X. Wang (Kwai)]

See notes above under W0070.

[JVET-W0172](https://jvet-experts.org/doc_end_user/current_document.php?id=11003) On conformance point of high throughput profile [K. Kawamura, K. Unno (KDDI)] [late]

This contribution proposes an additional conformance point for high throughput conditions in operation-range extensions profile. The current buffer model with hypothetical reference decoder in VVC is well defined for both theoretical and implementation point-of-view. When the VVC is used for mezzanine compression and/or video editing, decoding latency may be unstable due to the CABAC throughput for some implementation. In this contribution, an additional conformance point for the high throughput cases is proposed based on a bit allocation of each frame.

It is commented that the motivation needs more clarification. For example, for low latency applications, peak values of throughput would need to be considered rather than a sliding window potentially covering several frames. Which length of N? Could be 5 or 6 frames. Which maximum bit rate? Depends on compression ratio.

Furthermore, delay would be introduced.

It is commented that the proposal basically tries to limit the throughput which turns into a bit rate limitation. This would not work with a high throughput mode.

It is further pointed out that the approach limits the CPB size. Furthermore, it appears that this buffer model would mostly be appropriate for CBR not VBR

Currently still more at the conceptual level. A text description is announced but not available yet. That would be beneficial for better understanding the details. Further study encouraged.

## AHG11: Neural network-based video coding (18)

### General (2)

Contributions in this area were discussed in session 21 at 2200–2300 UTC on Wednesday 14 July 2021 (chaired by JRO).

[JVET-W0181](https://jvet-experts.org/doc_end_user/current_document.php?id=11012) AHG11: Small Ad-hoc Deep-Learning Library [F. Galpin, T. Dumas, P. Bordes, P. Nikitin, F. Le Leannec, E. Francois (InterDigital)] [late]

This contribution proposes a C++ source code library allowing inference of neural networks. The goal of the library is to provide a simple framework, without dependencies and compatible with JVET software development policy. The library can be interfaced with VTM or ECM and would allow to move forward exploration work on NN based technologies in AhG11.

The table below summarize the framework characteristics.

|  |  |
| --- | --- |
| Language | Pure C++, header only. |
| Footprint | ~3200 LOC, library ~200kB, no dependency |
| Optimization | Some SIMD at hot spots (best effort) |
| Compatibility | TF 1.x, 2.x, PyTorch converters |
| Layer Supports | constants, add, maxPool, matMul, reshape, ReLU,conv2D, mul, concat, max, leakyReLU |
| Type support | float, int32, int16, int8 |
| Quantization | Support adaptive quantizer per layer |

It is claimed that computation using the library, in case of small networks, can be faster than using optimized TensorFlow.

What is the current status? Is the software available? It is planned to make an update of the contribution with software attached.

It is reported that the contributors converted (beyond their own proposals) some other proposals, and it seemed to be simple.

For larger networks, the library implementation is 2-4x slower than Tensorflow.

Several experts expressed that this is a very interesting contribution, which would also support the investigation methodologies discussed in the BoG JVET-W0182. Looking forward towards software code becoming available for further investigation for possible future use.

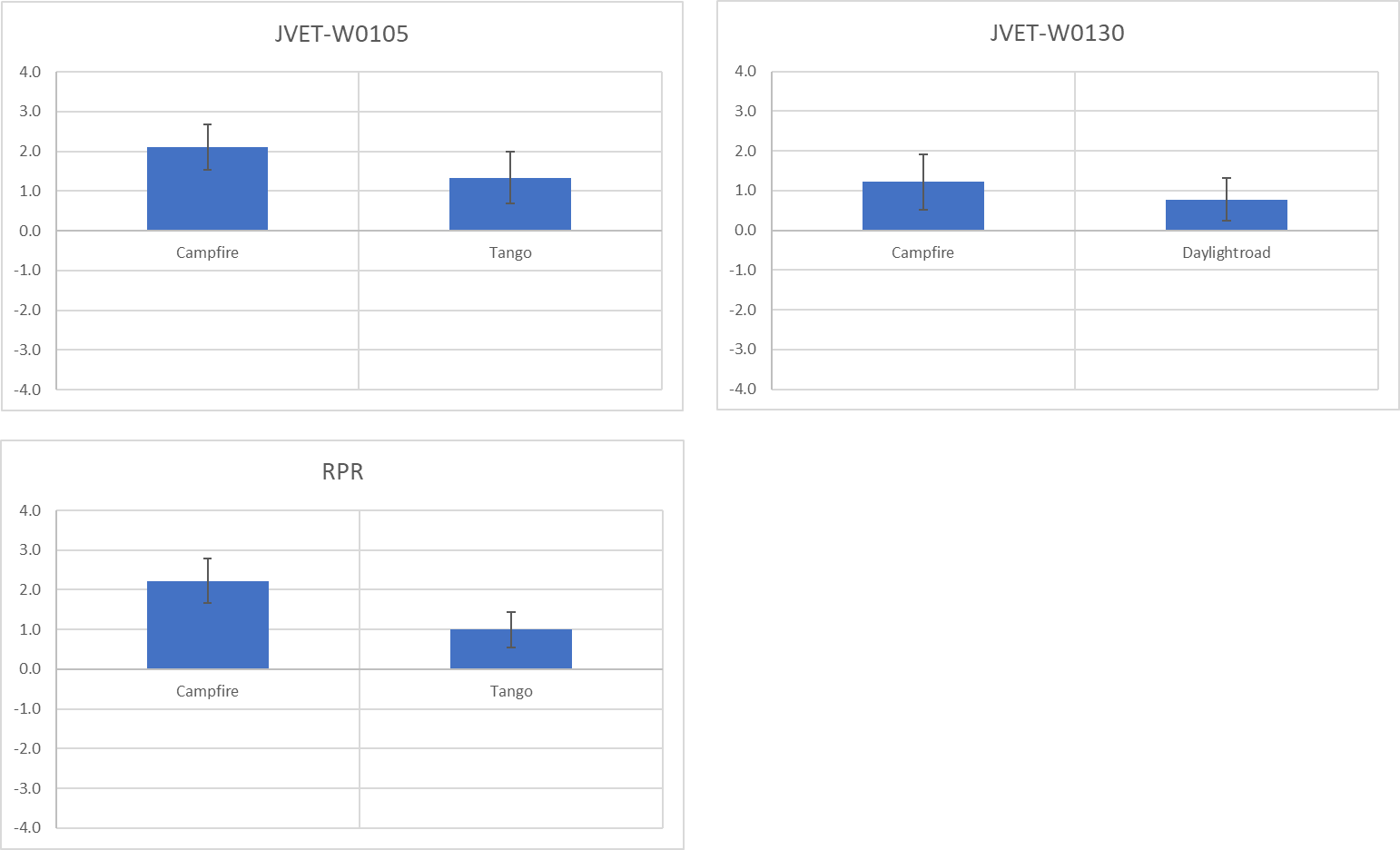
[JVET-W0186](https://jvet-experts.org/doc_end_user/current_document.php?id=11017) EE1-related: Report on results of remote viewing session [M. Wien (RWTH Aachen University), V. Baroncini, A. Segall]

Remote expert viewing tests were conducted during the 23rd JVET meeting for the exploration activity on DNN-based coding tools on UHD test sequences as defined in EE1. Calls for participation in the subjective viewing were issued on the JVET reflector and verbally during JVET sessions. Overall, 22 experts volunteered in participating in the tests. The testing procedure followed the draft SC29/AG5 guidelines for remote experts viewing, providing an A/B comparison of sequences under test.

The results reveal performance improvements for the proposals compared to the VTM anchor. The plots consistently reveal a visual benefit of the proposals over the VTM anchor. In the super-resolution category, the visual benefit of JVET-W0105 appears to be comparable to the visual benefit observed for the RPR configuration under test. For the Campfire sequence, this visual benefit appears to be more substantial than for the Tango sequence. The results for the loopfilter category suggest some visual benefit of JVET-W0130 over the VTM anchor.

Multiple participants remarked that the visual quality of the Campfire sequence was very low. Further, the fact that the sequences under test had different length was commented. It is suggested to align the duration of the video sequences to be presented by cropping a suitable 5 sec part out of 30Hz sequences for the visual tests.

Some results:



Some participating experts expressed that the quality was extremely low in case of Campfire.

Generally speaking, the tested cases are still quite limited and should be extended. However, it is interesting to note that JVET-W0130 (a loop filter) provides visual improvements over VTM (confidence interval above the zero line).

It is also commented that Campfire and Tango are relatively smooth and therefore benficial for reduced resolution coding. Herein, RPR and NN based superresolution perform approximately identical. It is suggested to include sequences with higher amount of detail.

### EE1 contributions: Neural network-based video coding (6)

Contributions in this area were discussed in session 8 at 0015–0120 UTC on Thursday 8 July 2021 (chaired by JRO).

[JVET-W0023](https://jvet-experts.org/doc_end_user/current_document.php?id=10976) EE1: Summary of Exploration Experiments on Neural Network-based Video Coding [E. Alshina, S. Lui, W. Chen, F. Galpin, Y. Li, Z. Ma, H. Wang]

This contribution provides a summary report for the Exploration Experiment 1 on Neural Network-based Video Coding.In total 6 tests have been completed within this EE between the JVET V and W meetings to study and evaluate NNVC technologies, analyze their performance and complexity aspects. Several variants of NN-based in-loop filters with complexity 20...600 kMAC/pxl providing average bit-saving in a range 1.1...8.4% (in random access configuration) have been demonstrated in this round of Exploration Experiment. Several variants of super-resolution (from available in VVC standard functionality to NN-based with signaled parameters) demonstrate 6...11% bit-saving in random access configuration for 4K content.

Test in this Exploration Experiment follows AhG11 tets conditions and complexity assessment methodology. Anchor is VTM11.0 with improved GOP-based temporal filter enabled. Comparison of test results from last and this EE rounds shows that there is very tiny overlap between technologies under study in this EE and MCTF. 5 quality levels (corresponding QP=22, 27, 32, 37, 42) are used for tests in in-loop filters cathegory. For super resolution category even lower bit-rates diapason is used (corresponding QP= 27, 32, 37, 42, 47). Anchors were generated by AhG11.

Unfortunately not all EE participants used the most recent version of results reporting template.

Most of the technologies under test in this EE are unchaged relatively to the last meeting. For them only Random Access configuration results have been up-dated, since new MCTF was enabled for both anchor and test. One participant was not able to provide results for JVET-V meeting, but provides results this meeting (only AI configuration was tested due to the limited computational resources).

EE part on NN based loop filters:

Test results for proposals in this category are shown in Table 1.

Following observations can be done. NN solutions trained using DIV-2K (images) have lower performance in test configurations with motion compentsation. Content adaptivity by signaling scaling factor for residual or/and switching between different NN-based filters models on picture or/and CTU level is source of performance improvement. Proponents keep searching for better complexity-performance trade-off in EE-related contribution [5].

Table 1. Test results in NN-based in-loop filters cathegory.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **#** | **Performance** | **Dec.T.** | **Features** | **#params** | **kMAC /pxl** |
| [JVET-W0062](https://jvet-experts.org/doc_end_user/current_document.php?id=10877) | MSSSIM:  AI: **4.6**%(Y); 6%(Ch)  L1:  AI: **5.8**%(Y); 13%(Ch)  Anchor: VTM-12.0 w/o SAO and ALF | - | **8** Residual Blocks, normalized QP map as additional input, **2 SE** (squeeze-and-excitation) blocks, 3×3 kernels, 64 channels (SAO and ALF are disabled in anchor and test)  2 tests: MSSSIM and L1 training. | 0.13M | 125 |
| [JVET-W0130](https://jvet-experts.org/doc_end_user/current_document.php?id=10947) | RA: **8.4**%(Y); 22% (Ch)  AI: **7.6**%(Y); 19%(Ch)  Anchor : VTM-11-nnvc | ×352  ×203  (CPU) | Input 10 planes: 4\*Y+U+V+QP+3\*Bsinfo, **72×72**, #channels of all hidden layers are M=216 or K=72 ; **14 filter blocks**, Residual scalling (per color, per picture, signalled); **LongActivation** = CONV1×1,K×M + LeakyReLu+CONV1×1,M×K ; | 8×1M | 624 |
| [JVET-W0151](https://jvet-experts.org/doc_end_user/current_document.php?id=10981) | RA:**1.1**%(Y); 4% (Ch)  AI: **1.6**%(Y); 5%(Ch)  Anchor : VTM-11-nnvc | ×32  ×31  (CPU) | Input YUV444, 4 planes + QP map, 64×64, DepthwiseSeparableConvolutions (DSC) in **4** Residual Blocks | 23K | 23 |

NN-based filter architectures tested in this EE are illustrated on Fig. 1, 2 and 3.

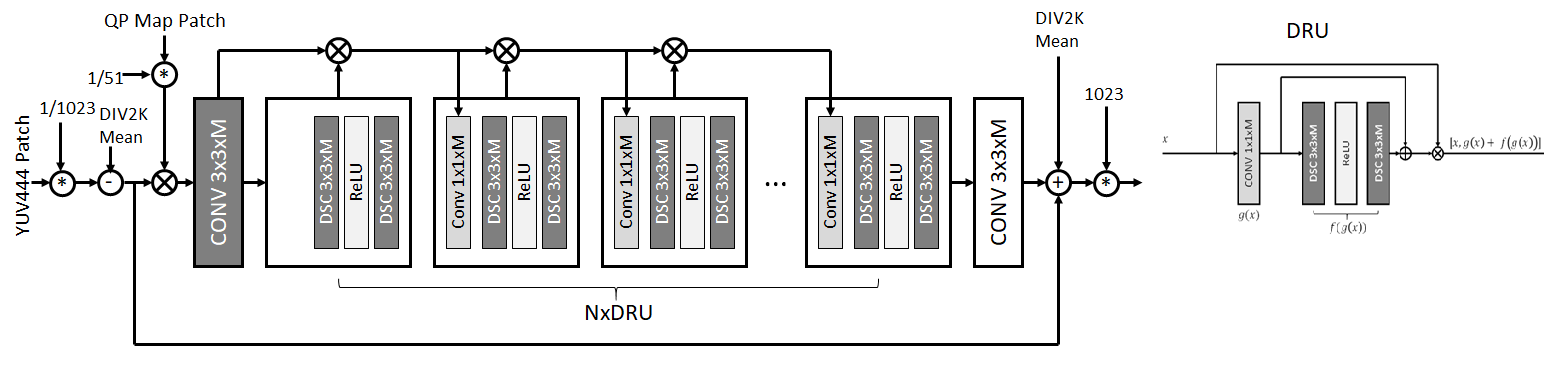


Fig. 1. Low-complexity NN-based in-loop filter with DSC from [JVET-V0137](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10798).





Fig. 2. NN-based in-loop with SE blockd from [JVET-W0062](https://jvet-experts.org/doc_end_user/current_document.php?id=10877).



Fig. 3. "NN-based de-block" with long activation function from [JVET-W0130](https://jvet-experts.org/doc_end_user/current_document.php?id=10947).

For the training proponents used DIV2K [3-4] and BVI-DVC [5-6].

It is noted that the relatively larger gains for chroma are reflecting an improvement in chroma PSNR

EE part on NN based superresolution

Proposals in this category were tested at lower bit-rates, corresponding QP=27,..., 47. Most of propponents reported results only for classes A1 and A2 (4K). For the comparison with NN-based Super Resolution existing in VVC functionality (RPR) was test (results generated by EE coordinators and attached to this report).

Quality metrics computed unsing reconstructed videos in full resolution (so-called “PSNR2” in VTM SW). It needs to be noticed that “MS-SSIM2” computation is not supported by VTM. So, MS-SSIM was not reported for tests in this category.

Table 2. Test results of superresolution-like technologies in EE1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **#** | **Performance** | **Dec.T.** | **Features** | **#params** | **kMAC /pxl** |
| [JVET-W0063](https://jvet-experts.org/doc_end_user/current_document.php?id=10878) | RA(4K):**5.6**%(Y), ~ 27% drop (Ch) RA(4K):**6.1**%(Y), ~ 10% drop (Ch)  Anchor : VTM-11-nnvc | ×2.3 ×2.3 | Test1: SuperRes CNN with const.param Test2: SuperRes CNN with **signalled** param NCTM5.0 used for CNN param compression; Simplified [ESRGAN](https://arxiv.org/pdf/1809.00219.pdf) | 0.5M | 122 |
| [JVET-W0105](https://jvet-experts.org/doc_end_user/current_document.php?id=10921) | RA(4K):**8.1**%(Y), ~20% drop (Ch) AI(4K): **11**% (Y), ~24% drop (Ch)  Anchor : VTM-11-nnvc | -  (×90 CPU) | RPR2 (↓VVC­↑) YUV444 CNN with const.param 2 models (high / low QP) Simplified [EDSR](https://openaccess.thecvf.com/content_cvpr_2017_workshops/w12/papers/Lim_Enhanced_Deep_Residual_CVPR_2017_paper.pdf) (multi-scale) | 2.7M (2×1.3M) | 344 |
| [JVET-W0125](https://jvet-experts.org/doc_end_user/current_document.php?id=10941) | RA(4K):**11.2**%(Y), ~15% drop (Ch) LD(4K):**18**%(Y),  Anchor : VTM-11-nnv | no info | DCS-based Video Coding:  Intra = spatial texture frames are coded in orignal resolution, Inter = temporal motion frames frames are coded ×2 ↓, synthesised with Motion Compensation & Texture Transfer Networks | 8.5M  (2×4.25M) | no info |
| VVC RPR cfg#1 | RA(4K):**2.3**%(Y), ~31% drop (Ch)  AI(4K):**4.3**%(Y), ~34% drop (Ch) | ×0.3  ×0.4 | All frames are coded downsampled by factor 2 vertically and horizontaly | NA | NA |
| VVC RPR cfg#3 | RA(4K):**8.2**%(Y), ~23% drop (Ch)  AI(4K):**4.7**%(Y), ~35% drop (Ch) | ×0.3  ×0.4 | Adaptive selection between full size and downsampled size coding (curves overlap) | NA | NA |
| VVC RPR cfg#4 | RA(4K):**6.1**%(Y), ~22% drop (Ch)  AI(4K):**5.7**%(Y), ~11% drop (Ch) | ×0.3  ×0.4 | Adaptive selection between full size and downsampled size coding (no curves overlap) | NA | NA |

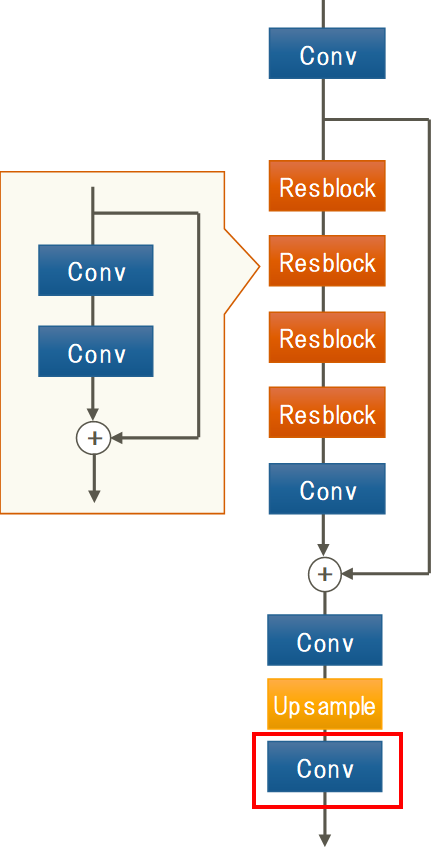


Fig. 4. NN-based Super resolution based on ESRGAN from [JVET-W0063](https://jvet-experts.org/doc_end_user/documents/23_Teleconference/wg11/JVET-W0063-v1.zip).

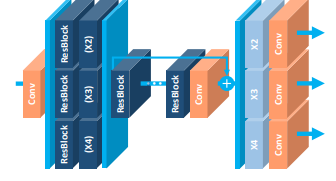


Fig. 5. NN-based Super resolution based on [EDSR](https://openaccess.thecvf.com/content_cvpr_2017_workshops/w12/papers/Lim_Enhanced_Deep_Residual_CVPR_2017_paper.pdf) from [JVET-W0105](https://jvet-experts.org/doc_end_user/current_document.php?id=10921) (Fig. from original paper profponent refers to).

图示

描述已自动生成

Fig. 6. DCS-based Video Coding from [JVET-W0125](https://jvet-experts.org/doc_end_user/current_document.php?id=10941).

Needs to be noticed that rate-distortion curves often overlaps for tests in SuperRes category, so BD-rate measurement is no 100% reliable (Fig. 7). In additional test (performed by EE coordinatos) for VVC RPR with adaptive selection betrween full size and downsampled size coding 2 variants (with and w/o rate-distortion curves overlap) were tested. Obsevred influence of rate-distoprtion curves overlap was ~2%.

|  |  |
| --- | --- |
|  |  |
| Fig. 7. Typical examples of rate-distortion curves overlap in SuperResolution category. | |

Proponents of [JVET-W0063](https://jvet-experts.org/doc_end_user/current_document.php?id=10878) and [JVET-W0105](https://jvet-experts.org/doc_end_user/current_document.php?id=10921) reported that they used BVI-DVC, proponents of [JVET-W0125](https://jvet-experts.org/doc_end_user/current_document.php?id=10941) used SJTU 4K sequences for training.

Viewing planning

Proponents were invited to suggest sequences and rate-point for viewing to demonstrate benefits of their proposal. By time of this document preparation following suggestions for the viewing have been done.

Among them are typical representative of NN based in-loop filter, NN-based super resolution and VVC RPR. All sequences are 5 -10 sec , all same resolution (4K).

In-loop filter category.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sequence, configuration | Anchor | | Test | | |
| QP | Bit-rate | Doc # | QP | Bit-rate |
| Campfilre, RA | 42 | 1651 | [JVET-W0130](https://jvet-experts.org/doc_end_user/current_document.php?id=10947) | 42 | 1651 |
| DayLightRoad, RA | 42 | 1093 | [JVET-W0130](https://jvet-experts.org/doc_end_user/current_document.php?id=10947) | 42 | 1094 |
| FoodMarket, RA | 42 | 1101 | [JVET-W0130](https://jvet-experts.org/doc_end_user/current_document.php?id=10947) | 42 | 1096 |

Super resolution category.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sequence, configuration | Anchor | | Test | | |
| QP | Bit-rate | Doc # | QP | Bit-rate |
| Campfire, RA | 45 | 1267 | [JVET-W0105](https://jvet-experts.org/doc_end_user/current_document.php?id=10921) | 37 | 1341 |
| FoodMarket, RA | 45 | 867 | [JVET-W0105](https://jvet-experts.org/doc_end_user/current_document.php?id=10921) | 37 | 925 |
| Tango, RA | 45 | 915 | [JVET-W0105](https://jvet-experts.org/doc_end_user/current_document.php?id=10921) | 37 | 928 |
| Campfilre, RA | 45 | 1267 | VVC RPR cfg#4 | 37 | 1332 |
| FoodMarket, RA | 45 | 867 | VVC RPR cfg#4 | 37 | 921 |
| Tango, RA | 45 | 915 | VVC RPR cfg#4 | 37 | 924 |

Except for JVET-W0062, all technology is more or less unchanged relative to last meeting, but reporting of results is now more consistent.

Viewing to be conducted, and prepare presentation of the results in a BoG (M. Wien, E. Alshina)

[JVET-W0062](https://jvet-experts.org/doc_end_user/current_document.php?id=10877) EE1: SSIM based CNN model for in-loop filtering [T. Ouyang, Y. Guo, Z. Chen (Wuhan Univ.), X. Xu, S. Liu (Tencent)]

This contribution reports the experimental results of Exploration Experiment 1 on Neural Network based filtering for video coding of JVET-U0074. In the test, the impact of loss function adopted during the training is investigated via the MS-SSIM results.

Two sets of experiments are set up, which are implemented on VTM-12.0. The experimental results under All Intra (AI) configuration with 5 QPs recommended by AHG11 are reported in this contribution. From which the inconsistency between the MS-SSIM and PSNR is observed. The model trained with MS-SSIM loss function is proven to have more BD-rate savings in MS-SSIM, while the L1 loss function has more gain over PSNR. The experimental results show that the model trained with MS-SSIM based loss function shows superiority in preserving subjective quality, while the model trained with L1 loss performs better in saving more BD-Rate in PSNR.

It is that SAO and ALF were disabled both in training and testing (in other loop filter proposals, these were turned on; W0105 turned deblocking off, as their loop filter also targets deblocking).

[JVET-W0063](https://jvet-experts.org/doc_end_user/current_document.php?id=10878) EE1.2.2: NN-based super resolution (JVET-V0073) [T. Chujoh, Y. Yasugi, T. Ikai (Sharp)]

[JVET-W0105](https://jvet-experts.org/doc_end_user/current_document.php?id=10921) EE1-2.3: Neural Network-based Super Resolution [A. M. Kotra, K. Reuzé, J. Chen, H. Wang, M. Karczewicz, J. Li (Qualcomm)]

[JVET-W0125](https://jvet-experts.org/doc_end_user/current_document.php?id=10941) EE1: Tests on Decomposition, Compression and Synthesis (DCS)-based Technology (JVET-V0149) [Ming Lu, Zhan Ma (NJU), Zhenyu Dai, Dong Wang (OPPO)] [late]

[JVET-W0130](https://jvet-experts.org/doc_end_user/current_document.php?id=10947) EE1-1.4: Test on Neural Network-based In-Loop Filter with Large Activation Layer [H. Wang, J. Chen, K. Reuzé, A.M. Kotra, M. Karczewicz (Qualcomm)]

[JVET-W0151](https://jvet-experts.org/doc_end_user/current_document.php?id=10981) EE1: neural network based in-loop filter using depthwise separable convolution and regular convolution [L. Wang, S. Lin, X. Xu, S. Liu, C. Auyeung, X. Li (Tencent)] [late]

### EE1 related contributions: Neural network-based video coding (6)

Contributions in this area were discussed in session 5 at 0500–0710 UTC on Thursday 8 July 2021 (chaired by JRO).

