|  |  |
| --- | --- |
| **Joint Video Experts Team (JVET)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29**  22nd Meeting, by teleconference, 20–28 Apr. 2021 | Document: JVET-V\_Notes\_d7 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Meeting Report of the 22nd Meeting of the Joint Video Experts Team (JVET), by teleconference, 20–28 April 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-second meeting during 20–28 April 2021 as an online-only meeting. It had previously been planned to be in Geneva, CH, 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 third 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 1310 hours UTC on Tuesday 20 April 2021. Meeting sessions were held on all days except the weekend days of Saturday and Sunday 17 and 18 April 2021, until the meeting was closed at approximately 1740 hours UTC on Wednesday 28 April 2021. Approximately XXX people attended the JVET meeting, and approximately XX 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 coordinated fashion with a teleconference meeting of SG16 – one of the two parent bodies of the JVET, under whose auspices this JVET meeting was held. 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-first JVET meeting in producing the following documents:

* JVET-U1000 Meeting report
* JVET-U1004 Errata report items for VVC, HEVC, AVC, Video CICP, and CP usage TR
* JVET-U1100 Common Test Conditions for HM Video Coding Experiments
* JVET-U2002 Algorithm description for Versatile Video Coding and Test Model 12 (VTM 12)
* JVET-U2005 New level and additional SEI messages for VVC (Draft 2)
* JVET-U2006 Additional SEI messages for VSEI (Draft 2)
* JVET-U2008 Conformance testing for versatile video coding (Draft 6)
* JVET-U2009 Reference software for versatile video coding (Draft 2)
* JVET-U2012 JVET common test conditions and evaluation procedures for 360° video
* JVET-U2016 Common Test Conditions and evaluation procedures for neural network-based video coding technology
* JVET-U2017 Common Test Conditions and evaluation procedures for enhanced compression tool testing
* JVET-U2018 Common Test Conditions for High Bit Depth and High Bit Rate Coding
* JVET-U2021 VVC verification test plan (Draft 5)
* JVET-U2022 CE on Entropy Coding for High Bit Depth and High Bit Rate Coding
* JVET-U2023 Exploration Experiment on Neural Network-based Video Coding
* JVET-U2024 Exploration Experiment on Enhanced Compression beyond VVC capability

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 XX output documents from the current meeting (update):

* JVET-U1004 Errata report items for VVC, HEVC, AVC, Video CICP, and CP usage TR
* JVET-U1100 Common Test Conditions for HM Video Coding Experiments
* JVET-U2002 Algorithm description for Versatile Video Coding and Test Model 12 (VTM 12)
* JVET-U2005 New level and additional SEI messages for VVC (Draft 2)
* JVET-U2006 Additional SEI messages for VSEI (Draft 2)
* JVET-U2008 Conformance testing for versatile video coding (Draft 6)
* JVET-U2009 Reference software for versatile video coding (Draft 2)
* JVET-U2012 JVET common test conditions and evaluation procedures for 360° video
* JVET-U2016 Common Test Conditions and evaluation procedures for neural network-based video coding technology
* JVET-U2017 Common Test Conditions and evaluation procedures for enhanced compression tool testing
* JVET-U2018 Common Test Conditions for High Bit Depth and High Bit Rate Coding
* JVET-U2021 VVC verification test plan (Draft 5)
* JVET-U2022 CE on Entropy Coding for High Bit Depth and High Bit Rate Coding
* JVET-U2023 Exploration Experiment on Neural Network-based Video Coding
* JVET-U2024 Exploration Experiment on Enhanced Compression beyond VVC capability

For the organization and planning of its future work, the JVET established XX “ad hoc groups” (AHGs) to progress the work on particular subject areas. At this meeting, X Core Experiment (CE) and X Exploration Experiments (EE) were defined. The next eight JVET meetings were planned for Wed. 7 – Fri. 16 July 2021, online under ISO/IEC SC 29 auspices, during Fri. 8 – Fri. 15 October 2021 as a mixed-mode meeting under ISO/IEC SC 29 auspices in Antalya, TR, during January 2022 under ITU-T SG16 auspices in Geneva, CH, during Fri. 22 – Fri. 29 April 2022 under ISO/IEC SC 29 auspices, location t.b.d., during Fri. 15 – Fri. 22 July 2022 under ISO/IEC SC 29 auspices in Cologne, DE, during October 2022 under ITU-T SG16 auspices in Geneva, CH, during January 2023 under ISO/IEC SC 29 auspices, location t.b.d., and during April 2023 under ISO/IEC SC 29 auspices, location t.b.d.

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 20–28 April 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 third 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_04_V_Virtual/>.

## Primary goals

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

* JVET-U1004 Errata report items for VVC, HEVC, AVC, Video CICP, and CP usage TR
* JVET-U1100 Common Test Conditions for HM Video Coding Experiments
* JVET-U2002 Algorithm description for Versatile Video Coding and Test Model 12 (VTM 12)
* JVET-U2005 New level and additional SEI messages for VVC (Draft 2)
* JVET-U2006 Additional SEI messages for VSEI (Draft 2)
* JVET-U2008 Conformance testing for versatile video coding (Draft 6)
* JVET-U2009 Reference software for versatile video coding (Draft 2)
* JVET-U2012 JVET common test conditions and evaluation procedures for 360° video
* JVET-U2016 Common Test Conditions and evaluation procedures for neural network-based video coding technology
* JVET-U2017 Common Test Conditions and evaluation procedures for enhanced compression tool testing
* JVET-U2018 Common Test Conditions for High Bit Depth and High Bit Rate Coding
* JVET-U2021 VVC verification test plan (Draft 5)
* JVET-U2022 CE on Entropy Coding for High Bit Depth and High Bit Rate Coding
* JVET-U2023 Exploration Experiment on Neural Network-based Video Coding
* JVET-U2024 Exploration Experiment on Enhanced Compression beyond VVC capability

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.

### Late and incomplete document considerations

The formal deadline for registering and uploading non-administrative contributions had been announced as Tuesday, 13 April 2021. Any documents uploaded after 1159 hours Paris/Geneva time on Wednesday 14 April 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-V0133 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. Also, all cross-check reports were uploaded late.

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-V0XXX (a proposal on …), uploaded 04-XX after deadline.
* ...

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-V0XXX (a document on …), uploaded 04-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-V0154.

“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-U1000, the Errata report items for VVC, VSEI, HEVC, AVC, Video CICP, and CP usage TR JVET-U1004, the Common Test Conditions for HM Video Coding Experiments JVET-U1100, the Algorithm description for Versatile Video Coding and Test Model 12 (VTM 12) JVET-U2002, the New level and additional SEI messages for VVC (Draft 2) JVET-U2005, the Additional SEI messages for VSEI (Draft 2) JVET-U2006, the Conformance testing for VVC (Draft 6) JVET-U2008, the Reference software for versatile video coding (Draft 2) JVET-U2009, the JVET common test conditions and evaluation procedures for 360° video JVET-U2012, the Common Test Conditions and evaluation procedures for neural network-based video coding technology JVET-U2016, the Common Test Conditions and evaluation procedures for enhanced compression tool testing JVET-U2017, the Common test conditions for high bit depth and high bit rate video coding JVET-U2018, the VVC verification test plan (Draft 5) JVET-U2021, the Description of the CE on Entropy Coding for High Bit Depth and High Bit Rate Coding JVET-U2022, the Description of the EE on Neural Network-based Video Coding JVET-U2023, and the Description of the EE on Enhanced Compression beyond VVC capability JVET-U2024 had been completed and were approved. The software implementations of VTM (versions 12.0 and 12.1) , HM 16.23, and HDRTools (versions 0.20 and 0.21) were also approved.

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

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.

* 1300–1500 1st “afternoon” session [break after 2 hours]
* 1520–1720 2nd “afternoon” session
* [“evening” break – nearly 4 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 Code of Conduct reminder***

Participants were reminded of the ISO Code of Conduct, found at [Ed. (GJS): Add IEC code of conduct]

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

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)
* <http://www.itscj.ipsj.or.jp/sc29/29w7proc.htm> (JTC 1/‌SC 29 Procedures)

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 1300. Furthermore, the JCT-VC email list currently had 1293 subscribers (as of 20 April 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)
* **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 Tuesday 20 April at 1300 UTC (chaired by GJS and JRO) were as follows.

* Timing and organization of online meetings, calendar
* Standards publication status – to be updated (with dates)
  + Working practices TR published in ISO/IEC and ITU-T
  + HEVC Amd.1 (additional SEI) and Amd.2 (shutter interval) in final publication stage in ISO/IEC
  + CICP usage 3rd ed. clarify publication status
  + The following freely available standards published here in ISO/IEC:  
    <https://standards.iso.org/ittf/PubliclyAvailableStandards/index.html>
    - ISO/IEC 23002-7:2021, 1st edition
    - ISO/IEC 23008-2:2020, 4th edition
    - ISO/IEC 23090-3:2021, 1st edition
    - ISO/IEC 23091-2:2019, 1st edition
* Draft standards progression status
  + VVC conformance – DIS ballot to be started soon, FDIS in October
  + VVC reference SW – DIS ballot to be started soon, FDIS in October
  + AVC additional SEI – CDAM this meeting, new draft 3
  + VSEI extensions – CDAM this meeting, new draft
  + VVC operation range extensions – CDAM this meeting, new draft
  + CICP v2 – FDIS in ISO/IEC and ITU-T consent this meeting (include errata items)
  + Further consent items for ITU-T:
    - HEVC shutter interval SEI and errata items
    - AVC annotated regions and shutter interval SEI messages, and errata items
    - CICP usage TR 3rd ed.
  + 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 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 and Amd 2?
    - ISO/IEC DIS 23091-2, 2nd edition
    - ISO/IEC 23002-7:2021/Amd 1 – could be later
    - ISO/IEC 23090-3:2021/Amd 1 – could be later
* The meeting logistics, agenda, working practices, policies, and document allocation were reviewed.
  + The meeting is 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 were reviewed.
* There was somewhat less of a problem of late non-cross-check documents and no “placeholders” (see section 2.4.2).
* The primary goals of the meeting were
  + Errata
  + Conformance and software for VVC & VSEI
  + Verification test planning
  + 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: Thank resolution, resolution calling for funding wrt upcoming tests.
* Conformance: Thank resolution to companies who provided bitstreams
* Liaisons?
* Number of documents increased, but if possible not conduct sessions in parallel
* Scheduling was discussed
* 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.

* 1300–1500 1st “afternoon” session [break after 2 hours]
* 1520–1720 2nd “afternoon” session
* [“evening” break – nearly 4 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:

* Tue. 20 April, 1st day
  + Session 1:
    - 1310–1350 Opening remarks, review of practices, agenda, IPR reminder
    - 1400–1510 Reports of AHGs 1–5
  + Session 2:
    - 1530–1720 Reports of AHGs 6–12
  + Session 3:
    - 2100–2300 Review of CE and related
  + Session 4:
    - 2320–0120 Review of CE and related
* Wed. 21 April, 2nd day
  + Session 5:
    - 1430-1540 Review of CE and related
  + Session 6:
    - 1600-1720 Review of CE and related
  + Session 7:
    - 2100-2300 Review of EE1 and related
  + Session 8:
    - 2320-0120 Review of EE1 and related
* Thu. 22 April, 3rd day
  + Session 9:
    - 1300-1500 HLS
  + Session 10:
    - 1520-1720 HLS
  + Session 11:
    - 2100-2300 Review of EE2 and related
  + Session 12:
    - 2320-0120 Review of EE2 and related
* Fri. 23 April, 4th day
  + Session 13:
    - 1300-1500 Review of CE/EE1 remaining
  + Session 14:
    - 1520-1720 Review of EE1/2 remaining
  + Session 15a:
    - 2100-2300 Review of EE2 remaining
  + Session 15b:
    - 2100-2300 HLS
  + Session 16a:
    - 2320-0120 Review of EE2 remaining
  + Session 16b:
    - 2320-0120 HLS
* Mon. 26 April, 5th day
  + Session 17:
    - 1300-1505 JVET plenary: Planning, Review 4.1, 4.2
  + Session 18a:
    - 1525-1720 Review 4.3, 4.8, 4.9
  + Session 18b:
    - 1525-1720 HLS
  + Session 19:
    - 2100-2310 Remaining docs 4.11, 4.5, 4.6
  + Session 20:
    - 2330-0130 Revisits, further planning, in particular
      * NN viewing results
      * CE revisits
* Tue. 27 April, 6th day
  + Session 21:
    - 1300-1400 Joint meeting with VCEG and MPEG WG2/WG3
    - 1400-1500 Joint meeting with AG5: Verification test, quality metrics
  + Session 22:
    - 1520-1620 EE1 planning (E. Alshina)
    - 1620-1720 EE2 planning (V. Seregin)
  + Session 23:
    - 2100-2310 Revisits, further planning
* Wed. 28 April, 7th day
  + Session 24:
    - 1300-1515 Plenary: Output Doc review
  + Session 25:
    - 1535-1720 Plenary: AHGs, recommendations, planning
  + Session 26:
    - 2100-2300 Plenary: Output Doc review, AHGs, recommendations, planning

## 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 of standards (3)
  + Text development and errata reporting (2)
  + Test conditions (1)
  + Verification testing (6) – 2 TBP
  + Test Material (1)
  + Quality assessment (1)
  + Conformance test development (0)
  + Software development (1)
  + Implementation studies (5)
  + Complexity analysis (0)
  + Encoder optimization (5)
  + Profile/tier/level specification (0)
* Low-level tool technology proposals (section 5) with subtopics
  + AHG8: High bit depth and high bit rate coding (27) (section 5.1) – 1 TBP?
  + AHG11: Neural network-based technology (18) (section 5.2)
  + AHG12: Enhanced compression beyond VVC capability (24) (section 5.3)
* High-level syntax (HLS) proposals (section 6) with subtopics
  + AHG9: SEI message studies and proposals (11) (section 6.1)
  + Non-SEI HLS aspects (4) (section 6.2)
* Joint meetings, plenary discussions, BoG reports (0), 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 Tuesday 20 April 2021 during 1400–1510 UTC and 1530-1720 UTC (chaired by GJS & JRO)

[JVET-V0001](https://jvet-experts.org/doc_end_user/current_document.php?id=10779) 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 an increased number of input documents (as compared to the previous two meetings) 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_01_U_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-U1000) [Posted 2021-02-12]
* Errata report items for VVC, HEVC, AVC, Video CICP, and CP usage TR (JVET-U1004) [Posted 2021-03-05]
* Common Test Conditions for HM Video Coding Experiments (JVET-U1100) [Posted 2021-02-08, last update 2021-02-16]
* Algorithm description for Versatile Video Coding and Test Model 12 (VTM 12) (JVET-U2002) [Posted 2021-04-12]
* New level and additional SEI messages for VVC (Draft 2) (JVET-U2005) [Posted 2021-03-05]
* Additional SEI messages for VSEI (Draft 2) (JVET-U2006) [Posted 2021-03-05]
* Conformance testing for versatile video coding (Draft 6) (JVET-U2008) [Posted 2021-03-31]
* Reference software for versatile video coding (Draft 2) (JVET-U2009) [Posted 2021-04-20]
* JVET common test conditions and evaluation procedures for 360° video (JVET-U2012) [Posted 2021-03-16]
* Common Test Conditions and evaluation procedures for neural network-based video coding technology (JVET-U2016) [Posted 2021-02-03, last update 2021-02-06]
* Common Test Conditions and evaluation procedures for enhanced compression tool testing (JVET-U2017) [Posted 2021-01-29]
* Common Test Conditions for High Bit Depth and High Bit Rate Coding (JVET-U2018) [Posted 2021-01-16, last update 2021-01-29]
* VVC verification test plan (Draft 5) (JVET-U2021) [Posted 2021-03-02]
* CE on Entropy Coding for High Bit Depth and High Bit Rate Coding (JVET-U2022) [Posted 2021-01-15, last update 2021-03-26]
* EE on Neural Network-based Video Coding (JVET-U2023) [Posted 2021-01-15, last update 2021-01-29]
* EE on Enhanced Compression beyond VVC capability (JVET-U2024) [Posted 2021-01-16, last update 2021-02-17]

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, two meetings of AHG4 (for preparing the verification tests), and two meetings of AHG11 (for discussing the exploration experiment on neural network-based video coding), were 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 100 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 22nd meeting had been made publicly available on the ITU-hosted ftp site <http://wftp3.itu.int/av-arch/jvet-site/2021_04_V_Virtual/>.

[JVET-V0002](https://jvet-experts.org/doc_end_user/current_document.php?id=10780) 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]

Output documents produced:

JVET-U1004 Errata report items for VVC, VSEI, 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-U meeting:

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

o Some minor editorial corrections and improvements (JVET-U0073, JVET-U0085, and one item from Hendry)

o Fixes for tickets #1416, #1428, #1432, #1454, and #1469

• For AVC: Added a reference to JVET-T1006-v1 for the errata changes, and removed the actual errata text changes from this document. (JVET-U0049)

JVET-U2002 Algorithm description for Versatile Video Coding and Test Model 12 (VTM 12)

The VVC Test Model 12 (VTM12) algorithm description and encoding method document was produced, with the following additions compared to the previous release:

• Refinement of high precision (1/16 pel) motion compensation and motion vector storage

• Added reference picture resampling

JVET-U2005 New level and additional SEI messages for VVC (Draft 2)

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

Draft 2 incorporated items:

• Text changes for clarification of 1) the exact meaning of a parameter set being referenced and 2) that the decoding of non-VCL NAL units in a PU after the last VCL NAL unit of the picture in decoder is deferred after all slices of the picture are decoded (JVET-U0073-v3)

• Addition of SEI payload type values 165 (alpha\_channel\_info), 177 (depth\_representation\_info), 179 (multiview\_acquisition\_info), and 205 (scalability\_dimension\_info) (JVET-U0082

• Addition of SEI payload type value 206 (extended\_drap\_indication) and text on the use of the extended dependent random access point (EDRAP) indication SEI message (JVET-U0084)

• Text changes for inclusion of the decoded picture hash SEI message to the list VclAssociatedSeiList, not imposing the same content requirement for repeated user-defined SEI messages, and rules for inclusion of SEI messages in prefix and suffix SEI NAL units (JVET-U0085-v2)

• Fix for ticket #1448

JVET-U2006 Additional SEI messages for VSEI (Draft 2)

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, and the extended dependent random access point (DRAP) indication SEI message. The draft text also includes text changes for some technical corrections and editorial improvements.

Draft 2 incorporated items:

• Addition of the alpha channel information SEI message, the depth representation information SEI message, the multiview acquisition information SEI message, and the scalability dimension information SEI message (JVET-U0082)

• Addition of the extended DRAP indication SEI message (JVET-U0084)

• Some minor editorial corrections and improvements (JVET-U0086)

• Fix for ticket #1412: change the coding of ffi\_display\_elemental\_periods\_minus1 from u(4) to u(8)

Related input contributions

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

• JVET-V0072 AHG2: Proposal to remove some RPL constraints (Nokia, LGE, Bytedance)

• JVET-V0111 AHG2: On Decoding Unit Information for VVC Version 1 (Sharp)

Recommendations

The AHG recommends to:

• Approve JVET-U1004, JVET-U2002, JVET-U2005, and JVET-U2006 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 and act on them if found to be necessary.

[JVET-V0003](https://jvet-experts.org/doc_end_user/current_document.php?id=10781) JVET AHG report: Test model software development (AHG3) [F. Bossen, X. Li, K. Sühring, K. Sharman, V. Seregin]

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

• VTM 12.1 (Mar. 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 0.21 (Jan. 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.

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.

Only SHM and HTM are still located in subversion repositories. It is suggested to convert and move these repositories to GitLab as well.

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 11.1 and 11.2 were tagged on Jan. 15 and Jan. 18, and VTM version 12.0 was tagged on Feb. 16. VTM version 12.1 was tagged on Mar 31. VTM 12.2 is expected during the 22nd JVET meeting.

VTM 11.1 was tagged on Jan. 15, 2021. Changes include:

• JVET-R0264: IRAP constraint

• JVET-T0053: Adding support for Annotated Regions SEI message

• Fix #1422: dpb parameters inference when subLayerInfoFlag = 0

• Fix access to data that might be deleted

• Fix memory allocation when decoding a stream changing bitdepth between CVS

• Fix #1439: GOP32 configuration for larger intra periods

• Fix #1442: Fix for PTL signalling in VPS

• Fix handling of suffix APS NAL units

• Fix #1438: Decoder crashes when decoding multi-layer bitstream

• Fix handling of suffix APS NAL units in RASL skipped pictures.

• Fix #1411: Decoder crash when decoding a CRA picture following an EOS

• Fix #1419: apply reference pic marking before getting a new pic buffer

• Update README file to include git retrieval

• Update copyright header to include year 2021

• Fix for minimum functionality of StreamMergeApp with VTM 11.0

VTM 11.2 was tagged on Jan. 18, 2021. Changes include:

• Add bound checks and use correct index for m\_vpsMaxTidIlRefPicsPlus1

• Fix #1394: correctly determine whether a NAL unit is a VCL unit

• Remove unused code and fix indentation and braces

• Remove macros from previous cycle

VTM 12.0 was tagged Feb. 16, 2021. Changes include:

• JVET-U0081: ALF filter optimization with filter strength target

• JVET-U0103: SIMD implementation for high bit depth coding

• JVET-U2018: Updated configuration file for HBD CTC

• JVET-U2018: Update sequence configuration files

• Fix #1451: Align SW with spec for RPR/4:4:4 combination

• Fix #1452: use "true" original for HDR metrics

• Fix #1453: use 64-bit variables for RPR to prevent overflow

• Fix #1456: Incorrect DPB flush when mixed NALU types is enabled

• Fix #1457: rplIdx variable for list 1 is not always set in parseSliceHeader()

• Cleanup: fix return type of PPS::getMixedNaluTypesInPicFlag

• Fix compilation when JVET\_R0351\_HIGH\_BIT\_DEPTH\_ENABLED is set to 1

• Fix: #1460: init MTS coeff Constraint in MTS loop during Intra Search

• Fix for ticket #1458: printout wPSNR for hbd ctc

• Enable processing of low bit depth content with JVET\_U0103\_HIGH\_BIT\_DEPTH\_SIMD

• Fix #1459: Modify ALF AVX2 code from JVET-U0103

• Refactor: rename ALF APS ID related functions and variables

• remove unused code and fix indentation and braces

VTM 12.1 was tagged on Mar 31, 2021. Changes include:

• Fix #1466: fix check on SPS virtual boundaries constraint.

• Fix #1465: skip mixed nalu leading pictures when beginning a stream

• Fix #1464: Non-conforming values of sn\_subpic\_id\_len\_minus1

• Fix #1467: Replace std::sort() with std::stable\_sort() to avoid cross-platform performance mismatch

• Fix #1468: bugfix parcat: remove duplicate prefix sei for IDR

• Add decoded picture hash SEI writing for subpictures in subpic merger app

• Fix call to initSEIScalableNesting

• Fixes for tickets #961, #1455 and #1447: Decoding and verifying of GDR conformance streams

• Fix tickets #1461, #1462, #1463: chroma dQP bug fixes and configuration parameters improvements.

• Fix #1471: DebugBitstream feature

• Add verification of subpicture based decoded picture hashes

• Fix !2018: ignore slice-level setting when feature is not used

• Reduced the memory required by EncModeCtrl by ~1GB

• Fix ticket #1377: remove ENABLE\_SPLIT\_PARALLELISM

• Fix bad bitstream generation: wrong parameter to rpcSlice->setUseChromaQpAdj() in initEncSlice().

• fix initialization of m\_initialCpbRemovalOffset in SEIBufferingPeriod class

• Fix non\_ref\_pic compliance check

• Fix CHECK statements

• Fix naming and error messages for VLC length calculation

• Fix #1475: Use SMultiValueInput<uint32\_t> instead of SMultiValueInput<uint8\_t>

• Fix braces and indentation and remove trailing spaces

VTM 12.2 is expected to be tagged during the 22nd JVET meeting. 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

The following tables show **VTM 12.1** performance over **HM 16.23**:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | **All Intra Main10** |  |  |
|  |  |  | **Over HM-16.23** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 | -29,04% | -32,17% | -34,07% | 1560% | 171% |
| Class A2 | -29,29% | -23,92% | -21,06% | 2503% | 182% |
| Class B | -21,73% | -26,96% | -30,76% | 2767% | 185% |
| Class C | -22,54% | -18,95% | -22,70% | 3869% | 199% |
| Class E | -25,76% | -25,91% | -24,46% | 2251% | 173% |
| **Overall** | -25,06% | -25,37% | -26,85% | 2574% | 183% |
| Class D | -18,47% | -13,31% | -13,42% | 4424% | 187% |
| Class F | -39,33% | -39,73% | -42,22% | 4761% | 178% |
|  |  |  |  |  |  |
|  |  |  | **Random access Main10** |  |  |
|  |  |  | **Over HM-16.23** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 | -39,63% | -39,43% | -46,23% | 682% | 163% |
| Class A2 | -43,41% | -41,01% | -40,23% | 757% | 174% |
| Class B | -36,50% | -49,27% | -47,74% | 751% | 164% |
| Class C | -32,82% | -35,20% | -37,22% | 1020% | 174% |
| Class E |  |  |  |  |  |
| **Overall** | -37,52% | -41,90% | -43,13% | 801% | 168% |
| Class D | -30,74% | -31,79% | -31,42% | 1138% | 167% |
| Class F | -45,76% | -49,18% | -50,10% | 561% | 147% |
|  |  |  |  |  |  |
|  |  |  | **Low delay B Main10** |  |  |
|  |  |  | **Over HM-16.23** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  |  |  |
| Class A2 |  |  |  |  |  |
| Class B | -29,24% | -34,80% | -32,41% | 750% | 149% |
| Class C | -25,89% | -17,42% | -17,95% | 924% | 179% |
| Class E | -28,73% | -33,03% | -26,38% | 372% | 146% |
| **Overall** | -28,00% | -28,56% | -26,08% | 675% | 158% |
| Class D | -25,01% | -12,57% | -11,79% | 959% | 190% |
| Class F | -40,20% | -41,56% | -41,87% | 495% | 142% |
|  |  |  |  |  |  |
|  |  |  | **Low delay P Main10** |  |  |
|  |  |  | **Over HM-16.23** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  |  |  |
| Class A2 |  |  |  |  |  |
| Class B | -33,97% | -37,79% | -34,99% | 695% | 162% |
| Class C | -27,68% | -17,28% | -18,05% | 844% | 192% |
| Class E | -32,32% | -36,86% | -30,30% | 367% | 155% |
| **Overall** | -31,46% | -30,72% | -28,17% | 632% | 169% |
| Class D | -26,32% | -11,99% | -10,87% | 893% | 199% |
| Class F | -39,97% | -41,10% | -41,48% | 527% | 146% |

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 12.1** performance compared to **VTM 11.0**:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | **All Intra Main10** |  |  |
|  |  |  | **Over VTM-11.10** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 | 0,00% | 0,00% | 0,00% | 100% | 101% |
| Class A2 | 0,00% | 0,00% | 0,00% | 100% | 99% |
| Class B | 0,00% | 0,00% | 0,00% | 99% | 102% |
| Class C | 0,00% | 0,00% | 0,00% | 99% | 102% |
| Class E | 0,00% | 0,00% | 0,00% | 100% | 101% |
| **Overall** | 0,00% | 0,00% | 0,00% | 100% | 101% |
| Class D | 0,00% | 0,00% | 0,00% | 100% | 98% |
| Class F | 0,00% | 0,00% | 0,00% | 93% | 103% |
|  |  |  |  |  |  |
|  |  |  | **Random access Main10** |  |  |
|  |  |  | **Over VTM-11.10** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 | 0,00% | 0,00% | 0,00% | 100% | 99% |
| Class A2 | 0,00% | 0,00% | -0,01% | 100% | 99% |
| Class B | 0,00% | 0,00% | 0,00% | 100% | 98% |
| Class C | 0,00% | 0,00% | -0,01% | 100% | 99% |
| Class E |  |  |  |  |  |
| **Overall** | 0,00% | 0,00% | 0,00% | 100% | 98% |
| Class D | 0,00% | 0,00% | 0,00% | 100% | 100% |
| Class F | 0,00% | 0,00% | 0,00% | 98% | 100% |
|  |  |  |  |  |  |
|  |  |  | **Low delay B Main10** |  |  |
|  |  |  | **Over VTM-11.10** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  |  |  |
| Class A2 |  |  |  |  |  |
| Class B | 0,01% | 0,01% | -0,34% | 98% | 95% |
| Class C | -0,01% | 0,35% | -0,30% | 100% | 98% |
| Class E | 0,05% | -0,38% | -0,11% | 101% | 100% |
| **Overall** | 0,01% | 0,03% | -0,27% | 99% | 97% |
| Class D | -0,05% | 0,09% | -0,09% | 100% | 100% |
| Class F | 0,00% | -0,26% | 0,06% | 99% | 101% |
|  |  |  |  |  |  |
|  |  |  | **Low delay P Main10** |  |  |
|  |  |  | **Over VTM-11.10** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  |  |  |
| Class A2 |  |  |  |  |  |
| Class B | 0,01% | -0,03% | -0,03% | 99% | 96% |
| Class C | -0,01% | 0,10% | 0,08% | 99% | 101% |
| Class E | -0,02% | -0,03% | 0,15% | 101% | 98% |
| **Overall** | -0,01% | 0,02% | 0,05% | 100% | 98% |
| Class D | 0,02% | -0,03% | -0,01% | 101% | 100% |
| Class F | -0,08% | -0,24% | -0,14% | 99% | 100% |

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

Issues in VTM 12.x affecting conformance

The following issues in VTM master branch (Apr 20, 2021) affect conformance:

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

• The macro GDR\_LEAK\_TEST was disabled due to causing mismatches with conformance streams (MR2050).

• Missing HLS features (see sections below)

However, there are no known issues in VTM that affect processing of current JVET conformance bitstreams.

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)

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

• S0156 aspect 3 (JVET-S0139 item 21)

• JVET-S0139 item 26 (no source listed)

• 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.

• JVET-T0055 aspect 4

• JVET-T0055 item 2

Status of proposals of the 21st JVET meeting (Online)

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

• JVET-U0082 (Scalability dimension SEI message and three HEVC SEI messages)

• JVET-U0084 (Cross RAP referencing (CRR) SEI message)

• JVET-U0085 (VVC Errata)

• JVET-U0086 (VSEI Errata)

It was clarified during the AHG report that U0085 and U0086 do not affect the software. The missing new SEI message SW is to be considered in the context of the VSEI amendment.

HM related activities

HM 16.23 was tagged on Mar. 24, 2021. Changes include:

• Typo corrected (SEIPreferredTransferCharacteristics)

• Enable TemporalFilter for random access configurations, except class F

• Backport of SEIRemovalApp from VTM

• Merge VTM readme file with build instructions.

• Fix segment coding mismatch for HDR (delta qp)

• Updated HDR config files

The following actions have yet to be included:

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

• JCTVC-AJ0028 (Encoder-only Supplemental Motion Vector Estimation for Point cloud Coding content).

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

The contributors og AJ0028 had been informed about the unsuitable quality of the software, but did not respond within the past few meeting cycles. It is suggested during the AHG report to remove this item from the list. Merge requests on the other two items are available, but merging is pending final review.

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 old,

• 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.

SCM related activities

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

SHM related activities

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

HTM related activities

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

HDRTools related activities

HDRTools version 0.21 was tagged on Jan. 19, 2021. Changes include:

• Fix Y4M frame rate issue

A few days earlier version 0.20 was tagged with the following changes:

• Add output support for Y4M

• Add AOM CTC downscaling/upscaling filters (for Adaptive Streaming conditions)

• Support Gaussian Window SSIM

• XYB color format support

• Fixed WT distortion metric

A new release is intended to be released during the meeting which will include the following updates:

• 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)

JM, JSVM, JMVM related activities

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

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

Please file all issues related to the VVC reference software into the 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.

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.

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.

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.

It is noted that there is a document JVET-U0079 that proposes substantial cleanup of the software.

[JVET-V0004](https://jvet-experts.org/doc_end_user/current_document.php?id=10782) 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]

AHG meetings were held related to the preparation of the VVC verification tests in the SDR HD, 360° video and HDR categories on 2021-2-18 and 2021-3-25, respectively. The reports of these meetings are provided in the input documents JVET-V0042 and JVET-V0044.

The status of the discussion in the SDR HD and 360° video categories was reflected in the verification test plan document JVET-U2021. In the SDR HD low delay category, a set of 6 sequences was selected including three test sequences representing the conversational application scenario and three test sequences representing the gaming-type application scenario. Out of the four test sequences under consideration in the conversational application scenario, the verification test coordinators selected three as per request of the AHG meeting. The resulting set considered for subjective assessment includes the test sequence Beatriz, OfficeWalkAtWall, and OfficeWalkCeiling. For the SDR HD and 360° video categories, bitstreams for the rate points and configurations defined in JVET-U2021 were generated and verified for HM-16.22 and VTM-11.0 by JVET members in an off-line activity. For the SDR RA category, additional bitstreams for VVenC-0.3 bitstreams were provided as an additional test case.

The verification test coordinators are in contact with laboratories who offered support in contributing to the testing procedure. However, due to the COVID-19 situation, the activities for formal subjective assessments have been strongly reduced in many countries which impacts the progress of the VVC verification tests. A recent update on the local conditions in Italy opened the possibility for performing visual tests which is currently ongoing.

In the HDR category, efforts have been extended to identify suitable test sequences for both, PQ and HLG, including experts viewing sessions for determination of a QP range suitable for covering the intended subjective quality range.

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).

Copyright licenses have been updated for the following HDR PQ content located at ftp://jvet@ftp.ient.rwth-aachen.de, in folder /ctc/hdr:

• Market3\_1920x1080p\_50\_10b\_pq\_709\_ct2020\_420\_rev1.zip

• SunRise\_1920x1080p\_25\_10b\_pq\_709\_ct2020\_420\_rev1.7z

Related contributions

The following related contribution is submitted.

JVET-V0041 “AHG4: Status Report on SDR HD and 360 Video Verification Test Preparation” [M. Wien, V. Baroncini, Y. Ye (AHG coordinators)]

JVET-V0042 “AHG4: Agenda and report of the AHG meeting on the SDR HD and 360 video verification test preparation on 2021-02-18” [M. Wien, V. Baroncini, Y. Ye (AHG coordinators)]

JVET-V0043 “ AHG4: Status Report on HDR Video Verification Test Preparation” [A. Segall, M. Wien, V. Baroncini, K. Andersson]

JVET-V0044 “AHG4: Agenda and report of the AHG meeting on the HDR verification test preparation on 2021-03-25” [A. Segall, M. Wien, V. Baroncini (AHG coordinators)]

JVET-V0045 “Aerial 4K HDR (HLG) Test Data” [Pankaj Topiwala (FastVDO)]

The AHG reports JVET-V0042 and JVET-V0044 include a summary of the discussion of JVET-V0041, JVET-V0043. The contributions related to the verification tests should be further discussed during the meeting.

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.

• To create a BoG to progress the update of the verification test plan.

• 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.

Some expert viewing may be necessary during the meeting.

It is further suggested to merge the available test data into one ftp site (including previous JCT-VC test sets).

[JVET-V0005](https://jvet-experts.org/doc_end_user/current_document.php?id=10783) 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

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

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

Activities and Discussion

The AHG activities are on schedule with the preliminary timeline shown in section 2.

JVET-V2008 “Conformance testing for Versatile Video Coding (Draft 6)” was published on 31 Mar 2021. DIS in document WG05 N00037 was submitted based on JVET-V2008, but at the time of preparation of this report the MPEG DIS ballot has not been issued.

All provided bitstreams can be decoded using VTM12.1, although issues related to picture output remain in multilayer bitstreams (see VTM tickets 1444 and 1449). It is expected that VTM12.2 will correctly process all bitstreams.

Since JVET-V2008 was published, 28 revised bitstreams have been provided in 5 categories, as follows:

- OPI\_B\_Nokia: fixed incorrect aud\_pic\_type, from ticket 1477

- OLS\_\*\_Tencent: Added missing picture hash SEI messages

- VPS\_C\_ERICSSON: fixed incorrect aud\_pic\_type, from ticket 1477

- 8b422\_\*\_Sony

- 10b422\_\*\_Sony

There are not currently any known issues with the provided conformance bitstream packages, except for one (see below). If no issues are raised in the future, the expectation is that the bitstream providers will not need to provide any additional updated versions of the bitstreams.

Several conformance tickets were opened regarding provided conformance bitstreams (see “Conformance tickets” tab on https://jvet.hhi.fraunhofer.de/trac/vvc/ ). Revised bitstreams were submitted to address all tickets, except ticket 1446 related to SLICES\_A\_HUAWEI\_2.

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.

There are no related input contributions.

The procedure to exchange the bitstream (ftp cite, bitstream files, etc.) is specified in Sec 2 “Procedure” of [JVET-R2008](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=8861). 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:

• Thank all conformance bitstream providers for contributing bitstream packages

• Encourage conformance bitstream providers to review and revise text descriptions of provided bitstreams for the Conformance specification

• Encourage all JVET participants to review submitted bitstreams, provide feedback, and consider if the flexibility of the tested tool is sufficiently exercised, including consideration of input contribution JVET-U0108

• Output an updated conformance draft specification from this meeting

It is noted that there was an email by Christian Herglotz of FAU pointing out that some block sizes were missing in some bitstreams (based on information extracted from FAU’s bitstream analyser). This should be further clarified.

Resolution thanking the companies (Jill Boyce provides list).

[JVET-V0006](https://jvet-experts.org/doc_end_user/current_document.php?id=10784) JVET AHG report: 360° video coding, software and test conditions (AHG6) [J. Boyce, Y. He, K. Choi, J.-L. Lin, Y. Ye]

There is no update for 360Lib. The latest version is 360Lib-12.0 released on Dec. 29, 2020.

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-12.0 can be found at:

https://jvet.hhi.fraunhofer.de/svn/svn\_360Lib/tags/360Lib-12.0/

360Lib-12.0 testing results with VTM-12.0 can be found at:

ftp.ient.rwth-aachen.de/ahg/testresults/360Lib-12.0/VTM-12.0-360Lib-12.0\_CTC.xlsm

360Lib bug tracker

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

Table 1 is for the projection formats comparison using VTM-12.0 according to 360-degree video CTC (JVET-U2012) compared to that using VTM-11.0. Table 2 compares generalized cubemap (GCMP) coding and padded equi-rectangular projection (PERP) coding using VTM-12.0.

Table 3 is for PERP coding comparison between VTM-12.0 and HM-16.16. Table 4 is to compare GCMP coding with VTM-12.0 with and CMP coding with HM-16.16.

**Table 1. VTM-12.0 vs VTM-11.0 (VTM-11.0 as anchor)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **PERP: VTM-12.0 over VTM-11.0** | | | | | |
|  | **End-to-end**  **WS-PSNR** | | | **End-to-end**  **S-PSNR-NN** | | |
|  | Y | U | V | Y | U | V |
| Class S1 | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Class S2 | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **Overall** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |

**Table 2. VTM-12.0 PHEC vs PERP (PERP as anchor)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **PHEC Over PERP** | | | | | |
|  | **End-to-end WS-PSNR** | | | **End-to-end S-PSNR-NN** | | |
|  | Y | U | V | Y | U | V |
| Class S1 | -11.41% | -5.70% | -6.34% | -11.39% | -5.64% | -6.29% |
| Class S2 | -3.67% | 0.69% | 0.85% | -3.66% | 0.78% | 0.92% |
| **Overall** | -8.32% | -3.14% | -3.46% | -8.30% | -3.07% | -3.41% |

**Table 3. VTM-12.0 PERP vs HM-16.16 PERP (HM-16.16 PERP as anchor)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **VTM-12.0 PERP - Over HM-16.16 PERP** | | | | | |
|  | **End-to-end WS-PSNR** | | | **End-to-end S-PSNR-NN** | | |
|  | Y | U | V | Y | U | V |
| Class S1 | -30.20% | -37.96% | -40.49% | -30.19% | -37.98% | -40.46% |
| Class S2 | -36.18% | -36.80% | -39.08% | -36.16% | -36.83% | -39.14% |
| **Overall** | -32.59% | -37.50% | -39.93% | -32.58% | -37.52% | -39.93% |

**Table 4. VTM-12.0 GCMP vs HM-16.16 CMP (HM-16.16 CMP as anchor)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **VTM-12.0 GCMP - Over HM-16.16 CMP** | | | | | |
|  | **End-to-end WS-PSNR** | | | **End-to-end S-PSNR-NN** | | |
|  | Y | U | V | Y | U | V |
| Class S1 | -33.93% | -39.31% | -41.57% | -33.86% | -39.30% | -41.54% |
| Class S2 | -37.68% | -38.05% | -40.15% | -37.67% | -38.05% | -40.18% |
| **Overall** | -35.43% | -38.81% | -41.00% | -35.38% | -38.80% | -41.00% |

HM is without padding (but the version with padding should already be in the software, as it was developed in context of verification test). This should also be included in CTC. It was later confirmed that this can be done via modification of the config file, does not require a new version of the CTC document.