[JVET-W0131](https://jvet-experts.org/doc_end_user/current_document.php?id=10960) EE1-related: Neural Network-based in-loop filter with constrained computational complexity [H. Wang, J. Chen, K. Reuze, A.M. Kotra, M. Karczewicz (Qualcomm)]

This EE related contribution studies the performance of the network structure tested in EE (JVET-W0130) with a different complexity configuration. The proposed configuration is roughly determined based on extrapolating computational capabilities of potential future processing units. As done in the EE test, one NN-based filter can be selected out of four candidates as an in-loop filter for each picture. The deblocking boundary strength information of VVC is also used as additional input to the NN filter but the actual deblocking filtering process is bypassed. The total number of parameters in each model is ~108K, which result in roughly 18 times lower MAC/pixel compared to the models tested in EE. The simulation results reportedly show 5.78%, 6.80% and 5.03% BD rate saving for RA and 5.16%, 5.93%, 5.19% BD rate saving for AI, for Y, Cb and Cr components respectively.

Similar to EE1 1.4, but target lower complexity

Filter position before SAO, disable deblocking, use BS from deblocking as input, filtering on basis of CTU with overlap (144x144 4:2:0), also QP as input

kMAC/pixel reduced from 624 to 34

no more separate models for luma and chroma, coding gains more uniform for luma and chroma compared to EE results

Also tested 16 bit integer implementation, roughly same result (only AI finished). Not optimized yet in terms of run time.

Q: Was 8 bit tested? Not yet.

It was pointed out that an analysis of complexity vs. compression benefit of different loop filter proposals (similar as usually done for tool investigation) would be desirable.

[JVET-W0059](https://jvet-experts.org/doc_end_user/current_document.php?id=10874) AHG11: A Deep In-Loop Filter Method [X. Zhang, C. Fang, D. Jiang, J. Lin (Dahua)]

This contribution presents a convolutional neural network-based in-loop filtering method with QP based models. Compared with VTM-11.0-NNVC, the proposed method reportedly shows on average {%}, {%}, and {%} BD-rate reductions for {Y, Cb, Cr}, under AI configuration.

Separate networks for luma and chroma

Luma input size is 160x160 (CTU plus 16 samples from neighboring blocks), chroma size half (420 downsampling)

Between SAO and ALF

Number of operations is reported in GMAC/CTU (and only for luma), should be converted to KMAC/Pixel for comparability with other proposals

Only partial results available by the time of presentation:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **All Intra Main10** | | | | |
|  | **Over VTM-11.0-NNVC** | | | | |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  | #NUM! | #NUM! |
| Class A2 |  |  |  | #NUM! | #NUM! |
| Class B |  |  |  | #NUM! | #NUM! |
| Class C | -4.55% | -13.21% | -15.06% | 1731% | 412921% |
| Class E | -5.64% | -14.76% | -15.05% | 4044% | 618997% |
| **Overall** |  |  |  | #NUM! | #NUM! |
| Class D | -5.23% | -12.27% | -15.56% | 2751% | 303132% |
| Class F | 1.45% | -2.37% | 2.54% | 2956% | 609187% |

[JVET-W0099](https://jvet-experts.org/doc_end_user/current_document.php?id=10915) AHG11: CNN-based Super Resolution for Video Coding Using Decoded Information [C. Lin, L. Zhang, K. Zhang, Y. Li (Bytedance)]

This contribution presents a convolutional neural network-based super-resolution for video coding. In the proposed design, the decoded information is fed into the up-sampling network. In addition, different up-sampling networks are designed for the luma and chroma components. Compared with VTM-11.0 with new MCTF enabled, the proposed method reportedly shows on average {-12.00%, 14.12%, -4.78%} BD-rate changes for {Y, U, V} components under all intra configurations, respectively.

It is proposed to feed the decoded information into the up-sampling network and separate the networks for luma and chroma components. To be more specific, for the luma up-sampling network, the sequence-level QP and the prediction signal are used as auxiliary information, while for the chroma up-sampling network, the luma information (not upsampled) is utilized to boost the performance of the up-sampling CNN for chroma components.

For RA, no results yet

The prediction signal might require additional buffer.

Roughly 5.8 MMAC/pixel

The chroma loss observed in all other SR contributions is somewhat decreased. Results show that the chroma loss/gain is quite inhomogeneous over sequences (>100% loss for U component of Campfire, for example, whereas other sequences have decent chroma gain).

[JVET-W0100](https://jvet-experts.org/doc_end_user/current_document.php?id=10916) AHG11: Deep In-Loop Filter with Adaptive Model Selection and External Attention [Y. Li, K. Zhang, L. Zhang (Bytedance)]

This contribution presents a convolutional neural network-based in-loop filtering method wherein adaptive model selection is introduced. The proposed Deep in-loop filter with Adaptive Model selection (DAM) method is developed from the prior contribution JVET-V0100, introducing a new network structure to the code base of VTM-11.0+NewMCTF. Similar to JVET-V0100, residual blocks are utilized as the basic module and stacked several times to construct the final network. As a further improvement from JVET-V0100, external attention mechanism is introduced in this contribution, leading to an increased representation capability with a similar model size. In addition, to deal with different types of content, individual networks are trained for different types of slices and quality levels. Compared with VTM-11.0+NewMCTF, the proposed method reportedly shows on average {9.12%, 22.39%, 22.60%}, {12.32%, 27.48%, and 27.22%}, and {10.61%, 28.19%, and 27.91%} BD-rate reductions for {Y, Cb, Cr}, under AI, RA, and LDB configurations, respectively.

Model selection at slice or CTU level

SAO and deblocking disabled/replaced

Attention map is computed from prediction and residual block

1.43 MMACs/pixel (more complex than the highest in EE, but also better performance); CPU decoder run time increased by approximately 1000x

A reduced complexity version is also presented.

It is pointed out that the above analysis seems to be only for luma, chroma should be included.

It is pointed out that the models are switched at finer granularity (32x32), which also has complexity impact in terms of reloading model parameters.

24 models in total, but at 32x32 granularity only 3 can be selected from a candidate list (based on QP).

In training, same augmentations of original blocks were used as in previous proposal V0100 (rotation, flip)

[JVET-W0101](https://jvet-experts.org/doc_end_user/current_document.php?id=10917) AHG11 & AHG12: Deep In-Loop Filter with Adaptive Model Selection and External Attention for Enhanced Compression Beyond VVC Capability [Y. Li, K. Zhang, L. Zhang (Bytedance)]

This contribution evaluates the deep in-loop filter presented in JVET-W0100 on top of ECM-1.0. Compared with ECM-1.0, the proposed method reportedly shows on average {6.74%, 19.53%, 20.16%}, {9.60%, 23.58%, 23.59%}, and {9.00%, 24.34%, 23.05%} BD-rate reductions for {Y, Cb, Cr}, under AI, RA, and LDB configurations, respectively. Compared with VTM-11.0 + new MCTF, a combination of the NN-based coding methods and the non-NN-based coding methods beyond VVC reportedly shows on average {12.16%, 26.17%, 26.68%}, {20.91%, 33.64%, 34.06%}, and {17.81%, 31.59%, 30.67%} BD-rate reductions for {Y, Cb, Cr}, under AI, RA, and LDB configurations, respectively.

SAO and deblocking are replaced, ALF and bilateral filter are still enabled. Deep filter before bilateral filter.

[JVET-W0132](https://jvet-experts.org/doc_end_user/current_document.php?id=10961) AHG11: 1.5x/2.0x Upsample method for NN-Based Super-Resolution Post-Filters [Y. Yasugi, T. Chujoh, T. Ikai (Sharp)]

This contribution presents several experimental results of NN-based super-resolution post-filter. In VVC, RPR (Reference Picture Re-sampling) has been introduced, and in several 4K sequences with low bit-rate, it was reported that there are some coding gains by changing the resolution of the whole sequence. In this experiment, the performance of NN-based filters with different network structures are compared. As shown in experimental results, the use of interpolation for upscaling in feature space can reduce the number of parameters more than 30% without significant loss of performance.

The super-resolution 1.5x post-filter used in this experiment is based on a simplified ESRGAN

- Configuration A: 2x, PixelShuffle

- Configuration B: 2x, Bicubic interpolation(2.0x)

- Configuration C: 1.5x, PixelShuffle + Bicubic interpolation(0.75x)

- Configuration D: 1.5x, Bicubic interpolation(1.5x)

The number of parameters for configurations A-D were 483597, 335884, 483597, and 335884, respectively.

**Table 1 Experimental results of configuration A (2.0x, PixelShuffle)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Random access Main10** | | | | |
|  | **BD-rate Over VTM-11.0&JVET-V0056 RPR20 QP=22,...,42** | | | | |
|  | Y-PSNR | U-PSNR | V-PSNR | EncT | DecT |
| Class A1 | -1.81% | -2.83% | -4.56% | 100% | 100% |
| Class A2 | -6.31% | -4.49% | -3.25% | 100% | 100% |

**Table 2 Experimental results of configuration B (2.0x, Bicubic interpolation)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Random access Main10** | | | | |
|  | **BD-rate Over VTM-11.0&JVET-V0056 RPR20 QP=22,...,42** | | | | |
|  | Y-PSNR | U-PSNR | V-PSNR | EncT | DecT |
| Class A1 | -1.64% | -2.13% | -4.67% | 100% | 100% |
| Class A2 | -6.03% | -4.68% | -2.08% | 100% | 100% |

**Table 3 Experimental results of configuration C (1.5x, PixelShuffle+Bicubic interpolation)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Random access Main10** | | | | |
|  | **BD-rate Over VTM-11.0&JVET-V0056 QP=27,...,47** | | | | |
|  | Y-PSNR | U-PSNR | V-PSNR | EncT | DecT |
| Class A1 | -1.43% | 1.76% | -0.12% | 100% | 100% |
| Class A2 | -3.33% | 0.95% | 2.74% | 100% | 100% |

**Table 4 Experimental results of configuration D (1.5x, Bicubic interpolation)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Random access Main10** | | | | |
|  | **BD-rate Over VTM-11.0&JVET-V0056 QP=27,...,47** | | | | |
|  | Y-PSNR | U-PSNR | V-PSNR | EncT | DecT |
| Class A1 | -1.71% | 0.67% | -1.62% | 100% | 100% |
| Class A2 | -3.39% | 0.99% | 3.03% | 100% | 100% |

Question: Why is there slight loss for 2x, slight gain for 1.5x? May need further analysis

The behaviour may also be different as different QP ranges are used for th two cases.

Question: Do the PSNR/rate graphs cross? Needs to be checked

Results comparing against full resolution anchor (as in EE CTC), and also other classes would be interesting to compare against other proposals.

BoG (A. Segall) to perform further analysis about the complexity/compression tradeoffs of the various loop filter and super resolution proposals, align the reports with AHG11 reporting conditions, and refine as necessary. The unbalance between luma and chroma gains should also be further analysed in detail. Also suggest candidates for upcoming EE.

### Other (4)

Contributions in this area were discussed in session 14 at 2150–2300 UTC on Monday 12 July 2021 (chaired by XXX).

[JVET-W0057](https://jvet-experts.org/doc_end_user/current_document.php?id=10872) AHG11: Content-adaptive neural network post-processing filter [M. Santamaria, Y.-H. Lam, J. Lainema, F. Cricri, R. Ghaznavi-Youvalari, H. Zhang, A. Zare, H. R. Tavakoli, M. Hannuksela (Nokia)]

This document proposes a content-adaptive post-processing filter based on Convolutional Neural Networks (CNNs). The work is based on the filter described earlier in JVET-V0075 by the same authors. The proposed filter is first trained offline on a large dataset, and then finetuned for each test video sequence. Each slice includes a scaling factor selected by the encoder and signalled in the bitstream. This scaling factor controls what portion of the filter’s impact is applied to the output. There is further Coding Tree Unit (CTU) level control allowing the encoder to switch the filter on and off separately for luma and chroma. The bitrate overhead caused by finetuned Neural Network (NN) is considered into the BD-rate calculations. The proposed filter was implemented on top of VVC Test Model (VTM) 11.0 and evaluated under Random Access (RA) configuration. It is reported that the proposed content-adaptive method outperforms both VTM 11.0 and the filter described earlier in JVET-V0075.

Roughly 3% Y bit rate reduction for classes B/C/D, more for U and V

Comparable in terms of complexity with the “lowest complexity” ones from EE (in terms of parameters), but decoder run time quite large (670x for class B)

Bias weight update roughly 1.3 … 2.3 kbit/s

Investigate in EE, in particular the impact of sequence adaptation (bias value)

[JVET-W0081](https://jvet-experts.org/doc_end_user/current_document.php?id=10897) AHG11: BD-rate gains vs complexity of NN-based intra prediction [T. Dumas, F. Galpin, P. Bordes, F. Le Léannec (InterDigital)]

This contribution presents a low complexity version of the neural network-based intra prediction proposed in last meeting.

Mean BD-rate reductions of -2.88% -2.33% -2.42% and -1.44% -0.52% -0.97% in AI and RA configurations respectively are reported when comparing VTM-11.0 including the low complexity version of the neural network-based intra prediction mode w.r.t VTM-11.0 (not VTM-11.0-NNVC). The complexity (in MAC/pixel) of the low complexity version of the neural network-based intra prediction mode is about 12 times smaller than its regular version.

Low complexity via sparse training of weights (enforcing zero weights)

8 different networks for different block sizes

Question: For which block sizes most efficient? Medium size, such as 8x8

Could this approach of complexity reduction also be applied to CNN? Likely would not work, as properties of filters, and their effect on dedicated spatial positions might get lost.

How is the training for sparsity implemented? How to deal with the fact that the non-zero weights could be randomly positioned in the network? Try to start at certain positions, and increase threshold gradually.

Also investigated version with 16 bit integer, almost no difference in results

Investigate in EE.

[JVET-W0111](https://jvet-experts.org/doc_end_user/current_document.php?id=10927) AHG11: neural network based cross-component prediction model [L. Wang, S. Lin, R. Chang, X. Xu, S. Liu (Tencent)]

This contribution presents a neural network based cross-component prediction (NNCCP) model. As an alternative mode of intra chroma prediction, NNCCP can further reduce the cross-component redundancy. Based on the AHG11 anchor (VTM11+V0056), it is reported that -0.34%, -0.78% and -0.35% BD-rate savings are achieved for Y, U and V components with the AI configuration, respectively.

Input is approximately same as in conventional CCLM

One bit used for signalling. Has it been tried to replace CCLM rather than adding an additional mode?

Network size approx. 166 kMAC/pixel

Only used for 8x8, 16x16 and 32x32 blocks

Further study recommended. Currently gain is relatively low compared to complexity.

[JVET-W0113](https://jvet-experts.org/doc_end_user/current_document.php?id=10929) AHG11: neural network based in-loop filter [L. Wang, W. Jiang, X. Xu, S. Liu (Tencent)]

In this contribution, a neural network based in-loop filter is exploited to replace the Deblocking and SAO modules. Compared with EE1 anchor (VTM11+V0056), the proposed method reports {4.73%, 10.73, 11.51%} and {3.54%, 10.99%, 10.24%} BD-rate savings with AI and RA configurations, respectively.

In order to improve the quality of reconstructed image, neural network models are trained for I slice and B slice, separately. In the in-loop filter process, Deblock and SAO are replaced by the proposed NN filter.

Operated before ALF

Similar architecture as in W0111, but larger (621 kMAC/pixel)

Performance is better for AI than RA, which is different from other proposals.

An interesting aspect could be the 5x5 convolution in the parallel path. Would be interesting if that is an element which gives higher intra gain.

Investigate in EE.

### NN related HLS signalling (0)

Contributions in this area were discussed in session X at XXXX–XXXX UTC on XXday X July 2021 (chaired by XXX).

## AHG12: Enhanced compression beyond VVC capability (30)

### General (2)

Contributions in this area were discussed in session 22 at 0610–0700 UTC on Thursday 15 July 2021 (chaired by JRO).

[JVET-W0049](https://jvet-experts.org/doc_end_user/current_document.php?id=10864) AHG12: on the status of the ECM software [Y. Ye (Alibaba), M. Karczewicz (Qualcomm), Y.-W. Huang (MediaTek), P. Yin (Dolby), D. Wang (OPPO), X. Wang (Kwai), J. Ström (Ericsson), F. Le Leannec (InterDigital), L. Zhang (Bytedance), S.-H. Kim (LGE), M. Hannuksela (Nokia), P. Wu (ZTE)]

This contribution proposes to establish the ECM software branch as an independent software base decoupled from the VTM and accompanied by an algorithm description output document. It is asserted that such an arrangement would follow a similar process for the ECM as was done in the past with the JEM work. In view of the proponents, the JEM work fostered significant progress in improving coding efficiency and broadening functionality beyond HEVC’s capability, and was instrumental to jumpstarting the successful development and finalization of the VVC standard. It is proposed that the ECM should follow a similar arrangement, such that it could serve to provide a common platform focused on coding technologies that could lead to a next generation standard beyond VVC down the road.

From the discussion:

It is noted that ECM or parts thereof might also end up in an extension of VVC. This cannot be foreseen at the current time.

The exploration activity is about exploring capabilities for higher compression. It is part of this activity to find out how much is possible with backward capability to VVC, without restricting to that.

Actually, ECM is already in a different repository. The change would be to establish ECM as a master branch.

ECM is based on VTM10. It is a matter of practicality not following each change in VTM. However, an important aspect would be that encoder improvements would likewise be applied to ECM and the VTM anchor compared against (which should be the most recent VTM), and obvious bug fixes found in VTM should also be fixed in ECM.

Basically, the current practice in the EE has already been somewhat like that.

It is agreed that this is an appropriate move. Also, the suggestion of establishing an algorithm description is agreed.

[JVET-W0102](https://jvet-experts.org/doc_end_user/current_document.php?id=10918) Preliminary draft of algorithm description for Enhanced Compression Model 1 Software (ECM 1) [M. Coban (Qualcomm), F. Le Léannec (InterDigital), J. Ström (Ericsson), Y. Ye (Alibaba)]

JVET-W0049 proposes to establish the ECM software branch as an independent software base decoupled from the VTM and accompanied by an algorithm description output document. To facilitate this effort, in this contribution, proponents of the coding tools in ECM 1 provide a preliminary draft of the suggested algorithm description output document.

It is pointed out that some encoder speedups of ECM are not described, which would be desirable to be added.

Decision: Use JVET-W0102 as basis for ECM algorithm description (new output document). This should furthermore include the new adoption for ECM2, and appropriate description of encoder optimizations (different section) as necessary for understanding the software.

### EE2 contributions: Enhanced compression beyond VVC capability (12)

Contributions in this area were discussed in session 6 at 0730–0940 UTC on Thursday 8 July 2021 (chaired by JRO).

[JVET-W0024](https://jvet-experts.org/doc_end_user/current_document.php?id=10988) EE2: Summary Report on Enhanced Compression beyond VVC capability [V. Seregin, J. Chen, S. Esenlik, F. Le Léannec, L. Li, J. Ström, M. Winken, X. Xiu, K. Zhang]

This document provides a summary report of Exploration Experiment on Enhanced Compression beyond VVC capability. The tests are categorized as partitioning, intra prediction, inter prediction, transform, in-loop filtering, and combination tests.

The software basis for this EE is ECM-1.0, released at <https://vcgit.hhi.fraunhofer.de/ecm/VVCSoftware_VTM/-/tags/ECM-1.0>

Software for EE tests is released in the corresponding branches at <https://vcgit.hhi.fraunhofer.de/ecm/jvet-v-ee2/VVCSoftware_VTM/-/branches>

ECM-1.0 is used as an anchor in the (tool-off) tests, and VTM-11.0 with the updated MCTF from JVET-V0056 is used as an anchor in the tool-on tests.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Tests** | **Tester** | **Cross-checker** |
| **1 Partitioning** | | | |
| 1.1 | ABT | InterDigital  [JVET-W0084](https://jvet-experts.org/doc_end_user/current_document.php?id=10900) |  |
| 1.2 | ABT with non-normative restrictions on partition splits | InterDigital  [JVET-W0084](https://jvet-experts.org/doc_end_user/current_document.php?id=10900) |  |
| 1.3.a | UQT, high efficiency configuration | Bytedance  [JVET-W0086](https://jvet-experts.org/doc_end_user/current_document.php?id=10902) | InterDigital  [JVET-W0137](https://jvet-experts.org/doc_end_user/current_document.php?id=10966) |
| 1.3.b | UQT, low complexity configuration | Bytedance  [JVET-W0086](https://jvet-experts.org/doc_end_user/current_document.php?id=10902) | InterDigital  [JVET-W0137](https://jvet-experts.org/doc_end_user/current_document.php?id=10966) |
| 1.4.a | UQT+UBT, high efficiency configuration | Bytedance  [JVET-W0086](https://jvet-experts.org/doc_end_user/current_document.php?id=10902) | InterDigital  [JVET-W0137](https://jvet-experts.org/doc_end_user/current_document.php?id=10966) |
| 1.4.b | UQT+UBT, low complexity configuration | Bytedance  [JVET-W0086](https://jvet-experts.org/doc_end_user/current_document.php?id=10902) | InterDigital  [JVET-W0137](https://jvet-experts.org/doc_end_user/current_document.php?id=10966) |
| 1.5a | UQT+ABT (based on test 1.4) | Bytedance, Interdigital  [JVET-W0087](https://jvet-experts.org/doc_end_user/current_document.php?id=10903) | Qualcomm  [JVET-W0171](https://jvet-experts.org/doc_end_user/current_document.php?id=11002) |
| 1.5b | ABT (based on test 1.4) | Bytedance, Interdigital  [JVET-W0087](https://jvet-experts.org/doc_end_user/current_document.php?id=10903) | Qualcomm  [JVET-W0171](https://jvet-experts.org/doc_end_user/current_document.php?id=11002) |
| **2 Intra prediction** | | | |
| 2.1 | TIMD | Bytedance  [JVET-W0053](https://jvet-experts.org/doc_end_user/current_document.php?id=10868) | Qualcomm  [JVET-W0157](https://jvet-experts.org/doc_end_user/current_document.php?id=10987) |
| 2.2 | Additional blending DIMD modes | LGE  [JVET-W0054](https://jvet-experts.org/doc_end_user/current_document.php?id=10869) |  |
| **3 Inter prediction** | | | |
| 3.1 | ARMC with template matching | Bytedance  [JVET-W0090](https://jvet-experts.org/doc_end_user/current_document.php?id=10906) | Alibaba  [JVET-W0144](https://jvet-experts.org/doc_end_user/current_document.php?id=10973) |
| 3.2 | ARMC with bilateral matching instead of TM | Bytedance  [JVET-W0090](https://jvet-experts.org/doc_end_user/current_document.php?id=10906) | Alibaba  [JVET-W0144](https://jvet-experts.org/doc_end_user/current_document.php?id=10973) |
| 3.3 | GPM with MMVD | Bytedance, Kwai  [JVET-W0088](https://jvet-experts.org/doc_end_user/current_document.php?id=10904) | Qualcomm  [JVET-W0153](https://jvet-experts.org/doc_end_user/current_document.php?id=10983) |
| 3.4 | GPM with template matching | Alibaba  [JVET-W0065](https://jvet-experts.org/doc_end_user/current_document.php?id=10880) | Kwai  [JVET-W0158](https://jvet-experts.org/doc_end_user/current_document.php?id=10989) |
| 3.5 | GPM with template matching and CU-level signaling | Qualcomm  [JVET-W0065](https://jvet-experts.org/doc_end_user/current_document.php?id=10880) | Kwai  [JVET-W0158](https://jvet-experts.org/doc_end_user/current_document.php?id=10989) |
| 3.7 | GPM with MMVD 3.3 and MV list construction from 3.5 | Bytedance, Kwai, Qualcomm  [JVET-W0089](https://jvet-experts.org/doc_end_user/current_document.php?id=10905) | Alibaba  [JVET-W0143](https://jvet-experts.org/doc_end_user/current_document.php?id=10972) |
| 3.6 | Extension of template matching to CIIP, GPM, Affine merge modes, and boundary sub-blocks | Qualcomm | withdrawn |
| 3.8 | Combination of 3.4 and 3.5  (GPM TM) | Alibaba, Qualcomm  [JVET-W0065](https://jvet-experts.org/doc_end_user/current_document.php?id=10880) | Kwai  [JVET-W0158](https://jvet-experts.org/doc_end_user/current_document.php?id=10989) |
| 3.9 | Combination of 3.3 and 3.8 (GPM MMVD and GPM TM), encoder option #1 | Kwai, Alibaba, Bytedance, Qualcomm  [JVET-W0097](https://jvet-experts.org/doc_end_user/current_document.php?id=10913) | InterDigital |
| 3.10 | Combination of 3.3 and 3.8 (GPM MMVD and GPM TM), encoder option #2 | Kwai, Alibaba, Bytedance, Qualcomm  [JVET-W0097](https://jvet-experts.org/doc_end_user/current_document.php?id=10913) | InterDigital |
| **4 Transform** | | | |
| 4.1 | Extended intra MTS | Qualcomm  [JVET-W0103](https://jvet-experts.org/doc_end_user/current_document.php?id=10919) | LGE  [JVET-W0164](https://jvet-experts.org/doc_end_user/current_document.php?id=10995)  Tencent  [JVET-W0170](https://jvet-experts.org/doc_end_user/current_document.php?id=11001) |
| 4.2 | Extended LFNST | Qualcomm  [JVET-W0103](https://jvet-experts.org/doc_end_user/current_document.php?id=10919) | LGE  [JVET-W0164](https://jvet-experts.org/doc_end_user/current_document.php?id=10995) |
| 4.3 | Extended LFNST - method 2 | Qualcomm | withdrawn |
| 4.4 | Extended intra MTS + extended LFNST | Qualcomm  [JVET-W0103](https://jvet-experts.org/doc_end_user/current_document.php?id=10919) | LGE  [JVET-W0164](https://jvet-experts.org/doc_end_user/current_document.php?id=10995) |
| 4.5 | Extended LFNST with large kernel – method A | LGE  [JVET-W0119](https://jvet-experts.org/doc_end_user/current_document.php?id=10935) | Qualcomm  [JVET-W0120](https://jvet-experts.org/doc_end_user/current_document.php?id=10936) |
| 4.6 | Extended LFNST with large kernel – method B | LGE  [JVET-W0119](https://jvet-experts.org/doc_end_user/current_document.php?id=10935) | Qualcomm  [JVET-W0120](https://jvet-experts.org/doc_end_user/current_document.php?id=10936) |
| 4.7 | Replace the existed LFNST with the proposed method A | LGE  [JVET-W0119](https://jvet-experts.org/doc_end_user/current_document.php?id=10935) | Qualcomm  [JVET-W0120](https://jvet-experts.org/doc_end_user/current_document.php?id=10936) |
| 4.8 | Replace the existed LFNST with the proposed method B | LGE  [JVET-W0119](https://jvet-experts.org/doc_end_user/current_document.php?id=10935) | Qualcomm  [JVET-W0120](https://jvet-experts.org/doc_end_user/current_document.php?id=10936) |
| 4.9 | Combination of 4.7 and 4.1 | LGE  [JVET-W0119](https://jvet-experts.org/doc_end_user/current_document.php?id=10935) |  |
| 4.10 | Combination of 4.8 and 4.1 | LGE  [JVET-W0119](https://jvet-experts.org/doc_end_user/current_document.php?id=10935) |  |
| **5 In-loop filtering** | | | |
| 5.1a | Cross-component SAO with separate clipping | Kwai  [JVET-W0066](https://jvet-experts.org/doc_end_user/current_document.php?id=10881) | Qualcomm  [JVET-W0146](https://jvet-experts.org/doc_end_user/current_document.php?id=10975) |
| 5.1b | Cross-component SAO with joint clipping | Kwai  [JVET-W0066](https://jvet-experts.org/doc_end_user/current_document.php?id=10881) | Qualcomm  [JVET-W0146](https://jvet-experts.org/doc_end_user/current_document.php?id=10975) |

Results on partitioning:



Not all results available yet

From the available results, gain ranges between 0.2% (for 16% enc runtime increase) and 1% (for 70% runtime increase) on top of ECM1 (tool-off test), in RA configuration. These do not appear reasonable tradeoffs

1.5b is an approach to reduce encoder run time, increasing 15-20% encoding time for 0.7% bit rate reduction.