It is further noted that virtual boundary is not enabled in CTC for VTM in GCMP case.

The AHG recommends:

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

[JVET-V0007](https://jvet-experts.org/doc_end_user/current_document.php?id=10785) JVET AHG report: Coding of HDR/WCG material (AHG7) [A. Segall, E. François, W. Husak, S. Iwamura, D. Rusanovskyy]

The primary activity of the AhG was related to the mandates of (i) comparing the performance of the VTM for HDR/WCG content, (ii) study of objective metrics, (iii) coordinating with AHG8 on preparing HDR material in 12 bit 4:2:0 format, and (iv) coordinating with AHG4 in preparation for verification testing for HDR video content. This work is described in the following subsection.

Anchor Generation

The AhG generated CTC anchors for the VTM according to JVET-T2011. A summary of the performance is provided below, and more detailed information may be found in the included XLS data.

VTM 12.0 versus VTM 11.0

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **All Intra** | | | | | | | | | |
|  | **Over VTM-10.0** | | | | | | | | | |
|  |  |  | **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% | 99% | 97% |
| Class H2 |  |  |  |  |  | 0.00% | 0.00% | 0.00% | 98% | 95% |
| **Overall** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 99% | 96% |
|  |  |  |  |  |  |  |  |  |  |  |
|  | **Random Access** | | | | | | | | | |
|  | **Over VTM-10.0** | | | | | | | | | |
|  |  |  | **wPSNR** | | | **PSNR** | | |  |  |
|  | DE100 | PSNR-L100 | Y | U | V | Y | U | V | EncT | DecT |
| Class H1 | 8.27% | 6.16% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100% | 99% |
| Class H2 |  |  |  |  |  | 0.00% | 0.00% | 0.00% | 100% | 98% |
| **Overall** | 8.27% | 6.16% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 100% | 99% |

2.2 Objective Metrics

During the interim period, it was determined that there was an error in the calculation of the HDR metrics when the temporal prefilter was enabled (using the prefiltered as original). This affected the calculation of DE100 and PSNR-L100 in VTM11.0.

A fix was created and integrated into VTM-12.0 to address this error, and the effect can be seen in the results above. More information can be found at https://jvet.hhi.fraunhofer.de/trac/vvc/ticket/1452.

The AHG coordinated with AHG8 to review the configuration files used to generated the 12-bit HLG 4:2:0 content.

The AHG coordinated with AHG4 to prepare for the HDR verification test dry run. It is specifically highlighted that AHG4 received and studied three, new HDR sequences during the AHG period. AHG7 provided support in preparing and converting the sequences.

There are six contributions related to HDR video coding.

|  |  |  |
| --- | --- | --- |
| [JVET-V0007](file:///C:\Users\ohm\AppData\Local\Temp\Temp1_JVET-V0007-v1.zip\current_document.php?id=10785) | JVET AHG report: Coding of HDR/WCG material (AHG7) | A. Segall, E. François, W. Husak, S. Iwamura, D. Rusanovskyy |
| [JVET-V0043](file:///C:\Users\ohm\AppData\Local\Temp\Temp1_JVET-V0007-v1.zip\current_document.php?id=10689) | AHG4: Status Report on HDR Video Verification Test Preparation | [A. Segall](mailto:asegall@sharplabs.com), [M. Wien](mailto:wien@lfb.rwth-aachen.de), [V. Baroncini](mailto:baroncini@gmx.com), [K. Andersson](mailto:kenneth.r.andersson@ericsson.com) |
| [JVET-V0044](file:///C:\Users\ohm\AppData\Local\Temp\Temp1_JVET-V0007-v1.zip\current_document.php?id=10691) | AHG4: Agenda and report of the AHG meeting on the HDR verification test preparation on 2021-03-25 | [A. Segall](mailto:asegall@sharplabs.com), [M. Wien](mailto:wien@lfb.rwth-aachen.de), [V. Baroncini (AHG coordinators)](mailto:baroncini@gmx.com) |
| [JVET-V0045](file:///C:\Users\ohm\AppData\Local\Temp\Temp1_JVET-V0007-v1.zip\current_document.php?id=10692) | Aerial 4K HDR (HLG) Test Data | [Pankaj Topiwala (FastVDO)](mailto:pankajtva@gmail.com) |
| JVET-V0078 | AHG10: QP control for very smooth blocks | K. Andersson, J. Enhorn, J. Ström (Ericsson) |
| [JVET-V0107](file:///C:\Users\ohm\AppData\Local\Temp\Temp1_JVET-V0007-v1.zip\current_document.php?id=10755) | Editorial suggestion for JVET CTC on HDR/WCG | [D. Rusanovskyy (Qualcomm)](mailto:dmytror@qti.qualcomm.com), [T.Hashimoto (Sharp)](mailto:tomonori.hashimoto@sharp.co.jp) |

The AHG recommends the following:

• Review all input contributions

[JVET-V0008](https://jvet-experts.org/doc_end_user/current_document.php?id=10786) 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]

The AHG used the main JVET reflector, jvet@lists.rwth-aachen.de, with [AHG8] in message headers. A message was sent to the reflector with respect to the conversion of HLG content within the CTC on 27th January, a CE change notification on 25th March, and a formal kick-off message was sent on 29th March. In addition to the messages on the reflector, the the co-chairs received a question about the calendar for establishing a 12-bit profile for VVC.

The primary activities of the AhG were the updates to the high bit depth CTC (JVET-U2018) and the core experiment on entropy coding for high bit depth and high bit rate coding (JVET-V0022).

As a part of refining the CTC, AhG conducted HDR 12bit test material preparation. This included generation of HDR PQ/HDR HLG test sequences in 4:2:0 format and agreed conversion UHD to HD resolution for HLG sequences. AhG8 would like to thank coordinators of AhG7 for discussion on HDR conversion practices and on HDR metric specification and implementation.

In total, there are 45 high bit depth related contributions, but none related to high frame rate. The following section lists these contributions.

Contributions

The contributions can be split into CE and CE-related (28 contributions), and others (17 contributions).

CE and CE-related

JVET-V0022, “CE: Summary Report on Entropy Coding for High Bit Depth and High Bit Rate Coding”, A.Browne, T. Hashimoto, H. Jhu, D. Rusanovskyy, K. Kawamura, T. Zhou (CE co-ordinators)

JVET-V0046, “CE-1.7: Rice parameter derivation for high bit-depth coding”, T. Hashimoto, T. Ikai (Sharp)

JVET-V0047, “CE-3.1 and CE-3.2: Transform coefficients range extension for high bit-depth coding”, T. Zhou, T. Chujoh, T. Ikai (Sharp)

JVET-V0048, “CE-4.1: Combination of CE-3.1, CE-1.5 and CE-2.1”, T. Zhou, T. Chujoh, T. Ikai (Sharp)

JVET-V0049, “CE-4.2: Combination of CE-3.2, CE-1.5 and CE-2.1”, T. Zhou, T. Chujoh, T. Ikai (Sharp)

JVET-V0050, “CE-1.5, CE-1.6, CE-2.2 and CE-2.3: Rice parameter selection for high bit depths”, A. Browne, S. Keating, K. Sharman (Sony)

JVET-V0051, “CE-4.3 and CE-4.4: Combinations of CE-1.6 and CE-2.3 with CE-3.2”, A. Browne, S. Keating, K. Sharman (Sony)

JVET-V0052, “CE-1.1, CE-1.2 and CE-1.4: On the Rice parameter derivation for high bit-depth coding”, D. Rusanovskyy, M. Karczewicz, L. P. Van, M. Coban (Qualcomm)

JVET-V0053, “CE-4.5, CE-4.6 and CE-4.7: Combinations of RRC tests CE-1.1, CE-1.2 and CE-1.4 with CE-3.1/CE-3.2”, D. Rusanovskyy, M. Karczewicz, L. P. Van, M. Coban (Qualcomm)

JVET-V0054, “CE-2.1: Slice based Rice parameter selection for transform skip residual coding”, H.-J. Jhu, X. Xiu, Y.-W. Chen, W. Chen, C.-W. Kuo, X. Wang (Kwai Inc.)

JVET-V0084, “CE-related: On Rice parameter selection for regular residual coding (RRC) at high bit depths”, A. Browne, S. Keating, K. Sharman (Sony)

JVET-V0085, “CE-related: On Rice parameter selection for transform skip residual coding (TSRC) at high bit depths”, A. Browne, S. Keating, K. Sharman (Sony)

JVET-V0106, “CE-related: On history-enhanced method of Rice parameter derivation for regular residual coding (RRC) at high bit depths”, D. Rusanovskyy, L. Pham Van, M. Coban, M. Karczewicz (Qualcomm)

JVET-V0123, “CE-related: On prefix code length of remaining level coding for high bit depth and high bit rate coding”, Y. Yu, Z. Xie, F. Wang, H. Yu, D. Wang (OPPO)

JVET-V0134, “Crosscheck of CE-1.4 from JVET-V0052 (CE-1.1, CE-1.2 and CE-1.4: On the Rice parameter derivation for high bit-depth coding)”, A. Browne (Sony)

JVET-V0135, “Crosscheck of CE-4.5 and CE-4.7 from JVET-V0053 (CE-4.5, CE-4.6 and CE-4.7: Combinations of RRC tests CE-1.1, CE-1.2 and CE-1.4 with CE-3.1/CE-3.2)”, A. Browne (Sony)

JVET-V0136, “Crosscheck of JVET-V0047 (CE-3.1 and CE-3.2: Transform coefficients range extension for high bit-depth coding)”, A. Browne (Sony)

JVET-V0138, “Crosscheck of CE-1.2 from JVET-V0052 (CE-1.1, CE-1.2 and CE-1.4: On the Rice parameter derivation for high bit-depth coding)”, T. Hashimoto (Sharp)

JVET-V0139, “Cross-check report for CE1.5 and CE1.6 of JVET-V0050 and CE4.3 of JVET-V0051”, D.Rusanovskyy (Qualcomm)

JVET-V0140, “Cross-check report for CE-1.7 of JVET-V0046 and CE-4.2 of JVET-V0049”, D.Rusanovskyy (Qualcomm)

JVET-V0142, “Crosscheck of JVET-V0048 (CE-4.1: Combination of CE-3.1, CE-1.5 and CE-2.1)”, H.-J. Jhu (Kwai Inc.)

JVET-V0143, “Crosscheck of CE-2.2 and CE-2.3 from JVET-V0050 (CE-1.5, CE-1.6, CE-2.2 and CE-2.3: Rice parameter selection for high bit depths)”, H.-J. Jhu (Kwai Inc.)

JVET-V0144, “Crosscheck of CE-4.4 from JVET-V0051 (CE-4.3 and CE-4.4: Combinations of CE-1.6 and CE-2.3 with CE-3.2)”, H.-J. Jhu (Kwai Inc.)

JVET-V0145, “Crosscheck of CE-4.6 from JVET-V0053 (CE-4.5, CE-4.6 and CE-4.7: Combinations of RRC tests CE-1.1, CE-1.2 and CE-1.4 with CE-3.1/CE-3.2)”, T. Zhou (Sharp)

JVET-V0155, “Crosscheck of CE-1.1 from JVET-V0052 (CE-1.1, CE-1.2 and CE-1.4: On the Rice parameter derivation for high bit-depth coding)”, M. G. Sarwer (Alibaba)

JVET-V0156, “Crosscheck of JVET-V0106 (CE-related: On history-enhanced method of Rice parameter derivation for regular residual coding (RRC) at high bit depths)”, M. G. Sarwer (Alibaba)

JVET-V0157, “Crosscheck of JVET-V0084 (CE-related: On Rice parameter selection for regular residual coding (RRC) at high bit depths)”, M. G. Sarwer (Alibaba)

JVET-V0161, “Crosscheck of JVET-V0123 (CE-related: On prefix code length of remaining level coding for high bit depth and high bit rate coding)”, H.-J. Jhu (Kwai Inc.)

Other contributions

JVET-V0059, “AHG8: CABAC-bypass alignment for high bit-depth coding”, M.G. Sarwer, J. Chen, Y. Ye, R. -L. Liao (Alibaba)

JVET-V0066, “AHG8: Encoder improvements to palette coding for high bit depth”, T. Tsukuba, M. Ikeda, T. Suzuki (Sony)

JVET-V0067, “AHG8: A constraint of max transform size for high bit depth and high bit rate coding”, K. Kondo, M. Ikeda (Sony)

JVET-V0068, “AHG8: A constraint of max CTU size for high bit depth and high bit rate coding”, K. Kondo, M. Ikeda (Sony)

JVET-V0121, “AHG8: on 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-V0122, “AHG8: a full-bypass mode in residual coding for high bit depth and high bit rate extensions”, F. Wang, Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO)

JVET-V0124, “AHG8: Weighted Prediction for VVC High Bit Depth Extension”, Y. Yu, H. Yu, Z. Xie, F. Wang, D. Wang (OPPO)

JVET-V0131, “AHG8: Fixing the forward transform matrices for high bit depth coding”, K. Naser, F. Galpin, F. F. Le Léannec, P. De Lagrange (InterDigital)

JVET-V0133, “AHG8: Content Adaptive Transform Precision for High Bit Depth Coding”, K. Naser, F. Galpin, F. Le Léannec, P. De Lagrange (InterDigital)

JVET-V0141, “Cross-check report of AHG8: CABAC-bypass alignment for high bit-depth coding (JVET-V0059)”, D.Rusanovskyy (Qualcomm)

JVET-V0150, “AHG8: A study on bin rate of VTM-12.0 for high bit depth coding”, T. Tsukuba, M. Ikeda, T. Suzuki (Sony)

JVET-V0158, “Crosscheck of JVET-V0121 (AHG8: on coding of last significant coefficient position for high bit depth and high bit rate extensions)”, M. G. Sarwer (Alibaba)

JVET-V0159, “Crosscheck of JVET-V0067 (AHG8: A constraint of max transform size for high bit depth and high bit rate coding)”, T. Hashimoto (Sharp)

JVET-V0160, “Crosscheck of JVET-V0068 (AHG8: A constraint of max CTU size for high bit depth and high bit rate coding)”, T. Hashimoto (Sharp)

JVET-V0162, “Crosscheck of JVET-V0066 (AHG8: Encoder improvements to palette coding for high bit depth)”, H.-J. Jhu (Kwai Inc.)

JVET-V0163, “Crosscheck of JVET-V0122 (AHG8: a full-bypass mode in residual coding for high bit depth and high bit rate extensions)”, A. Browne (Sony)

JVET-V0166, “Cross-check report on AHG8: Content Adaptive Transform Precision for High Bit Depth Coding (JVET-V0133)”, D. Rusanovskyy (Qualcomm)

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.

VTM 12.0 versus HM16.23

Low QP range

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HDR PQ** |  |  | **AI** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -5.42% | -6.08% | -6.69% | 2996% | 175% |
| PQ422 | -8.17% | -12.30% | -12.82% | 2550% | 169% |
| **Overall** | -6.79% | -9.19% | -9.75% | 2773% | 172% |
|  |  |  |  |  |  |
|  |  |  | **LDB** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -6.00% | -4.18% | -5.75% | 300% | 164% |
| PQ422 | -7.21% | -11.17% | -11.80% | 394% | 164% |
| **Overall** | -6.61% | -7.67% | -8.77% | 347% | 164% |
|  |  |  |  |  |  |
|  |  |  | **RA** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -6.14% | -4.32% | -5.56% | 343% | 164% |
| PQ422 | -7.47% | -11.50% | -11.72% | 474% | 164% |
| **Overall** | -6.81% | -7.91% | -8.64% | 409% | 164% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HDR HLG** |  |  | **AI** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| HLG444 | -3.39% | -5.40% | -5.68% | 3198% | 171% |
| HLG422 | -3.97% | -6.73% | -6.70% | 2943% | 167% |
| **Overall** | -3.68% | -6.06% | -6.19% | 3070% | 169% |
|  |  |  |  |  |  |
|  |  |  | **LDB** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| HLG444 | -5.36% | -7.49% | -6.79% | 401% | 158% |
| HLG422 | -4.92% | -8.86% | -7.23% | 495% | 162% |
| **Overall** | -5.14% | -8.18% | -7.01% | 448% | 160% |
|  |  |  |  |  |  |
|  |  |  | **RA** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| HLG444 | -5.92% | -8.08% | -7.19% | 449% | 157% |
| HLG422 | -5.51% | -9.19% | -7.44% | 612% | 162% |
| **Overall** | -5.72% | -8.64% | -7.32% | 531% | 160% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SVT RGB** |  |  | **AI** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrG | psnrB | psnrR | EncT | DecT |
| SVT16 | 36.32% | 34.27% | 34.20% | 4934% | 144% |
| SVT12 | 0.13% | 0.68% | 0.58% | 5230% | 138% |
| **Overall** | 18.22% | 17.48% | 17.39% | 5082% | 141% |
|  |  |  |  |  |  |
|  |  |  | **LDB** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrG | psnrB | psnrR | EncT | DecT |
| SVT16 | 20.21% | 21.49% | 21.55% | 584% | 144% |
| SVT12 | -2.91% | -0.57% | -0.55% | 629% | 141% |
| **Overall** | 8.65% | 10.46% | 10.50% | 606% | 143% |
|  |  |  |  |  |  |
|  |  |  | **RA** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrG | psnrB | psnrR | EncT | DecT |
| SVT16 | 17.97% | 20.35% | 20.33% | 722% | 144% |
| SVT12 | -3.26% | -0.83% | -0.82% | 784% | 140% |
| **Overall** | 7.36% | 9.76% | 9.76% | 753% | 142% |

Lossless

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PQ** | **All Intra** | | | **Low delay B** | | | **Random Access** | | |
| **ratio** | | bit-rate savings | **ratio** | | bit-rate savings | **ratio** | | bit-rate savings |
| HM16.23 | VTM12.0 | HM16.23 | VTM12.0 | HM16.23 | VTM12.0 |
| PQ444 | 2.5 | 2.5 | 3.01% | 3.1 | 3.0 | 4.21% | 3.1 | 3.0 | 3.83% |
| PQ422 | 2.3 | 2.2 | 2.86% | 2.9 | 2.8 | 3.13% | 2.9 | 2.8 | 2.78% |
| **Overall** | **2.4** | **2.4** | **2.93%** | **3.0** | **2.9** | **3.67%** | **3.0** | **2.9** | **3.30%** |
| Enc Time[%] | 7413% | | | 878% | | | 867% | | |
| Dec Time[%] | 158% | | | 139% | | | 140% | | |
|  |  |  |  |  |  |  |  |  |  |
| **HLG** | **All Intra** | | | **Low delay B** | | | **Random Access** | | |
| **ratio** | | bit-rate savings | **ratio** | | bit-rate savings | **ratio** | | bit-rate savings |
| HM16.23 | VTM12.0 | HM16.23 | VTM12.0 | HM16.23 | VTM12.0 |
| HLG444 | 1.8 | 1.7 | 6.56% | 1.9 | 1.9 | 2.10% | 2.0 | 1.9 | 1.87% |
| HLG422 | 1.7 | 1.6 | 7.55% | 1.9 | 1.9 | 2.08% | 1.9 | 1.9 | 1.83% |
| **Overall** | **1.7** | **1.6** | **7.05%** | **1.9** | **1.9** | **2.09%** | **1.9** | **1.9** | **1.85%** |
| Enc Time[%] | 8301% | | | 765% | | | 722% | | |
| Dec Time[%] | 138% | | | 116% | | | 117% | | |
|  |  |  |  |  |  |  |  |  |  |
| **SVT** | **All Intra** | | | **Low delay B** | | | **Random Access** | | |
| **ratio** | | bit-rate savings | **ratio** | | bit-rate savings | **ratio** | | bit-rate savings |
| HM16.23 | VTM12.0 | HM16.23 | VTM12.0 | HM16.23 | VTM12.0 |
| SVT16 | 1.2 | 0.7 | 73.47% | 1.2 | 0.7 | 80.20% | 1.2 | 0.7 | 80.25% |
| SVT12 | 1.3 | 1.0 | 31.02% | 1.3 | 1.0 | 29.88% | 1.3 | 1.0 | 29.30% |
| **Overall** | **1.2** | **0.8** | **52.24%** | **1.3** | **0.8** | **55.04%** | **1.3** | **0.8** | **54.78%** |
| Enc Time[%] | 7727% | | | 831% | | | 802% | | |
| Dec Time[%] | 133% | | | 151% | | | 152% | | |

VTM CE anchor versus HM16.23

The VTM CE anchor consisted of VTM12.0 with the RRC and TSRC rice coding fixes in the initial versions of CE-1.1 and CE-2.1 respectively.

Low QP range

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HDR PQ** |  |  | **AI** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -5.82% | -6.50% | -7.13% | 3068% | 175% |
| PQ422 | -8.66% | -12.78% | -13.32% | 2562% | 169% |
| **Overall** | -7.24% | -9.64% | -10.23% | 2815% | 172% |
|  |  |  |  |  |  |
|  |  |  | **LDB** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -6.06% | -4.23% | -5.79% | 301% | 165% |
| PQ422 | -7.30% | -11.23% | -11.86% | 397% | 164% |
| **Overall** | -6.68% | -7.73% | -8.82% | 349% | 164% |
|  |  |  |  |  |  |
|  |  |  | **RA** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -6.26% | -4.44% | -5.69% | 347% | 165% |
| PQ422 | -7.61% | -11.64% | -11.85% | 482% | 164% |
| **Overall** | -6.93% | -8.04% | -8.77% | 414% | 164% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **HDR HLG** |  |  | **AI** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| HLG444 | -3.75% | -5.67% | -5.96% | 3188% | 171% |
| HLG422 | -4.44% | -7.11% | -7.10% | 2973% | 167% |
| **Overall** | -4.09% | -6.39% | -6.53% | 3080% | 169% |
|  |  |  |  |  |  |
|  |  |  | **LDB** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| HLG444 | -5.34% | -7.53% | -6.84% | 406% | 159% |
| HLG422 | -4.90% | -8.88% | -7.29% | 503% | 164% |
| **Overall** | -5.12% | -8.20% | -7.06% | 454% | 162% |
|  |  |  |  |  |  |
|  |  |  | **RA** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| HLG444 | -5.93% | -8.12% | -7.24% | 455% | 157% |
| HLG422 | -5.54% | -9.24% | -7.51% | 616% | 162% |
| **Overall** | -5.73% | -8.68% | -7.37% | 535% | 159% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SVT RGB** |  |  | **AI** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrG | psnrB | psnrR | EncT | DecT |
| SVT16 | 0.70% | 0.82% | 0.79% | 4842% | 119% |
| SVT12 | -2.54% | -1.95% | -2.03% | 5252% | 137% |
| **Overall** | -0.92% | -0.57% | -0.62% | 5047% | 128% |
|  |  |  |  |  |  |
|  |  |  | **LDB** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrG | psnrB | psnrR | EncT | DecT |
| SVT16 | -0.52% | 1.13% | 1.13% | 572% | 132% |
| SVT12 | -3.31% | -1.05% | -1.02% | 647% | 142% |
| **Overall** | -1.91% | 0.04% | 0.05% | 609% | 137% |
|  |  |  |  |  |  |
|  |  |  | **RA** |  |  |
|  |  |  | **Over HM16.23** |  |  |
|  | psnrG | psnrB | psnrR | EncT | DecT |
| SVT16 | -0.86% | 0.91% | 0.91% | 714% | 132% |
| SVT12 | -3.72% | -1.32% | -1.29% | 802% | 142% |
| **Overall** | -2.29% | -0.21% | -0.19% | 758% | 137% |

Lossless

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PQ** | **All Intra** | | | **Low delay B** | | | **Random Access** | | |
| **ratio** | | bit-rate savings | **ratio** | | bit-rate savings | **ratio** | | bit-rate savings |
| HM16.23 | VTM\_U\_CE | HM16.23 | VTM\_U\_CE | HM16.23 | VTM\_U\_CE |
| PQ444 | 2.5 | 2.5 | 0.89% | 3.1 | 3.0 | 4.80% | 3.1 | 3.0 | 4.47% |
| PQ422 | 2.3 | 2.3 | -0.18% | 2.9 | 2.8 | 3.40% | 2.9 | 2.8 | 3.13% |
| **Overall** | **2.4** | **2.4** | **0.36%** | **3.0** | **2.9** | **4.10%** | **3.0** | **2.9** | **3.80%** |
| Enc Time[%] | 7206% | | | 976% | | | 983% | | |
| Dec Time[%] | 152% | | | 147% | | | 149% | | |
|  |  |  |  |  |  |  |  |  |  |
| **HLG** | **All Intra** | | | **Low delay B** | | | **Random Access** | | |
| **ratio** | | bit-rate savings | **ratio** | | bit-rate savings | **ratio** | | bit-rate savings |
| HM16.23 | VTM\_U\_CE | HM16.23 | VTM\_U\_CE | HM16.23 | VTM\_U\_CE |
| HLG444 | 1.8 | 1.7 | 3.27% | 1.9 | 1.9 | 4.25% | 2.0 | 1.9 | 4.08% |
| HLG422 | 1.7 | 1.6 | 2.37% | 1.9 | 1.9 | 3.66% | 1.9 | 1.9 | 3.49% |
| **Overall** | **1.7** | **1.7** | **2.82%** | **1.9** | **1.9** | **3.96%** | **1.9** | **1.9** | **3.79%** |
| Enc Time[%] | 7782% | | | 817% | | | 767% | | |
| Dec Time[%] | 132% | | | 128% | | | 126% | | |
|  |  |  |  |  |  |  |  |  |  |
| **SVT** | **All Intra** | | | **Low delay B** | | | **Random Access** | | |
| **ratio** | | bit-rate savings | **ratio** | | bit-rate savings | **ratio** | | bit-rate savings |
| HM16.23 | VTM\_U\_CE | HM16.23 | VTM\_U\_CE | HM16.23 | VTM\_U\_CE |
| SVT16 | 1.2 | 1.2 | 1.46% | 1.2 | 1.2 | 1.18% | 1.2 | 1.2 | 1.11% |
| SVT12 | 1.3 | 1.2 | 3.12% | 1.3 | 1.3 | 2.71% | 1.3 | 1.3 | 2.55% |
| **Overall** | **1.2** | **1.2** | **2.29%** | **1.3** | **1.2** | **1.95%** | **1.3** | **1.2** | **1.83%** |

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.

No study on HFR so far.

Generally, the gains over HEVC are much lower. For 16 bit, losses are observed even still with the bug-fixed CE anchor.

Target of 12 bit profile? 4:2:0, 4:2:2 or 4:4:4? Needs further consideration, and no contributions on this yet. If 4:2:0 might be interesting for consumer applications, is the low QP range investigation appropriate for that?

One expert mentions that also higher bit depth profiles (14, 16) would be interesting to support emerging types of content.

[JVET-V0009](https://jvet-experts.org/doc_end_user/current_document.php?id=10787) 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]

At total of 16 contributions are identified relating to AHG9, of which:

• 8 contributions relate to the mandate on the potential needs for additional SEI messages;

• 5 contributions relate to the mandate on SEI messages previously adopted or specified in AVC, HEVC, VVC, or VSEI;

• 1 contribution relates to the mandate on possible need of mandatory post processing in the context of SEI messages; and

• 2 contributions are crosschecks of proposed software that relate to the mandate on showcase and usage information for SEI messages

The following is a list of contributions related to AHG9.

New SEI messages

Post-filtering

JVET-V0058 AHG9: On post-filter SEI [M. M. Hannuksela, E. B. Aksu, F. Cricri, H. R. Tavakoli (Nokia)]

JVET-V0091 AHG9/AHG11: SEI messages for carriage of neural network information for post-filtering [B. Choi, Z. Li, W. Wang, W. Jiang, X. Xu, S. Wenger, S. Liu (Tencent)]

Decoder initialization

JVET-V0081 AHG9: On the decoder initialization information [Hendry, S. Lee, S. Kim (LGE), Y.-K. Wang, K. Zhang, L. Zhang, Y. Wang, J. Xu, Z. Deng (Bytedance)]

JVET-V0112 AHG9: On Bitstream Properties Signalling for Decoder Initialization [S. Deshpande (Sharp)]

Other

JVET-V0062 AHG9: Picture quality metrics SEI message [Y. He, M. Coban, M. Karczewicz (Qualcomm)]

JVET-V0069 AHG9: Decoded GDR clean area hash SEI message [L. Wang, S. Hong, K. Panusopone, M. M. Hannuksela (Nokia)]

JVET-V0071 AHG9: Temporal sublayer information SEI message [R. Sjöberg, M. Pettersson, M. Damghanian, J. Ström (Ericsson)]

JVET-V0108 AHG9: Colour Transform Information SEI message [E. François, M. Radosavljevic, P. de Lagrange, F. Leleannec (InterDigital)]

SEI messages previously adopted in AVC, HEVC, VVC, and VSEI

JVET-V0061AHG9: Display orientation information SEI message [Y. He, M. Coban, M. Karczewicz (Qualcomm), J. Boyce (Intel)]

JVET-V0063 AHG9: On the scalability dimension information SEI message [Y. Wang, Y.-K. Wang, L. Zhang (Bytedance)]

JVET-V0064 AHG9: On the MAI, DRI, and ACI SEI messages and their interactions with the SDI SEI message [Y.-K. Wang, Y. Wang, L. Zhang (Bytedance)]

JVET-V0065 AHG9: On the DRAP and EDRAP indication SEI messages [Y.-K. Wang, L. Zhang, Y. Wang (Bytedance)]

JVET-V0093 AHG9: Film grain estimation and film grain synthesis for VVC – SEI message characteristics, film grain estimation and film grain synthesis modules [Miloš Radosavljević, Edouard François (InterDigital)]

On possible mandatory post processing in the context of SEI messages

JVET-V0113 Thoughts on SEI messages [Walt Husak]

Crosschecks

JVET-V0151 Crosscheck of JVET-V0093 (AHG9: Film grain estimation and film grain synthesis for VVC Film grain characteristics SEI message , film grain estimation and film grain synthesis modules) [P. Yin, S.McCarthy (Dolby)]

JVET-V0152 Crosscheck of JVET-V0108 (AHG9: Colour Transform Information SEI message) [F. Pu (Dolby)]

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 besides the AHG kickoff message sent to the JVET reflector during the AHG period.

Recommendations

The AHG recommends to:

• Review all related contributions;

• Continue SEI messages studies.

[JVET-V0010](https://jvet-experts.org/doc_end_user/current_document.php?id=10788) JVET AHG report: Encoding algorithm optimization (AHG10) [A. Duenas, A. Tourapis, A. Norkin, R. Sjöberg]

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

JVET-V0056 [AHG10] GOP-based temporal filter improvements

This contribution proposes a set of changes to the GOP-based temporal filter in the VTM that aim at improving the BD-rate performance. These changes were first proposed in JVET-U0056.

The average Y/U/V BD-rate improvements using the VTM-12 CTC are reported to be equal to −1.4%/−1.0%/−1.1% for RA for the 3 colour components respectively. All BDR numbers were computed using the unfiltered source sequences as the reference.

In addition, the same changes were implemented in the HM where the BD-rate improvements were found to be equal to -1.0%/-1.0%/-1.0% under the same conditions.

The VTM proposal itself is unchanged since JVET-U0056, except for fixing a bug regarding motion compensation of blocks along the edge a picture. The code for the original proposal has been available since 2021-01-22 on the VTM Github.

Subjective viewing has been carried out by the proponents, where the anchor and test were found to have similar quality for all viewed sequences.

Finally, an HM version of the algorithm has been implemented, with similar performance. While the algorithm is the same for the HM implementation, the multiplications for the weights have been slightly modified which is explained in the contribution.

We do observe a relationship of the QP used with both the performance of this scheme as well as its impact on encoder and decoder complexity. In particular, for the lower QP values, i.e. QP values of 22 and 27, quite often a considerable bitrate reduction is observed. For the higher QP values an increase in bitrate for some material was also observed. Although it is not clear how much such numbers could be trusted, at the higher QP values an increase in encoding and decoding time was also observed that might not be reflected in the averages.

The proposal was implemented on top of VTM-12.0. The method is applied for RA coding only. The BD-rate result relative to VTM-12.0 is reported below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Over VTM-12.0** | | | | |
|  | Y | U | V | EncT | DecT |
| Class A1 | -1.83% | -1.86% | -2.55% | 93% | 95% |
| Class A2 | -1.07% | -0.78% | -0.79% | 97% | 104% |
| Class B | -1.40% | -0.95% | -0.98% | 95% | 93% |
| Class C | -1.19% | -0.57% | -0.30% | 98% | 109% |
| Class E |  |  |  |  |  |
| **Overall** | -1.36% | -1.00% | -1.08% | 96% | 100% |
| Class D | -1.48% | -0.09% | -0.15% | 99% | 106% |
| Class F | 0.00% | 0.00% | 0.00% | 96% | 119% |

It is pointed out that the encoding time may not be reliable, but includes runtime of MCTF.

In addition, these are the results when compared to VTM-12.0 without temporal filtering:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Over VTM-12.0 with MCTF = 0** | | | | |
|  | Y | U | V | EncT | DecT |
| Class A1 | -4.54% | -8.89% | -8.66% | 90% | 81% |
| Class A2 | -6.21% | -10.83% | -10.59% | 92% | 88% |
| Class B | -7.66% | -9.17% | -9.39% | 94% | 86% |
| Class C | -2.82% | -4.37% | -3.86% | 98% | 95% |
| Class E |  |  |  |  |  |
| **Overall** | -5.45% | -8.17% | -8.01% | 94% | 88% |
| Class D | -0.81% | -3.53% | -3.47% | 98% | 92% |
| Class F | 0.00% | 0.00% | 0.00% | 99% | 100% |

Finally, the contribution presents also results using HM-16.23 compared to the CTC:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Random Access Main 10** | | |
|  | Y | U | V |
| Class A1 | -1.4% | -1.5% | -1.8% |
| Class A2 | -1.1% | -1.1% | -1.2% |
| Class B | -1.6% | -1.8% | -1.7% |
| Class C | -1.0% | -0.9% | -1.0% |
| Class D | -1.0% | -0.8% | -0.6% |
| Class E |  |  |  |
| Class F | 0.0% | 0.0% | 0.0% |
| **Overall** | -1.0% | -1.0% | -1.0% |
| Enc Time[%] | 107% | | |
| Dec Time[%] | 100% | | |

It is interesting to note that the proposed method, and ignoring the per QP variations mentioned earlier, shows an average decoder complexity reduction on the results. Nevertheless, it is indicated in JVET-U0056 that there seems to be no discernible impact on encoding complexity.

JVET-V0057 [AHG10] Block importance mapping

This contribution proposes a new algorithm for the VTM to signal CTU level QP delta values in pictures that will be used for reference. The QP value to use for each CTU is based on the estimated importance of a given CTU for future pictures and the QP selected is in the range of −2 to +2 relative to the picture QP.

The method was reportedly implemented and tested together with the temporal filter changes first presented in JVET-U0056 under the VTM-12.0 RA test configurations. The method is not proposed to be used for AI or LD configurations.

The average CTC RA Y/U/V BD-rate improvements for the proposed changes together with JVET-V0056 compared to VTM-12 with JVET-V0056 are reported to be −1.4%/−3.9%/−3.8% for RA. Comparing the proposed changes together with JVET-V0056 to the regular VTM-12, the average BD-rates are −2.8%/−5.0%/−4.9%.

It is proposed by the contribution to adopt the algorithm into the VTM software.

Version 2 of this document provides a statement that the results for the HM implementation will be added later on.

The estimation of the difference between pictures is based on calculations performed in the temporal filter. The QP is only modified for frames where the temporal filter is active, i.e., where the POC is divisible by eight.

The contribution proposes that rather than using a fixed QP for each temporal layer, individual QP values for each CTU is set based on how important it is calculated the CTU will be to future pictures. This was proposed to estimate this importance is estimated by measuring the difference between motion compensated pictures.

The contribution indicates that there seems to be no discernible impact on complexity.

The proposal was implemented on top of VTM-12.0 with JVET-V0056 [2]. The method is applied for RA coding only. The BD-rate result relative to VTM-12.0 with JVET-V0056 is reported below. The contribution says that encoding and decoding times are not accurate.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Over VTM-12.0 with V0056** | | | | |
|  | Y | U | V | EncT | DecT |
| Class A1 | -0.66% | -0.84% | -2.41% | 108% | 114% |
| Class A2 | -1.09% | -4.11% | -2.93% | 108% | 111% |
| Class B | -2.02% | -4.97% | -5.10% | 107% | 116% |
| Class C | -1.56% | -4.88% | -3.94% | 132% | 126% |
| Class E |  |  |  |  |  |
| **Overall** | -1.44% | -3.95% | -3.82% | 114% | 117% |
| Class D | -1.08% | -4.69% | -4.35% | 125% | 152% |
| Class F | -2.61% | -3.48% | -2.13% | 105% | 107% |

In addition, these are the results when comparing the proposal together with JVET-V0056 to VTM-12.0:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Over VTM-12.0** | | | | |
|  | Y | U | V | EncT | DecT |
| Class A1 | -2.52% | -2.76% | -4.95% | 101% | 109% |
| Class A2 | -2.22% | -4.98% | -3.80% | 105% | 116% |
| Class B | -3.47% | -5.95% | -6.11% | 101% | 108% |
| Class C | -2.76% | -5.50% | -4.30% | 129% | 138% |
| Class E |  |  |  |  |  |
| **Overall** | -2.84% | -5.00% | -4.94% | 109% | 117% |
| Class D | -2.61% | -4.81% | -4.53% | 123% | 161% |
| Class F | -2.61% | -3.48% | -2.13% | 101% | 114% |

The method was also tested on HM-16.23, results will be provided at a later stage.

JVET-V0078 AHG 10: QP control for very smooth blocks

Large transforms are excellent in reducing spatial redundancy. However, when source samples are very smooth the subjective quality can drop very quickly in such areas with too coarse quantization. It is therefore suggested to reduce the QP according to a linear model for very smooth regions to prolong the breakdown of the subjective quality. Since smooth regions are relatively easy to code it is claimed that the overhead in bitrate can be manageable.

The impact on BDR has been tested on HDR:

VTM-12.0 HDR RA CTC (luma/Cb/Cr): 0.14%/0.14%/0.02%

HM-16.22 (TF on) HDR RA CTC (luma/Cb/Cr): 0.49%/-0.36%/-0.28%

It is asserted that the proposal can improve subjective quality of smooth regions for HDR and it is suggested to update the VTM and HM with this functionality. The proposal leads to an objective BDR loss by reducing the QP for smooth blocks. In contrast, it is indicating that the subjective quality can be improved in smooth areas.

Objective results

The method has been tested for HDR RA CTC. It can be noted that there is a drop in BDR by the proposed method and that the drop is slightly larger for the HM than for the VTM. The encoding and decoding times are not reliable.

Table 1: VTM-12.0 for CTC HDR RA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **PSNR** | | |  |  |
|  | Y | U | V | EncT | DecT |
| Class H1 | 0,17% | 0,19% | 0,07% | 100% | 102% |
| Class H2 | 0,09% | 0,04% | -0,08% | 104% | 122% |
| **Overall** | 0,14% | 0,14% | 0,02% | 101% | 109% |

Table 2: HM-16.22 (TF on) for CTC HDR RA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **PSNR** | | |  |  |
|  | Y | U | V | EncT | DecT |
| Class H1 | 0,71% | -0,48% | -0,23% | 85% | 85% |
| Class H2 | 0,11% | -0,15% | -0,38% | 84% | 84% |
| **Overall** | 0,49% | -0,36% | -0,28% | 85% | 85% |

No additional objective results requested for H1 sequence class on the HDR RA CTC were presented.

Subjective video quality is best evaluated by playout on an HDR capable TV. To illustrate the subjective quality improvement by the proposed method the contribution contains a snapshot from the 4K HDR PQ Netflix sequence SparksTruckPack (H1 sequence class) and SunsetBeach (H2 sequence class).

2.5 JVET-V0095 AHG10: Using original samples for SAO and ALF optimization

In VTM-12.0, motion compensated temporal pre-filtering (MCTF) is used for the random access (RA) configuration. Samples filtered by the MCTF are used for sample adaptive offset (SAO) and adaptive loop filter (ALF) optimizations. This contribution proposes to use original samples for SAO and ALF optimization as an encoder only modification. Compared to VTM-12.0, the average Y/U/V BD-rates reportedly are -0.22%/-0.09%/-0.08% for RA with similar running time.

In VTM-12.0, the encoder minimizes the rate\*lambda + distortion cost metric for the sample adaptive offset (SAO) and adaptive loop filter (ALF) methods. When motion compensated temporal pre-filtering (MCTF) is enabled for the random access (RA) configuration, it was observed that the SAO and ALF processes minimize the cost for samples filtered by the MCTF. On the other hand, the final PSNR computation that is also used for the BD-rate computation is calculated using original unfiltered input samples.

In the proposal, instead of using samples filtered by the MCTF, the original samples are used to calculate the distortion for SAO and ALF for parameters estimation.