It is noted that for some cases the runtime is lower than the anchor. Would similar encoder speedups be applicable to MTT of anchor? Seems to be the case, so this should be considered in the comparison.

It is remarked that typically (in past codec development efforts) the gain due to new partitioning schemes was significantly larger.

Should this be considered as an extension of the existing partitioning scheme? Then it might be looked at (in terms of benefit) as any other tool.

Further study of 1.5b in EE, using similar methods of encoder optimization for the ECM anchor.

Complete results of 1.5a and 1.5b were reviewed later (Wed. 15). It was concluded that both methods are interesting to investigate in next EE, with different tradeoffs complexity vs. bitrate gain. Similar complexity points should be compared.

Results on intra prediction:



2.2 does not provide sufficient gain for justifying the additional complexity

It was pointed out that 2.1 has interesting gain (considering that the ECM in AI only gains 5.6% over VTM currently), but also has considerable run time increase. It is mentioned that EE related proposals (e.g., W0123 which extends 2.1) which have a better tradeoff. As later JVET-W0123 was adopted (which is a variant of CE2-2.1 with additional weighting), no further decision is necessary here.

Question: What would be the benefit of 2.1 without doubling the number of angles? Gain would be lower by roughly 0.1% in AI, with minor impact on run time.

Results on inter prediction:



Decision: Adopt JVET-W0090 ARMC with template matching, version EE-3.1 to ECM 2

Decision: Adopt JVET-W0097 encoder option 2 (EE-3.10) to ECM 2

Results on transform:



For 4.1, the contribution reports about supplemental results with an additional encoder optimization without runtime increase in RA.

LFNST large kernel method B uses longer kernel than method A. There is already a modification of LFNST in the ECM, but the proposed methods (in particular method B) seem to provide better compression, and decrease encoder runtime (might be more complex in hardware implementation, though).

4.7 and 4.8 also change the LFNST signalling and encoder, which however does not seem to be beneficial when comparing 4.5/4.7, and 4.6/4.8. Complexity in terms of transform computation is identical.

Among the LFNST proposals, 4.6 provides the best tradeoff.

4.1 also provides reasonable tradeoff compression vs. complexity, and it seems to be additive with the LFNST gains (as can be seen from the combinations, e.g. 4.10).

Decision: Adopt JVET-W0103 (EE 4.1)

Decision: Adopt JVET-W0119 (EE 4.6)

Results on in-loop filtering



Generally, gain in chroma. In AI, slight loss in luma. It is asserted by the proponents that this is due to some additional signalling. Also in RA, various sequences should have luma losses, as Campfire has luma gain, and on average the luma gain is zero.

5.2 (joint clipping) is somewhat simpler implementation-wise (parallelizable).

Cross-checker reports having seen slight subjective improvements in Campfire (which is the sequence with most objective gain). No subjective difference in other sequences.

If some of the gain would be shifted from chroma to luma, gain is interesting in particular for LDB and RA.

Decision: Adopt JVET-W0066 (method with joint clipping called 5.1b in the EE report identical to “5.2” in the table above).

[JVET-W0053](https://jvet-experts.org/doc_end_user/current_document.php?id=10868) EE2-2.1: Results for template-based intra mode derivation using MPMs [Y. Wang, L. Zhang, K. Zhang, Z. Deng, N. Zhang (Bytedance)]

[JVET-W0157](https://jvet-experts.org/doc_end_user/current_document.php?id=10987) Crosscheck of JVET-W0053 (EE2-2.1: Results for template-based intra mode derivation using MPMs) [K. Cao (Qualcomm)] [late]

[JVET-W0054](https://jvet-experts.org/doc_end_user/current_document.php?id=10869) EE2 Test 2.2: DIMD with multiple blending modes [J. Zhao, S. Paluri, S. Kim (LGE)]

[JVET-W0065](https://jvet-experts.org/doc_end_user/current_document.php?id=10880) EE2: Results of Test 3.4 and Test 3.5 [R.-L. Liao, Y. Ye, J. Chen, X. Li (Alibaba), Y.-J. Chang, H. Huang, V. Seregin, C.-C. Chen, M. Karczewicz (Qualcomm)]

[JVET-W0158](https://jvet-experts.org/doc_end_user/current_document.php?id=10989) Crosscheck of JVET-W0065: EE2: Results of Test 3.4 and Test 3.5 [X. Xiu (Kwai)] [late]

[JVET-W0066](https://jvet-experts.org/doc_end_user/current_document.php?id=10881) EE2-5.1: Cross-component Sample Adaptive Offset [C.-W. Kuo, X. Xiu, Y.-W. Chen, H.-J. Jhu, W. Chen, X. Wang (Kwai)]

[JVET-W0146](https://jvet-experts.org/doc_end_user/current_document.php?id=10975) Crosscheck of JVET-W0066 (EE 2-5.1: Cross-Component Sample Adaptive Offset) [A. M. Kotra (Qualcomm)] [late]

[JVET-W0084](https://jvet-experts.org/doc_end_user/current_document.php?id=10900) EE2-1.1 and EE2-1.2: Asymmetric Binary Tree partitioning [F. Le Léannec, K. Naser, T. Dumas, A. Robert, F. Galpin, E. François (InterDigital)]

[JVET-W0086](https://jvet-experts.org/doc_end_user/current_document.php?id=10902) EE2-1.3/EE2-1.4: Unsymmetric partitioning methods in video coding [K. Zhang, L. Zhang, Z. Deng, N. Zhang, Y. Wang (Bytedance)]

[JVET-W0137](https://jvet-experts.org/doc_end_user/current_document.php?id=10966) Cross-check of JVET-W0086 "EE2-1.3/EE2-1.4: Unsymmetric partitioning methods in video coding" [F. Le Léannec (InterDigital)] [late]

[JVET-W0087](https://jvet-experts.org/doc_end_user/current_document.php?id=10903) EE2-1.5: A combining test of EE2-1.2 and EE2-1.4b [K. Zhang, L. Zhang, Z. Deng, N. Zhang, Y. Wang (Bytedance), F. Le Léannec, K. Naser, T. Dumas, A. Robert, F. Galpin, E. François (InterDigital)]

[JVET-W0171](https://jvet-experts.org/doc_end_user/current_document.php?id=11002) Crosscheck of JVET-W0087 (EE2-1.5: A combining test of EE2-1.2 and EE2-1.4b) [V. Seregin (Qualcomm)] [late]

[JVET-W0088](https://jvet-experts.org/doc_end_user/current_document.php?id=10904) EE2-3.3: GPM with MMVD (JVET-V0103 and JVET-V0125) [Z. Deng, K. Zhang, L. Zhang, N. Zhang, Y. Wang (Bytedance), X. Xiu, C.-W. Kuo, X. Wang (Kwai)]

[JVET-W0153](https://jvet-experts.org/doc_end_user/current_document.php?id=10983) Crosscheck of JVET-W0088 (EE2-3.3: GPM with MMVD) [Y.-J. Chang, H. Huang (Qualcomm)] [late]

[JVET-W0166](https://jvet-experts.org/doc_end_user/current_document.php?id=10997) Crosscheck of JVET-W0088 (EE2-3.3: GPM with MMVD) [R.-L. Liao (Alibaba)] [late]

[JVET-W0089](https://jvet-experts.org/doc_end_user/current_document.php?id=10905) EE2: Combined test of EE2-3.3 and EE2-3.5 [Z. Deng, K. Zhang, L. Zhang, N. Zhang, Y. Wang (Bytedance), X. Xiu, C.-W. Kuo, X. Wang (Kwai), Y.-J Chang, H. Huang, V. Seregin, M. Karczewicz (Qualcomm)]

[JVET-W0143](https://jvet-experts.org/doc_end_user/current_document.php?id=10972) Crosscheck of JVET-W0089 (EE2: Combined test of EE2-3.3 and EE2-3.5) [R.-L. Liao (Alibaba)] [late]

[JVET-W0090](https://jvet-experts.org/doc_end_user/current_document.php?id=10906) EE2-3.1/EE2-3.2: Adaptive Reordering of Merge Candidates with Template/Bilateral Matching [N. Zhang, K. Zhang, L. Zhang, H. Liu, Z. Deng, Y. Wang (Bytedance)]

[JVET-W0144](https://jvet-experts.org/doc_end_user/current_document.php?id=10973) Crosscheck of JVET-W0090 (EE2-3.1/EE2-3.2: Adaptive Reordering of Merge Candidates with Template/Bilateral Matching) [R.-L. Liao (Alibaba)] [late]

[JVET-W0103](https://jvet-experts.org/doc_end_user/current_document.php?id=10919) EE2: Enhanced intra MTS and LFNST (tests 4.1, 4.2, and 4.4) [B. Ray, M. Coban, V. Seregin, H. Egilmez, M. Karczewicz (Qualcomm)]

[JVET-W0164](https://jvet-experts.org/doc_end_user/current_document.php?id=10995) Crosscheck of JVET-W0103 (EE2: Enhanced intra MTS and LFNST (tests 4.1, 4.2, and 4.4)) [M. Koo (LGE)] [late]

[JVET-W0170](https://jvet-experts.org/doc_end_user/current_document.php?id=11001) Crosscheck of JVET-W0103 (EE2: Enhanced intra MTS and LFNST (Test 4.1)) [Cheung Auyeung (Tencent)] [late]

[JVET-W0119](https://jvet-experts.org/doc_end_user/current_document.php?id=10935) EE2: LFNST extension with large kernel (tests 4.5, 4.6, 4.7, and 4.8) [M. Koo, J. Zhao, J. Lim, S. Kim (LGE)]

[JVET-W0120](https://jvet-experts.org/doc_end_user/current_document.php?id=10936) Crosscheck of JVET-W0119 (EE2: LFNST extension with large kernel (tests 4.5, 4.6, 4.7, and 4.8)) [M. Coban (Qualcomm)] [late]

### EE2 related contributions: Enhanced compression beyond VVC capability (11)

Contributions in this area were discussed in session 10 at 0745–0920 UTC on Friday 9 July 2021, and in session 15 2325–0045 UTC on Monday 12 July 2021 (chaired by JRO).

[JVET-W0067](https://jvet-experts.org/doc_end_user/current_document.php?id=10882) EE2-related: Implicit derivation of DIMD blend modes [X. Li, R.-L. Liao, J. Chen, Y. Ye (Alibaba)]

In this contribution, a method of implicitly deriving the DIMD blend mode is proposed. In the multiple DIMD blend modes method which is studied in EE2 test 2.2, three DIMD blend modes are allowed and a DIMD index is signaled in the bitstream to indicate which DIMD blend mode is to be used. It is proposed to use the DIMD histogram information to implicitly select two of the three DIMD blend modes. It is reported that on top of ECM-1.0, the overall coding performance impact for {Y, U, V, EncT, DecT} is {-0.10%, 0.00%, -0.03%, 107%, 101%} for AI and {-0.05%, -0.05%, -0.06%, 101%, 100%} for RA.

The encoding time is decreased, but additional processing is necessary at the decoder (up to 67 additions in worst case).

It is asked if an alternative could be avoiding RDO decisions at the encoder.

Further study recommended.

[JVET-W0155](https://jvet-experts.org/doc_end_user/current_document.php?id=10985) Crosscheck of JVET-W0067 (EE2-related: Implicit derivation of DIMD blend modes) [Y. Wang (Bytedance)] [late]

[JVET-W0068](https://jvet-experts.org/doc_end_user/current_document.php?id=10883) EE2-related: A combination of CIIP and DIMD/TIMD [X. Li, R.-L. Liao, J. Chen, Y. Ye (Alibaba)]

In CIIP mode, the prediction samples are generated by weighting an inter prediction signal predicted using regular merge mode and an intra prediction signal predicted using planar mode. In this contribution, it is proposed to replace planar mode in CIIP with other intra prediction mode which is implicitly derived by DIMD or TIMD method. When using DIMD to derive the intra prediction mode, it is reported that on top of ECM-1.0, the overall coding performance impact for {Y, U, V, EncT, DecT} is {-0.05%, -0.11%, -0.15%, 100%, 100%} for RA and {-0.19%, -0.35%, -0.06%, 100%, 101%} for LDB. On the other hands, when using TIMD to derive the intra prediction mode, the overall coding performance impact over ECM-1.0+EE2 tset2.1(TIMD) is {-0.06%, -0.08%, -0.06%, 101%, 100%} for RA and {-0.21%, -0.18%, -0.20%, 101%, 101%} for LDB.

TIMD mode is mapped to one of the “normal” 67 intra prediction modes, which is then used instead of planar in CIIP

Interesting gain in LDB, with almost no increase of encoder/decoder run time.

Possible candidate for next EE.

[JVET-W0156](https://jvet-experts.org/doc_end_user/current_document.php?id=10986) Crosscheck of JVET-W0068 (EE2-related: A combination of CIIP and DIMD/TIMD) [Y. Wang (Bytedance)] [late]

[JVET-W0097](https://jvet-experts.org/doc_end_user/current_document.php?id=10913) EE2-related: Combination of EE2-3.3, EE2-3.4 and EE2-3.5 [X. Xiu, C.-W. Kuo, X. Wang (Kwai), R.-L. Liao, Y. Ye, X. Li, J. Chen (Alibaba), Z. Deng, K. Zhang, L. Zhang, N. Zhang, Y. Wang (Bytedance), Y.-J. Chang, H. Huang, V. Seregin, C.-C. Chen, M. Karczewicz (Qualcomm)]

Option 2 was adopted in EE context

[JVET-W0175](https://jvet-experts.org/doc_end_user/current_document.php?id=11006) Cross-check of JVET-W0097: EE2-related: Combination of EE2-3.3, EE2-3.4 and EE2-3.5 [A. Robert, F. Le Léannec (InterDigital)] [late]

[JVET-W0098](https://jvet-experts.org/doc_end_user/current_document.php?id=10914) Non-EE2: Bilateral Inloop Filter on Chroma [W. Yin, K. Zhang, L. Zhang (Bytedance)]

This contribution extends Bilateral Inloop Filter (BIF) in ECM to chroma components. The BIF-chroma keeps the same design as BIF-luma in ECM. A diamond 5×5 filtering kernel and a CTU level on/off control mechanism are used. The filtering coefficients are retrained for chroma components and the filtering strength for chroma is decided based on the chroma block size as well as the corresponding luma block size.

On top of the EE code of ECM-1.0, simulation results of the proposed method are reported as below:

Tool-on test (VTM-11 with improved MCTF as the anchor):

AI: 0.02%, -0.82%, -0.89%, 101%, 104%;

RA: 0.03%, -0.65%, -0.49%, 101%, 102%;

LB: 0.03%, -1.07%, -1.45%, 100%, 101%.

Tool-off test (ECM-1.0 as the anchor):

AI: 0.00%, -0.77%, -1.01%, 101%, 101%;

RA: -0.01%, -0.55%, -0.49%, 100%, 100%;

LB: -0.06%, -1.19%%, -1.30%, 99%, 99%.

It is mentioned that the current BIF of ECM comes with a slight loss in chroma (probably due to sligt increase in bit rate). This would be more than compensated by this proposal, some additional bit rate reduction seems to be achieved.

Candidate for EE.

[JVET-W0106](https://jvet-experts.org/doc_end_user/current_document.php?id=10922) EE2-related: Bilateral matching AMVP-merge mode [Z. Zhang, H. Huang, C.-C. Chen, V. Seregin, M. Karczewicz (Qualcomm)]

In this contribution, bilateral matching AMVP-merge mode is proposed to combine AMVP and merge inter predictors. The final motion vector of the merge inter predictor is selected from the merge candidate list and is refined based on the minimum bilateral matching error between a reference block that is generated by AMVP candidate and a reference block that is generated by the merge candidate. The proposed method was implemented on top of ECM-1.0, and it reports the BD-rate -0.11% (Y), -0.05% (U), -0.12% (V) for random access configuration with 106% encoding and 99% decoding run time. In addition, it reports the BD-rate -0.48% (Y), -0.44% (U), -0.38% (V) with template matching disabled in both the ECM-1.0 anchor and the test for random access configuration with 110% encoding and 100% decoding run time.

The coding gain of 0.1% is relatively low, considering the 6% encoder run time increase.

It is noted that the configuration with TM disabled would have roughly 1.5% coding loss compared to ECM1, but that loss would be 2% without the method from W0106.

It is commented that template matching might have some difficult implementation aspects at the decoder side. In the long term, it may be desirable to remove. This aspect should be further studied, and experts are encouraged to investigate possible replacements which would retain the attractive gain that TM provides. In general, we are still at a relatively early phase of this exploration, and investigating the potential margin for additional compression performance is of highest priority.

The cross-checker reports that he believes the encoder is not fully optimized yet.

Candidate for EE, where investigating versions with reduced encoding time and similar compression benefit would be of importance.

[JVET-W0177](https://jvet-experts.org/doc_end_user/current_document.php?id=11008) Crosscheck of JVET-W0106: EE2-related: Bilateral matching AMVP-merge mode [C.-W. Kuo (Kwai)] [late]

[JVET-W0107](https://jvet-experts.org/doc_end_user/current_document.php?id=10923) EE2-related: Adaptive decoder side motion vector refinement [H. Huang, Z. Zhang, V. Seregin, W.-J. Chien, C.-C. Chen, M. Karczewicz (Qualcomm)]

This contribution presents an adaptive decoder side motion vector refinement method. The proposed method was implemented on top of ECM-1.0, and it reports -0.23%, -0.23%, -0.26% BD rate for luma and chroma components, respectively, with 115% encoder and 101% decoder run time in random access configurations of common test condition. It also reports -0.82%, -0.75%, -0.78% BD rate for luma and chroma components, respectively, with 117% encoder and 106% decoder run time in random access configurations with template matching tools disabled in both anchor and test.

See notes under W0106. Further study on possible competitive replacements of template matching is strongly encouraged. In this context, compression and runtime should also be reported relative to CTC anchor (with TM on).

Tradeoff compression vs. encoder runtime is not attractive in the current version. Proponents believe that they would be able to reduce the runtime.

Candidate for EE, where investigating versions with reduced encoding time and similar compression benefit would be of importance.

The EE might also provide supplemental results comparing against a configuration with TM off. Question: What does “TM off” mean? Needs to be clarified. This might in the future be an “AHG13”-like effort.

[JVET-W0108](https://jvet-experts.org/doc_end_user/current_document.php?id=10924) EE2-related: Low complexity sign prediction [C. Auyeung, X. Li, S. Liu (Tencent)]

Sign prediction is a part of ECM-1.0. It was found to be one of the hot spot in ECM-1.0. This contribution proposes a method to reduce the computation of sign prediction in ECM-1.0. Using ECM-1.0 as reference and the common test conditions in JVET-V2017, the BDR of the proposed method for AI/RA/LB is x.xx%/y.yy%/0.04% The encoder time for AI/RA/LB is x.xx%/y.yy%/98%,. The decoder time for AI/RA/LB is x.xx%/y.yy%/97%.

Rather small reduction in run time, incurring a small loss. Does not seem worthwhile to consider at this stage of the exploration.

[JVET-W0159](https://jvet-experts.org/doc_end_user/current_document.php?id=10990) Crosscheck of JVET-W0108: EE2-related: Low complexity sign prediction [X. Xiu (Kwai)] [late]

[JVET-W0122](https://jvet-experts.org/doc_end_user/current_document.php?id=10938) EE2-related: On spatial MV propagation and neighboring template block access for template matching and multi-pass DMVR [C.-C. Chen, C.-T. Hsieh, H. Huang, V. Seregin, W.-J. Chien, Y.-J. Chang, Z. Zhang, Y. Zhang, M. Karczewicz (Qualcomm)]

This contribution proposed a bounding-box approach to limit the number of reference samples fetched from reference pictures within a deterministic size and a replacement method for current block templates by using their corresponding prediction samples through inferred motion vectors for template matching (TM) and multi-pass decoder-side motion vector refinement (MPDMVR) in ECM-1.0. These two were tested on top of ECM-1.0 with the common test condition (JVET-V2024). Experimental results of Class B sequences with random access test condition are summarized, as follows.

* Test 1 (Bounding-box approach) : 0.99 % (Y), 0.99% (U), 1.06% (V), 89% (EncT), 82%(DecT)
* Test 2 (Test 1+Template replacement): 2.04% (Y), 2.20% (U), 2.30% (V), 90% (EncT), 79% (DecT)
* Test 3 (TM off+MPDMVR off) : 3.97% (Y), 4.01% (U), 4.25% (V), 80% (EncT), 60% (DecT)

It is asked if the memory bandwidth might be increased by the method due to the prefetch. The proponents answer that this would not be the case, as the area has the same size as in current DMVR.

It is further discussed that the benefit may be hardware dependent.

As a benefit, it is claimed that it not necessary to rely on motion compensation of the neighbor CU, only motion derivation of the neighbor is needed and causes less latency.

Some of the loss (0.2-0.3%) comes by disabling TMMVP

Has it been considered using unrefined MVs? No

At the current stage of exploration, exploring the potential of additional compression gain is still the highest importance, rather than re-designing tools for practicality of implementation. The loss incurred by the given proposal is not negligible, even in the simplest version “Test 1” which does not seem to resolve the problems of DMVR entirely.

Further study recommended – would be premature for EE.

[JVET-W0123](https://jvet-experts.org/doc_end_user/current_document.php?id=10939) EE2-related: Fusion for template-based intra mode derivation [K. Cao, N. Hu, V. Seregin, M. Karczewicz (Qualcomm), Y. Wang, K. Zhang, L. Zhang (Bytedance)]

This contribution proposes a fusion of intra modes derived from the template-based intra mode derivation (TIMD) method tested in EE2-2.1. Two modes are derived using TIMD method and are fused with the weights computed from the sum of absolute transformed differences, which is used as the cost function.

This method was implemented on top EE2-2.1 software, it reportedly provides Y, U, V BD rate reduction with encoder and decoder runtime as follows:

AI: {-0.47%, -0.34%, -0.37%} runtime: {124%, 111%}

RA:{-0.19%, -0.08%, -0.23%} runtime: {105%, 102%}

Straightforward extension of EE2-2.1, derivation of the weights for fusion is done additionally, but the tradeoff between encoder runtime and compression benefit is improved (same runtime approx. 0.35%->0.47% for AI). It is however mentioned that the additional gain is less in case of RA.