The proposed method was implemented on top of VTM-12.0 and evaluated under the common test conditions for RA and LDB.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Random access Main10** | | | | |
|  | **Over VTM-12.0** | | | | |
|  | Y | U | V | EncT | DecT |
| Class A1 | -0.15% | -0.02% | -0.17% | 101% | 101% |
| Class A2 | -0.02% | -0.04% | 0.15% | 100% | 99% |
| Class B | -0.10% | 0.18% | 0.12% | 100% | 100% |
| Class C | -0.59% | -0.51% | -0.43% | 100% | 99% |
| Class E |  |  |  |  |  |
| **Overall** | **-0.22%** | **-0.09%** | **-0.08%** | **100%** | **100%** |
| Class D | -0.90% | -0.65% | -0.52% | 98% | 98% |
| Class F | 0.00% | 0.00% | 0.00% | 101% | 101% |
|  |  |  |  |  |  |
|  | **Low delay B Main10** | | | | |
|  | **Over VTM-12.0** | | | | |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  |  |  |
| Class A2 |  |  |  |  |  |
| Class B | 0.00% | 0.00% | 0.00% | 100% | 101% |
| Class C | 0.00% | 0.00% | 0.00% | 100% | 103% |
| Class E | 0.00% | 0.00% | 0.00% | 101% | 100% |
| **Overall** | **0.00%** | **0.00%** | **0.00%** | **100%** | **101%** |
| Class D | 0.00% | 0.00% | 0.00% | 100% | 98% |
| Class F | 0.00% | 0.00% | 0.00% | 101% | 102% |

JVET-V0066 AHG8: Encoder improvements to palette coding for high bit depth

This contribution proposes two encoder modifications to palette coding for high bit depth:

• Proposal#1 uses on-the-fly bit calculation functions instead of pre-defined lookup tables to reduce memory usage and has no impact to coding performance and can reduce memory size.

• Proposal#2 fixes QP offset derivation used for an error limit to improve coding performance of palette coding.

Proposal#2, it is reported that average bd-rate changes for YUV at normal QP are (-8.70%, -10.86%, -11.66%) for AI and (-6.09%, -8.43%, -9.16%) for RA, and average bd-rate changes for RGB are (-8.51%, -8.65%, -8.32%) for AI and (-5.64%, -5.86%, -5.73%) for RA; average bd-rate changes for YUV at low QP are (-1.80%, -2.08%, -2.09%) for AI and (-1.34%, -1.27%, -1.31%) for RA, and average bd-rate changes for RGB are (-1.87%, -1.93%, -1.96%) for AI and (-1.19%, -1.33%, -1.27%) for RA, respectively.

At the encoder, to retrieve errorLimit from the lookup table, the QP needs to be adjusted to an 8-bit domain QP by a QP offset. However, in the VTM, the QP offset always takes a fixed value of “-12”. This contribution proposes to fix the QP offset based on BitDepth, i.e. considering BdOffset as below.

The QPs used for testing are listed as follow:

 Normal QP (QP=22, 27, 32, 37)

 Low QP (QP=-13, -8, -3, 2, 7, 12)

This contribution reports encoder improvements to palette coding by using on-the-fly bit calculation functions instead of the pre-defined lookup tables and fixing QP Offset derivation used for an error limit. It is recommended to adopt proposed software fixes to the next VTM.

JVET-V0127 An optimized VVC encoder implementation

This contribution reports the performance of an optimized VVC encoder, BVC, developed by Bytedance. A 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 the VTM, the corresponding global BD-rate gain and speedup factors under the Random Access configuration are summarized as follows:

• Fast preset (6 threads): 34.43%, 17x (HM), 131x (VTM)

• Slow preset (6 threads): Will be presented on an update of the contribution.

Simulation Results

For those simulations, 6 threads are used for the BVC encoder. It could be seen that with the fast preset, the BD-rate gain over HM is still approx. 34.43% on average, but a speedup of 17 times and 131 times over the HM and the VTM, respectively is observed.

The slow preset, which is targeting at close performance to VTM, the BD-rate gain over HM is no yet available, but a speedup of more than 7 times over VTM was reported.

**Coding performance of the BVC encoder @ fast preset**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Class** | **RA** | | | | | |
| PSNR BD-rate relative to HM | | | | Speedup | |
| Y (%) | U (%) | V (%) | YUV (%) | vs HM | vs VTM |
| A1 | -35.76 | -43.56 | -48.57 | -38.33 | 22.36 | 148.45 |
| A2 | -36.93 | -45.53 | -45.83 | -39.12 | 18.40 | 133.37 |
| B | -34.12 | -54.24 | -50.68 | -38.70 | 18.05 | 130.91 |
| C | -20.04 | -30.53 | -30.50 | -22.66 | 11.78 | 118.43 |
| **Overall** | **-31.26** | **-44.04** | **-43.91** | **-34.43** | **16.88** | **131.20** |

[JVET-V0011](https://jvet-experts.org/doc_end_user/current_document.php?id=10789) JVET AHG report: Neural network-based video coding (AHG11) [S. Liu, A. Segall, Y. Ye, E. Alshina, J. Chen, F. Galpin, J. Pfaff, S. S. Wang, M. Wien, P. Wu, J. Xu]

The AHG used the main JVET reflector, jvet@lists.rwth-aachen.de, for email exchange with AHG11 included in the subject lines. 16 emails were exchanged on the reflector.

2.1 Technical Evaluation

The AHG made meaningful progress on the mandate to evaluate and quantify potential NN based video coding technologies. This was done in both the context of the EE and non-EE contributions.

A summary of the performance of technology provided as input to the 22nd meeting is provided below:

**EE Technology**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **#** | **Title** | **JVET-U** | **Category** | **Performance** | **Dec.T.** |
| JVET-V0073 | EE1.2: Additional experimental results of NN-based super resolution (JVET-U0053) | JVET-U0053 | SupeRes | RA(4K):**2.5**%(Y) RA(4K):**5.8**%(Y) RA(4K):**6.5**%(Y) | ×0.3 ×1.5 ×1.5 |
| JVET-V0096 | EE1-2.3: Neural Network-based Super Resolution | JVET-U0099 | SupeRes | RA(4K):**8.1**%(Y) AI(4K): **10.8**% (Y) | ×0.3 ×0.4 |
| JVET-V0114 | EE1-1.3: Test on Neural Network-based In-Loop Filter with No Deblocking Filtering stage | JVET-U0115 | NN deblock | RA:**7.2**%(Y); 13% (Ch) RA:**7.5**%(Y); 14% (Ch) | ×137 ×146 |
| JVET-V0115 | EE1-1.4: Test on Neural Network-based In-Loop Filter with Large Activation Layer | JVET-U0104 | NN deblock | RA:**8.3**%(Y); 20% (Ch) RA:**8.5**%(Y); 22% (Ch) | ×337 ×344 |
| JVET-V0137 | EE1-1.1: neural network based in-loop filter using depthwise separable convolution and regular convolution (JVET-U0061) | JVET-U0061 | CNN in-loop | RA:**1.0**%(Y); 6% (Ch) | ×50 |
| JVET-V0149 | EE1: Tests on Decomposition, Compression and Synthesis (DCS)-based Technology (JVET-U0096) | JVET-U0096 | SupeRes | RA(4K):**9.6**%(Y) LD(4K):**17.5**%(Y) | no info |

**Non-EE Technology (Performance)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Title** | **Category** | **Performance** | **Dec.T.** |
| [JVET-V0074](https://jvet-experts.org/doc_end_user/current_document.php?id=10722) | AHG11: Separate density attention network for loop filtering | NN In-loop filter | RA: **5**% (Y) 14% (Ch.), in VTM10.0 | ×100 |
| [JVET-V0075](https://jvet-experts.org/doc_end_user/current_document.php?id=10723) | AHG11: Content-adaptive neural network post-processing filter | NN post-filter | RA(C&D classes): **2.5**%(Y), 5%(Ch.) in VTM.11.0 | ×450 |
| [JVET-V0076](https://jvet-experts.org/doc_end_user/current_document.php?id=10724) | AHG11: Deep-learning based inter prediction blending | NN BIO | RA: **0.6** % / **0.9** % (Y) 14% (Ch.), in VTM10.0 | ×11 |
| [JVET-V0090](https://jvet-experts.org/doc_end_user/current_document.php?id=10738) | AHG11: Neural network based temporal processing | NN In-loop filter | RA(C&D classes): **6**%(Y), 6%(Ch.) in VTM.11.0 | ×4000 |
| [JVET-V0091](https://jvet-experts.org/doc_end_user/current_document.php?id=10739) | AHG9/AHG11: SEI messages for carriage of neural network information for post-filtering | HLS |  |  |
| [JVET-V0092](https://jvet-experts.org/doc_end_user/current_document.php?id=10740) | AHG11: Replacing SAO in-loop filter with Neural Networks | NN SAO | K=5 RA: **2.8** % (Y) AI: **3.5** % (Y), in VTM11.0 K=9 RA: **3.5** % (Y) | ×95 (RA) ×350(AI) ×125 (RA) |
| [JVET-V0100](https://jvet-experts.org/doc_end_user/current_document.php?id=10748) | AHG11: Deep In-Loop Filter with Adaptive Model Selection | NN In-loop filter | RA (classes BCD) :**11** % (Y)  AI: **8.8** % (Y), in VTM11.0 | ×290 (RA)  ×143(AI) |
| [JVET-V0101](https://jvet-experts.org/doc_end_user/current_document.php?id=10749) | AHG11: Conditional In-Loop Filter with Parameter Selection | NN In-loop filter | ~JVET-V0100 | ×250 (RA) |
| [JVET-V0102](https://jvet-experts.org/doc_end_user/current_document.php?id=10750) | AHG11 & AHG12: Deep In-Loop Filter with Adaptive Model Selection for Enhanced Compression Beyond VVC Capability | NN In-loop filter | RA (classes BCD) :**9** % (Y) | ×70 (RA) |
| [JVET-V0105](https://jvet-experts.org/doc_end_user/current_document.php?id=10753) | AHG11: neural network-based intra prediction: updated signaling | NN Intra | RA : **1.7** % (Y)  AI: **3.5** % (Y), in VTM11.0 | ×5.5...8.4 (RA)  ×36..63(AI) |

**Non-EE Technology (Features and Characteristics)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Features** | **#params** | **kMAC /pixel** | **NN structure** |
| [JVET-V0074](https://jvet-experts.org/doc_end_user/current_document.php?id=10722) | frame level residual scaling, CTU on/off, input: 128x128 YUV444, before SAO | **1 M** | ? | 26\*RB+¯3\*RB­+2\*¯RB2\*­ |
| [JVET-V0075](https://jvet-experts.org/doc_end_user/current_document.php?id=10723) | 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) | ? | CONV&B+4\*RB+CONV\*B+LongSkipConnection |
| [JVET-V0076](https://jvet-experts.org/doc_end_user/current_document.php?id=10724) | N=5 and N=6 (different depth of NN), 16bits parameters; input: 16x16 | **6/8K** (13/17KB ) | 11.3 / 16.6 | N\*CONV+LongSkipConnection |
| [JVET-V0090](https://jvet-experts.org/doc_end_user/current_document.php?id=10738) | Different CNN filters for output and for future refPic, NN post != NN in-loop filters | **100M** (5 x 80.9MB ) | 2000 | PCD (Pyramid, Cascading and Deformable) +TSA (Temporal and Spatial Attention) |
| [JVET-V0091](https://jvet-experts.org/doc_end_user/current_document.php?id=10739) | Good to Understand if parameters are sufficient to describe all filters structure |  |  |  |
| [JVET-V0092](https://jvet-experts.org/doc_end_user/current_document.php?id=10740) | Offset are signaled, K masks --> out of CNN, Input: 128x128; 16 bits and 32 bits precision for NN parameters tested --> almost no difference | 1.3M / 2.5M | K\*322 | 10\*CONV |
| [JVET-V0100](https://jvet-experts.org/doc_end_user/current_document.php?id=10748) | Picture on/off control + CTU level model selection (not in AI) New: RB number extended 16-->32, number of channels reduced 128-->96; models "dictionary" depends on QP and resolution | **528**M (24\*22M) | 1400 | ¯32\*RB­ (69 CONV in total) |
| [JVET-V0101](https://jvet-experts.org/doc_end_user/current_document.php?id=10749) | Conditional (QP map is auxiliary input) | **90**M (4\*22M) | 1400 | ¯8\*(RB+FС)­ (69 CONV+ 10 FC total) |
| [JVET-V0102](https://jvet-experts.org/doc_end_user/current_document.php?id=10750) | JVET-V0100 but in top JVET JVET-U0100 | **528**M (24\*22M) | 1400 | ¯32\*RB­ (69 CONV in total) |
| [JVET-V0105](https://jvet-experts.org/doc_end_user/current_document.php?id=10753) | same as JVET-T0073 + new syntax; float and 16bit integer versions, 16bits integer is ~2 times faster than float-point | ~**10**M | 100 | different number of CONV and FC layers depending on Intra block size |

The AHG finalized, conducted and discussed the EE on NN based video coding. This was accomplished mainly with two teleconferences, which were held on January 22nd and April 6th.

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

|  |  |  |
| --- | --- | --- |
| JVET-V0023 | EE1: Summary of Exploration Experiments on Neural Network-based Video Coding | E. Alshina, S. Liu, W. Chen, Y. Li, R.-L. Liao, Z. Ma, H. Wang |

CTC Refinement and Support

The AHG refined and released the CTC test conditions on April 6th, 2021.

Additionally, to better support the calculation of the MS-SSIM metric according to the test conditions, the AHG did the following:

1. HDRTools was updated to enable support for the MS-SSIM calculuation that is present in VTM 11.0. This was made available in the 0.22-dev branch.

2. A patch for VTM 11.0 was created to enable the output of the MS-SSIM data with high precision. This patch was incorporated into the branched used for the EE tests.

The AHG11 coordinates would like to thank Alexis and Timofey for their generous contributions.

Anchors for the NN-based video coding activity were created and released on March 2, 2021. The anchors were announced on the reflector and 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.

The AHG coordinated with SC29/AHG5 to prepare a viewing procedure for EE contributions. Sequences were selected for five proposals included in the EE, with each proponent selecting two to three sequences/QP pair that illustrated the visual benefit of the proposal. These selections could include results from either the RA and LDB configurations.

The decoded sequences were encapsulated in an MP4 container and uploaded to the JVET FTP upload directory. With close coordination with SC29/AHG5, 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.

While not explicitly in the AHG mandates, the AHG did receive a comment regarding licensing terms and conditions (e.g., for training materials). The comment was posted on the reflector on April 6, 2021 and is highlighted in case it is useful for further discussion.

There were 18 input contributions related to the AHG mandates. Seven of the contributions are directly related to the EE activity, while the remaining 11 contributions are related to AHG11 but not directly part of the EE. The list of input contributions is provided below.

EE Related Input Contributions

|  |  |  |
| --- | --- | --- |
| [JVET-V0073](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10721) | EE1.2: Additional experimental results of NN-based super resolution (JVET-U0053) | [T. Chujoh](mailto:chujoh.takeshi@sharp.co.jp), T. Ikai (Sharp) |
| [JVET-V0096](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10744) | EE1-2.3: Neural Network-based Super Resolution | [A. M. Kotra](mailto:akotra@qti.qualcomm.com), K. Reuzé, [J. Chen](mailto:cjianle@qti.qualcomm.com), H. Wang, M. Karczewicz, J. Li (Qualcomm) |
| [JVET-V0114](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10763) | EE1-1.3: Test on Neural Network-based In-Loop Filter with No Deblocking Filtering stage | [H. Wang](mailto:hongtaow@qti.qualcomm.com), [J. Chen](mailto:cjianle@qti.qualcomm.com), [A. M. Kotra](mailto:akotra@qti.qualcomm.com), [K. Reuzé](mailto:kreuz@qti.qualcomm.com), [M. Karczewicz (Qualcomm)](mailto:martak@qti.qualcomm.com) |
| [JVET-V0115](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10764) | EE1-1.4: Test on Neural Network-based In-Loop Filter with Large Activation Layer | [H. Wang](mailto:hongtaow@qti.qualcomm.com), [J. Chen](mailto:cjianle@qti.qualcomm.com), [A. M. Kotra](mailto:akotra@qti.qualcomm.com), [K. Reuzé](mailto:kreuz@qti.qualcomm.com), [M. Karczewicz (Qualcomm)](mailto:martak@qti.qualcomm.com) |
| [JVET-V0137](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10798) | EE1-1.1: neural network based in-loop filter using depthwise separable convolution and regular convolution (JVET-U0061) | [Z. Li](mailto:zeqiangli@tencent.com), C. Auyeung, [X. Xu](mailto:xiaozhongxu@tencent.com), W. Wang, X. Li, S. Liu (Tencent) |
| [JVET-V0148](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10809) | Crosscheck of JVET-V0096 (EE1-2.3: Neural Network-based Super Resolution) | [T. Ikai](mailto:ikai.tomohiro@sharp.co.jp), [T. Chujoh (Sharp)](mailto:chujoh.takeshi@sharp.co.jp) |
| [JVET-V0149](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10810) | EE1: Tests on Decomposition, Compression and Synthesis (DCS)-based Technology (JVET-U0096) | [Ming Lu](mailto:luming@smail.nju.edu.cn), [Zhan Ma (NJU)](mailto:mazhan@nju.edu.cn), Zhenyu Dai, Dong Wang (OPPO) |

Non-EE contributions

|  |  |  |
| --- | --- | --- |
| Reporting | | |
| [JVET-V0011](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10789) | JVET AHG report: Neural network-based video coding (AHG11) | S. Liu, A. Segall, Y. Ye, E. Alshina, J. Chen, F. Galpin, J. Pfaff, S. S. Wang, M. Wien, P. Wu, J. Xu |
| Loop Filtering | | |
| [JVET-V0055](https://jvet-experts.org/doc_end_user/current_document.php?id=10703) | Improved end to end deep intra frame compression with cross channel context model and loop filter | [C. Ma](mailto:changyue.mcy@alibaba-inc.com), [Z. Wang](mailto:baixiu.wz@alibaba-inc.com), [R.-L. Liao](mailto:ruling.lrl@alibaba-inc.com), [Y. Ye (Alibaba)](mailto:yan.ye@alibaba-inc.com) |
| [JVET-V0074](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10722) | AHG11: Separate density attention network for loop filtering | [Z. Wang](mailto:baixiu.wz@alibaba-inc.com), C. Ma, R.-L. Liao, Y. Ye (Alibaba) |
| [JVET-V0075](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10723) | AHG11: Content-adaptive neural network post-processing filter | Y. Lam, M. Santamaria, J. Lainema, F. Cricri, R. Ghaznavi-Youvalari, A. Zare, H. Zhang, H. R. Tavakoli, M. M. Hannuksela (Nokia) |
| [JVET-V0092](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10740) | AHG11: Replacing SAO in-loop filter with Neural Networks | [P. Bordes](mailto:philippe.bordes@interdigital.com), [F. Galpin](mailto:franck.galpin@interdigital.com), [T. Dumas](mailto:thierry.dumas@interdigital.com), [P. Nikitin (InterDigital)](mailto:pavel.nikitin@interdigital.com) |
| [JVET-V0100](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10748) | AHG11: Deep In-Loop Filter with Adaptive Model Selection | [Y. Li](mailto:yue.li@bytedance.com), [L. Zhang](mailto:lizhang.idm@bytedance.com), [K. Zhang (Bytedance)](mailto:zhangkai.video@bytedance.com) |
| [JVET-V0101](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10749) | AHG11: Conditional In-Loop Filter with Parameter Selection | [Y. Li](mailto:yue.li@bytedance.com), [L. Zhang](mailto:lizhang.idm@bytedance.com), [K. Zhang (Bytedance)](mailto:zhangkai.video@bytedance.com) |
| [JVET-V0102](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10750) | AHG11 & AHG12: Deep In-Loop Filter with Adaptive Model Selection for Enhanced Compression Beyond VVC Capability | [Y. Li](mailto:yue.li@bytedance.com), [L. Zhang](mailto:lizhang.idm@bytedance.com), [K. Zhang (Bytedance)](mailto:zhangkai.video@bytedance.com) |
| Inter Prediction | | |
| [JVET-V0076](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10724) | AHG11: Deep-learning based inter prediction blending | [F. Galpin](mailto:franck.galpin@interdigital.com), [P. Bordes](mailto:philippe.bordes@interdigital.com), [T. Dumas](mailto:thierry.dumas@interdigital.com), [A. Robert](mailto:antoine.robert@interdigital.com), [P. Nikitin](mailto:pavel.nikitin@interdigital.com), [F. Le Leannec (InterDigital)](mailto:Fabrice.LeLeannec@interdigital.com) |
| [JVET-V0090](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10738) | AHG11: Neural network based temporal processing | [B. Choi](mailto:bdchoi@tencent.com), [Z. Li](mailto:zeqiangli@tencent.com), [W. Wang](mailto:rickweiwang@tencent.com), [W. Jiang](mailto:vwjiang@tencent.com), [X. Xu](mailto:xiaozhongxu@tencent.com), [S. Liu (Tencent)](mailto:shanl@tencent.com) |
| High Level Syntax | | |
| [JVET-V0091](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10739) | AHG9/AHG11: SEI messages for carriage of neural network information for post-filtering | [B. Choi](mailto:bdchoi@tencent.com), [Z. Li](mailto:zeqiangli@tencent.com), [W. Wang](mailto:rickweiwang@tencent.com), [W. Jiang](mailto:vwjiang@tencent.com), [X. Xu](mailto:xiaozhongxu@tencent.com), [S. Wenger](mailto:swenger@tencent.com), [S. Liu (Tencent)](mailto:shanl@tencent.com) |
| Intra Prediction | | |
| [JVET-V0105](file:///\\Users\asegall\Documents\Standards%20Proposals\2021_04_Online\JVET-V0011\current_document.php%3fid=10753) | AHG11: neural network-based intra prediction: updated signaling | [T. Dumas](mailto:thierry.dumas@interdigital.com), [F. Galpin](mailto:franck.galpin@interdigital.com), [P. Bordes](mailto:philippe.bordes@interdigital.com), [P. Nikitin (InterDigital)](mailto:pavel.nikitin@interdigital.com) |

The AHG recommends:

* Review all input contributions
* Further discuss and define of common test conditions and exploration experiments as appropriate
* Perform viewing of the neural network-based video coding tools during the meeting, in coordination with SC29/AHG5.
* Continue investigating neural network-based video coding tools, including coding performance and complexity;

[JVET-V0012](https://jvet-experts.org/doc_end_user/current_document.php?id=10790) JVET AHG report: Enhanced compression beyond VVC capability (AHG12) [M. Karczewicz, Y. Ye, 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-U2024). The first indication of the achievable improvement over VVC are the results of this experiment. For example, the combined improvement of the package proposed in JVET-U0100 and bilateral filter over VTM 11 for AI and RA configurations are:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | All Intra Main10 | | | | |
|  | Y | U | V | EncT | DecT |
| Class A1 | -5.72% | -9.69% | -10.65% | 183% | 227% |
| Class A2 | -5.43% | -7.88% | -5.91% | 179% | 205% |
| Class B | -4.96% | -7.14% | -7.55% | 214% | 175% |
| Class C | -5.97% | -7.43% | -7.67% | 212% | 168% |
| Class E | -6.13% | -7.97% | -6.80% | 219% | 193% |
| Overall | -5.58% | -7.89% | -7.69% | 203% | 189% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Random Access Main 10 | | | | |
|  | Y | U | V | EncT | DecT |
| Class A1 | -12.10% | -13.42% | -15.89% | 245% | 442% |
| Class A2 | -13.44% | -14.36% | -14.21% | 229% | 490% |
| Class B | -11.49% | -12.16% | -11.57% | 217% | 388% |
| Class C | -13.30% | -13.15% | -13.04% | 215% | 365% |
| Class E |  |  |  |  |  |
| Overall | -12.49% | -13.12% | -13.35% | 224% | 411% |

In addition, Intra Template Matching method also evaluated in this EE provides average {Y, U, V} BD-rate savings of {6.47%, 6.13%, 6.16%} for TMG class in AI configuration.

In addition to six EE2 contributions, 20 related contributions were received which can be subdivided as follows:

Partitioning

JVET-V0083, “EE2 related: asymmetric binary tree splitting on top of VVC”, F. Le Léannec, K. Naser, T. Dumas, A. Robert, F. Galpin, E. Francois (InterDigital)

JVET-V0097, “AHG12: Unsymmetric partitioning methods in video coding”, K. Zhang, L. Zhang, Z. Deng, N. Zhang, Y. Wang (Bytedance)

In Loop Filters

JVET-V0080, “AHG12: Extensions to CCALF”, S. Keating, A. Browne, K. Sharman (Sony)

JVET-V0104, “EE2-related: TU-level adaptive self-guided filter”, W. Yin, K. Zhang, L. Zhang, Y. Wang, H. Liu (Bytedance)

JVET-V0102, “AHG11 & AHG12: Deep In-Loop Filter with Adaptive Model Selection for Enhanced Compression Beyond VVC Capability”, Y. Li, L. Zhang, K. Zhang (Bytedance)

JVET-V0153, “AHG12: Cross-component Sample Adaptive Offset”, C.-W. Kuo, X. Xiu, Y.-W. Chen, H.-J. Jhu, W. Chen, X. Wang (Kwai)

Intra

JVET-V0098, “EE2-related: Template-based intra mode derivation using MPMs”, Y. Wang, L. Zhang, K. Zhang, Z. Deng, N. Zhang (Bytedance)

JVET-V0110, “AHG12: Two additional reference sample sets for CCLM”, C.-Y. Teng, Y.-C. Yang (FG Innovation)

JVET-V0087, “EE2-Related: Improvements of Decoder-Side Intra Mode Derivation” J. Zhao, S. Paluri, S. Kim (LGE)

Inter

JVET-V0086, “EE2-related: OBMC fixes and updates”, A. Robert, F. Galpin, F. Le Léannec, T. Poirier (InterDigital)

JVET-V0089, “EE2-related: Inter coding modes modifications”, A. Robert, F Le Léannec, F. Galpin, T. Poirier (InterDigital)

JVET-V0099, “AHG12: Adaptive Reordering of Merge Candidates with Template Matching”, N. Zhang, K. Zhang, L. Zhang, H. Liu, Z. Deng, Y. Wang (Bytedance)

JVET-V0103, “AHG12: Geometric prediction mode with motion vector differences”, Z. Deng, K. Zhang, L. Zhang, N. Zhang, Y. Wang (Bytedance)

JVET-V0117, “EE2-related: Combination of GPM and template matching”, R.-L. Liao, Y. Ye, X. Li, J. Chen (Alibaba)

JVET-V0118, “EE2-related: Extension of template matching to Affine, CIIP, GPM merge modes, and boundary sub-blocks”, C.-C. Chen, V. Seregin, Y.-J. Chang, Z. Zhang, H. Huang, Y. Zhang, M. Karczewicz (Qualcomm)

JVET-V0125, “AHG12: Evaluation of GPM with MMVD for coding efficiency improvement over VVC”, X. Xiu, H.-J. Jhu, C.-W. Kuo, W. Chen, Y.-W. Chen, X. Wang (Kwai)

JVET-V0126, “EE-related: Modified OBMC”, Y. Kidani, K. Kawamura, K. Unno (KDDI)

JVET-V0129, “AHG12: 3D Geometry for Global Motion Compensation”, H. Golestani, C. Rohlfing, M. Wien (RWTH Aachen)

Transforms

JVET-V0116, “Enhanced Intra MTS and LFNST for compression beyond VVC”, B. Ray, L. Kerofsky, M. Coban, M. Karczewicz, H. Egilmez, V. Seregin (Qualcomm)

JVET-V0119, “AHG12: LFNST extension with large kernel”, M. Koo, J. Zhao, J. Lim, S. Kim (LGE)

Recommendations

The AHG recommends:

• To review all the related contributions.

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

It is suggested to clarify at some point the purpose of these explorations, and how rigid JVET has to be in investigating the practicality of the technology. This depends on the question how “long term” a real standards development would be.

Several experts suggested to act at the current moment in an “inclusive” manner to be able judging the potential benefit, but also investigate the complexity vs. benefit tradoff.

It is commented that visual quality should also be considered.

# Project development (21)

## Deployment of standards (3)

Contributions in this area were discussed in session 17 at 1330–1400 UTC on Monday 26 April 2021 (chaired by JRO).

[JVET-V0020](https://jvet-experts.org/doc_end_user/current_document.php?id=10760) Deployment status of the HEVC standard [G. J. Sullivan]

This information contribution contains a survey of deployed products and services using the HEVC standard and the formal specifications in which it is supported, along with a brief introduction to the standard written for broad readership. Revision marking is included to show changes relative to JVET-U0020-v1 of January 2021.

New developments:

As of March 2021, a survey of 509 “broadcast and video streaming professionals representing more than 40 countries … in a variety of organizations from broadcast/OB services to network operators and OTT/VOD services” (conducted in November and December 2020) by Haivision reported that “HEVC is clearly becoming the most important codec for the future of broadcast video”, with:

a. 50% of broadcast and video streaming professionals currently using HEVC

b. 82% of broadcast and video streaming professionals planning to be using it in one year

In April 2021, an industry market study focused on entertainment devices with remote controls concluded that “a shift in technology from AVC (Advanced Video Coding) to HEVC (High Efficiency Video Coding) to improve video quality for entertainment and media industry, has increased the sales of various entertainment devices including remote control”.

Huawei’s P10 phone supports HEVC with 4K resolution, as of March 2017

Google’s Android 5.0 (December 2014) includes APIs for HEVC (including UHD support), and Google’s Chromecast Ultra media player includes HEVC support (November 2016). Android core app quality guidelines issued in February 2021 state that “If the app encodes video, it should do so using the HEVC video compression standard”. In March 2021, it was announced that although Google reportedly “urges developers to include HEVC support for their apps”, Android 12 will include compatible media transcoding to support HEVC for apps made for the older AVC standard in order to ease the transition to HEVC support.

Google’s YouTube service, as of March 2021, accepts HEVC as a supported format.

[JVET-V0021](https://jvet-experts.org/doc_end_user/current_document.php?id=10690) Deployment status of the VVC standard [G. J. Sullivan]

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

Revision marking is included to show changes relative to JVET-V00210-v3 of January 2021.

New develoments:

- Speedup of VVEnc

- KDDI Research announced a real-time VVC encoder with 4K @60 fps capability

A new version will include developments from section 4.9

[JVET-V0167](https://jvet-experts.org/doc_end_user/current_document.php?id=10830) Updated information on the TV 3.0 project Call for Proposals from the Brazilian Digital Terrestrial TV Forum [M. Raulet, T. Biatek, T. Guionnet (Ateme), B. Bross (Fraunhofer HHI), P. de Lagrange, R. Schaefer, M. Kerdranvat, E. François (InterDigital)] [late]

This contribution is a follow-up of JVET-U0128, which presented the Call for Proposals (CfP) issued by the Brazilian Digital Terrestrial TV Forum (SBTVD Forum) in July 2020, and reported that several companies (including Ateme, F. Fraunhofer HHI, and InterDigital) proposed VVC as video coding format.

It provides updated information and the schedule, and additional tests related to spatial scalability, including visual testing, with a final report due in October 2021.

The following points in the SBTVD schedule may interact with standard bodies or JVET schedule, or may be difficult to reach:

• The HDR verification tests for HDR PQ and HLG availability before October 29, 2021 to enable submitting them to SBTVD,

• The amendment for Level 6.3 (8K/4K scalable profile) published before end of 2021. A draft would need to be shared with SBTVD by June 25th,

• VVC scalability visual quality assessment before October 29, 2021, with preliminary test description on June 25th.

HDR verification should be available by October

DIS status of VVC v2 in October, but no published standard yet by end of year

The contributors would be willing to prepare sequences for a scalable verification test – discuss in joint meeting with AG5 Tuesday.

## Text development and errata reporting (2)

Contributions in this area were discussed in session 17 at 1405–1505 UTC on Monday 26 April 2021 (chaired by JRO and GJS).

[JVET-V0072](https://jvet-experts.org/doc_end_user/current_document.php?id=10720) AHG2: Proposal to remove some RPL constraints [A. Hallapuro, M. M. Hannuksela (Nokia), Hendry (LGE), Y.-K. Wang (Bytedance)]

The contribution is a follow-up of JVET-U0076, which was left for further study in the 21st JVET meeting. This contribution asserts that constraints in the VVC standard state that an RPL shall not have entries that represent (sub)pictures that precede previous IRAP (sub)picture in decoding or output order. The contribution alleges that these constraints prevent certain prediction possibilities when IRAP and non-IRAP subpictures are mixed in a picture and can lead to coding efficiency of the bitstream being lower than it could be. It is asserted that it has been an oversight to specify the constraints and therefore the contribution proposes to remove them.

VVC specification contains the following constraints related to RPL:

* When the current picture follows an IRAP picture having the same value of nuh\_layer\_id in both decoding order and output order, there shall be no picture referred to by an active entry in RefPicList[ 0 ] or RefPicList[ 1 ] that precedes that IRAP picture in output order or decoding order.
* When the current subpicture follows an IRAP subpicture having the same value of nuh\_layer\_id and the same value of subpicture index in both decoding and output order, there shall be no picture referred to by an active entry in RefPicList[ 0 ] or RefPicList[ 1 ] that precedes the picture containing that IRAP subpicture in output order or decoding order.
* When the current picture follows an IRAP picture having the same value of nuh\_layer\_id and all the leading pictures, if any, associated with that IRAP picture in both decoding order and output order, there shall be no picture referred to by an entry in RefPicList[ 0 ] or RefPicList[ 1 ] that precedes that IRAP picture in output order or decoding order.
* When the current subpicture follows an IRAP subpicture having the same value of nuh\_layer\_id and the same value of subpicture index and all the leading subpictures, if any, associated with that IRAP subpicture in both decoding and output order, there shall be no picture referred to by an entry in RefPicList[ 0 ] or RefPicList[ 1 ] that precedes the picture containing that IRAP subpicture in output order or decoding order.

It is suggested removing the latter two constraints. This would for example allow the following combination of two subpictures:



The previous contribution was single-company. Now, two additional companies are supporting the proposal as co-authors, and several other experts believe it is no problem. Other experts expressed that there might be some impact if an already implemented decoder believes that a bitstream without the constraint would be illegal and e.g. tries running some error resilience mechanism.

The proponents believe that the proposal makes encoder implementation (or bitstream merging) simpler, and would also slightly help compression by allowing N+4 to take reference to N.

Not clear how large that benefit in compression would be.

No unanimous agreement that such a change should be made. No action at this time. More analysis would be needed on the potential compression benefit and possible impact on decoders. This might best be illustrated by real bitstream examples.

[JVET-V0111](https://jvet-experts.org/doc_end_user/current_document.php?id=10759) AHG2: On Decoding Unit Information for VVC Version 1 [S. Deshpande (Sharp)]

Following is proposed as bug-fixes for asserted issues related to dui\_dpb\_output\_du\_delay:

* It is proposed to fix the inference rule for the dui\_dpb\_output\_du\_delay when not present.
* It is proposed to add a bitstream conformance constraint related to dui\_dpb\_output\_du\_delay\_present\_flag.
* It is proposed to remove the note in C3.3.
* It is proposed to clarify the condition for the existing constraint on the value of dui\_dpb\_output\_du\_delay.
* Ye-kui, Karsten, Yago, Miska and Virginie to review, confirm there is a problem, and if yes, the solution is appropriate. It was later confirmed (session 23 at 2145 and 2245) that the problem exists. In terms of the best solution, there was additional offline discussion. Finally, the experts converged in their opinion that the following solution is appropriate:It is proposed to fix the inference rule for the dui\_dpb\_output\_du\_delay when not present – agreed to use “option 1” of the proposal
* It is proposed to add a bitstream conformance constraint related to dui\_dpb\_output\_du\_delay\_present\_flag – agreed to use the first constraint of the proposal
* It is proposed to remove the note in C3.3
* It is proposed to clarify the condition for the existing constraint on the value of dui\_dpb\_output\_du\_delay.

The proponent will provide a specification text (see v3 of the zip package) with the finally agreed version for the solution.

Decision (BF/text): Adopt JVET-V0111 (text in version 3)

## Test conditions (1)

Contributions in this area were discussed in session 18a at 1525–1535 UTC on Monday 26 April 2021 (chaired by JRO).

[JVET-V0107](https://jvet-experts.org/doc_end_user/current_document.php?id=10755) Editorial suggestion for JVET CTC on HDR/WCG [D. Rusanovskyy (Qualcomm), T. Hashimoto (Sharp)] [late]

This contribution informs on inaccuracies identified in the text of JVET CTC on HDR/WCG video coding related to wPSNR calculation. Text improvement is proposed to address identified inaccuracies and enabling HDR/WCG metric interpretation for higher bit depths.

Agreed to make this change, use the text version 1 which does not require software change

## Verification testing (6)

Contributions in this area were discussed in session X at XXXX–XXXX UTC on XXday 2X April 2021 (chaired by XXX).

[JVET-V0041](https://jvet-experts.org/doc_end_user/current_document.php?id=10687) AHG4: Status Report on SDR HD and 360 Video Verification Test Preparation [M. Wien, V. Baroncini, Y. Ye]

No need to present – was in included in AHG report.

[JVET-V0042](https://jvet-experts.org/doc_end_user/current_document.php?id=10688) AHG4: Agenda and report of the AHG meeting on the SDR HD and 360 video verification test preparation on 2021-02-18 [M. Wien, V. Baroncini, Y. Ye]

No need to present – was in included in AHG report.

[JVET-V0043](https://jvet-experts.org/doc_end_user/current_document.php?id=10689) AHG4: Status Report on HDR Video Verification Test Preparation [A. Segall, M. Wien, V. Baroncini, K. Andersson]

No need to present – was in included in AHG report.

[JVET-V0044](https://jvet-experts.org/doc_end_user/current_document.php?id=10691) AHG4: Agenda and report of the AHG meeting on the HDR verification test preparation on 2021-03-25 [A. Segall, M. Wien, V. Baroncini]

No need to present – was in included in AHG report.

[JVET-V0174](https://jvet-experts.org/doc_end_user/current_document.php?id=10837) Report on VVC compression performance verification testing in the SDR HD and 360 Video categories [V. Baroncini (VABtech), M. Wien (RWTH Aachen University)]

This document reports the second set of verification test results comparing VVC to its predecessor, HEVC. The compression performance capabilities of the two specifications are compared based on the HEVC reference software HM-16.22, and the VVC reference software VTM-11.0. In the SDR HD random access category, the open-source VVC implementation VVenC-0.3 is included in the evaluation. The results of a visual assessment of VVC compared to HEVC performing formal subjective assessment evaluation (with the participation of naïve test subjects) is reported. The assessment this time was made at HD resolution content and included three Categories: SDR encoded in random-access configuration, SDR encoded in Low Delay configuration and the 360° video category assessed through viewports. The measured MOS figures indicate a significant improvement of all the tested configurations over HEVC. A huge number video material was tested at this time, namely 60 test points for SDR RA, 60 test points for SDR SD and 80 test points for the 360° video category. This huge testing activities were conducted at Vabtech and GBtech labs, regardless the severe restrictions imposed by COVID pandemic, with the participation of more than 60 non-expert viewers.

Preliminary version was presented Tuesday 27 April 1430 UTC (during joint meeting with AG5).

It was reported that MCTF had not been enabled in the SDR HD RA tests for HM (but for VTM it was on). Tests need to be re-done which could be performed quickly, and test run during next weekend.

It was concluded not delaying the report due to this oversight, but rather include data with both cases MCTF on and off.

Graphs per sequence were presented

Estimate Average BD/MOS rate gain over HM:

- 360 approx. 50% for PERP, 56% GCMP

- HD LD roughly 40% both gaming content and conversational content

- HD RA roughly 50% for VVENC, around 47% for VTM

[JVET-V0176](https://jvet-experts.org/doc_end_user/current_document.php?id=10839) AHG4: On ITU-R BT.2245 HDR Test Content [A. Segall, M. Wien, K. Andersson] [late]

As part of the VVC HDR verification test activity, three test sequences were acquired from the ITE for evaluation. These sequences are included in the ITU-R BT.2245 recommendation. This document describes how the sequences were converted to raw YUV files and provides initial performance analysis.­

Presented Tuesday 27 April 2330

Swim race seems to benefit significantly by MCTF enabling.

Initial QP settings for verification test have been defined.

## Test material (1)

Contributions in this area were discussed in session 19 at 2250–2300 UTC on Monday 26 April 2021 (chaired by JRO and GJS).

[JVET-V0045](https://jvet-experts.org/doc_end_user/current_document.php?id=10692) Aerial 4K HDR (HLG) Test Data [Pankaj Topiwala (FastVDO)]

This contribution offers some drone-captured aerial HDR video with 4K resolution, and HLG transfer function, for potential use in HDR coding and other standards-related investigations.

Data were compressed at 100 Mbit/s with rate control. Relatively slow motion.

Data were converted to 60 fps. 30 fps originals should be provided.

Copyright given in the context of standards development.

For studying the properties of the sequences, RD plots, and analysis of the colour gamut would be helpful.