Crosschecker also confirms that the method is straightforward.

Decision: Adopt JVET-W0123.

[JVET-W0142](https://jvet-experts.org/doc_end_user/current_document.php?id=10971) Crosscheck of JVET-W0123 (EE2-related: Fusion for template-based intra mode derivation) [X. Li (Alibaba)] [late]

[JVET-W0124](https://jvet-experts.org/doc_end_user/current_document.php?id=10940) EE2-related: Template based intra most probable modes sorting [K. Cao, N. Hu, V. Seregin, M. Karczewicz (Qualcomm)]

This contribution proposes to apply sorting method on intra most probable modes (MPMs) list. For each intra mode in the list, the sum of absolute transformed differences between prediction and reconstruction samples of a template is computed, based on which MPMs are sorted.

This method was implemented on top ECM-1.0, it reportedly provides Y, U, V BD rate reduction with encoder and decoder runtime as follows:

AI: {-0.26%, -0.30%, -0.32%} runtime: {103%, 122%}

RA: {-0.11%, -0.03%, -0.11%} runtime: {101%, 101%}

Question: Is this competing with W0123, or are the gains additive? Not known, likely some overlap, but someadditional gain might be retained.

The encoder runtime is largely reduced in comparison to EE2-2.1, but the decoder run time increases due to the sorting.

One expert suggests that it might be interesting to not always perform sorting for MPM mode.

Investigate in EE.

[JVET-W0174](https://jvet-experts.org/doc_end_user/current_document.php?id=11005) Cross-check of JVET-W0124: EE2-related: Template based intra most probable modes sorting [F. Urban (InterDigital)] [late]

[JVET-W0126](https://jvet-experts.org/doc_end_user/current_document.php?id=10942) EE2 Related – DIMD with implicitly derived multiple blending modes [J. Zhao, S. Paluri, S. Kim (LGE)] [late]

In this contribution, a method to implicitly derive the DIMD blending modes is presented. It utilizes techniques in both EE2-2.2 -- DIMD with multiple blending modes and EE2-2.1 – template-based decoder mode derivation (TIMD). Instead of selecting the blending mode by full RDO and explicitly signaling it, now the blending mode is derived by template-based approach as in TIMD at both encoder and decoder.

On top of ECM-1.0, the coding performance is

AI: -0.04%, -0.04%, -0.08%, 101% (Enc), 112% (Dec)

RA: -0.01%, 0.00%, -0.12%, 100% (Enc), 100% (Dec).

High decoding time increase, probably due to template matching.

Small gain not justifying additional complexity.

Further study

### Technology elements beyond EE2 (4)

Contributions in this area were discussed in session 15 at 0050–0120 UTC on Tuesday 13 July 2021 (chaired by JRO).

[JVET-W0069](https://jvet-experts.org/doc_end_user/current_document.php?id=10885) [AHG12] On Intra TMP Boundary Conditions [K. Naser, T. Poirier, F. Le Léannec, G. Martin-Cocher (InterDigital)]

Intra Template Matching Prediction (TMP) is a recent adopted tool in ECM mainly targeting screen content coding. The prediction block is generating by matching the L-shaped template of the current block with the prediction block’s template. It is asserted that when the template of the current block is not available, e.g. when it exceeds the frame boundary, the decoding process is undefined. This contribution proposes to fix this bug with a similar implementation of intra prediction, where only available reference samples are used for template matching. The luma BD-rate change on top of ECM-1.0 are:

- AI:

o Class F -0.01% with enc/dec time 99% / 98%

o Class TGM -0.04 with enc/dec time 99% / 101%

- RA

o Class F 0.03% with enc/dec time 97% / 92%

o Class TGM -0.02% with enc/dec time 97% /97%

Decision(BF/SW): Adopt JVET-W0069

[JVET-W0154](https://jvet-experts.org/doc_end_user/current_document.php?id=10984) Cross-check on JVET-W0069: [AHG12] On Intra TMP Boundary Conditions [T. Tsukuba (Sony)] [late]

[JVET-W0079](https://jvet-experts.org/doc_end_user/current_document.php?id=10895) AHG12: CTB level filter shape selection of CCALF [M. G. Sarwer, R. -L. Liao, J. Chen, Y. Ye, X. Li (Alibaba)]

This contribution proposes CTB level filter shape selection of CCALF process. Two filter shapes are proposed. In the proposed method, in each APS, up-to 16 filters and their associated shapes and coefficients are signalled. For each CTB, an index is signalled to the decoder to specify which filter shape and coefficients are used for that CTB. In addition to multiple filter shapes, this contribution also proposes to remove the power of 2 constraint of filter coefficient values.

Following results are reported, as compared to ECM-1.0:

* All Intra: 0.05% (Y), -2.17%(U), -2.70%(V), 101% (EncT), 101%(DecT)
* Random access: x.xx% (Y), - x.xx %(U), - x.xx %(V), xxx% (EncT), xxx%(DecT)
* Low delay B: -0.01% (Y), -4.14%(U), -3.68%(V), 100% (EncT), 100%(DecT)

Gain in chroma. YUV BD rate is roughly 0.4%

It is pointed out that the additional gain is significantly lower than the gain by CCALF was before included in VTM.

The complexity is increased. Memory requirements are identical with current CCALF.

Question: Is the filter shape adaptation really needed, or could similar gain be achieved by increasing the filter size in CCALF.

Investigate in EE.

[JVET-W0148](https://jvet-experts.org/doc_end_user/current_document.php?id=10978) Crosscheck of JVET-W0079 (AHG12: CTB level filter shape selection of CCALF) [N. Hu (Qualcomm)] [late]

[JVET-W0110](https://jvet-experts.org/doc_end_user/current_document.php?id=10926) AHG12: GPM with inter and intra prediction [Y. Kidani, H. Kato, K. Kawamura (KDDI)]

This contribution proposes geometric partitioning mode (GPM) with inter and intra prediction to enhance the coding performance beyond VVC. With the method, pre-defined intra prediction modes against geometric partitioning line can be selected in addition to motion vectors from a merge candidate list for each non-rectangular split region in the GPM-applied coding unit.

On top of the VTM-11.0+MCTF and ECM-1.0 which are defined as the anchor for tool-on and tool-off tests in EE2 of JVET-V2024, simulation results of the proposed method are reported as follows:

* Tool-on test (VTM-11.0+MCTF as the anchor):
  + RA: BD-rate YUV: -0.18%, -0.51%, -0.36%; EncT: 102%; DecT: 101%
  + LB: BD-rate YUV: -0.43%, -1.00%, -1.00%; EncT: 103%; DecT: 106%
  + LP: BD-rate YUV: -1.01%, -1.84%, -1.81%; EncT: 111%; DecT: 101%
* Tool-off test (ECM-1.0 as the anchor):
  + RA: BD-rate YUV: -0.04%, -0.29%, -0.18%; EncT: 101%; DecT: 100%
  + LB: BD-rate YUV: -0.20%, -0.87%, -0.48%; EncT: 101%; DecT: 100%
  + LP: BD-rate YUV: -0.38%, -0.93%, -0.92%; EncT: 107%; DecT: 101%

Intra prediction modes are selected such that their direction corresponds with the GEO separation line (parallel, perpendicular)

Question: Is the encoding time correct? The proponents assume it is due to the low number of pre-selected modes that are checked. For LP, the encoding time is more increased as in current CTC for LP GEO is not used at all (as it would require two MC ops).

Most likely, GEO itself contributes to the higher gain in LP.

Has it been tried with AI? No

How is the MV handled in the partition that is intra coded? Needs to be checked.

Investigate in EE.

[JVET-W0112](https://jvet-experts.org/doc_end_user/current_document.php?id=10928) AHG12: Diagonal MMVD with ARMC [Y. Kidani, K. Kawamura (KDDI)]

Withdrawn

[JVET-W0128](https://jvet-experts.org/doc_end_user/current_document.php?id=10945) AHG12: Alternative classifiers for ALF [N. Hu, V. Seregin, M. Karczewicz (Qualcomm)]

In adaptive loop filter process of ECM-1.0, when a signalled filter set is applied to a luma coding tree block, a classifier is applied to each 2x2 luma block. For each class, a specific filter is applied among the signalled filter set. Based on the classification, a geometric transformation can also be applied. In this contribution, a two-classifier method is proposed and five different second classifier options are tested. For a luma filter set, a flag is signalled to indicate whether the classifier in ECM-1.0 or the proposed alternative classifier is applied. On top of ECM-1.0 with common test condition, simulation results of the proposed methods are reported as below.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | All intra | | | Random access | | | Low delay B | | |
|  | BD-rate Y | EncT | DecT | BD-rate Y | EncT | DecT | BD-rate Y | EncT | DecT |
| Test 1 | -0.05% | 100% | 97% | -0.18% | 99% | 99% | -0.29% | 102% | 98% |
| Test 2 | -0.09% | 105% | 96% | -0.30% | 103% | 99% | -0.44% | 110% | 101% |
| Test 3 | -0.11% | 100% | 97% | -0.26% | 100% | 99% | -0.44% | 100% | 99% |
| Test 4 | -0.14% | 106% | 97% | -0.36% | 104% | 99% | -0.60% | 110% | 98% |
| Test4 fast | -0.14% | 102% | 97% | -0.35% | 101% | 100% | -0.61% | 102% | 100% |
| Test 5 | -0.22% | 101% | 96% | -0.43% | 100% | 99% | -0.73% | 101% | 96% |

Investigate in EE: Test4 fast and Test5, also report on impact on decoder complexity, in particular regarding the sample-wise switching between filters for Test5 (also including statistics, parallelization, etc.)

[JVET-W0163](https://jvet-experts.org/doc_end_user/current_document.php?id=10994) Cross-check report on JVET-W0128 (AHG12: Alternative classifiers for ALF) [M. G. Sarwer (Alibaba)] [late]

# High-level syntax (HLS) proposals (16)

## AHG9: SEI message studies and proposals (11)

Contributions in this area were discussed in session 7 at 2100–2315 and session 7 2335-0125 UTC on Thursday 8 July 2021 (chaired by JRO and GJS). Continued in session 12 2325-0045 UTC on Friday 9 July 2021

See also NB comments.

[JVET-W0085](https://jvet-experts.org/doc_end_user/current_document.php?id=10901) AHG9: Picture quality metrics SEI message [Y. He, M. Coban, D. Rusanovskyy, M. Karczewicz (Qualcomm)]

This contribution is a follow-up of JVET-V0062 and proposes the inclusion of a new SEI message, picture quality metrics SEI message, in the VSEI specification. Three quality metrics, PSNR, wPSNR and WS-PSNR, are specified in the new SEI message for SDR, HDR and 360 video content, respectively.

It was commented that the contribution has a number of editorial and clarity problems and it does not seem so clear how it would be used.

It was commented that normative referencing of a non-normative technical report is not appropriate, if that is the intended way to specify how the described metrics are to be computed.

Only luma measures are proposed. There was some discussion of what to do for, e.g., RGB video, and whether chroma should also be measured.

It was asked what to do for sequence averages if the picture resolution varies within the CVS. The presenting proponent said the sequence average would just be computed as an average of measurements taken across the different resolutions.

Another proponent said the encoder could use a different resolution for comparison. But a participant commented that the defined semantics should match what we believe an encoder should use.

A suggestion was to just define some quality scale (e.g. with a range of 0 to 10) without specifying exactly how to compute values.

It was asked what to do if new metrics are developed. Having an external registration authority was suggested for providing a way to identify different metrics.

It was pointed out that there might be some problems in the proposed text, e.g. referencing 23002-8 in terms of defining some metrics which is a TR and not a normative spec.

Use cases? It could for example be used for post processing.

It was commented that the selection of WPSNR and WS-PSNR may not be well justified by market needs, as those are mainly used internally in JVET.

The previous proposal also included SSIM and MS-SSIM, which were removed (as it had been commented in the last meeting that it might be difficult properly specifying them). On the other hand, those metrics are well established.

It is further mentioned that VQEG is currently running a similar activity (which had also been pointed out in the last meeting regarding a communication with AG5). They intend to establish a registration procedure for mew metrics

What about the case when the “original” video (e.g. UGC) is already distorted?

From the discussion, various aspects were pointed out that need to be further studied:

- Are the proposed metrics those which would be relevant in the market?

- What are the use cases?

- Is it enough to specify luma PSNR?

- Relationship with VQEG activities

- Properties of the source (e.g. was the resolution changed, was it already compressed?)

- Which precision of the metrics is needed?

- Editorial quality, in particular regarding the normative description of the derivation of the metrics

- How could new metrics be handled?

Further study recommended, taking the issues above into consideration.

[JVET-W0071](https://jvet-experts.org/doc_end_user/current_document.php?id=10887) AHG9: Green Metadata SEI message for VVC [C. Herglotz, M. Kränzler, A. Kaup (FAU), E. François, M. Radosavljevic, E. Reinhard (InterDigital), X. Ducloux (Harmonic), D. Menard (INSA), Y. He, M. Coban, D. Rusanovskyy, M. Karczewicz (Qualcomm)]

This contribution proposes an initial update of the ISO\_IEC\_23001-11 (Green Metadata) specification with the definition of a new Green Metadata SEI message for VVC. The update specifically addresses syntax signalling for quality recovery after low power encoding and decoder complexity metrics. For the former, it is proposed to adopt the same syntax as already used for HEVC and AVC and add syntax for signalling the quality based on SSIM, and VMAF, wPSNR and WS-PSNR, which have become widely used quality metrics throughout the last years. For the latter, the update covers decoder complexity metrics for VVC, which are not covered by existing SEI messages for AVC and HEVC. The proposed syntax is based on extensive decoding complexity measurements performed on a large video dataset using VTM and VVdeC decoding. The measurement results are analyzed in detail using Tool On/Off tests and complexity modeling. As for the AVC and HEVC Green metadata SEI messages, the new SEI message is defined in ISO/IEC 23001-11, and a reference to ISO/IEC 23001-11 is inserted in the VSEI specification for the definition of Green metadata SEI message of VVC.

It was asked what is the status of work on green metadata development in SC 29/WG 3 (MPEG Systems) and whether an approach similar to that in AVC and HEVC would be appropriate.

It was suggested that a receiving system could use the decoder complexity information to decide which resolution to decode.

It was asked whether this should be written into VSEI or into VVC instead. In the past cases of AVC and HEVC, the definition of the green metadata has depended on the video coding context, which also seems to be the case in the current proposal. This question should be considered. If the data is customized specifically for VVC rather than being generic, perhaps the reference to green metadata should be defined within VVC rather than in VSEI. Basically all that is needed is a definition of the payloadType and a reference to the green metadata standard, and payloadType is already a VVC matter rather than a VSEI matter. The proposal suggests to use the same payloadType value that was used in AVC and HEVC, which seemed appropriate.

Usage for decoder complexity and video quality metrics was proposed. For video quality metrics, some specific metrics were proposed.

As the green metadata for VVC is customized for VVC, the link to 23001-11 might better be included in VVC rather than VSEI. This is agreed.

The syntax and semantics for linking to 23001-11 are constructed similar as it was done in AVC and HEVC. However, it would be premature to already establish this in the ongoing V2 of VVC, as the Green Metadata for VVC in 23001-11 would probably not exist yet when v2 is finalized. This could use the same payload type as HEVC and AVC, as it is uniquely associated with a specific standard.

Further coordinate with WG3 about timelines.

[JVET-W0076](https://jvet-experts.org/doc_end_user/current_document.php?id=10892) AHG9: Independently coded region output SEI message [B. Choi, S. Wenger, X. Li, S. Liu (Tencent)]

A new SEI message “Independently coded region output SEI message” that carries useful information for individual access and output of an independently coded region is proposed. The proposed SEI message can be used for cropping and outputting one or more specific rectangular regions which may consists of one or more subpictures. When a decoder system tries to early access and use one or more specific rectangular regions by discarding other regions, the information carried by the proposed SEI message can be used for cropping and outputting the rectangular regions which may consist of one or more subpictures. It is asserted that the use cases of the proposed SEI message include immersive media (VR360, point cloud coding) and game/e-sport contents.

This relates to the video decoder interface design, which has been discussed at recent meetings and is further discussed in W0168 and W0169.

It was commented that subpictures can be used for a type of similar functionality. This proposal has alternative approaches, such as indicating overlapping independently decodable regions.

This is proposed for VSEI rather than VVC, and is proposed in a way that does not depend on subpictures.

It is further commented that the same functionality might be implemented using subpictures. It is confirmed by othr experts that it is possible to extract multiple subpictures and merge them into a bitstream. This could achieve a similar functionality.

One expert comments that transcoding may be necessary to achieve this when a user requests a certain region.

After review of W0168/W0169, it can be concluded that various aspects of this SEI message would rather relate to VDI. It is however claimed by the proponents that the aspect of multiple cropping windows would still remain specific to the video decoder and its output.

The proponent said it would be desirable to specify this in VSEI in a way that would not depend on VVC and thus could also apply to other video coding designs such as AVC and HEVC.

Further study recommended in context of VDI, and with regard to the multiple cropping window aspect.

[JVET-W0077](https://jvet-experts.org/doc_end_user/current_document.php?id=10893) AHG9: Comments on multiview-related SEI messages in VSEI [B. Choi, S. Wenger, S. Liu (Tencent)]

The following modifications related to scalability dimension information (SDI) SEI message and multiview acquisition information (MAI) SEI message in VSEI are proposed:

Proposal-1: Defining the variable ViewId in the semantics of SDI SEI message.

Proposal-2: Clarifying the target layer that the MAI SEI message is applied to.

[Group with NB input]

Essentially editorial fixes for draft new SEI messages.

Adopt both aspects, at least in principle - exact wording delegated to the editor.

Both option-1 and option-2 of proposal-2 would be viable

Decision(edBF): Adopt – implementation to discretion of editors.

[JVET-W0078](https://jvet-experts.org/doc_end_user/current_document.php?id=10894) AHG9: Multiview view position SEI message [B. Choi, S. Wenger, S. Liu (Tencent)]

Multiview View Position (MVP) SEI message is proposed for VSEI. The MVP SEI message is useful for indicating the relative view position along a horizontal axis of 3D rendering view, especially for glassless 3D display as well as multiview-based 3D rendering. The MVP SEI message was adopted into HEVC. To adopt the same SEI message into VSEI, minor semantics changes are also proposed.

This SEI message is beneficial when the physical view position deviates from the view position in coding/decoding order. It is reasonable to be included in VSEI for supporting multiview applications.

The suggested semantics changes appear appropriate for alignment with VVC/VSEI expressions.

Decision: Adopt JVET-W0078.

[JVET-W0080](https://jvet-experts.org/doc_end_user/current_document.php?id=10896) AHG9: some errata and clarification items for Additional SEI messages for VSEI [E. François, M. Radosavljevic (InterDigital)]

This contribution reports some possible errata and clarification items in the Additional SEI messages for VSEI.

1.1.a

In the Annotated regions SEI message syntax, the following line:

for( i = 0; i  <=  ar\_num\_object\_updates; i++ ) {

is proposed to be changed into:

for( i = 0; i  <  ar\_num\_object\_updates; i++ ) {

since ar\_num\_object\_updates is defined as follows.

**“ar\_num\_object\_updates** indicates the number of object updates to be signalled. ar\_num\_object\_updates shall be in the range of 0 to 255, inclusive.”

The following aspects were later clarified and are reflected in the new draft text.

1.1.b – agreed this needs to be fixed.

1.2a – was checked by Jill Boyce and clarified/agreed in coordination with other editors

1.2b – purely editorial, might have been modified in last call text, and NB comments could exist.

1.2c –was checked by Jill Boyce and clarified/agreed in coordination with other editors

1.2d – purely editorial, agreed.

1.2e – editorial clarification, agreed.

1.2f – purely editorial, agreed.

1.3x – obvious typos

[JVET-W0083](https://jvet-experts.org/doc_end_user/current_document.php?id=10899) AHG9: Bug fixes for some SEI messages in the VSEI amendment [Y.-K. Wang, Y. Wang, L. Zhang (Bytedance)]

This contribution proposes some changes to the latest draft of the VSEI amendment in JVET V2006, for fixing some asserted bugs. Involved SEI messages are the annotated regions SEI message, the depth representation information (DRI) SEI message, and the extended DRAP indication SEI message.

It is asserted that some of issues for the annotated regions SEI message and the DRI SEI message also apply for HEVC and/or AVC.

The proposed changes are summarized in the contribution and the exact text changes for VSEI are provided in an attachment to this contribution. Text changes for HEVC and AVC are not provided, but once decided those are asserted to be the same as the corresponding VSEI changes.

The following issues for the annotated regions SEI message, the depth representation information SEI message, and the EDRAP indication SEI message have been observed:

1) The syntax table for the annotated regions SEI message has the column "C", which is for AVC syntax table but not for VSEI syntax table, thus should be removed.

2) For the annotated regions SEI message, the value range of the ue(v)-coded syntax element ar\_object\_label\_idx[ ar\_object\_idx[ i ] ] is missing.

3) For the DRI SEI message, the descriptor (i.e., coding method) of the depth\_nonlinear\_representation\_model[ i ] syntax element is unspecified.

4) For the DRI SEI message, the value ranges for the ue(v)-coded syntax elements depth\_representation\_type, disparity\_ref\_view\_id, depth\_nonlinear\_representation\_num\_minus1, and depth\_nonlinear\_representation\_model[ i ] are missing.

5) In the semantics of the u(5)-coded syntax element da\_mantissa\_len\_minus1, the specification of the value range of 0 to 31, inclusive, is redundant.

6) For the EDRAP indication SEI message, the semantics of the edrap\_leading\_pictures\_decodable\_flag syntax element is missing, and in the syntax table there is a typo of "minsu1" that should be "minus1".

The suggested corrections appear largely appropriate and should be included in draft for v2 of VSEI (some are identical with comments from W0080, and also clarification with Jill Boyce may be necessary on items relating to annotated regions). Some of the issues also apply to AVC and HEVC versions of the related SEI messages.

These corrections were later enrolled into the ITU consent texts (due Tuesday 13 July) – see also notes under JVET-W0187.

[JVET-W0072](https://jvet-experts.org/doc_end_user/current_document.php?id=10888) AHG9: Enhancement of film grain parameter estimation for different intensity intervals [M. Radosavljević, E. François (InterDigital), W. Hamidouche, T. Amestoy, G. Gautier (INSA)] [late]

This contribution is an update of the previous contribution JVET-V0093. JVET-V0093 proposes a complete VTM software implementation of the film grain feature, including film grain analysis and parameter estimation module at encoder side, and film grain synthesis module at decoder side.

An enhanced *film grain analysis and parameter estimation module* is proposed in this contribution. It is proposed to estimate scaling factor, or scaling factors if more than one intensity interval is estimated, as part of the film grain parametrized model, in order to adjust the intensity of the film grain per image intensity range. In that way, it is possible to adapt film grain intensity to the local characteristics of the image in more flexible way. The implementation is based on the SMPTE RDD 5 Film Grain Technology specification, with some adaptations as proposed in JVET-W0095, and it is compatible with the current SEI message specification.

The proponents suggest harmonization with other proposals, would provide software (currently, software is in a VTM fork, which is still the previous version). It is suggested by the proponents to move the new version of the software into the main branch.

Software was not yet provided, but the proponent said it should be available in a few days

See further discussion below

[JVET-W0095](https://jvet-experts.org/doc_end_user/current_document.php?id=10911) AHG 9: Bit-accurate grain blending process for film grain characteristics SEI message [S. McCarthy, P. Yin, W. Husak, F. Pu, T. Lu, T. Chen (Dolby), E. François, M. Radosavljević (InterDigital), V. G R, K. Patankar, S. Kadaramandalgi, Ajayshyam (Ittiam)]

This contribution proposes to add semantics to specify a bit-accurate grain blending process to the film grain characteristics (FGC) SEI message in the next version of VSEI (ITU-T H.274 | ISO/IEC 23002-7). No change in syntax is proposed.

The proposed bit-accurate grain blending process is identical to the process specified in SMPTE RDD 5, Film Grain Technology, except for two modifications: 1) use of a transform already specified in VVC instead of the tramsform specified in SMPTE RDD 5; and 2) increase in grain block size for resolutions greater than 1920x1080.

This contribution proposes an addition / change to the semantics of the existing film grain characteristics SEI message, to specify a particular bit-accurate film grain synthesis algorithm to be performed after decoding.

Relative to what was previously described, it uses the VVC inverse transform design and uses larger block sizes. The prior described method uses an 8x8 inverse transform.

It was commented that the proposal seems to be trying to extend or replace semantics of an already documented SEI message and somewhat deprecate a prior well known interpretation method.

It was commented that the proposal is appearing to try to establish a conformance requirement to obtain a particular exact result for the post-processing to be performed in response to an SEI message, which is not what has been historically done and seems especially problematic after the fact.

It was commented that post-processing is sometimes done in a separate stage of the system where there may not be synergy with already having support for the VVC inverse transform in the decoder.

It was commented that if we wish to specify a normative method, it may not be clear that this would be exactly the method we would choose.

It was commented that rephrasing this as an example could be worth considering.

It is pointed out that instead it could be referenced in an external reference, e.g. to SMPTE RDD5.

Furthermore, the wording “…all decoders … will produce …” is not appropriate, suggesting normative decoder behaviour.

Discussion stopped here Friday July 9 0125 – further discussion was held, see e.g. notes under 7.1 and 7.3.

In a short follow-up discussion during session 12, it was pointed out that a better name than “bit accurate” would be desirable to reflect that this process is not mandatory.

Another version was uploaded in response to the initial discussion.

Potential development of a technical report was suggested by the proponent.

[JVET-W0179](https://jvet-experts.org/doc_end_user/current_document.php?id=11010) Cross-check of JVET-W0095: Bit-accurate grain blending process for film grain characteristics SEI message (AHG9) [C. Fogg (MovieLabs)] [late]

[JVET-W0096](https://jvet-experts.org/doc_end_user/current_document.php?id=10912) AHG9: Demonstration of AVC FGC SEI in real-time bit-accurate grain blending process on smartphone [V. G R, J. Shingala, S. Kadaramandalgi, Ajayshyam (Ittiam), S. McCarthy, P. Yin, W. Husak, F. Pu, T. Lu, T. Chen (Dolby)]

This contribution describes an optimized software implementation of a real-time bit-accurate grain blending process and AVC decode on a smartphone where the grain blending parameters are signalled in the film grain characteristics SEI message in an AVC bitstream. The bit-accurate grain blending process is as described in JVET-W0095. On a CortexA57 based ARM processor [verified on mid-tier smartphone with Qualcomm SDM730 chipset] for 8-bit 1920x1080 YUV 4:2:0 video content, the single-threaded implementation of the grain blending process completes in 4.5 ms/picture for 8x8 grain blocks and 4 ms/picture for 16x16 grain blocks. For 4-thread implementation, the grain blending process completes in 1.2 ms/picture and 1.1 ms/picture for 8x8 and 16x16 grain blocks, respectively.

Version 2 of this contribution provides additional examples, clarifications, and some editorial corrections.

Version 3 of this contribution adds Figure 3 showing a picture of the real-time grain blending process running on a smartphone.

Smartphone with 4 threads was used for the demo shown during the presentation, run at 1.8 GHz. SIMD optimization

The table in the contribution shows only the computational resource spent for the film grain synthesis

[JVET-W0104](https://jvet-experts.org/doc_end_user/current_document.php?id=10920) AHG9: Resampling SEI message [T. Poirier, G. Martin-Cocher, F. Le Léannec, K. Naser (InterDigital)]

Was withdrawn – remove.

[JVET-W0169](https://jvet-experts.org/doc_end_user/current_document.php?id=11000) [AHG 9] VDI binding for VVC SEI message [A. Gabriel, E. Potetsianakis (TNO), Y. Lim (Samsung)] [late]

This contribution proposes the addition of an SEI message for functionalities relevant to the Video Decoder Interface for Immersive Media (VDI), as specified by the VDI working group of MPEG. The SEI message will reference the ISO/IEC 23090-13 specification where the SEI syntax, semantics and persistence scope are defined.

The status of the of the SEI message in 23090-13 is currently working draft, not approved work item. The concept of establishing the link to the video and VSEI specs is similar, green metadata, i.e. reserving a codepoint for something that is defined elsewhere.

Need to clarify if the link to this SEI would better be established in VVC rather than VSEI. It could be foreseen to be carried in the video stream or in systems stream. Based on current understanding what is planned to be included in this SEI message, it appears that the majority of these aspects would rather relate to systems operations and carriage in the video stream might not be necessary.

For example, possible applications could be subpicture extraction, or substream extraction. However, sub-bitstream extraction is not directly a task of a video decoder. It would be performed by a bitstream extractor which converts it into an elementary stream to be decoded by a decoder.

It is commented that it would be necessary to learn more about foreseen use case, and if the information carried in this SEI needs to be specifically aligned to VVC. If it would be universally applicable to any standard, it might better be hosted in VSEI.

Further action would need to be taken when the specification is maturing in WG3, e.g. a mature WD prior to CD.

This contribution was discussed in session 12 at 0000 on Sat 10 July.

It was asked what is the current status of the VDI specification development. The proponent said there was some drafted content but it is not yet an approved work item.

It was commented that we should consider whether it is more appropriate to put such a binding into VVC versus VSEI.

It was commented that there should be discussion of whether this data needs to be communicated within the video layer.

## Non-SEI HLS aspects (5)

Contributions in this area were discussed in session 12 at 0045–0125 UTC on Saturday 10 July 2021, and in session 19 at 2320–0120 UTC on Tuesday 13 July 2021, and in session 23 at 0725–XXXX on Thursday 15 July 2021 (chaired by JRO and GJS).

[JVET-W0073](https://jvet-experts.org/doc_end_user/current_document.php?id=10889) ALF parameters for RPR [P. Bordes, F. Galpin, F. Le Léannec (InterDigital)]

In VVC, RPR allows coding some frames at lower resolution to dynamically adapt to the available transport bit rate. For low bit rates, RPR allows competing with VVC at full resolution while reducing encoder and decoder complexity. However, its performance may depend on the video content and the chroma distortion may be significant compared to full-resolution coding. It is proposed to add an ALF post-filtering stage applied on the reconstructed up-sampled pictures when RPR is enabled. The BD-rate performance of RPR is improved in chroma by about 15% in RA, LDB and AI.

See also W0074 and W0104.

Currently, RPR is applied after post/loop filtering. The proposal is to add another stage of post filters (ALF, CCALF) after resolution conversion. In this case, also the upsampling is out of loop

It was commented that also the upsampling could be done with adaptive filters

It is further mentioned that the ALF stage before upsampling could be omitted

It is also pointed out that the reported results may not be related to RPR, as RPR is conducted in the loop, and here everything is out-of-loop.

The gain in chroma in this proposal is likely due to the additional CCALF

It is also mentioned that the loss in chroma for RPR upsampling may be due to chroma phase misalignment in upsampling (which would not cause visual artifacts, but bad PSNR). It is mentioned that it could be expected that also spatial scalability using RPR filters might have a bad compression performance due to chroma phase misalignment.

The proponents suggest to study this in EE2, but that may be premature without understanding the reason for the chroma loss in RPR upsampling.

Discussion stopped here Saturday 10 July 0130 – see further notes under W0074

[JVET-W0074](https://jvet-experts.org/doc_end_user/current_document.php?id=10890) HLS for ALF parameters for RPR [P. Bordes, F. Galpin, K. Naser, F. Le Léannec (InterDigital)]

See also W0073.

The contribution proposes an HLS compatible with VVC for signaling additional ALF post-filter parameters for reducing distortion of the up-sampled RPR pictures with original full-size pictures.

Method 1: SEI and new APS, or SEI and re-use of APS.

A new SEI message contains SPS, PH,SH and CTU ALF parameters for the up-sampled pictures, and a new APS type ALF\_RPR\_APS contains the filters coefficients *alf\_data()*. To avoid creating a new APS type, one can re-use existing ALF\_APS but since the range value of *aps\_adaptation\_parameter\_set\_id* is [0,..7] the encoder should reserve apsId values for the regular ALF and others for the new ALF.

Method 2: use scalable VVC.

The HLS for scalable VVC allows for specifying another layer with regular ALF parameters. Enhancement layer would only consist of CUs coded in skip mode

It is commented that method 2 would even be more generic, as it would also allow coding some CUs in skip mode, others not.

Method 2 could be mimicking post processing using RPR for upsampling and ALF for improving, if all enhancement CUs re coded in skip mod and never used as reference. This would however require a decoder which is conformant with a multilayer profile.

Method 1 could be used with a single-layer decoder, but it would be expected that there is some postprocessing entity which is able to interpret the SEI message. If such a mechanism is not there, the decoder would just decode the low resolution.

Question: Why is a new APS parameter type needs? A version 1 single layer decoder would just drop it.

It is also pointed out that from the contribution it is not obvious how large the benefit in terms of compression would be.

One expert points out that potentially some of the existing high-level flag could be used to turn off ALF in the base layer.

It is suggested to further study whether such a “light weight scalability” approach would be needed, and particularly what the advantages / disadvantages would be in terms of complexity vs. compression benefit.

Anyway, method 2 would be available already now in mult-layer decoders.

[JVET-W0075](https://jvet-experts.org/doc_end_user/current_document.php?id=10891) Subpicture-wise reference picture resampling and scalability for Game/E-sport contents streaming [B. Choi, S. Wenger, X. Li, S. Liu (Tencent)]

For Game/E-sport contents streaming use cases including,

* Local enhancement of visual quality or resolution
* Customized layout of the played video
* Update of layout within a coded video sequence,

we propose the following modifications of VVC high-level syntax:

1. Signaling a scaling window per subpicture, when more than one subpicture is present  
   : For subpicture-wise reference picture resampling in a layer or across layers, scaling windows need to be signaled in PPS for each subpicture that has a different scaling ratio with the reference subpicture.
2. Signaling the updated subpicture partitioning information in PPS  
   : To enable the picture-level updates of subpicture partitioning information, the subpicture partitioning information needs to be signaled in PPS, optionally. When the subpicture partitioning information is updated, some subpictures with particular subpicture IDs may appear or disappear, and the spatial resolution and the vertical and/or horizontal offset values of each subpicture may be changed.
3. Permitting that each layer may have a different subpicture partitioning layout.  
   : In VVC version 1, all layers belonging to the same dependency tree shall have the same subpicture partitioning layout. In VVC next versions, it is proposed to relax this constraint to support the argued use cases that need different subpicture layouts across layers.
4. Permitting the combination of subpicture partitioning and reference picture resampling  
   : In VVC version 1, the combination of subpicture partitioning and reference picture resampling is not supported. We propose to relax this constraint to support region-wise quality/resolution enhancement.

Many of these aspects might fall into areas which could better be handled at systems level (see discussion under JVET-W0169). The approach of RPR/scalability at subpicture level would definitely require changes in VVC (it was by purpose left out of V1 due to complexity consideration). However, evidence would be required that this provides benefit e.g. in terms of compression performance, and that the additional complexity wold be justified.

It is further commented that this is not purely high-level syntax, but rather functionality with non-negligible impact on low-level implementation.

Further study recommended.

[JVET-W0133](https://jvet-experts.org/doc_end_user/current_document.php?id=10962) Constrained RASL encoding for bitstream switching [R. Skupin, C. Bartnik, A. Wieckowski, K. Sühring, Y. Sanchez, B. Bross (HHI)] [late]

Document JVET-V0060 presented a constrained RASL encoding method that allows for resolution or quality switching in HTTP streaming with open GOP coding structures while avoiding the severe drift artefacts occurring when using unconstrained open GOP encoding for such switching.

The presented constraints entail disabling BDOF, PROF, DVMR and CCLM for RASL pictures as well as restricting the choice of the collocated reference pictures to pictures not preceding the associated CRA picture. It is reported that the constrained RASL encoding method retains most of the unconstrained open GOP coding gains with BD-rate gains of up to -8.57% (64 pictures RAP period) relative to closed GOP coding.

This document presents updated results based on the publicly available VVenC version 0.3.1 and proposes to add a GCI flag to VVC to clearly specify and indicate the necessary constraints and serve as a conformance point for related systems to allow for interoperability.

In response to comments received for JVET-V0060, a note was added to the proposal explaining the constraints intent.

Colour artifacts caused by CCLM were observed, which are resolved when the method is implemented in stream switching.

It is pointed out that an encoder can implement this method already in version 1. Why are additional constraints and a flag needed, and what would it help if only defined in a later version?

It is argued that the receiving system would act different in bitstream switching when knowing that the encoder has properly produced the bitstream.

It is pointed out by one expert that the proposal of the constraint flag is aligned with a specific implementation, but there may be other ways to resolve the problem. It is further pointed out that some restrictions may be too strong, e.g. if the RASL picture is an IDR picture it would not nbe necessary to disable CCLM.

An encoder might also have other mechanisms to limit the drift, without totally disabling certain tools.

Several experts expressed the opinion that other less restrictive ways might exist to resolve the problem. Further study is recommended.

The proponents announced to provide an update incorporating some of the suggestions that were made in the discussions. This was presented in session 23 0900 UTC on Thursday 15 July.

It was pointed out that PRO may not need to be restricted, as this just refines the affine prediction and does not rely on missing reference pictures.

Two options are included in the new version.

Option 2 uses a gci flag, and adds the suggested restrictions as an informative note. It is commented that this stays somewhat vague and might not help.

Option 1 also uses gci flag, but would somewhat specify more “normative encoder behaviour”.

Could not an encoder intended for a domain like stream switching prevent such behaviour without specific signalling? Answer by proponents: In an ideal word, yes, but in stream switching the source may be coming from different encoders, including some which do not take prevention mechanisms.

It is argued that the packager might have a chance not doing stream switching at RASL positions where it is not signalled that it is “safe”. It would be difficult to find out otherwise from the bitstream itself that this is the case. Actually, there are packagers which try avoiding switching in open GOP.

Could this alternatively be handled in the systems level, e.g. DASH?

It is commented that this would be a somewhat uncommon usage of GCI.

Looking at the intent that it is more meant for usage at systems level rather than decoder, signalling by an SEI message may be a better solution.

The SEI solution (based on option 1) appears to be agreeable to other experts. This could even be an SEI message without syntax, just indicating by its presence that the encoder has observed certain constraints beneficial for open GOP stream switching.

The proponents are asked to propose a version based on an SEI signalling solution. It is to be checked if it is really necessary to restrict PROF.

The new version was presented in the plenary session 0500 UTC on Friday 16 July.

BDOF and PROF are not disabled (for further study if that might be necessary).

Decision: Adopt JVET-W0133v3 option 3 (for VVC V2). Further editorial refinements of the note appear necessary

[JVET-W0168](https://jvet-experts.org/doc_end_user/current_document.php?id=10999) [AHG 9] Early access to samples for low latency applications [A. Gabriel, Y. Shieferaw, E. Potetsianakis (TNO)] [late]

This contribution proposes a SEI message that will indicate that partial output of the bitstream is possible in the current bitstream. Partial output means that samples once finalized are made available to an application without having to have all the samples in a frame decoded. This is beneficial for low latency applications where post-decoding operations occur before display as can be the case for cloud rendering and gaming (ex. overlays, rotations and up/downscaling). The earlier release of samples can allow post decoding operations to run in parallel on parts of a frame and allow for reduced latency or more operations in the same time.

A modified VTM decoder is provided implementing the output of partially decoded pictures. A frame is output every time a subpicture is fully decoded and it is assumed that each subpicture is fully independently decodable.

It is pointed out that the assumption a decoder would output one subpicture after the other might not be realistic. Decoders typically operate in raster scan order, and write the result into the output buffer. Also displays often operate line by line. Therefore, this additional signalling would not help for low latency.

Another comment is made that certain decoders (in particular software decoders) could implement decoding and output subpicture-wise, and take this information as a hint that they should do so. There would however be no guarantee that a decoder would do so.

Would the flag impose any constraint on the encoding of the subpictures?

As an alternative, it is suggested to put subpictures into separate bitstreams, and provide synchronisation and composition at systems level. This should not introduce significant delay.

It is also asked what would be the overall delay in the envisioned application, in particular if composition of different players as subpictures into a picture is done in the cloud, and how relevant the decoding/output delay is in that context. The proponents answers that any saving of latency is relevant in the given application.

It is not fully obvious what the advantage of a subpicture-based approach in comparison to a systems-level implementation would be.

Further study recommended in the context of VDI.

# Plenary meetings, joint meetings, BoG reports, and liaison communications

## JVET plenaries

Some of the discussions and actions at plenary sessions are noted in this section (especially those of Monday 12 July 0750–XX00):

* BoG reporting
* Scheduling / further BoG sessions?
* Conformance V2 planning (BoG D. Rusanovskyy)
* Planning of output documents: Standard parts & DoCs, CE & EE descriptions, verification test plan/report
  + Prepare integrated version texts of VVC V2 and VSEI V2, issue resolutions to ballot DIS of 2nd edition rather than DAM texts.
  + Add Gary Sullivan as editor, nominate editors for new edition (revisit: resolution)
* Whether to hold a hybrid meeting in October 2021: An informal poll gave 5% who would consider travelling to participate physically, 90% who would participate remote, and 5% who were not wanting to participate.

Wed. 14 Jul. 0820:

* Report from NN BoG:
  + Further work on summary analysis (Excel sheet attached to report)
  + Updates on future contributions’ reporting requirements
  + EE1 planning (JVET-W2023): proposals identified for next round
  + More emphasis on cross-checking, training should be cross-checked as well
  + No need to meet again; offline activities on preparing EE and CTC documents
  + Recommendations of BoG (as per v3 of W0182, further editing cleanups in v4) were approved
  + Plan Liaison to JPEG on NN activities (E. Alshina and A. Segall to prepare)
* Report from BoG on VVC V2 conformance (W0188)
  + Identified 65 bitstreams
  + Plan to issue draft 1 of VVC V2 conformance
  + Clarify 9 new profiles, new tools not enabled for Main12, Main12SP, Main12Intra
  + Add chairs to AHG5
  + Make a request for ISO amendment? Revisit
* EE2 (JVET-W2024):
  + It was agreed to also investigate the previous 1.5a in the next round, together with 1.5b
  + EE document to be prepared offline (previous coordinators)
* CE on high bit depth to be discontinued
* New CE on film grain: Investigate the method from JVET-W0095 versus the RDD-5, in terms of complexity and quality; also investigate the impact of the block sizes, and whether larger block sizes than 8x8 would be necessary (JVET-W2022); it is commented that it would be desirable having a precise description of the algorithms, as useful for a TR later.

## Information sharing meetings

In addition to the joint meetings listed below, information sharing sessions with other WGs of the MPEG community were held on Monday 12 July 0500–0730, Wednesday 14 July 0500–0630, and Friday 16 July 2100–XX00. The status of the work in the MPEG WGs was reviewed at these information sharing sessions.

## Joint meeting with Q6/16 (VCEG) and WG2 MPEG Requirements 0730–0845 Tuesday 13 July

The following topics were discussed in this joint session. See also the notes recorded on these topics in other sections of this document.

* VVC V2 Profiles (Tuesday 0730-0800 UTC with JVET and MPEG Requirements on VVC v2 profiles)
* [JVET-W0136](https://jvet-experts.org/doc_end_user/current_document.php?id=10965) Suggested initial profile text for VVC operation range extension [T. Ikai (Sharp)] [late]
  + Monochrome profiles proposed
  + All-intra video profiles proposed (with loop filter support required)
  + 10, 12, and 16 b profiles proposed
  + Monochrome, 4:2:2, 4:4:4
  + Inter-predictive profiles proposed (unlike in HEVC)
  + Main 12 Still Picture, Main 12 4:4:4 Still Picture proposed (not in HEVC)
  + ~16 profiles
  + Discussion:
    - Tiers and levels roughly per HEVC (OK for now - W0183?)
    - Do we need 4:2:2 profiles? (maybe not)
    - Do we need monochrome profiles? (maybe not)
    - Loop filter support requirement? (OK)
    - Palette mode and new coding tools proposed to be supported except in 10/12 b 4:2:0 & 10/12 b Mono profiles (OK)
      * (Some tools are already conditioned on b > 10)
* Film grain (optional mandatory, additional profile(s), or as alternative VVC-based example)
  + JVET-W0095, Proposed 8x8 HD, 16x16 4K, 32x32 8K
  + Possible tech report
  + Reference software
  + SMPTE relationship
  + VSEI vs VVC
  + Not necessarily for pre-filter removal with post-process restoration of grain
  + Current proposal is inclusion in VSEI as alternative example

## Joint meeting with MPEG WG3 0630–0700 Wednesday 14 July on Systems metadata

Topics of discussion:

Proposal m57457 is to use a single code for any system metadata rather than multiple SEI messages as in JVET-W0169. Systems doesn’t have a strong opinion on that.

Should this go in VSEI or in a video coding standard directly? Tentatively directly from the video coding standard.

What should it reference? Should it reference multiple systems standards (initially two)? The tentative answer is yes.

Should it be extensible? Yes.

Timeline? VDI planning to go to CD at this meeting, green metadata following shortly after.

Not planning to put it into VVC v2 DAM text; it should follow a bit later.

Conclusions:

* Codepoints for Green Metadata and VDI: Better define separate ones in VVC, and refer the corresponding systems specs directly
* WG3 would have the primary responsibility what is going in
* JVET should start a VVC amendment in October or January to align with the progression in WG3 (expecting this could be final at DAM stage)

## Joint meeting with with WGs 4, 5 and 7 0720–0820 Wednesday 14 July

* Presentation of m57393 on remote testing coordination of proposals by J. Jung
* Further work was expected to study various aspects
* The relationship with ITU subjective assessment Recommendations was discussed; nearly all aspects are well aligned – perhaps the forced-choice A-B comparison is less well established (5-grade comparison vs. 4-grade comparison)
* Not intended for formal subjective assessment, but rather for informal testingwas asked how well verified it is that a remote viewing test can match the results of a lab test; this is a major question for study, e.g., in VQEG. However, since testing in labs is currently not feasible, such studies are not immediately planned.
* Also, expert viewing tests would not necessarily always match naïve subject viewing in any case.

[Insert notes from M. Wien and summary info from document]

Presentation of draft verification test guidelines 57463 by M. Wien

* There was discussion of types of encoder optimization, depending on what is intended to be tested and demonstrated.
  + Having similar types of optimization vs independent optimization of each encoder separately
  + Similarity of configurations - e.g., intra period and GOP structures – in some cases a different GOP structure is intended to be part of what is being tested, and in other cases it is not
* A list was included of what prior verification tests have been conducted in recent years

[Further notes from M. Wien and summary info from document]

## BoGs (3)

The following break-out groups were established at this meeting to conduct discussion and develop recommendations on specific topics.

[JVET-W0180](https://jvet-experts.org/doc_end_user/current_document.php?id=11011) BoG report on high-throughput video coding [[M. G. Sarwer](mailto:m.sarwer@alibaba-inc.com)]

BoG reported Mon 12 July 0815 UTC

This is a report of activities from the BoG on high-throughput video coding. The BoG held meetings at the following times during the 23rd JVET meeting:

• July 9th 23:20 – July 10th 1.30

1 Mandates

The BoG was established with the following mandates:

- Discuss methods and criteria to analyse the entropy coding throughput.

- Discuss how to measure/analyse the worst-case throughput

- Discuss a common/standard method to measure throughput

2 Recommendation

BoG recommended to measure the required clock-rate using following method.

Compute clock rate for parsing:

• For context coded bins: 1.8/1.0

• For bypass coded bins without bypass alignment: 4

• For bypass coded bins with pass alignment: 12

Compute clock rate for debinarization:

• Compute the clock rate using non-flag symbols rate. Currently, it is assumed that 1 non-flag symbol can be processed per cycle and 1 flag symbol can be processed nearly at zero cycle.

Select the maximum clock rate between parsing and debinarization as the required clock rate.

BoG recommend attached excel template to report the result.

Further study is recommended to identify how many bypass coded bins can be processed in a cycle when the bypass alignment method is applied. It is expected to estimate the number of bypass bin per cycle before and after the bypass alignment. This can be study further for more accurate estimation.

Recommended action item:

The current VTM software does not support the output of symbol rate. The symbol rate of VTM was reported in JVET-W0052. It is asked the proponent of JVET-W0052 to provide the number of both flag and non-flag symbol rate for both VTM-13.0 and CABAC skip-mode.

Discussion about common methodology:

An excel template was presented, where the required clock rate is computed from the following statistics:

* The peak context coded bin rate (peakCtxBinRate)
* The peak bypass bin (i.e. EP) rate and (peakEPBinRate)
* The pick bitrate of a coded sequence.

The clock cycle is estimated as follows:

Clk Cycle in GHz = (peakCtxBinRate/CtxW + peakEPBinRate /EpW)/1000000000

Where,

* CtxW = Number of context coded bins can be parsed in one cycle
* EpW = Number of EP bins can be parsed in one cycle

It was commented (assumption) that a single core CABAC engine can process 1 GHz. It is expected that high bit depth professional application may need to have multiple CABAC process in parallel. One way to increase the throughput is using tile etc.

Another participant commented the benefits of having multiple tile-based technique to improve the throughput.

It was commented that the current VTM software does not support the output of symbol rate. The symbol rate of VTM was reported in JVET-W0052. It is asked the proponent of JVET-W0052 to provide the number of both flag and non-flag symbol rate for both VTM-13.0 and CABAC skip-mode.

De-binarization, Symbol rate, how to measure, set-upper limit?

One way can be to set maximum value based on the symbol rate?

One expert commented that debinarization is not so straight forward to measure. It is required to conduct more study on this. It is required to answer how many symbols per cycle can be parse in each cycle.

One participant commented to set a maximum value of GHz using the symbol rate.

It is also commented that when the proposals are evaluated it is also need to consider bin-to-bit ratio.

It was commented that there is a multiplication process in VVC CABAC (whereas HEVC CABAC there is no multiplication) which can also impact the throughput.

Measure the clock-rata using following method:

Compute clock rate for parsing:

* For context coded bins: 1.8/1.0
* For bypass coded bins without bypass alignment: 4
* For bypass coded bins with pass alignment: 12

Compute clock rate for debinarization:

* Compute the clock rate using non-flag symbols rate. Assume, 1 non-flag symbol can be processed per cycle.

Select the maximum clock rate between parsing and debinarization as the required clock rate.

Based on the Excel sheet, data are currently collected from the different proposals.

Further discussed in session 16, Tue 13 July 0510, after analysis of proposals had been compiled.