## Quality assessment (1)

Contributions in this area were discussed in session 19 at 2300–2310 UTC on Monday 27 April 2021 (chaired by JRO and GJS).

[JVET-V0168](https://jvet-experts.org/doc_end_user/current_document.php?id=10831) AHG4: An approach to objective video quality assessment for both full and no reference cases [P. Topiwala, W. Dai, J. Pian (FastVDO)] [late]

This contribution offers an approach to objective video quality assessment, based on feature extraction and learning based assessment, which can be applied to both the full reference and no reference cases, and it is asserted that the performance in each domain is quite competitive, at least for a key measure (SRCC). Our approach strives to both minimize the computationally burden of the feature extraction stage and, using advanced yet fast machine-learning based assessment stage, to achieve high performance. Performance is measured in terms of the common Pearson and Spearman Rank Correlation Coefficients (PCC and SRCC, respectively). Asserted advantages in the full reference case are against the well-known Video Multi-Algorithm Fusion (VMAF) algorithm, where it is reported that our method can exceed 90% accuracy according to SRCC, even without prior training. In the no reference case, comparisons are provided against the current state of the art algorithm, VidEval, using over 3K video clips, where it is asserted that our method, claimed to be of substantially lower complexity, yet exceeds VidEval performance and achieves mid-80% accuracy in SRCC, again without prior training (but tested against a fully trained VidEval algorithm). All our results are obtained by training and testing (on a sequestered portion) on the data at hand. It is hoped to achieved a usable, fast-computable algorithm, in both FR and NR cases, which serve both compression research and development as well as video services applications at scale.

For the purpose of standards development, typically full reference testing would be done.

No specific action requested. Software could be provided.

## Conformance test development (0)

Contributions in this area were discussed in session X at XXXX–XXXX UTC on XXday 2X April 2021 (chaired by XXX).

## Software development (1)

See also JVET-V0093 on film grain analysis and synthesis in section 4.11.

Contributions in this area were discussed in session 18 at 1535–1650 UTC on Monday 26 April 2021 (chaired by JRO).

[JVET-V0079](https://jvet-experts.org/doc_end_user/current_document.php?id=10727) Reference software cleaning proposal [F. Galpin, T. Poirier, P. De Lagrange, F. Le Léannec, E. Francois (InterDigital)]

This contribution proposes a set of cleanings on top of the VTM reference software. The main motivation for this cleaning is to ease the development and testing of new tools inside the reference software. As the proposed cleaning involves significant files refactoring, it is not compatible with a series of merge requests (might also require a new management approach for the repository). A first version of cleaned software is also proposed as a basis for discussions and further improvements.

HLS not cleaned up yet.

Current version does not have speedup compared to VTM.

Currently VTM11 – one major change is the update of GDR in VTM12 – would that be a difficulty? Proponents believe it is not.

It is pointed out by software coordinators that various other issues would be desirable to be cleaned. It might be difficult however to get exactly same results of VTM encoder then.

Switching to another package may also be causing new bugs.

The motivation for the new software (as per proponents) is about better support for explorations. Even though, having a compatibility with VTM / conformance with VVC is highly desirable.

Would a speedup be more important for explorations?

Question: If the software is more flexible, should it not be simple to get same results as VTM encoder?

Could the current EE2 software easily be ported? This would give an evidence of advantages of the software package.

Experts are asked to study the software for possible advantages of being used in EEs. By the next meeting, it should be decided if a switch of the software has an advantage for progress in the EEs. It is also pointed out that in the past a good point of switching software is by the time when a new standardization is started.

## Implementation studies (5)

Contributions in this area were discussed in session X at 1655–1800 UTC on Monday 26 April 2021 (chaired by JRO).

[JVET-V0070](https://jvet-experts.org/doc_end_user/current_document.php?id=10718) VVC Software Decoder for Mobile Platforms [W.-L. Feng, F.-L. Luo, Y.-S. He, Z.-H. Liu, S.-M. Meng, X. Wen (Kwai Inc.)]

This contribution presents **K266Dec,** a VVC software decoder designed for ARM platforms. The decoder is designed from scratch following the VVC specification and can support all coding tools specified in VVC main profile.

For Android platform, K266Dec can decode 2K 8-bit CTC bitstreams at 33 fps with single thread on Huawei P40, which is 4.11x of VTM-11.0 decoder. In addition, we also test CTC bitstreams and UGC videos 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 4.78x compared to VTM-11.0 reference decoder.

Support for most features in VVC Main profile (ALF not yet, and disabled for CTC sequences).

[JVET-V0088](https://jvet-experts.org/doc_end_user/current_document.php?id=10736) Multi-thread VTM decoder: information update [F. Hiron, R. Jullian, F. Urban, P. de Lagrange (InterDigital)]

This contribution presents updates to the implementation of the multi-thread VTM based VVC decoder proposed in JVET-T0061. The design takes advantage of task and data parallelization to accelerate VVC decoding on a general-purpose multi-core processor. The decoder has been upgraded to VTM-12.0, and performance has been improved by various scheduling and low-level changes.

* Clean-ups, less duplicate code by splitting functions
* New task scheduling and thread pool: now both CTU wavefront and CTU pipeline with several stages. Jobs are queued and can start as soon as their dependencies are met.
* New inverse transform task (finer-grained pipeline)
* Final reconstruction moved to intra stage because of LMCS and CIIP dependencies
* DBF rework: split in two tasks, boundary strength computation, and actual filter, which have different dependencies, thus triggered at different stages.
* More SIMD. Not all SIMD improvements from VVdeC have been ported, and performance improvement from SIMD does not seems as significant as in VVdeC. This is still under investigation.

Approximately 10% speedup.

Real time for class B (and classes with lower resolution as well). At least on average, not sure if for all sequances and QP points.

[JVET-V0127](https://jvet-experts.org/doc_end_user/current_document.php?id=10776) An optimized VVC encoder implementation [J. Cui, Y. Fan, Y. He, X. Jiang, H. Liu, H. Shi, H. Yang, H. Yang, H. Yin, J. Zhang, L. Zhang (Bytedance)]

This contribution reports the performance of an optimized VVC encoder, BVC, developed by Bytedance. 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 HM and VTM, the corresponding global BD-rate gain and speedup factors under the Random Access configuration are summarized as follows:

* Fast preset (6 threads): 34.13%, 17x (HM), 131x (VTM)
* Slow preset (6 threads): 41.63%, 1x (HM), 8x (VTM)

Internal 8 bit coding. Slice, tiles subpictures, wraparound MC not supported yet.

Above numbers are measured via YUV PSNR. Slow preset approx. 1% less BR reduction than VTM.

Rate control also supported, but results above are constant QP.

What are differences between fast and slow? Level of optimization of coding tool decisions.

Not planned making SW available.

[JVET-V0128](https://jvet-experts.org/doc_end_user/current_document.php?id=10777) Performance of a VVC software decoder – BVC [Y. He, L. Li, Y. Li, H. Yin, J. Zhang, L. Zhang, Y. Zhang (Bytedance)]

A VVC software decoder, BVC decoder, fully developed by Bytedance is introduced. 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 10x 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 8.8x faster than the VTM-11.0 reference decoder.

No results on Andrid available in the contribution

Power consumption, e.g. comparing against VVC decoder? Not analysed

Delay due to frame level parallelism? Only due to GOP structure, according to proponents

Bitstreams were generated using VTM11.0

Speed dependency on bit rate? Not analysed.

[JVET-V0132](https://jvet-experts.org/doc_end_user/current_document.php?id=10793) VVC software decoder implementation for mobile devices [L. Yu, Z. Cui, L. Wang, J. Cheng, Z. Huang, L. Xu (Alibaba)]

This input document is a follow-up of JVET-U0088 which introduces Ali266, a software VVC decoder optimized specifically for the mobile platform by Alibaba. In JVET-U0088 Ali266 was claimed to occupy only 30 MB of memory when decoding 720p video in real time and to achieve real-time decoding for coded bit rates up to 7 Mbps for the 4K content; and for 1080p content, Ali266 achieves real-time decoding for most cases using only two threads. Ali266 performs and real-time decoding of e-commerce content with one or two threads on a wide variety of phones. However, the claimed performance was achieved by disabling ALF or CCALF and setting internal bit depth to 8.

In this document, additional results are provided with both ALF and CCALF on, the updated decoder requires 33 MB memory occupation and binary size of less than 1 MB, both are slightly increased compared to the decoder reported in JVET-U0088. Meanwhile the decoding speed decreases for 33.42% on average for CTC sequences in order to support the ALF and CCALF. The new Ali266 reaches 15x speed-up compared to VTM-11.0 on Android P40 with 8 threads on the CTC bitstreams of classA and class B, and still performs real-time decoding of e-commerce content on most phones with 3 threads.

Presentation deck to be provided.

Single thread about 4x faster than VTM

ALF/CCALF speedup by NEON

Power consumption? Not analysed yet.

## Complexity analysis (0)

Contributions in this area were discussed in session X at XXXX–XXXX UTC on XXday 2X April 2021 (chaired by XXX).

## AHG10: Encoding algorithm optimization (5)

Contributions in this area were discussed in session 19 at 2100–2250 UTC on Monday 26 April 2021 (chaired by JRO and GJS), except otherwise noted.

[JVET-V0093](https://jvet-experts.org/doc_end_user/current_document.php?id=10741) AHG9: Film grain estimation and film grain synthesis for VVC – SEI message characteristics, film grain estimation and film grain synthesis modules [Miloš Radosavljević, Edouard François (InterDigital)]

This contribution was initially discussed as an SEI message contribution, and the notes were moved here as the contribution was considered to contain a substantial enhancement of software functionality.

This contribution was discussed in sesssion 15b at 22100 on Friday 23 April 2021 (chaired by GJS).

This contribution proposes a complete VTM software implementation of the film grain feature, including an estimation module at encoder side and a synthesis module at decoder side. The implementation is based on the SMPTE RDD 5 Film Grain Technology specification, with some adaptations.

The software implementation illustrates the use of the film grain for VVC. Specifically, the software illustrates:

Film grain characteristics SEI message for VVC specified in Rec. ITU-T H.SEI | ISO/IEC 23002-7;

Film grain synthesis module based on frequency filtering model specified in SMPTE RDD 5;

A film grain analysis (parameter estimation) module used to estimate film grain parameters according to SMPTE RDD 5 model.

It is reported that the visual quality benefits from the synthesized film grain and the algorithmic simplicity for both estimation and synthesis part makes this approach good candidate for future use in VVC and VTM reference software.

The (integerized) DCT in RDD is replaced with the VVC transform.

Sample images were shown, including illustration of how the film grain synthesis can mask coding artefacts.

The software was offered but not uploaded with the contribution.

It is proposed to include the software in the fork of VTM for further study

The proponent suggested to consider a mandatory support specification.

About potentially including it in the VTM main branch - this might be possible, although some further work is needed.

It was commented that this is a substantial contribution.

The contribution was welcomed, and the availability of the software was encouraged.

There is HM support for RDD 5 synthesis, contributed by Dolby. This contribution also provides the encoder-side analysis and filtering using MCTF functionality.

In our spec, RDD 5 is just an example of a way to operate the post-processing.

An analogy with rate control was mentioned.

Further study for experimentation with the software, evaluation of the quality of the software and its potential incorporation into the VTM was encouraged.

Providing the software in a public fork of the VTM was agreed for further study and experimentation.

Further potential action e.g. for normative standardization action, is yet to be considered.

[JVET-V0151](https://jvet-experts.org/doc_end_user/current_document.php?id=10812) Crosscheck of JVET-V0093 (AHG9: Film grain estimation and film grain synthesis for VVC – SEI message characteristics, film grain estimation and film grain synthesis modules) [P. Yin, S.McCarthy (Dolby)] [late]

[JVET-V0056](https://jvet-experts.org/doc_end_user/current_document.php?id=10704) [AHG10] GOP-based temporal filter improvements [P. Wennersten, C. Hollmann, J. Ström (Ericsson)]

This contribution proposes a set of changes to the GOP-based temporal filter in VTM that aim at improving the BD-rate performance. These changes were first proposed in JVET-U0056.

The average Y/U/V BD-rates for VTM-12 CTC are reported to be −1.4%/−1.0%/−1.1% for RA. All BDR numbers were computed using unfiltered source sequences as reference.

In addition, the same changes were implemented in HM where the BD-rates are -1.0%/-1.0%/-1.0%.

Visual inspection of the resulting sequences was reportedly carried out with no problems reported.

It is proposed to adopt the proposed changes into the VTM and HM software.

Version 2 of this document adds results for HDR.

Changes: Block error weighting (stronger filter when MC error larger), Block frequency weighting (stronger filter when higher frequencies present, Reference frame weighting.

Also a bug fix regarding MC at picture boundaries.

Subjective viewing: Equal or better quality was found by proponents

Question: Would there be more headroom for subjective improvement? A: Likely by still stronger filtering.

Question: What happens in case of illumination changes? Might be appropriate applying a weight in the MC stage.

Decision (SW/CTC): Adopt (both for VTM and HM), and also use in the EEs

The other items above for further study.

It was further suggested to investigate potential need for changing the filter weights in the EEs. A first step on this could be to investigate if the gain that MCTF gives for VTM is similarly preserved for the EE configuration.

[JVET-V0057](https://jvet-experts.org/doc_end_user/current_document.php?id=10705) [AHG10] Block importance mapping [P. Wennersten, C. Hollmann, J. Ström (Ericsson)]

This contribution proposes a new algorithm for VTM to signal CTU QP delta values in pictures that will be used for reference. The QP value to use for each CTU is based on the estimated importance of a given CTU for future pictures and the QP selected is in the range of −2 to +2 relative to the picture QP.

The method was reportedly implemented and tested together with the temporal filter changes first presented in JVET-U0056 under VTM-12.0 RA test configurations. The method is not proposed to be used for AI or LD configurations.

The average CTC RA Y/U/V BD-rates for the proposed changes together with JVET-V0056 compared to VTM-12 with JVET-V0056 are reported to be −1.4%/−3.9%/−3.8% for RA. Comparing the proposed changes together with JVET-V0056 to regular VTM-12, the average BD-rates are −2.8%/−5.0%/−4.9%.

It is proposed to adopt the algorithm into the VTM software.

Version 2 of this document provides a statement that the results for the HM implementation will be added later on.

Version 3 of this document adds HDR and HM results.

Question: Why difference in performance per sequence, particularly class A has less gain? A: One reason could be saving bits in sequences with more changes, where the QP of I pictures is increased relative to the current CTC setting.

HM shows losses in some of the class A sequences.

Generally a very interesting approach for encoder optimization, it is mentioned that similar algorithms are often applied in practical coders. Before adopting it, further study is recommended to better understand the behaviour for specific sequences and investigate the reason for losses in HM.

Not investigated for low delay, as it requires lookahead which introduces additional delay.

[JVET-V0078](https://jvet-experts.org/doc_end_user/current_document.php?id=10726) AHG 10: QP control for very smooth blocks [K. Andersson, J. Enhorn, J. Ström (Ericsson)]

Large transforms are excellent in reducing spatial redundancy. However, when source samples are very smooth the subjective quality can drop very quickly in such areas with too coarse quantization. It is therefore suggested to reduce the QP according to a linear model for very smooth regions to prolong the breakdown of the subjective quality. Since smooth regions are relatively easy to code it is claimed that the overhead in bitrate can be manageable.

The impact on BDR has been tested on HDR:

VTM-12.0 HDR RA CTC (luma/Cb/Cr): 0.14%/0.14%/0.02%

HM-16.22 (TF on) HDR RA CTC (luma/Cb/Cr): 0.50%/-0.42%/-0.14%

It is asserted that the proposal can improve subjective quality of smooth regions for HDR and it is suggested to update VTM and HM with this functionality.

Cross-checker confirms that subjective benefit can be observed particularly for HM at QP37.

There is a small loss in BD rate. Tradeoff subjective against objective quality, or better only optimize for subjective?

Could this be used in the HDR verification test? Perhaps better not, as it would need to be explained in the report why such method was chosen. Even if HM benefits at lower QPs than VTM, it might be necessary to perform more individual optimization for each codec.

Question: Is there more effect on bright areas? A: Perhaps, but in particular for PQ the reshaping should take care of this. Could be different for HLG.

For further study by AHG 10. This goes into perceptual quantization, for which various other mechanisms exist, e.g. quant matrices, chroma QP adjustment, etc.

Decision (SW HM & VTM): Include as an option in conf file, but off in CTC.

[JVET-V0172](https://jvet-experts.org/doc_end_user/current_document.php?id=10835) Crosscheck of JVET-V0078 (AHG10: QP control for very smooth blocks) [A. Segall (Sharp)] [late]

[JVET-V0095](https://jvet-experts.org/doc_end_user/current_document.php?id=10743) AHG10: Using original samples for SAO and ALF optimization [N. Hu, V. Seregin, M. Karczewicz (Qualcomm)]

In VTM-12.0, motion compensated temporal pre-filtering (MCTF) is used for random access (RA) configuration. Samples filtered by MCTF are used for sample adaptive offset (SAO) and adaptive loop filter (ALF) optimizations. This contribution proposes to use original samples for SAO and ALF optimization as encoder only modification. Compared to VTM-12.0, the average Y/U/V BD-rates reportedly are -0.22%/-0.09%/-0.08% for RA with similar running time.

Would it be useful to make it configurable? The proponents believe not, as it gives some gain on all sequences. However, in a practical encoder implementation, this might require an additional buffer (or more as MCTF operates over a sequence of pictures). Therefore, it may be desirable being able to test both options.

It was remarked that this already used in EE2.

Was subjective quality investigated? No

It is further suggested to study if it has some impact in HM for SAO.

Also study in AHGs if it should be implemented in HDR and high bit depth CTC.

Further study on these aspects before including in CTC

Decision (SW): Include this as a switchable option. Consider in EE2 to switch this on in the VTM anchor.

[JVET-V0175](https://jvet-experts.org/doc_end_user/current_document.php?id=10838) Crosscheck of JVET-V0095 [P. Wennersten, C. Hollmann, J. Ström (Ericsson)] [late]

## Profile/tier/level specification (0)

Contributions in this area were discussed in session X at XXXX–XXXX UTC on XXday 2X April 2021 (chaired by XXX).

# Low-level tool technology proposals (69)

## AHG8: High bit rate and high bit depth coding for VVC (27)

### General (1)

Contributions in this area were discussed in session 5 at 1430–1500 UTC on Wednesday 21 April 2021 (chaired by JRO and GJS).

[JVET-V0150](https://jvet-experts.org/doc_end_user/current_document.php?id=10811) AHG8: A study on bin rate of VTM-12.0 for high bit depth coding [T. Tsukuba, M. Ikeda, T. Suzuki (Sony)] [late]

This contribution reports the coding performance of VTM-12.0 relative to HM-16.22 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 saving is -17.17% for AI and -31.12% for RA, respectively.
* BD-bin rate saving is -2.44% for AI and -23.78% 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 reduction for AI/LDB/RA is {-6.79%, -6.61%, -6.81%} for PQ, {-3.68%, -5.14%, -5.72%} for HLG, and {18.22%, 8.65%, 7.36%} for SVT, respectively.
* BD-bin rate for AI/LDB/RA is increased by {10.26%, 7.11%, 7.17%} for PQ, {6.05%, 4.65%, 4.19%} for HLG, and {9.88%, 4.51%, 4.14%} for SVT, respectively.
* The unweighted bin to bit ratios of VTM for AI/LDB/RA are a bit higher than HM by {6.6%, 5.28%, 3.59%} for PQ, {3,18%, 3.89%, 2.87%} for HLG, and {-5.63%, -5.63%, -5.99%} for SVT, respectively. The weighted bin to bit ratios of VTM for AI/LDB/RA are higher than HM by {18.25%, 17.78%, 17.14%} for PQ, {11.40%, 12.75%, 12.38%} for HLG and {-10.87%, -5.25%, -6.47%} for SVT, respectively.

In addition, the coding performance of VTM-12.0 CE Anchor for normal QP and low QP is reported as below.

In the normal QPs

* BD-bitrate saving is -17.17% for AI and -31.10% for RA, respectively.
* BD-bin rate saving is -2.43% for AI and -23.78% 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 reduction for AI/LDB/RA is {-7.24%, -6.68%, -6.93%} for PQ, {-4.09%, -5.12%, -5.73%} for HLG and {-0.92%, -1.91%, -2.29%} for SVT, respectively.
* BD-bin rate for AI/LDB/RA is {9.93%, 7.07%, 7.09%} for PQ, {5.58%, 4.76%, 4.27%} for HLG and {-3.36%, -2.50%, -2.32%} for SVT, respectively.
* The unweighted bin to bit ratios of VTM for AI/LDB/RA are a bit higher than HM by {6.70%, 5.40%, 3.69%} for PQ, {3.16%, 3.98%, 2.96%} for HLG, and {-5.13%, -5.15%, -5.52%} for SVT, respectively. The weighted bin to bit ratios of VTM for AI/LDB/RA are higher than HM by {18.47%, 18.12%, 17.58%} for PQ, {11.49%, 13.05%, 12.74%} for HLG and {-2.01%, -1.92%, -2.74%} 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-16.22.