The brief description of all proposals is given below:

|  |  |
| --- | --- |
| **Proposal** | **Short Description** |
| **CE3.1** | Bypass alignment in CG level + sb\_coded\_flag bypass after maximum Ctx limit |
| **CE3.2** | Bypass alignment in TB level + bypass coding of all transform levels |
| **W0051** | CE3.2 + disable last position coding + bypass coding of TS flag + alignment in TU level |
| **W0052** | Fully skip the CABAC engine. |
| **W0114** | CE3.2 + CE1.1 + bypass coding of last position |
| **W0117** | CE3.2 + a slice level enable/disable flag to enable/disable the high throughput mode |
| **W0118** | CE3.2 + CE1.1 |
| **W0060-t1** | WPP: 1 |
| **W0060-t2** | WPP: 1 and CTU size: 64 |
| **W0060-t3** | WPP: 1 and Tile: 2x1 |

Table 1: Simulation results for 12 bits data, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
|  | wY | wU | wV | Y | U | V | Aver.GBR | G | B | R |
| **AI** | **CE3.1** | 1.13% | 1.18% | 1.14% | 0.87% | 0.93% | 0.94% | 0.56% | 0.48% | 0.59% | 0.60% |
| **CE3.2** | 10.33% | 12.57% | 12.70% | 6.34% | 8.09% | 8.59% | 1.63% | 1.51% | 1.70% | 1.68% |
| **W0051** | 17.07% | 20.69% | 20.51% | 8.17% | 9.75% | 10.20% | 1.64% | 1.76% | 1.74% | 17.07% |
| **W0052** | 12.05% | 13.80% | 13.96% | 7.61% | 8.17% | 8.82% | 3.09% | 3.40% | 2.94% | 2.92% |
| **W0114** | 12.43% | 14.37% | 14.66% | 7.14% | 8.42% | 8.94% | 1.76% | 1.76% | 1.76% | 1.75% |
| **W0117** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.22% | 0.17% | 0.24% | 0.24% |
| **W0118** | 10.31% | 12.55% | 12.66% | 6.15% | 7.9% | 8.37% | 1.32% | 1.18% | 1.41% | 1.39% |
| **W0060-t1** | 0.04% | 0.04% | 0.02% | 0.01% | 0.00% | 0.01% | 0.01% | 0.01% | 0.01% | 0.01% |
| **W0060-t2** | 0.12% | 0.02% | -0.06% | 0.07% | 0.02% | 0.02% | 0.02% | 0.03% | 0.02% | 0.01% |
| **W0060-t3** | 0.11% | 0.11% | 0.09% | 0.04% | 0.03% | 0.03% | 0.02% | 0.03% | 0.02% | 0.02% |
| **LDB** | **CE3.1** | 1.49% | 1.60% | 1.62% | 1.43% | 1.47% | 1.51% | 0.75% | 0.72% | 0.76% | 0.76% |
| **CE3.2** | 17.40% | 20.27% | 20.36% | 16.89% | 18.27% | 18.82% | 3.19% | 3.00% | 3.29% | 3.27% |
| **W0051** | 25.55% | 34.10% | 34.32% | 19.92% | 22.80% | 23.77% | 3.07% | 3.41% | 3.39% | 25.55% |
| **W0052** | 17.47% | 20.17% | 19.51% | 14.58% | 16.26% | 16.70% | 3.93% | 3.79% | 4.00% | 3.99% |
| **W0114** | 20.19% | 23.35% | 23.68% | 18.04% | 18.89% | 19.52% | 3.26% | 3.08% | 3.36% | 3.35% |
| **W0117** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.18% | 0.16% | 0.19% | 0.19% |
| **W0118** | 17.40% | 20.29% | 20.40% | 16.81% | 18.18% | 18.67% | 2.93% | 2.72% | 3.04% | 3.02% |
| **W0060-t1** | 0.09% | 0.09% | 0.03% | 0.08% | 0.04% | 0.00% | 0.01% | 0.02% | 0.01% | 0.01% |
| **W0060-t2** | 0.13% | 0.00% | -0.23% | 0.17% | 0.02% | -0.12% | 0.04% | 0.07% | 0.03% | 0.02% |
| **W0060-t3** | 0.22% | 0.21% | 0.12% | 0.12% | 0.10% | 0.01% | 0.03% | 0.04% | 0.02% | 0.02% |
| **RA** | **CE3.1** | 1.49% | 1.57% | 1.58% | 1.43% | 1.47% | 1.51% | 0.78% | 0.74% | 0.80% | 0.80% |
| **CE3.2** | 17.11% | 19.76% | 20.20% | 17.19% | 18.16% | 18.61% | 3.59% | 3.40% | 3.69% | 3.67% |
|  | **W0051** | 24.97% | 33.22% | 33.56% | 20.08% | 22.72% | 23.28% | 3.53% | 3.86% | 3.84% | 24.97% |
|  | **W0052** | 17.21% | 19.75% | 19.52% | 14.87% | 16.29% | 16.64% | 4.11% | 4.07% | 4.14% | 4.13% |
|  | **W0114** | 19.83% | 22.63% | 23.32% | 18.35% | 18.72% | 19.26% | 3.78% | 3.61% | 3.88% | 3.86% |
|  | **W0117** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.12% | 0.10% | 0.13% | 0.13% |
|  | **W0118** | 17.11% | 19.77% | 17.75% | 17.11% | 18.08% | 18.48% | 3.34% | 3.13% | 3.46% | 3.45% |
|  | **W0060-t1** | 0.11% | 0.08% | 0.02% | 0.08% | 0.05% | 0.01% | 0.01% | 0.02% | 0.01% | 0.01% |
|  | **W0060-t2** | 0.19% | 0.00% | -0.21% | 0.20% | 0.03% | -0.15% | 0.03% | 0.07% | 0.01% | 0.01% |
|  | **W0060-t3** | 0.24% | 0.21% | 0.11% | 0.10% | 0.08% | 0.03% | 0.03% | 0.04% | 0.03% | 0.03% |

Table 2: Simulation results for 16 bits data, HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **SVT16 RGB** | | | |
|  | Aver.GBR | G | B | R |
| **AI** | **CE3.1** | 0.31% | 0.28% | 0.33% | 0.33% |
| **CE3.2** | 0.07% | 0.06% | 0.09% | 0.07% |
| **W0051** | 0.05% | 0.06% | 0.05% | 0.03% |
| **W0052** | 1.92% | 2.13% | 1.84% | 1.79% |
| **W0114** | 0.19% | 0.26% | 0.17% | 0.15% |
| **W0117** | 0.08% | 0.06% | 0.09% | 0.07% |
| **W0118** | -0.21% | -0.23% | -0.19% | -0.20% |
| **W0060 test1** | 0.01% | 0.01% | 0.02% | 0.01% |
| **W0060 test2** | 0.03% | 0.03% | 0.03% | 0.02% |
| **W0060 test3** | 0.03% | 0.03% | 0.03% | 0.02% |
| **LDB** | **CE3.1** | 0.47% | 0.46% | 0.47% | 0.47% |
| **CE3.2** | 0.34% | 0.32% | 0.36% | 0.35% |
| **W0051** | 0.23% | 0.21% | 0.25% | 0.24% |
| **W0052** | 1.77% | 1.68% | 1.82% | 1.82% |
| **W0114** | 0.31% | 0.31% | 0.31% | 0.31% |
| **W0117** | 0.32% | 0.29% | 0.33% | 0.32% |
| **W0118** | 0.05% | 0.01% | 0.07% | 0.06% |
| **W0060 t1** | 0.02% | 0.01% | 0.02% | 0.02% |
| **W0060 t2** | 0.04% | 0.04% | 0.04% | 0.04% |
| **W0060 t3** | 0.03% | 0.03% | 0.03% | 0.03% |
| **RA** | **CE3.1** | 0.48% | 0.46% | 0.49% | 0.49% |
| **CE3.2** | 0.39% | 0.35% | 0.41% | 0.41% |
|  | **W0051** | 0.28% | 0.24% | 0.30% | 0.29% |
|  | **W0052** | 1.76% | 1.75% | 1.77% | 1.77% |
|  | **W0114** | 0.35% | 0.35% | 0.35% | 0.35% |
|  | **W0117** | 0.28% | 0.23% | 0.30% | 0.30% |
|  | **W0118** | 0.10% | 0.04% | 0.13% | 0.12% |
|  | **W0060 test1** | 0.01% | 0.02% | 0.01% | 0.01% |
|  | **W0060 test2** | 0.04% | 0.06% | 0.03% | 0.03% |
|  | **W0060 test3** | 0.03% | 0.04% | 0.03% | 0.03% |

Table 3: Reported run-time estimates for HBD/HBR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | **HDR HLG** | | **SVT12 RGB** | | **SVT16 RGB** | |
|  | Enc | Dec | Enc | Dec | Enc | Dec | Enc | Dec |
| **AI** | **CE3.1** | 100% | 98% | 100% | 99% | 101% | 99% | 101% | 93% |
| **CE3.2** | 91% | 88% | 84% | 83% | 86% | 81% | 85% | 76% |
| **W0051** | 96% | 75% | 86% | 72% | 84% | 78% | 83% | 73% |
| **W0052** | 87% | 82% | 89% | 70% | 82% | 63% | 80% | 59% |
| **W0114** | 93% | 87% | 87% | 78% | 83% | 76% | 81% | 71% |
| **W0117** | 101% | 102% | 101% | 102% | 94% | 94% | 85% | 80% |
| **W0118** | 93% | 91% | 88% | 84% | 84% | 80% | 81% | 74% |
| **W0060 t1** | 102% | 100% | 105% | 99% | 104% | 100% | 106% | 100% |
| **W0060 t2** | 101% | 99% | 103% | 98% | 105% | 99% | 106% | 99% |
| **W0060 t3** | 99% | 99% | 98% | 100% | 100% | 99% | 102% | 100% |
| **LDB** | **CE3.1** | 100% | 98% | 99% | 99% | 102% | 101% | 101% | 96% |
| **CE3.2** | 86% | 92% | 73% | 81% | 81% | 83% | 86% | 82% |
| **W0051** | 82% | 74% | 68% | 71% | 80% | 79% | 86% | 79% |
| **W0052** | 84% | 87% | 77% | 80% | 70% | 74% | 72% | 70% |
| **W0114** | 86% | 89% | 70% | 81% | 82% | 80% | 90% | 79% |
| **W0117** | 100% | 102% | 97% | 98% | 98% | 98% | 86% | 85% |
| **W0118** | 88% | 92% | 73% | 81% | 82% | 85% | 87% | 82% |
| **W0060 t1** | 102% | 100% | 107% | 99% | 108% | 101% | 106% | 100% |
| **W0060 t2** | 95% | 96% | 102% | 96% | 100% | 99% | 100% | 98% |
| **W0060 t3** | 95% | 100% | 99% | 99% | 101% | 99% | 102% | 99% |
| **RA** | **CE3.1** | 100% | 98% | 99% | 99% | 102% | 101% | 101% | 96% |
| **CE3.2** | 86% | 90% | 72% | 84% | 82% | 86% | 87% | 86% |
|  | **W0051** | 81% | 75% | 66% | 71% | 79% | 79% | 85% | 79% |
|  | **W0052** | 86% | 87% | 77% | 80% | 69% | 75% | 70% | 70% |
|  | **W0114** | 85% | 89% | 68% | 81% | 80% | 80% | 92% | 78% |
|  | **W0117** | 99% | 101% | 99% | 98% | 95% | 100% | 87% | 89% |
|  | **W0118** |  |  |  |  |  |  |  |  |
|  | **W0060 t1** | 103% | 99% | 105% | 99% | 103% | 99% | 98% | 98% |
|  | **W0060 t2** | 101% | 97% | 104% | 98% | 99% | 98% | 100% | 97% |
|  | **W0060 t3** | 96% | 100% | 100% | 100% | 101% | 99% | 101% | 98% |

Table 4: Simulation results for HBD/HBR CTC, lossless test configuration.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | bit-rate saving | | | | |
|  |  | HDR PQ | HDR HLG | SVT12 | SVT16 |
| **AI** | **CE3.1** | 0.55% | 0.38% | 0.29% | 0.20% |
| **CE3.2** | 2.91% | 0.93% | 0.42% | -0.29% |
| **W0051** | 3.04% | 0.93% | 0.37% | -0.32% |
| **W0052** | - | - | - | - |
| **W0114** | 3.10% | 1.19% | 0.28% | -0.11% |
| **W0117** | 0.22% | 0.86% | 0.42% | -0.29% |
| **W0118** | 2.54% | 0.59% | 0.16% | -0.47% |
| **W0060t1** |  |  |  |  |
| **W0060t2** |  |  |  |  |
| **W0060t3** |  |  |  |  |
| **LDB** | **CE3.1** | 0.67% | 0.47% | 0.39% | 0.26% |
| **CE3.2** | 6.11% | 1.29% | 0.39% | -0.28% |
| **W0051** | 7.10% | 1.16% | 0.29% | -0.35% |
| **W0052** | - | - | - | - |
| **W0114** | 6.17% | 1.33% | 0.39% | -0.29% |
| **W0117** | 0.01% | 1.27% | 0.39% | -0.28% |
| **W0118** | 5.69% | 0.85% | 0.11% | -0.48% |
| **W0060t1** |  |  |  |  |
| **W0060t2** |  |  |  |  |
| **W0060t3** |  |  |  |  |
| **RA** | **CE3.1** | 0.69% | 0.48% | 0.40% | 0.27% |
| **CE3.2** | 6.10% | 1.30% | 0.41% | -0.25% |
| **W0051** | 7.01% | 1.16% | 0.31% | -0.33% |
| **W0052** | - | - | - | - |
| **W0114** | 6.12% | 1.32% | 0.41% | -0.26% |
| **W0117** | 0.01% | 1.28% | 0.41% | -0.25% |
| **W0118** | 5.69% | 0.86% | 0.14% | -0.45% |
| **W0060t1** |  |  |  |  |
| **W0060t2** |  |  |  |  |
| **W0060t3** |  |  |  |  |

Table 5: Simulation results for 16 bits data, HBD/HBR CTC, LowQP test configuration, HM16.23 with High Throughput 4:4:4 Intra 16 profile.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **SVT16 RGB** | | | |
|  | Aver.GBR | G | B | R |
| **AI** | **CE3.1** | -1.82% | -2.40% | -1.53% | -1.55% |
| **CE3.2** | -2.05% | -2.61% | -1.76% | -1.80% |
| **W0051** | -2.61% | -1.80% | -1.84% | -2.61% |
| **W0052** | -0.26% | -0.61% | -0.05% | -0.12% |
| **W0114** | -1.94% | -2.42% | -1.69% | -1.72% |
| **W0117** | -2.05% | -2.61% | -1.76% | -1.80% |
| **W0118** | -2.33% | -2.89% | -2.03% | -2.06% |
| **W0060 t1** | -1.82% | -2.35% | -1.54% | -1.57% |
| **W0060 t2** | -1.81% | -2.34% | -1.53% | -1.55% |
| **W0060 t3** | -1.81% | -2.34% | -1.53% | -1.55% |

Table 6: Simulation results for CE3.x tests, 12 bits data, HBD/HBR CTC, LowQP test configuration, HM16.23 with High Throughput 4:4:4 Intra 16 profile

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
|  | wY | wU | wV | Y | U | V | Aver.GBR | G | B | R |
| **AI** | **CE3.1** | -7.03% | -9.37% | -10.06% | -4.62% | -6.50% | -6.67% | -2.27% | -4.03% | -2.86% | -2.94% |
| **CE3.2** | 1.15% | 0.51% | -0.21% | 0.54% | 0.16% | 0.42% | -2.24% | -3.04% | -1.79% | -1.89% |
| **W0051** | 7.32% | 7.72% | 6.62% | 2.32% | 1.82% | 2.02% | -2.92% | -1.73% | -1.83% | -2.92% |
| **W0052** | 2.71% | 1.59% | 0.91% | 1.70% | 0.18% | 0.58% | -0.88% | -1.31% | -0.61% | -0.71% |
| **W0114** | 3.06% | 2.12% | 1.52% | 1.30% | 0.48% | 0.74% | -2.81% | -1.73% | -1.83% | -2.81% |
| **W0117** | -8.05% | -10.41% | -11.06% | -5.44% | -7.36% | -7.54% | -3.60% | -4.32% | -3.20% | -3.28% |
| **W0118** | 1.14% | 0.49% | -0.23% | 0.36% | -0.01% | 0.21% | -2.54% | -3.36% | -2.07% | -2.17% |
| **W0060 test1** | -7.29% | -9.66% | -10.34% | -4.73% | -6.68% | -6.87% | -3.29% | -3.97% | -2.90% | -2.99% |
| **W0060 test2** | -7.22% | -9.67% | -10.38% | -4.67% | -6.66% | -6.85% | -3.27% | -3.95% | -2.89% | -2.98% |
|  | **W0060 test3** | -7.24% | -9.61% | -10.29% | -4.70% | -6.65% | -6.84% | -3.27% | -3.95% | -2.89% | -2.98% |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **Required clock rate** | | | | | | | |
|  | **1.8 Context coded bins per cycle** | | | | **1.0 Context coded bins per cycle** | | | |
|  | **PQ** | **HLG** | **SVT12** | **SVT16** | **PQ** | **HLG** | **SVT12** | **SVT16** |
|  | **VTM-13.0** | 0.592 GHz | 0.885 GHz | 0.935 GHz | 1.220 GHz | 0.887 GHz | 1.250 GHz | 1.236 GHz | 1.531 GHz |
|  | % Ratio as compared to VTM-13.0 | | | | % Ratio as compared to VTM-13.0 | | | |
| **AI** |  | PQ | HLG | SVT12 | SVT16 | PQ | HLG | SVT12 | SVT16 |
| **CE3.1** | 76% | 69% | 61% | 55% | 85% | 78% | 71% | 64% |
| **CE3.2** | 31% | 40% | 36% | 32% | 26% | 29% | 29% | 30% |
| **W0051** | 31% | 40% | 36% | 29% | 21% | 28% | 28% | 24% |
| **W0052** | 31% | 40% | 36% | 29% | 21% | 28% | 28% | 23% |
| **W0114** | 31% | 40% | 36% | 29% | 23% | 28% | 28% | 24% |
| **W0117** | - | - | - | 32% | - | - | - | 30% |
| **W0118** | 31% | 40% | 36% | 29% | 24% | 28% | 28% | 26% |
| **W0060 t1** | 100% / 25% | 100% / 25% | 100% / 25% | 100% / 25% | 100% / 25% | 100% / 25% | 100% / 25% | 100% / 25% |
| **W0060 t2** | 100% / 25% | 100% / 25% | 100% / 25% | 100% / 25% | 100% / 25% | 100% / 25% | 100% / 25% | 100% / 25% |
| **W0060 t3** | 100% / 25% | 100% / 25% | 100% / 25% | 100% / 25% | 100% / 25% | 100% / 25% | 100% / 25% | 100% / 25% |

The numbers above consider the throughput concerning the frame with maximum number of bins as found from the tests.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **%diff (Bin to bit ratio)** | | | |
| PQ | HLG | SVT12 | SVT16 |
| **AI** | **CE3.1** | 0% | 1% | 1% | 0% |
| **CE3.2** | -29% | -20% | -11% | -11% |
| **W0051** | -34% | -25% | -15% | -14% |
| **W0052** | -36% | -26% | -16% | -15% |
| **W0114** | -30% | -22% | -15% | -14% |
| **W0117** | - | - | - | -11% |
| **W0118** | -28% | -18% | -14% | -13% |
| **W0060 t1** | 0% | 0% | 0% | 0% |
| **W0060 t2** | -1% | 0% | 0% | 0% |
| **W0060 t3** | 0% | 0% | 0% | 0% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **16-bit 4:4:4: Worst case data, MinCR = 2**  **Level 5.1**  **Max luma sample rate: 534,773,760** | | | |
|  | **1.8 Context coded bins per cycle** | | **1.0 Context coded bins per cycle** | |
|  | **Required clock rate (GHz)** | **Ratio** | **Required clock rate (GHz)** | **Ratio** |
| **AI** | **VTM-13.0** | 4.565 | 100% | 5.725 | 100% |
| **CE3.1** | 2.513 | 55% | 3.692 | 64% |
| **CE3.2** | 1.882 | 41% | 1.882 | 33% |
| **W0051** | 1.882 | 41% | 1.882 | 33% |
| **W0052** | 1.882 | 41% | 1.882 | 33% |
| **W0114** | 1.882 | 41% | 1.882 | 33% |
| **W0117** | 1.882 | 41% | 1.882 | 33% |
| **W0118** | 1.882 | 41% | 1.882 | 33% |
| **W0060 test1** | 4.565 / 1.141 | 100% / 25% | 5.726 / 1.431 | 100% / 25% |
| **W0060 test2** | 4.565 / 1.141 | 100% / 25% | 5.726 / 1.432 | 100% / 25% |
| **W0060 test3** | 4.565 / 1.141 | 100% / 25% | 5.727 / 1.432 | 100% / 25% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **16-bit 4:4:4: Worst case data, MinCR = 2**  **Level 6.2**  **Max luma sample rate: 4,278,190,080** | | | |
|  | **1.8 Context coded bins per cycle** | | **1.0 Context coded bins per cycle** | |
|  | **Required clock rate (GHz)** | **Ratio** | **Required clock rate (GHz)** | **Ratio** |
| **AI** | **VTM-13.0** | 36.517 | 100% | 45.800 | 100% |
| **CE3.1** | 20.101 | 55% | 29.533 | 64% |
| **CE3.2** | 15.057 | 41% | 15.057 | 33% |
| **W0051** | 15.057 | 41% | 15.057 | 33% |
| **W0052** | 15.057 | 41% | 15.057 | 33% |
| **W0114** | 15.057 | 41% | 15.057 | 33% |
| **W0117** | 15.057 | 41% | 15.057 | 33% |
| **W0118** | 15.057 | 41% | 15.057 | 33% |
| **W0060 test1** | 36.521 / 9.130 | 100% / 25% | 45.808 / 11.452 | 100% / 25% |
|  | **W0060 test2** | 36.521 / 9.130 | 100% / 25% | 45.808 / 11.452 | 100% / 25% |
|  | **W0060 test3** | 36.523 / 9.131 | 100% / 25% | 45.814 / 11.453 | 100% / 25% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **12-bit 4:2:2: Worst case data, MinCR = 2**  **Level 5.1**  **Max luma sample rate: 534,773,760** | | | |
|  | **1.8 Context coded bins per cycle** | | **1.0 Context coded bins per cycle** | |
|  | **Required clock rate (GHz)** | **Ratio** | **Required clock rate (GHz)** | **Ratio** |
| **AI** | **VTM-13.0** | 3.498 | 100% | 4.919 | 100% |
| **CE3.1** | 2.378 | 68% | 3.823 | 78% |
| **CE3.2** | 1.517 | 43% | 1.517 | 31% |
| **W0051** | 1.517 | 43% | 1.517 | 31% |
| **W0052** | 1.517 | 43% | 1.517 | 31% |
| **W0114** | 1.517 | 43% | 1.517 | 31% |
| **W0117** | - | - | - | - |
| **W0118** | 1.517 | 43% | 1.517 | 31% |
| **W0060 test1** | 3.498 / 0.875 | 100% / 25% | 4.921 / 1.230 | 100% / 25% |
| **W0060 test2** | 3.441 / 0.860 | 98% / 25% | 4.837 / 1.209 | 98% / 25% |
| **W0060 test3** | 3.485 / 0.871 | 100% / 25% | 4.903 / 1.226 | 100% / 25% |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **12-bit 4:2:2: Worst case data, MinCR = 2**  **Level 6.2**  **Max luma sample rate: 4,278,190,080** | | | | | | | | |
|  | **1.8 Context coded bins per cycle** | | | | **1.0 Context coded bins per cycle** | | | | |
|  | **Required clock rate (GHz)** | **Ratio** | | | **Required clock rate (GHz)** | | | **Ratio** | |
| **AI** | **VTM-13.0** | 27.982 | | 100% | | 39.354 | | | | 100% |
| **CE3.1** | 19.024 | | | 68% | | 30.584 | 78% | | |
| **CE3.2** | 12.134 | | | 43% | | 12.134 | 31% | | |
| **W0051** | 12.134 | | | 43% | | 12.134 | 31% | | |
| **W0052** | 12.134 | | | 43% | | 12.134 | 31% | | |
| **W0114** | 12.134 | | | 43% | | 12.134 | 31% | | |
| **W0117** | - | | | - | | - | - | | |
| **W0118** | 12.134 | | | 43% | | 12.134 | 31% | | |
| **W0060 test1** | 27.988 / 6.997 | | | 100% / 25% | | 39.369 / 9.842 | 100% / 25% | | |
|  | **W0060 test2** | 27.525 / 6.881 | | | 98% / 25% | | 38.700 / 9.675 | 98% / 25% | | |
|  | **W0060 test3** | 27.883 / 6.971 | | | 100% / 25% | | 39.225 / 9.806 | 100% / 25% | | |

Significant losses are observed for some proposals, in particular for HDR 12 bit data.

Even for the proposals that completely skip CABAC for transform coefficient coding, the necessary clock rate reduction is limited, due to debinarization.

Question: What would be the throughput data for HEVC high throughput profile? Would require more analysis.

For W0060 tests, it is assumed that 4 processing threads are run in parallel when the lower numbers (righthand) were computed. This is a straightforward solution of the throughput problem. However, this would mean that the cost of equipment would be increased.

For W0117, the slice-level switching was not used in cases of 12 bit.

It is commented that it is surprising that CE3.2 and some other proposals have exactly the same numbers in terms of throughput. This however applies to worst case.

HEVC high throughput profiles are enabling WPP, except the 16 bit intra case, where bypass alignment method is mandatory.

Considering that for this range of applications, cost may not be of high concern, using WPP with multiple threads appears to be a straightforward solution. One concern in this context may be decoding latency. With 4 threads, the latency would come with 4 CTU rows, which is roughly 1/8 of picture height in case of 4K.

It is mentioned that low latency may be highly important for medical application.

Agreement: For a possible high throughput profile at this stage of definition (DIS), it would be most appropriate using the solution suggested in JVET-W0060 test 1 (making WPP mandatory).

[JVET-W0182](https://jvet-experts.org/doc_end_user/current_document.php?id=11013) BoG Report: Neural Networks Video Coding Analysis and Planning [A. Segall]

Initial report (verbally) Monday 12 July 0830, follow up presenting v2 of the report in session 16 0630 UTC

The BoG held the following meetings during the 23rd JVET meeting:

* July 9 – 23:20 – July 10 01:20
* July 12 – 23:20 – July 13 01:20

Recommendations from the BoG are summarized as:

* + Review the summary analysis prepared by the BoG
  + Include an indication of current hardware performance when summarizing performance in future EE reporting.
  + AhG mandate to study and collect information related to near-term and long-term architectures for neural-network video coding.
  + Encourage members of hardware companies to comment.

The BoG was established with the following mandates:

* Perform further analysis about the complexity/compression trade-offs of the various loop filter and super resolution proposals and align the reports with AHG11 reporting conditions
* Perform further analysis between luma and chroma gains in detail
* Suggest candidates for an upcoming EE
* Refine the AHG11 reporting conditions as necessary

The following questions were proposed to be considered in performing the analysis:

1. How should we align test results that did not use the AHG11 reporting conditions?
   1. Used VTM-11.0
   2. Used a different QP range
   3. Results were incomplete
2. How should we summarize the different luma and chroma performances that are being observed?
   1. One potential solution - use a weighted combination of the Y, Cb and Cr BD-rates. For example, AHG13 previously used 6:1:1

The BoG agreed that these were the main questions to be answered.

***Discussion***

**On results that used VTM-11.0 as an anchor**

One participant commented the use of VTM-11.0 as an anchor would not affect the all-intra results. Additionally, it was commented that the random-access performance would not be affected for all-intra tools in this case. A second participant agreed with the above.

After subsequent discussion (below), the group decided to remove the random-access results for proposals that used a VTM-11.0 anchor for reporting the summary results.

One participant commented that JVET-W0062 used VTM 12.0 with ALF and SAO disabled as an anchor. It was decided to re-compute the BD-Rate performance compared to the VTM-11.0-nnvc anchor.

**On the different QP range:**

One participant suggested to compute the BD-Rate performance of all proposals over the range of QP27, 32, 37, 42. This would just be for 4K only results, as all of the other results used a QP range of QP22, 27, 32, 37, 42.

One participant suggested to prepare an additional plot that attempts to compare the super-resolution and in-loop filtering technologies.

One participant suggested that super-resolution method should be required to bring all CTC results going forward. It was further suggested that the performance of super-resolution methods should be considered after in-loop filters.

The following plan was put forward for reporting:

1. Create a summary for all tools using the NNVC CTC. This would exclude proposals that did not provide results for all sequences and QP points in the CTC.
2. Create a summary of all tools for 4K using the NNVC CTC but with a QP range of QP27, 32, 37, 42. This could then include more of the proposals, as some of the super-resolution methods did not report results for QP22.
3. Create a summary of super-resolution tools for 4K using CTC but with a QP range of QP27, 32, 37, 42, 47

The group decided on the above strawman approach.

**On incomplete results**

The group decided to remove proposals from the summary that had incomplete results.

**On chroma results**

One participant recommended that the group should inspect the chroma BD-rate curves and identify cases where they were crossing. And, if the curves are crossing – the chroma results should be ignored.

Multiple participants expressed concern on reporting a metric that was averaged over the luma and chroma channels. It was proposed to report the luma and chroma results separately.

Multiple participants expressed support for reporting a combined metric to the group.

The group decided to report set of luma results and a set of chroma results, where the chroma results would be the average of the Cb/Cr channels. In addition, the group will check for cases where the chroma curves are crossing – and create a note in these cases that the results are unreliable.

**On complexity reporting**

One participant raised a question about complexity reporting in the case that a proposal used multiple models that are selectively enabled. The question was if the summary should use the worst-case performance or a different metric.

One participant expressed that worst-case performance should be reported.

Multiple participants expressed support for reporting the decoding complexity and average complexity in terms of run-time.

One participant expressed support for capturing more detailed information on complexity. This could include latency.

One participant recommended that the summary output of the BoG should also capture memory size.

One participant recommended that the summary capture if MAC operations use a floating point or integer operation.

Multiple participants recommended that the group study what is a realistic complexity and architecture for the near-term and long-term.

The group recommended that this should be a mandate of the NNVC AhG group. And, that the discussion should be incorporated into the EE design as much as possible.

The group decided to report worst-case complexity in the summary.

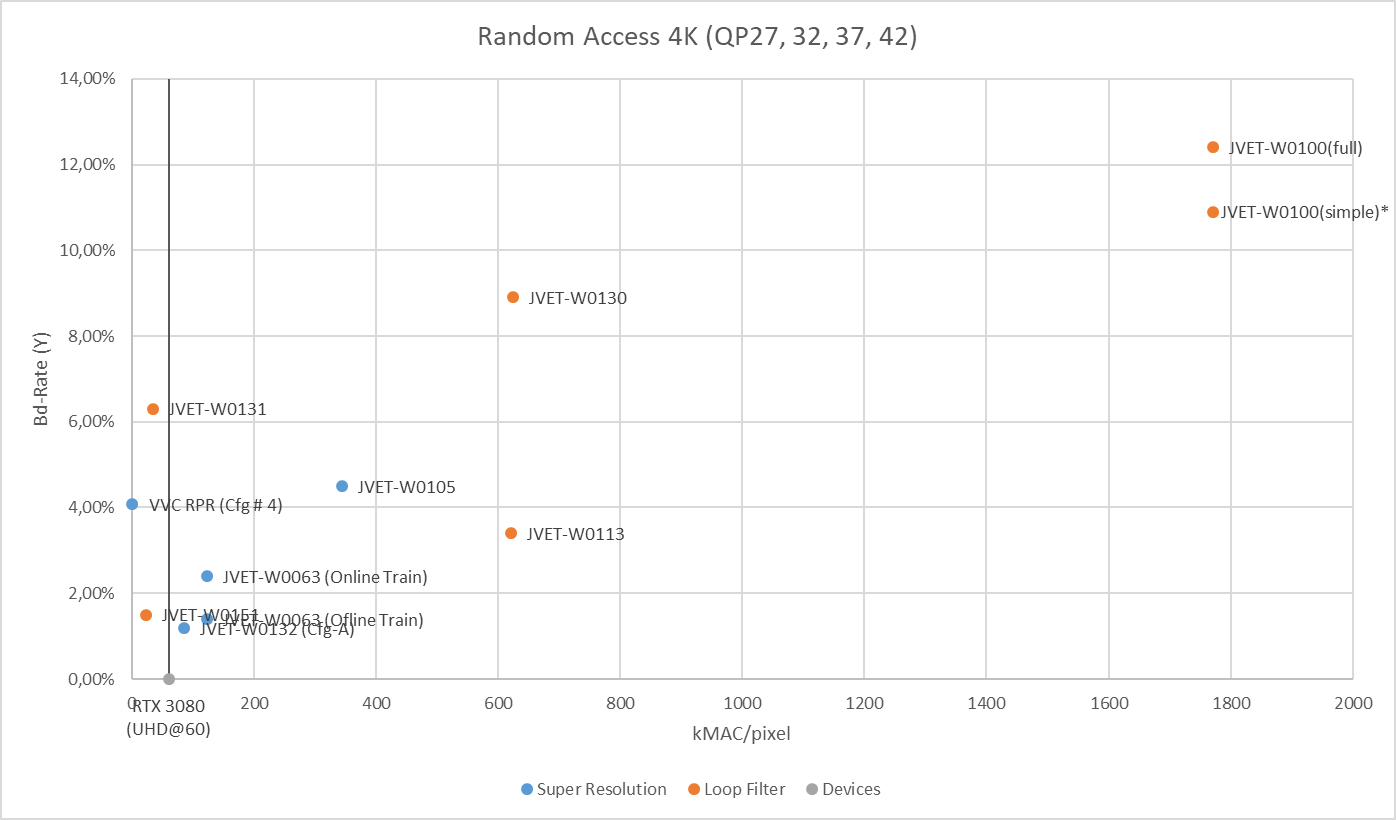
The group decided to capture the precision of the calculations in the summary.

The group decided to capture the memory size (number of parameters \* the precision of the parameters) of all models being used in the proposals.

Next step – off-line activity to begin preparing the aligned summary results. Following the activity, review and re-visit the summary plan.

Plan summary results and plan were revised in the July 12th meeting.

The BoG report also includes a detailed analysis of the different proposals in an attached xls (including plots on number of operations, amount of memory vs. compression gain, and chroma vs. luma). An example is given below (it is noted that this is just meant as an example from the initial presentation, and may be subject to further changes in the follow-up analysis of the BoG).



The benefit in terms of BD gain seems to increase roughly linearly with the number of operations. Goal would be to shift the points more to left/top (better tradeoff complexity vs. compression).

The chroma analysis for superresolution seems to indicate that the losses are quite non-uniform, e.g. most contributions seem to lose most in park running.

**Exploration Experiment Planning**

The following issues were proposed to be considered when developing the EE test:

1. Should the proposals in the EE be in integer precision?
2. What is the realistic complexity that should be targeted by the EE?
3. What is the near-term and long-term architecture that should be targeted by the EE?
4. How should the group work toward a common software?
5. What tests should be performed?
6. What technologies should be included in the test?
7. What is the timeline for the EE?

The group agreed that these were the main topic to be discussed.

**On integer precision**

One participant suggested that proposals in the EE should be implemented in integer precision.

Multiple participants agreed with the above the statement.

Multiple participants expressed concerns about making integer precision mandatory in the next EE cycle.

It was commented that porting a floating-point implementation to a 32-bit integer implementation may be straightforward.

One participant suggested that the memory size of the proposals should be reported and prioritized to encourage participants to transition to integer implementations.

**Realistic Complexity**

One participant suggested that the group identify the target resolution and frame rate of the activity.

Multiple participants recommended that UHD@60p and 8K@60p should be key targets.

A summary of some capabilities of available devices was provide in JVET-W0131:

All number is based on public information

•Snapdragon 865: peak 15 trillion operation per second (TOPS)

•Snapdragon 888: peak 26 TOPS

•A14 Bionic: 11 TOPS

•M1X: may manage around 5.2 TFLOPS.

•2080 TI (Single GPU core): around 11 TFLOPS,

•GeForce RTX 3080(Single GPU core): up to 30 TFLOPS

It was commented that one could expect that device performance will increase in the future. But, at the same time, it would be unlikely that the entire device capability would be available for decoding.

One participant commented that it could be beneficial to capture and plot the device performance as part of result reporting. Multiple participants agreed with this idea.

One participant proposed to capture the performance for UHD@60p and a second for 8K@60p.

The group recommended to include an indication of current hardware performance when summarizing performance in future EE reporting.

One participant introduced a paper that would be relevant for this work: Sun et. al., “Summarizing CPU and GPU Design Trends with Product Data “

One participant suggested that the memory size of current devices should also be captured and plotted when summarizing performance in future EE reporting. Here, the memory size would be the number of bytes needed to store the parameters.

**Near-term and long-term architectures**

One participant commented that it would be beneficial to discuss with other hardware companies if general GPU/TPU resources can be used in their decoding pipeline.

One participant commented that it was unlikely in their architecture to use GPU/TPU resources if the coding tool is in the loop.

One participant commented that technologies that are completely pre- and post-processing solutions could use GPU/TPU resources.

The group recommended the AhG study and collect information related to near-term and long-term architectures. And, that members of hardware companies be encouraged to comment.

**Common software**

It was commented that JVET-W0181 was related to this topic.

One participant suggested that having a technology that used the JVET-W0181 framework in the EE could be beneficial.

One participant suggested that the common software should be limited to inference.

One participant commented that common software should help with cross checking activities. But, a first issue in performing cross checks is the lack of integer implementations.

The BoG should meet again to refine the analysis and continue on defining next EE, including refinements of reporting and CTC.

Further BoG reporting Wed 0830 (JRO & GJS)

The BoG had met further. Some complexity estimates had been refined. Current hardware performance reporting was suggested by the BoG for further work; also reporting to require reports to use uniform reporting. EE planning with cross-checking using a two-step cross-checking had been conducted. There is expected to be more emphasis on cross-checking in the future. In-loop versus post-processing comparative result reporting is planned. About 10 proposals were planned to be tested in the next round of EE.

Recommendations from the BoG [insert from BoG report] were reviewed and approved.

Conformance planning for v2 profiles

Discussed 0850. Plans were presented (9 added profiles).

[JVET-W0188](https://jvet-experts.org/doc_end_user/current_document.php?id=11019) BoG Report: VVCv2 Conformance Testing [D. Rusanovskyy]

Was presented (see under 7.1)

## Liaison communications (update)

The JVET did not directly receive or send any liaison statements at its current meeting. However, there was some related liaison communication between ITU-T SG 16, the AGs and WGs of ISO/IEC JTC 1/‌SC 29, and ITU-T Study Group 12 that were coordinated with JVET.

This included exchange of general status information about JVET work and management arrangements for video coding collaboration between the parent bodies of WP 3/16 and SC 29. SC 29 had sent a liaison statement to SG16 [[TD535/Gen](http://www.itu.int/md/meetingdoc.asp?lang=en&parent=T17-SG16-210419-TD-GEN-0535) / [SC29 N 19023](file:///C:\GarySull2\OneDrive%20-%20Microsoft\2021-04-telecons\2021-04-SG16-telecon\Gen\ISO\IEC%20JTC1\SC29-N19023)], and SG 16 prepared a reply.

ITU-T SG12 had sent a liaison statement [m56363](https://dms.mpeg.expert/doc_end_user/current_document.php?id=78266&id_meeting=186) to ISO/IEC JTC 1/‌SC 29/‌AG 5 (in reply to SC 29 N 19285, MDS 19755 from October 2020). SG12 expressed their strong interest in exchanging information and collaborating on aspects of video quality assessment. SG12 also welcomed the information received about the status of the verification tests of Versatile Video Coding (ITU-T H.266 and ISO/IEC 23090-3 VVC) conducted with SG16 in the Joint Video Experts Team (JVET). In light of possible future extensions of existing video quality models and the development of new assessment methods, SG12 expressed their particular interest in further exchanges about this and similar assessment campaigns. SC 29/AG 5 (Visual Quality Testing) prepared a reply liaison statement to ITU-T SG12 and ITU-R WP6C as document AG 05 N 23 that included a description of recent verification testing of VVC and further plans for such testing.

SC 29/‌WG 1 had recently sent several liaison statements to ITU-T SG16 [[TD537/Gen](http://www.itu.int/md/meetingdoc.asp?lang=en&parent=T17-SG16-210419-TD-GEN-0537) / [SC 29/WG 1 N 89052 (SC 29 N 19202)](https://www.itu.int/ifa/t/2017/ls/isoiecjtc1sc29wg1/sp16-iso_iecjtc1_sc29_wg1-iLS-00022.zip), [TD581/Gen](http://www.itu.int/md/meetingdoc.asp?lang=en&parent=T17-SG16-210419-TD-GEN-0581) / [SC 29/WG 1 N 90081 (SC 29 N 19574)](http://handle.itu.int/11.1002/ls/sp16-iso_iecjtc1_sc29_wg1-iLS-00025.pdf), and [TD600/Gen](https://www.itu.int/md/meetingdoc.asp?lang=en&parent=T17-SG16-210419-TD-GEN-0600) / [SC 29/WG 1 N 91067](https://www.itu.int/ifa/t/2017/ls/isoiecjtc1sc29wg1/sp16-iso_iecjtc1_sc29_wg1-iLS-00026.docx)] that included discussion of neural network image coding, and Gary Sullivan and Elena Alshina of JVET were given the action item to prepare status information about the neural network video coding exploration in JVET to be included in a liaison letter reply from ITU-T SG16 to SC 29/WG 1.

ITU-T SG13 had sent a liaison statement [m56366](https://dms.mpeg.expert/doc_end_user/current_document.php?id=78269&id_meeting=186) to SC 29 with an invitation to review an Artificial Intelligence Standardization Roadmap and provide missing or updated information. SC 29/AG 3 (Liaison and Communications) prepared a reply as document AG 3 N 28 that included a mention of the JVET exploration experiment work on neural network based video coding.

# Project planning

## Software timeline (update)

ECM2 (including all adoptions) will be available 2 weeks after the meeting. F. Le Léannec and J. Chen will be additional coordinators (in addition to V. Seregin and K. Zhang).

VTM14 will be available on 08-09.

## Core experiment and exploration experiment planning

A CE on entropy coding for high bit depths and high bit rates was established, as recorded in output document JVET-U2022.

An EE on neural network-based video coding was established, as recorded in output document JVET-U2023.

An EE on enhanced compression technology beyond VVC capability using techniques other than neural-network technology was also established, as recorded in output document JVET-U2024.

Initial versions of these documents were presented and approved in the plenary on Friday 15 January.

## Drafting of specification text, encoder algorithm descriptions, and software

The following agreement has been established: the editorial team has the discretion to not integrate recorded adoptions for which the available text is grossly inadequate (and cannot be fixed with a reasonable degree of effort), if such a situation hypothetically arises. In such an event, the text would record the intent expressed by the committee without including a full integration of the available inadequate text.

## Plans for improved efficiency and contribution consideration

The group considered it important to have the full design of proposals documented to enable proper study.

Adoptions need to be based on properly drafted working draft text (on normative elements) and HM/VTM encoder algorithm descriptions – relative to the existing drafts. Proposal contributions should also provide a software implementation (or at least such software should be made available for study and testing by other participants at the meeting, and software must be made available to cross-checkers in EEs).

Suggestions for future meetings included the following generally-supported principles:

* No review of normative contributions without draft specification text
* VTM algorithm description text is strongly encouraged for non-normative contributions
* Early upload deadline to enable substantial study prior to the meeting
* Using a clock timer to ensure efficient proposal presentations (5 min) and discussions

The document upload deadline for the next meeting was planned to be Tuesday 13 April 2021.

As general guidance, it was suggested to avoid usage of company names in document titles, software modules etc., and not to describe a technology by using a company name.

## General issues for experiments

It was emphasized that those rules which had been set up or refined during the 12th JVET meeting should be observed. In particular, for some CEs of some previous meetings, results were available late, and some changes in the experimental setup had not been sufficiently discussed on the JVET reflector.

Group coordinated experiments have been planned as follows:

* “Core experiments” (CEs) are the coordinated experiments on coding tools which are deemed to be interesting but require more investigation and could potentially become part of a draft standard by the next meeting or in the near future.
* “Exploration experiments” (EEs) are also coordinated experiments. These are conducted on technology which is not foreseen to become part of a draft standard in near future. Investigating methodology for assessment of such technology can also be an important part of an EE. (Further general rules for EEs, as far as deviating from the CE rules below, should be discussed in a future meeting. For the current meeting, procedures as described in the EE description document are deemed to be sufficient)
* A CE is a test of a specific fully described technology in a specific agreed way. It is not a forum for thinking of new ideas (like an AHG). The CE coordinators are responsible for making sure that the CE description is complete and correct and has adequate detail. Reflector discussions about CE description clarity and other aspects of CE plans are encouraged.
* A description of each experiment is to be approved at the meeting at which the experiment plan is established. This should include the issues that were raised by other experts when the tool was presented, e.g., interference with other tools, contribution of different elements that are part of a package, etc. The experiment description document should provide the names of individual people, not just company names.
* Software for tools investigated in a CE will be provided in one or more separate branches of the software repository. Each CE will have a “fork” of the software, and within the CE there may be multiple branches established by the CE coordinator. The software coordinator will help coordinate the creation of these forks and branches and their naming. All JVET members will have read access to the CE software branches (using shared read-only credentials as described below).
* During the experiment, revisions of the experiment plans can be made, but not substantial changes to the proposed technology.
* The CE description must match the CE testing that is done. The CE description needs to be revised if there has been some change of plans.
* The CE summary report must describe any changes that were made in the process of finalizing the CE.
* By the next meeting it is expected that at least one independent cross-checker will report a detailed analysis of each proposed feature that has been tested and confirm that the implementation is correct. Commentary on the potential benefits and disadvantages of the proposed technology in cross-checking reports is highly encouraged. Having multiple cross-checking reports is also highly encouraged (especially if the cross-checking involves more than confirmation of correct test results). The reports of cross-checking activities may (and generally should) be integrated into the CE report rather than submitted as separate documents.

It is possible to define sub-experiments within particular CEs, for example designated as CEX.a, CEX.b, etc., where X is the basic CE number.

As a general rule, it was agreed that each CE should be run under the same testing conditions using one software codebase, which should be based on the group test model software codebase. An experiment is not to be established as a CE unless there is access given to the participants in (any part of) the CE to the software used to perform the experiments.

The general agreed common conditions for single-layer coding efficiency experiments for SDR video are described in the prior output document JVET-T2010.

Experiment descriptions should be written in a way such that it is understood as a JVET output document (written from an objective “third party perspective”, not a proponent perspective – e.g. not referring to methods as “improved”, “optimized”, etc.). The experiment descriptions should generally not express opinions or suggest conclusions – rather, they should just describe what technology will be tested, how it will be tested, who will participate, etc. Responsibilities for contributions to CE work should identify individuals in addition to company names.

CE descriptions contain a basic description of the technology under test, but should not contain excessively verbose descriptions of a technology (at least not unless the technology is not adequately documented elsewhere). Instead, the CE descriptions should refer to the relevant proposal contributions for any necessary further detail. However, the complete detail of what technology will be tested must be available – either in the CE description itself or in documents that are referenced in the CE description that are also available in the JVET document archive.

Any technology must have at least one cross-check partner to establish a CE – a single proponent is not enough. It is highly desirable have more than just one proponent and one cross-checker.

The CE development workflow is described at:

<https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/wikis/Core-experiment-development-workflow>

CE read access is available using shared accounts: One account exists for MPEG members, which uses the usual MPEG account data. A second account exists for VCEG members with account information available in the TIES system at:

<https://www.itu.int/ifa/t/2017/sg16/exchange/wp3/q06/vceg_account.txt>

Some agreements relating to CE activities were established as follows:

* Only qualified JVET members can participate in a CE.
* Participation in a CE is possible without a commitment of submitting an input document to the next meeting. Participation is requested by contacting the CE coordinator.
* All software, results, and documents produced in the CE should be announced and made available to JVET in a timely manner.
* A JVET CE reflector will be established and announced on the main JVET reflector. Discussion of logistics arrangements, exchange of data, minor refinement of the test plans, and preparation of documents shall be conducted on the JVET CE reflector, with subject lines prefixed by “[CEx: ]”, where “x” is the number of the CE. All substantial communications about a CE other than such details shall take place on main JVET reflector. In the case that large amounts of data are to be distributed, it is recommended to send a link to the data rather than the data itself, or upload the data as an input contribution to the next meeting.

General timeline for CEs

T1= 3 weeks after the JVET meeting: To revise the CE description and refine questions to be answered. Questions should be discussed and agreed on JVET reflector. Any changes of planned tests after this time need to be announced and discussed on the JVET reflector. Initially assigned description numbers shall not be changed later. If a test is skipped, it is to be marked as “withdrawn”.

T2 = Test model software release + 2 weeks: Integration of all tools into a separate CE branch of the VTM is completed and announced to JVET reflector.

* Initial study by cross-checkers can begin.
* Proponents may continue to modify the software in this branch until T3.
* 3rd parties are encouraged to study and make contributions to the next meeting with proposed changes

T3: 3 weeks before the next JVET meeting or T2 + 1 week, whichever is later: Any changes to the CE test branches of the software must be frozen, so the cross-checkers can know exactly what they are cross-checking. A software version tag should be created at this time. The name of the cross-checkers and list of specific tests for each tool under study in the CE plan description shall be documented in an updated CE description by this time.

T4: Regular document deadline minus 1 week: CE contribution documents including specification text and complete test results shall be uploaded to the JVET document repository (particularly for proposals targeting to be promoted to the draft standard at the next meeting).

The CE summary reports shall be available by the regular contribution deadline. This shall include documentation about crosscheck of software, matching of CE description and confirmation of the appropriateness of the text change, as well as sufficient crosscheck results to create evidence about correctness (crosscheckers must send this information to the CE coordinator at least 3 days ahead of the document deadline). Furthermore, any deviations from the timelines above shall be documented. The numbers used in the summary report shall not be changed relative to the description document.

CE reports may contain additional information about tests of straightforward combinations of the identified technologies. Such supplemental testing needs to be clearly identified in the report if it was not part of the CE plan.

New branches may be created which combine two or more tools included in the CE document or the VTM (as applicable).

It is not necessary to formally name cross-checkers in the initial version of the CE description document. To adopt a proposed feature at the next meeting, we would like see comprehensive cross-checking done, with analysis that the description matches the software, and recommendation of value of the tool given tradeoffs.

The establishment of a CE does not indicate that a proposed technology is mature for adoption or that the testing conducted in the CE is fully adequate for assessing the merits of the technology, and a favourable outcome of CE does not indicate a need for adoption of the technology into a standard.

Availability of spec text is important to have a detailed understanding of the technology and also to judge what its impact on the complexity of the spec will be. There must also be sufficient time to study it in detail. CE contributions without sufficiently mature draft spec text in the CE input document should not be considered for adoption.

Lists of participants in CE documents should be pruned to include only the active participants. Read access to software will be available to all members.

# Establishment of ad hoc groups

The ad hoc groups established to progress work on particular subject areas until the next meeting are described in the table below. The discussion list for all of these ad hoc groups was agreed to be the main JVET reflector ([jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de)).

Review of AHG plans was conducted in session 25 on Wednesday 28 April 2021.

|  |  |  |
| --- | --- | --- |
| **Title and Email Reflector** | **Chairs** | **Mtg** |
| **Project Management (AHG1)**  ([jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de))   * Coordinate overall JVET interim efforts. * Supervise AHG studies. * Report on project status to JVET reflector. * Provide a report to the next meeting on project coordination status. | J.-R. Ohm, G. J. Sullivan (co-chairs) | N |
| **Draft text and test model algorithm description editing (AHG2)**  ([jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de))   * Produce and finalize draft text outputs of the meeting (JVET-W2005 and JVET-W2006). * Collect reports of errata for the VVC, VSEI, HEVC, AVC, CICP, the codepoint usage TR specification and the published HDR-related technical reports and produce the JVET-W1004 errata output collection. * Produce and finalize JVET-W2002 VVC Test Model 14 (VTM 14) Algorithm and Encoder Descriptions. * Coordinate with the test model software development AhG to address issues relating to mismatches between software and text. * Collect and consider errata reports on the texts | B. Bross, J. Chen, C. Rosewarne (co-chairs), F. Bossen, J. Boyce, A. Browne, S. Kim, S. Liu, J.‑R. Ohm, G. J. Sullivan, A. Tourapis, Y.-K. Wang, Y. Ye (vice-chairs) | N |
| **Test model software development (AHG3)**  ([jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de))   * Coordinate development of test models (VTM, HM, SCM, SHM, HTM, MFC, MFCD, JM, JSVM, JMVM, 3DV-ATM, 360Lib, and HDRTools) software and associated configuration files. * Produce documentation of software usage for distribution with the software. * Enable software support for recently standardized additional SEI messages. * Discuss and make recommendations on the software development process. * Propose improvements to the guideline document for developments of the test model software. * Perform comparative tests of test model behaviour using common test conditions. * Suggest configuration files for additional testing of tools. * Investigate how to minimize the number of separate codebases maintained for group reference software. * Coordinate with AHG on Draft text and test model algorithm description editing (AHG2) to identify any mismatches between software and text, and make further updates and cleanups to the software as appropriate. * Prepare drafts of merged CTC documents for HM and VTM, as applicable. | F. Bossen, X. Li, K. Sühring (co-chairs), Y. He, K. Sharman, V. Seregin, A. Tourapis (vice‑chairs) | N |
| **Test material and visual assessment (AHG4)**  ([jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de))   * Produce the verification test report JVET-W2020, and consider plans for additional verification testing of VVC capability. * Maintain the video sequence test material database for testing the VVC and HEVC standards and potential future extensions, as well as exploration activities. * Study coding performance and characteristics in relation to video test materials, including new test materials. * Identify and recommend appropriate test materials for testing the VVC standard and potential future extensions, as well as exploration activities. * Identify missing types of video material, solicit contributions, collect, and make available a variety of video sequence test material. * Maintain and update the directory structure for the test sequence repository as necessary. * Collect information about test sequences that have been made available by other organizations. * Prepare and conduct remote expert viewing for purposes of subjective quality evaluation. * Prepare availability of viewing equipment and facilities arrangements for future meetings. | V. Baroncini, T. Suzuki, M. Wien (co-chairs), E. François, S. Liu, A. Norkin, A. Segall, P. Topiwala, S. Wenger, Y. Ye (vice-chairs) | N |
| **Conformance testing (AHG5)**  ([jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de))   * Study the JVET-U2008 draft conformance testing specification and investigate the need for extensions. * Study the JVET-W2026 draft conformance testing for operation rage extensions and investigate the need for improvements. * Study the requirements of VVC, HEVC, and AVC conformance testing to ensure interoperability. * Maintain and update the conformance bitstream database. * Study additional testing methodologies to fulfil the needs for VVC conformance testing. | J. Boyce and W. Wan (co-chairs), E. Alshina, F. Bossen, I. Moccagatta, K. Kawamura, D. Rusanovskyy, K. Sühring, X. Xu, T. Zhou (vice-chairs) | N |
| **ECM software development (AHG6)**  ([jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de))   * Coordinate development of the ECM software and associated configuration files. * Produce documentation of software usage for distribution with the software. * Prepare and deliver ECM-2.0 software version and the reference configuration encodings according to JVET-W2017 common test conditions. * Investigate encoder speedup and other encoder software optimization. * Coordinate with ECM algorithm description editors to identify any mismatches between software and text, make further updates and cleanups to the software as appropriate. | V. Seregin (chair), J. Chen, F. Le Léannec, K. Zhang (vice-chairs) | N |
| **Coding of HDR/WCG material (AHG7)**  ([jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de))   * Study and evaluate available HDR/WCG test content. * Study objective metrics for quality assessment of HDR/WCG material, including investigation of the correlation between subjective and objective results. * Compare the performance of the VTM and HM for HDR/WCG content. * Study the luma/chroma bit allocation in the HDR CTC, especially for HLG content. * Contribute to activities in merging HDR related CTC documents, in ccordination with AHG3. * Study additional aspects of coding HDR/WCG content. | A. Segall (chair), E. François, W. Husak, S. Iwamura, D. Rusanovskyy (vice-chairs) | N |
| **High bit depth, high bit rate, and high frame rate coding (AHG8)**  ([jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de))   * Study the benefits and characteristics of VVC coding tools for high bit depth, high bit rate, and high frame rate coding. * Study lossless coding characteristics of VVC. * Identify technologies for future extension of VVC to support such application usage. * Study the JVET-U2018 testing conditions for high bit depth, high bit rate, and high frame rate coding, and suggest improvements as applicable. * Contribute to the development of conformance testing for operation range extensions in coordination with AHG5. * Identify suitable test material for testing of high bit depth, high bit rate, and high frame rate coding in coordination with AHG4 and AHG7. * Study VVC entropy decoding throughput and latency in the cases of high bit depth, high bit rate, and high frame rate coding. | A. Browne and T. Ikai (co-chairs), D. Rusanovskyy, M. Sarwer, X. Xiu, Y. Yu (vice-chairs) | N |
| **SEI message studies (AHG9)**  ([jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de))   * Study the SEI messages in VSEI, VVC, HEVC and AVC. * Collect software and showcase information for SEI messages, including encoder and decoder implementations and bitstreams for demonstration and testing. * Identify potential needs for additional SEI messages. * Investigate the possible need of mandatory post processing in the context of SEI messages * Study SEI messages defined in HEVC and AVC for potential use in the VVC context. * Coordinate with AHG3 for software support of SEI messages. | J. Boyce, S. McCarthy (co-chairs), C. Fogg, P. de Lagrange, A. Luthra, G. J. Sullivan, A. Tourapis, Y.-K. Wang, S. Wenger (vice-chairs) | N |
| **Encoding algorithm optimization (AHG10)**  ([jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de))   * Study the impact of using techniques such as tool adaptation and configuration, and perceptually optimized adaptive quantization for encoder optimization. * Study the impact of non-normative techniques of pre processing for the benefit of encoder optimization. * Study encoding techniques of optimization for objective quality metrics and their relationship to subjective quality. * Study optimized encoding for reference picture resampling and scalability modes in VTM. * Consider neural network-based encoding optimization technologies for video coding standards. * Investigate other methods of improving objective and/or subjective quality, including adaptive coding structures and multi-pass encoding. * Study methods of rate control and rate-distortion optimization and their impact on performance, subjective and objective quality. * Study the potential of defining software configuration settings optimized for subjective quality, and coordinate such efforts with AHG3. | P. de Lagrange, R. Sjöberg and A. Tourapis (co-chairs) | N |
| **Neural network-based video coding (AHG11)**  ([jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de))   * Evaluate and quantify performance improvement potential of NN-based video coding technologies compared to existing video coding standards such as VVC, including both individual coding tools and novel architectures. * Finalize, conduct and discuss the EE on neural network-based video coding JVET-W2023. * Solicit input contributions on NN-based video coding technologies. * Refine the test conditions for NN-based video coding, and develop supporting software as needed. * Investigate technical aspects specific to NN-based video coding, such as encoding and decoding complexity of neural networks, design network representation, operation, tensor, on-the-fly network adaption (e.g. updating during encoding) etc; * Study the impact of training (including the impact of loss function) on the performance of candidate technology, and identify suitable materials for training. * Analyse complexity characteristics, perform complexity analysis, and develop complexity reductions of candidate technology. * Refine testing methods for assessment of the effectiveness and complexity of considered technology. * Study the impact of parameter quantization and fixed-point computations in NN-based video coding. * Study and collect information related to near-term and longer-term architectures for neural network-based video coding. * Review the outcome of the expert viewing conducted at the meeting, refine the methodology, and prepare viewing for the next meeting. * Generate and distribute anchor encodings and develop improvements of the JVET-W2016 common test conditions for NNVC technology. * Coordinate with other relevant groups, including SC29/AG5 on visual quality assessment. | E. Alshina, S. Liu, A. Segall, (co‑chairs), J. Chen, F. Galpin, J. Pfaff, S. S. Wang, Z. Wang, M. Wien, P. Wu, J. Xu (vice‑chairs) | N |
| **Enhanced compression beyond VVC capability (AHG12)**  ([jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de))   * Solicit and study non-neural-network video coding tools with enhanced compression capabilities beyond VVC. * Discuss and propose refinements to the ECM2 algorithm description JVET-W2025. * Study the performance and complexity tradeoff of these video coding tools. * Coordinate with AHG6 on ECM software development. * Refine test conditions in JVET-W2017, generate anchors, identify new test sequences to be added, especially high resolution ones in 8K, in coordination with AHG4. * Analyse the results of exploration experiments described in JVET-W2024 in coordination with the EE coordinators. * Coordinate with AHG11 to study the interaction with neural network-based coding tools. | M. Karczewicz, Y. Ye and L. Zhang (co-chairs), B. Bross, X. Li, K. Naser, H. Yang (vice chairs) | N |

It was confirmed that the rules which can be found in document ISO/IEC JTC 1/‌SC 29/‌AG 2 N010 “Ad hoc group rules for MPEG AGs and WGs” (available at <https://www.mpegstandards.org/adhoc/>), are consistent with the operation mode of JVET AHGs. It is however pointed out that JVET does not allow separate AHG reflectors, such that any JVET member is implicitly a member of any AHG. This shall be mentioned in the related WG Recommendations. The list above was also issued as a separate WG 5 document (ISO/IEC JTC 1/‌SC 29/‌WG 5 [N 45](https://sd.iso.org/documents/ui/#!/browse/iso/iso-iec-jtc-1/iso-iec-jtc-1-sc-29/iso-iec-jtc-1-sc-29-wg-5/library/2/List%20of%20AHGs%20established%20at%20the%202nd%20WG%205%20meeting)) in order to make it easy to reference.

# Output documents

The following documents were agreed to be produced or endorsed as outputs of the meeting. Names recorded below indicate the editors responsible for the document production. Where applicable, dates of planned finalization and corresponding parent-body document numbers are also noted.

It was reminded that in cases where the JVET document is also made available as a WG 5 output document, a separate version under the WG 5 document header should be generated. This version should be sent to GJS and JRO for upload.

The list of JVET ad hoc groups was also issued as a WG 5 output document [WG 5 N 45](https://sd.iso.org/documents/ui/#!/browse/iso/iso-iec-jtc-1/iso-iec-jtc-1-sc-29/iso-iec-jtc-1-sc-29-wg-5/library/2/List%20of%20AHGs%20established%20at%20the%202nd%20WG%205%20meeting), as noted in section 9.

[JVET-W1000](https://jvet-experts.org/doc_end_user/current_document.php?id=10845) Meeting Report of the 23rd JVET Meeting [G. J. Sullivan, J.-R. Ohm] [WG 5 N 64] (2021-08-13)

Initial versions of the meeting notes (d0 … d8) were made available on a daily basis during the meeting.

Remains valid – not updated: [JCTVC-H1001](http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=5095) HEVC software guidelines [K. Sühring, D. Flynn, F. Bossen (software coordinators)]

Remains valid – not updated: [JVET-V1002](https://jvet-experts.org/doc_end_user/current_document.php?id=10846) High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Encoder Description Update 15 [C. Rosewarne (primary editor), K. Sharman, R. Sjöberg, G. J. Sullivan (co-editors)]

Remains valid – not updated: [JVET-T1003](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10535) Revised coding-independent code points for video signal type identification (Draft 2) [G. J. Sullivan, T. Suzuki, A. Tourapis]

[JVET-W1004](https://jvet-experts.org/doc_end_user/current_document.php?id=10847) Errata report items for VVC, HEVC, AVC, Video CICP, and CP usage TR [C. Rosewarne, G. J. Sullivan, Y. Syed, Y.-K. Wang] (2021-09-30, near next meeting)

Remains valid – not updated [JVET-T1006](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10538) Annotated regions and shutter interval information SEI messages for AVC (Draft 2) [J. Boyce, S. McCarthy, Y.-K. Wang]

Was included in a new edition issued as WG 5 DIS N 66 (due 2021-07-30). Twin text already consented in ITU-T from the last meeting

Editors of the DIS will be J. Boyce, S. McCarthy, G. Sullivan, and Y.-K. Wang.

DoC of CDAM (WG5 N 65) was reviewed Thursday 15 July in session 22.

Remains valid – not updated: [JCTVC-V1007](http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=10312) SHVC Test Model 11 (SHM 11) Introduction and Encoder Description [G. Barroux, J. Boyce, J. Chen, M. M. Hannuksela, Y. Ye] [WG 11 N 15778]

Remains valid – not updated: [JCTVC-X1009](http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=10572) Common Test Conditions for SHVC [V. Seregin, Y. He]

Remains valid – not updated [JCTVC-O1010](http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=8511) Guidelines for Conformance Testing Bitstream Preparation [T. Suzuki, W. Wan]

No output: JVET-T1011 through JVET-T1013

Remains valid – not updated [JCTVC-V1014](http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=10316) Screen Content Coding Test Model 7 Encoder Description (SCM 7) [R. Joshi, J. Xu, R. Cohen, S. Liu, Y. Ye] [WG 11 N 16049]

Remains valid for HM – not updated: [JCTVC-Z1015](http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=10689) Common Test Conditions for Screen Content Coding [H. Yu, R. Cohen, K. Rapaka, J. Xu]

No output: JVET-T1016 through JVET-T1019

Remains valid for HM – not updated: [JCTVC-Z1020](http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=10692) Common Test Conditions for HDR/WCG Video Coding Experiments [E. François, J. Sole, J. Ström, P. Yin]

Remains valid for HM – not updated: [JVET-U1100](https://jvet-experts.org/doc_end_user/current_document.php?id=10675) Common Test Conditions for HM Video Coding Experiments [K. Sühring, K. Sharman] (2021-02-01)

Remains valid – not updated: [JVET-T2001](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10540) Versatile Video Coding Draft 10 [B. Bross, J. Chen, S. Liu, Y.-K. Wang]

[JVET-W2002](https://jvet-experts.org/doc_end_user/current_document.php?id=10848) Algorithm description for Versatile Video Coding and Test Model 14 (VTM 14) [A. Browne, J. Chen, Y. Ye, S. Kim] [WG 5 N 71] (2021-09-30, near next meeting)

Changes to encoding algorithms, include the tools for operation range extensions.

Remains valid – not updated: [JVET-N1003](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=6638) Guidelines for VVC reference software development [K. Sühring]

Remains valid – not updated: [JVET-T2004](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10542) Algorithm descriptions of projection format conversion and video quality metrics in 360Lib (Version 12) [Y. Ye, J. Boyce]

It was noted that this includes some “stale” formats are no longer subject of active investigation and had been moved to the last part. It was agreed to consider whether they should be removed when a new version is produced in the future. In the future, this could be converted into a technical report. This was added as a mandate to the previous AHG6. (remove this note?)

[JVET-W2005](https://jvet-experts.org/doc_end_user/current_document.php?id=10849) VVC operation range extensions (Draft 4) [F. Bossen, B. Bross, T. Ikai, D. Rusanovskyy, G. Sullivan, Y.-K. Wang] (2021-07-30)

Was integrated into a DIS of new edition issued as WG 5 N 70 (due 2021-07-30)

Editors are F. Bossen, B. Bross, T. Ikai, D. Rusanovskyy, G. Sullivan, Y.-K. Wang.

DoC on CDAM (WG 5 N 69) was reviewed Thursday 15 July in session 22.

[JVET-W2006](https://jvet-experts.org/doc_end_user/current_document.php?id=10850) Additional SEI messages for VSEI (Draft 4) [J. Boyce, G. Sullivan, Y.-K. Wang] (2021-07-30)

Was integrated into a DIS of new edition issued as WG 5 N 68 (due 2021-07-30)

Editors are J. Boyce, G. Sullivan, Y.-K. Wang.

DoC on CDAM (WG 5 N 67) was reviewed Thursday 15 July in session 22.

Remains valid – not updated: [JVET-S2007](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=9679) Versatile supplemental enhancement information messages for coded video bitstreams Draft 5 [J. Boyce, V. Drugeon, G. J. Sullivan, Y.-K. Wang]

Remains valid – not updated: [JVET-U2008](https://jvet-experts.org/doc_end_user/current_document.php?id=10679) Conformance testing for versatile video coding (Draft 6) [J. Boyce, E. Alshina, F. Bossen, K. Kawamura, I. Moccagatta, W. Wan] (2021-03-31)

Remains valid – not updated: [JVET-U2009](https://jvet-experts.org/doc_end_user/current_document.php?id=10680) Reference software for versatile video coding (Draft 2) [F. Bossen, K. Sühring, X. Li] (2021-03-31)

Remains valid – not updated: [JVET-T2010](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10545) VTM common test conditions and software reference configurations for SDR video [F. Bossen, J. Boyce, X. Li, V. Seregin, K. Sühring]

Remains valid – not updated: [JVET-V2011](https://jvet-experts.org/doc_end_user/current_document.php?id=10851) VTM common test conditions and evaluation procedures for HDR/WCG video [A. Segall, E. François, W. Husak, S. Iwamura, D. Rusanovskyy]

Remains valid – not updated: [JVET-U2012](https://jvet-experts.org/doc_end_user/current_document.php?id=10681) JVET common test conditions and evaluation procedures for 360° video [Y. He, J. Boyce, K. Choi, J.-L. Lin] (2021-03-31)

Remains valid – not updated: [JVET-T2013](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10546) VTM common test conditions and software reference configurations for non-4:2:0 colour formats [Y.-H. Chao, Y.-C. Sun, J. Xu, X. Xu]

Remains valid – not updated: [JVET-Q2014](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=9683) JVET common test conditions and software reference configurations for lossless, near lossless, and mixed lossy/lossless coding [T.-C. Ma, A. Nalci, T. Nguyen]

Remains valid – not updated: [JVET-Q2015](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=9684) JVET functionality confirmation test conditions for reference picture resampling [J. Luo, V. Seregin]

[JVET-W2016](https://jvet-experts.org/doc_end_user/current_document.php?id=10852) Common Test Conditions and evaluation procedures for neural network-based video coding technology [S. Liu, A. Segall, E. Alshina, R.-L. Liao] (2021-07-30)

Updates to template, and clarifications

[JVET-W2017](https://jvet-experts.org/doc_end_user/current_document.php?id=10853) Common Test Conditions and evaluation procedures for enhanced compression tool testing [M. Karczewicz and Y. Ye] (2021-07-30)

Clarifications, and editorial refinements

Remains valid – not updated: [JVET-U2018](https://jvet-experts.org/doc_end_user/current_document.php?id=10683) Common test conditions for high bit depth and high bit rate video coding [A. Browne, T. Ikai, D. Rusanovskyy, M. Sarwer, X. Xiu]

[JVET-W2019](https://jvet-experts.org/doc_end_user/current_document.php?id=10853) White paper on Versatile Video Coding [B. Bross and G. Sullivan] [SC 29 AG 3 N 35] (2021-09-14)

[JVET-W2020](https://jvet-experts.org/doc_end_user/current_document.php?id=10854) VVC verification test report for high dynamic range video content [V. Baroncini, M. Wien] [WG 5 N 72] (2021-08-31)

It is pointed out that it might be useful to release an integrated report of SDR, HDR, 360 in the future

No output: [JVET-?2021](https://jvet-experts.org/doc_end_user/current_document.php?id=10855)

Retain number for future purposes of planning possible additional verification testing.

[JVET-W2022](https://jvet-experts.org/doc_end_user/current_document.php?id=10843) Core Experiment on Film Grain Synthesis [S. McCarthy, M. Radosavljević, J. Shingala] [WG 5 N 74] (2021-07-30)

An initial draft was reviewed and approved.It is suggested to determine the runtime of the proposals relative to each other rather than taking VTM runtime as reference. VTM 13 is used, as there would be no difference in performance compared to VTM 14.

[JVET-W2023](https://jvet-experts.org/doc_end_user/current_document.php?id=10844) Exploration Experiment on Neural Network-based Video Coding (EE1) [E. Alshina, S. Liu, W. Chen, F. Galpin, Y. Li, Z. Ma, H. Wang] [WG 5 N 75] (2021-07-30)

An initial draft was reviewed and approved.

Three categories: Enhancement filters (loop and post), super resolution, intra prediction

It was suggested to use negative values for bit rate saving in the diagrams.

[JVET-W2024](https://jvet-experts.org/doc_end_user/current_document.php?id=10842) Exploration Experiment on Enhanced Compression beyond VVC capability (EE2) [ V. Seregin, J. Chen, S. Esenlik, F. Le Léannec, L. Li, J. Ström, M. Winken, X. Xiu, K. Zhang] [WG 5 N 76] (2021-08-06)

An initial draft was reviewed and approved. It was suggested to also test UQT standalone additionally to its combination with ABT.

[JVET-W2025](https://jvet-experts.org/doc_end_user/current_document.php?id=10848) Algorithm description of Enhanced Compression Model 2 (ECM 2) [M. Coban, F. Le Léannec, J. Ström] [WG 5 N 73] (2021-08-31)

See notes under JVET-W0102

JVET-W2026 Conformance testing for VVC operation range extensions (draft 1) [D. Rusanovskyy, J.-H. Jhu, I. Moccagatta, M. Sarwer, Y. Yu] (2021-08-31)

Initial version was reviewed and approved.

# Future meeting plans, expressions of thanks, and closing of the meeting

Future meeting plans were established according to the following guidelines:

* Meeting under ITU-T SG 16 auspices when it meets (ordinarily starting meetings on the Wednesday of the first week and closing it on the Wednesday of the second week of the SG 16 meeting – a total of 8 meeting days), and
* Otherwise meeting under ISO/IEC JTC 1/‌SC 29 auspices when its MPEG WGs meet (ordinarily starting meetings on the Friday prior to the main week of such meetings and closing it on the same day as other MPEG WGs – a total of 8 meeting days).

In cases where an exceptionally high workload is expected for a meeting, an earlier starting date may be defined. In case of online meetings, no sessions should be held on weekend days. This may imply an earlier starting date as well.

Some specific future meeting plans (to be confirmed) were established as follows:

* Wed. 6 – Fri. 15 October 2021, 24th meeting under ISO/IEC JTC 1/‌SC 29 auspices as a virtual meeting
* During January 2022, 25th meeting under ITU-T SG16 auspices, location t.b.d.
* Fri. 22 – Fri. 29 April 2022, 26th meeting under ISO/IEC JTC 1/‌SC 29 auspices in Alpbach, AT
* Fri. 15 – Fri. 22 July 2022, 27th meeting under ISO/IEC JTC 1/‌SC 29 auspices in Cologne, DE.
* During October 2022, 28th meeting under ITU-T SG16 auspices in Geneva, CH.
* During January 2023, 29th meeting under ISO/IEC JTC 1/‌SC 29 auspices, location t.b.d.
* During April 2023, 30th meeting under ISO/IEC JTC 1/‌SC 29 auspices, location t.b.d.
* During July 2023, 31st meeting under ITU-T SG16 auspices in Geneva, CH.

The agreed document deadline for the 24th JVET meeting was planned to be Wednesday 30 Sept. 2021.

Giacomo Baroncini and Vittorio Baroncini were thanked for conducting, and Mathias Wien was thanked for coordinating the VVC verification test HDR video. It is greatly appreciated that this testing was successfully completed ahead of time relative to the original plans, and despite requiring extraordinary efforts due the complications caused by the pandemic situation.

InterDigital was thanked for providing financial support for the VVC verification tests.

Mathias Wien was thanked for planning, organizing and conducting the remote expert viewing related to the exploration experiment on neural network-based video compression. It is emphasized that this effort was extremely helpful to better understand the benefits of this new technology in terms of subjective quality improvement.

The 23rd JVET meeting was closed at approximately XXXX hours UTC on Friday 16 July 2021.

# Annex A to JVET report: List of documents

# Annex B to JVET report: List of meeting participants

The participants of the twenty-third meeting of the JVET, according to the participation records from the Zoom sessions, (approximately XXX people in total), were as follows:

1. …

# Annex C to JVET report: Recommendations of the 4th meeting of ISO/IEC JTC 1/SC 29/WG 5 MPEG Joint Video Coding Team(s) with ITU-T SG 16

**ISO/IEC JTC 1/SC 29/WG 5 N XX**