It is commented that the VVC gain compared to HEVC is lower also in the case of 10 bit.

It is likely that the gain in the low QP range is not evenly distributed, this should be further investigated.

Also the question should be considered what a “normal” range of QP values would be e.g. in 12 bit consumer applications.

Is the higher bin-to-bit ratio critical? Further study is needed on this aspect, also relates to the possible need of a high throughput profile.

From the results (in particular for SVT16) it might be the case that VVC tends to go closer to the possible max limit of context coded bins for transform coefficients.

It is known that typically VVC requires a higher throughput capability of the entropy coding engine also for Main10. For 12 bit, this does not seem too much different. It should however be studied if the throuput problem becomes more serious for extremely high bit rates as in professional applications.

### CE contributions: Entropy Coding for High Bit Depth and High Bit Rate Coding (12)

Contributions in this area were discussed in session 3 at 2100–2310 UTC on Tuesday 20 April 2021 (chaired by JRO and GJS).

[JVET-V0022](https://jvet-experts.org/doc_end_user/current_document.php?id=10701) CE: Summary Report on Entropy Coding for High Bit Depth and High Bit Rate Coding [A. Browne, T. Hashimoto, H. Jhu, D. Rusanovskyy, K. Kawamura, T. Zhou]

This contribution provides a summary report for the Core Experiment on entropy coding for high bit depth and high bit rate coding. A total of 20 tests have been completed within the CE between the JVET U and V meetings to study and evaluate technologies related to Rice parameter derivation. The scope of this experiment covered tests on changes to the regular residual coding (RRC) (8 tests), transform skip residual coding (TSRC) (3 tests) and transform clipping (2 tests). In addition 7 tests have been completed in a second stage on combinations of RRC/TSRC tests with transform clipping tests. Results have been integrated on and tested against the Anchor CE software, which comprised integration of low complexity methods proposed in the CE of the U meeting, namely the RRC method of CE1.1 and the TSRC method of CE2.1.

In this report, coding performance and complexity are reported and analyzed. Test results against CE anchors are provided to show the coding efficiency and complexity trade-off of each tool. Where appropriate results are also provided for lossless conditions and screen content. Cross-check results for the performed tests are submitted as separate documents and their summaries are integrated in this contribution.

CE Anchor

The source code anchor for the CE was VTM12.0 modified with the lowest complexity rice coding modifications proposed in the CE of the U meeting. These changes consisted of the RRC modifications of CE1.1 and the TSRC modifications of CE2.1.

The anchor was generated using the high bit depth (HBD), high bit rate (HBR), high frame rate (HFR) CTC (JVET-U2018), which includes simulations within a standard QP range (QPs 22,27,32,37 for 12 bit YUV 4:2:0) and a low QP range (QPs 12,7,2,-3,-8,-13 for 12 bit YUV 4:2:2/4:4:4, -8,-13,-18,-23,-28,-33 for 16 bit RGB).

In addition to testing against VTM12.0, the CE anchor was also tested against HM16.23 configured for lossy and lossless coding using RExt test conditions (JVET-U1100) adjusted for the content and low QP range of JVET-U2018.

Table 1.1. Simulation results for CE Anchor and its RRC and TSRC components vs. VTM12.0, 12 bits data, HBD/HFR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
|  | wY | wU | wV | Y | U | V | Aver.GBR | G | B | R |
| **AI** | **CE Anchor** | -0.49% | -0.49% | -0.52% | -0.44% | -0.35% | -0.36% | -2.68% | -2.73% | -2.67% | -2.64% |
| **RRC** | -0.50% | -0.53% | -0.56% | -0.33% | -0.33% | -0.34% | -2.53% | -2.62% | -2.48% | -2.48% |
| **TSRC** | -0.10% | -0.06% | -0.07% | -0.21% | -0.15% | -0.15% | -1.55% | -1.45% | -1.62% | -1.57% |
| **LDB** | **CE Anchor** | -0.07% | -0.06% | -0.05% | 0.02% | -0.02% | -0.06% | -0.50% | -0.45% | -0.53% | -0.53% |
| **RRC** | -0.07% | -0.08% | -0.08% | -0.03% | -0.02% | -0.05% | -0.37% | -0.31% | -0.40% | -0.40% |
| **TSRC** | 0.00% | 0.00% | -0.01% | 0.05% | 0.00% | -0.04% | -0.40% | -0.38% | -0.42% | -0.40% |
| **RA** | **CE Anchor** | -0.13% | -0.13% | -0.14% | -0.01% | -0.04% | -0.06% | -0.54% | -0.51% | -0.56% | -0.54% |
| **RRC** | -0.14% | -0.15% | -0.15% | -0.04% | -0.04% | -0.05% | -0.42% | -0.40% | -0.43% | -0.43% |
| **TSRC** | 0.00% | 0.00% | -0.01% | 0.03% | 0.00% | -0.02% | -0.39% | -0.37% | -0.41% | -0.38% |

Table 1.2. Simulation results for CE Anchor and its RRC and TSRC components vs. VTM12.0, 16 bits data, HBD/HFR CTC, LowQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **SVT16 RGB** | | | |
|  | Aver.GBR | G | B | R |
| **AI** | **CE Anchor** | -24.70% | -25.62% | -24.25% | -24.24% |
| **AnchorRRC** | -24.35% | -25.14% | -23.95% | -23.95% |
| **AnchorTSRC** | -21.64% | -22.27% | -21.36% | -21.30% |
| **LDB** | **CE Anchor** | -16.38% | -16.99% | -16.06% | -16.09% |
| **RRC** | -15.83% | -16.66% | -15.29% | -15.55% |
| **TSRC** | -13.60% | -13.97% | -13.41% | -13.41% |
| **RA** | **CE Anchor** | -15.56% | -15.70% | -15.50% | -15.48% |
| **RRC** | -15.15% | -15.26% | -15.04% | -15.14% |
| **TSRC** | -12.80% | -12.67% | -12.89% | -12.84% |

Table 1.3. Run time estimates for CE Anchor and its RRC and TSRC components vs. VTM12.0, HBD/HFR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | **HDR HLG** |  | **SVT RGB** | |
|  | Enc | Dec | Enc | Dec | Enc | Dec |
| **AI** | **CE Anchor** | 101% | 95% | 100% | 96% | 98% | 89% |
| **RRC** | 102% | 99% | 101% | 100% | 99% | 89% |
| **TSRC** | 100% | 99% | 99% | 99% | 101% | 98% |
| **LDB** | **CE Anchor** | 101% | 100% | 99% | 98% | 99% | 94% |
| **RRC** | 102% | 103% | 102% | 100% | 99% | 95% |
| **TSRC** | 100% | 100% | 100% | 100% | 99% | 99% |
| **RA** | **CE Anchor** | 99% | 101% | 100% | 100% | 95% | 89% |
| **RRC** | 102% | 103% | 101% | 99% | 99% | 95% |
| **TSRC** | 100% | 99% | 100% | 100% | 98% | 98% |

Table 1.4. Simulation results for CE Anchor and its RRC and TSRC components vs. VTM12.0, HBD/HFR CTC, NormQP test configuration.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | | | **HDR HLG** | | |
|  | DE100 | PSNR-L100 | wY | wU | wV | Y | U | V |
| **AI** | **CE Anchor** | 0.02% | 0.00% | 0.00% | 0.01% | 0.04% | 0.01% | -0.02% | -0.06% |
| **RRC** | 0.02% | 0.00% | 0.00% | 0.01% | 0.04% | 0.00% | -0.02% | -0.06% |
| **TSRC** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| **RA** | **CE Anchor** | -0.07% | 0.01% | 0.01% | -0.04% | -0.16% | 0.04% | -0.16% | 0.02% |
| **RRC** | -0.09% | -0.01% | -0.01% | -0.06% | -0.18% | 0.03% | -0.17% | 0.00% |
| **TSRC** | 0.01% | 0.01% | 0.01% | 0.01% | 0.01% | 0.01% | 0.01% | 0.01% |

Table 1.5. Run time estimates for CE Anchor and its RRC and TSRC components vs. VTM12.0, HBD/HFR CTC, LowQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NormQP | **Test** | **HDR PQ** | | **HDR HLG** |  |
|  | Enc | Dec | Enc | Dec |
| **AI** | **CE Anchor** | 100% | 99% | 99% | 100% |
| **RRC** | 100% | 98% | 101% | 101% |
| **TSRC** | 100% | 100% | 100% | 100% |
| **RA** | **CE Anchor** | 102% | 102% | 101% | 101% |
| **RRC** | 99% | 99% | 100% | 100% |
| **TSRC** | 100% | 100% | 100% | 101% |

As the results are for camera-captured content, the benefit for TSRC is only margjnal.

Results compared to HM in the following

Table 1.7. Simulation results for CE Anchor vs. HM16.23, 12 bits data, HBD/HFR CTC for VTM , RExt CTC for HM, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
| wY | wU | wV | Y | U | V | Aver.GBR | G | B | R |
| **AI** | -7.24% | -9.64% | -10.23% | -4.09% | -6.39% | -6.53% | -2.17% | -2.54% | -1.95% | -2.03% |
| **LDB** | -6.68% | -7.73% | -8.82% | -5.12% | -8.20% | -7.06% | -1.79% | -3.31% | -1.05% | -1.02% |
| **RA** | -7.61% | -11.64% | -11.85% | -5.73% | -8.68% | -7.37% | -2.11% | -3.72% | -1.32% | -1.29% |

Table 1.8. Simulation results for CE Anchor vs. HM16.23, 16 bits data, HBD/HFR CTC for VTM, RExt CTC for HM, LowQP test configuration.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test** | **SVT16 RGB** | | | |
| Aver.GBR | G | B | R |
| **AI** | 0.77% | 0.70% | 0.82% | 0.79% |
| **LDB** | 0.58% | -0.52% | 1.13% | 1.13% |
| **RA** | 0.32% | -0.86% | 0.91% | 0.91% |

Table 1.9. Run time estimates for CE Anchor vs. HM16.23, HBD/HFR CTC for VTM, RExt CTC for HM, LowQP test configuration.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Test** | **HDR PQ** | | **HDR HLG** |  | **SVT RGB** | |
| Enc | Dec | Enc | Dec | Enc | Dec |
| **AI** | 2815% | 172% | 3080% | 169% | 5047% | 128% |
| **LDB** | 349% | 164% | 454% | 162% | 609% | 137% |
| **RA** | 414% | 164% | 535% | 159% | 758% | 137% |

Table 1.10. Simulation results for CE Anchor vs. HM16.23, lossless configuration, HBD/HFR CTC for VTM, RExt CTC for HM, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **PQ** | **All Intra** | | | **Low delay B** | | | **Random Access** | | |
| **ratio** | | bit-rate savings | **ratio** | | bit-rate savings | **ratio** | | bit-rate savings |
| HM16.23 | VTM\_U\_CE | HM16.23 | VTM\_U\_CE | HM16.23 | VTM\_U\_CE |
| PQ444 | 2.5 | 2.5 | 0.89% | 3.1 | 3.0 | 4.80% | 3.1 | 3.0 | 4.47% |
| PQ422 | 2.3 | 2.3 | -0.18% | 2.9 | 2.8 | 3.40% | 2.9 | 2.8 | 3.13% |
| **Overall** | **2.4** | **2.4** | **0.36%** | **3.0** | **2.9** | **4.10%** | **3.0** | **2.9** | **3.80%** |
| Enc Time[%] | 7206% | | | 976% | | | 983% | | |
| Dec Time[%] | 152% | | | 147% | | | 149% | | |
|  |  |  |  |  |  |  |  |  |  |
| **HLG** | **All Intra** | | | **Low delay B** | | | **Random Access** | | |
| **ratio** | | bit-rate savings | **ratio** | | bit-rate savings | **ratio** | | bit-rate savings |
| HM16.23 | VTM\_U\_CE | HM16.23 | VTM\_U\_CE | HM16.23 | VTM\_U\_CE |
| HLG444 | 1.8 | 1.7 | 3.27% | 1.9 | 1.9 | 4.25% | 2.0 | 1.9 | 4.08% |
| HLG422 | 1.7 | 1.6 | 2.37% | 1.9 | 1.9 | 3.66% | 1.9 | 1.9 | 3.49% |
| **Overall** | **1.7** | **1.7** | **2.82%** | **1.9** | **1.9** | **3.96%** | **1.9** | **1.9** | **3.79%** |
| Enc Time[%] | 7782% | | | 817% | | | 767% | | |
| Dec Time[%] | 132% | | | 128% | | | 126% | | |
|  |  |  |  |  |  |  |  |  |  |
| **SVT** | **All Intra** | | | **Low delay B** | | | **Random Access** | | |
| **ratio** | | bit-rate savings | **ratio** | | bit-rate savings | **ratio** | | bit-rate savings |
| HM16.23 | VTM\_U\_CE | HM16.23 | VTM\_U\_CE | HM16.23 | VTM\_U\_CE |
| SVT16 | 1.2 | 1.2 | 1.46% | 1.2 | 1.2 | 1.18% | 1.2 | 1.2 | 1.11% |
| SVT12 | 1.3 | 1.2 | 3.12% | 1.3 | 1.3 | 2.71% | 1.3 | 1.3 | 2.55% |
| **Overall** | **1.2** | **1.2** | **2.29%** | **1.3** | **1.2** | **1.95%** | **1.3** | **1.2** | **1.83%** |

Questions:

- Would the anchor change of Rice parameter improve VVC performance in lossless coding also for 10 bit data? Possibly, but was not investigated.

- What is the gain of anchor versus HM in normal QP range? JVET-V0150 might give an answer.

**Tests on** Regular **Residual Coding (RRC)**

This test category includes 8 tests studying proposed changes in the Regular Residual Coding (RRC) method of VVC for purpose of High Bit Depth Coding.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test** | **Tester** | **Description Document** | **Cross checker** | **Cross-chek document** |
| CE-1.1 | Dmytro Rusanovskyy  (Qualcomm) | [JVET-V0052](https://jvet-experts.org/doc_end_user/current_document.php?id=10699) | Mohammed Golam Sarwer (Alibaba) | JVET-V0156 |
| CE-1.2 | Dmytro Rusanovskyy  (Qualcomm) | [JVET-V0052](https://jvet-experts.org/doc_end_user/current_document.php?id=10699) | Tomonori Hashimoto  (Sharp) | JVET-V0138 |
| CE-1.3 | Widthrawn | | | |
| CE-1.4 | Dmytro Rusanovskyy  (Qualcomm) | [JVET-V0052](https://jvet-experts.org/doc_end_user/current_document.php?id=10699) | Adrian Browne  (Sony) | JVET-V0134 |
| CE-1.5 | Adrian Browne  (Sony) | [JVET-V0050](https://jvet-experts.org/doc_end_user/current_document.php?id=10697) | Dmytro Rusanovskyy  (Qualcomm) | [JVET-V0139](https://jvet-experts.org/doc_end_user/current_document.php?id=10800) |
| CE-1.6 | Adrian Browne  (Sony) | [JVET-V0050](https://jvet-experts.org/doc_end_user/current_document.php?id=10697) | Dmytro Rusanovskyy  (Qualcomm) | [JVET-V0139](https://jvet-experts.org/doc_end_user/current_document.php?id=10800) |
| CE-1.7 | Tomonori Hashimoto  (Sharp) | [JVET-V0046](https://jvet-experts.org/doc_end_user/current_document.php?id=10693) | Dmytro Rusanovskyy  (Qualcomm) | [JVET-V0140](https://jvet-experts.org/doc_end_user/current_document.php?id=10801) |
| CE-1.8 | Withdrawn | | | |

Table 2.2. Simulation results for RRC tests, 12 bits data, CE 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** | -0.04% | 0.00% | 0.00% | -0.04% | 0.07% | 0.07% | -0.24% | -0.35% | -0.19% | -0.19% |
| **CE1.2** | -0.09% | -0.06% | -0.07% | -0.10% | -0.03% | -0.04% | -0.41% | -0.45% | -0.39% | -0.39% |
| **CE1.3** |  |  |  |  |  |  |  |  |  |  |
| **CE1.4-A** | -0.33% | -0.24% | -0.24% | -0.50% | -0.30% | -0.32% | -1.43% | -1.64% | -1.32% | -1.34% |
| **CE1.4-B** | -0.33% | -0.23% | -0.25% | -0.50% | -0.30% | -0.32% | -1.43% | -1.64% | -1.32% | -1.33% |
| **CE1.5** | -0.22% | -0.12% | -0.13% | -0.42% | -0.23% | -0.23% | -1.36% | -1.60% | -1.24% | -1.24% |
| **CE1.6** | -0.22% | -0.11% | -0.12% | -0.41% | -0.22% | -0.22% | -1.34% | -1.58% | -1.21% | -1.22% |
| **CE1.7** | -0.09% | -0.07% | -0.07% | -0.08% | -0.03% | -0.03% | -0.37% | -0.42% | -0.34% | -0.34% |
| **CE1.8** |  |  |  |  |  |  |  |  |  |  |
| **LDB** | **CE1.1** | -0.01% | -0.03% | -0.02% | -0.03% | -0.01% | 0.00% | -0.05% | -0.05% | -0.05% | -0.05% |
| **CE1.2** | 0.03% | 0.03% | 0.02% | 0.02% | 0.05% | 0.06% | -0.19% | -0.20% | -0.19% | -0.19% |
| **CE1.3** |  |  |  |  |  |  |  |  |  |  |
| **CE1.4-A** | -0.06% | -0.07% | -0.08% | -0.02% | -0.01% | 0.01% | -0.66% | -0.71% | -0.63% | -0.63% |
| **CE1.4-B** | -0.06% | -0.08% | -0.09% | -0.06% | -0.01% | 0.00% | -0.66% | -0.71% | -0.63% | -0.63% |
| **CE1.5** | -0.02% | -0.03% | -0.06% | -0.03% | -0.01% | -0.01% | -0.62% | -0.69% | -0.58% | -0.58% |
| **CE1.6** | -0.03% | -0.04% | -0.06% | -0.04% | -0.02% | -0.01% | -0.60% | -0.68% | -0.57% | -0.56% |
| **CE1.7** | 0.01% | 0.00% | 0.02% | 0.02% | 0.04% | 0.06% | -0.15% | -0.16% | -0.14% | -0.14% |
| **CE1.8** |  |  |  |  |  |  |  |  |  |  |
| **RA** | **CE1.1** | -0.01% | -0.03% | -0.03% | -0.01% | 0.00% | 0.00% | -0.06% | -0.06% | -0.06% | -0.05% |
| **CE1.2** | -0.01% | -0.01% | -0.02% | 0.01% | 0.03% | 0.03% | -0.17% | -0.19% | -0.15% | -0.16% |
| **CE1.3** |  |  |  |  |  |  |  |  |  |  |
| **CE1.4-A** | -0.08% | -0.07% | -0.09% | -0.04% | -0.01% | -0.01% | -0.57% | -0.66% | -0.53% | -0.54% |
| **CE1.4-B** | -0.08% | -0.08% | -0.09% | -0.04% | -0.02% | -0.02% | -0.57% | -0.65% | -0.53% | -0.54% |
| **CE1.5** | -0.04% | -0.03% | -0.04% | -0.04% | -0.03% | -0.03% | -0.57% | -0.67% | -0.51% | -0.52% |
| **CE1.6** | -0.03% | -0.03% | -0.04% | -0.04% | -0.02% | -0.02% | -0.55% | -0.66% | -0.50% | -0.50% |
| **CE1.7** | 0.00% | 0.00% | -0.01% | 0.01% | 0.03% | 0.05% | -0.12% | -0.15% | -0.10% | -0.10% |
| **CE1.8** |  |  |  |  |  |  |  |  |  |  |

Table 2.3. Simulation results for RRC tests, 16 bits data, CE CTC, LowQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **SVT16 RGB** | | | |
|  | Aver.GBR | G | B | R |
| **AI** | **CE1.1** | -0.71% | -0.67% | -0.72% | -0.73% |
| **CE1.2** | -0.79% | -0.79% | -0.79% | -0.79% |
| **CE1.3** |  |  |  |  |
| **CE1.4-A** | -2.30% | -2.72% | -2.09% | -2.08% |
| **CE1.4-B** | -2.27% | -2.70% | -2.05% | -2.04% |
| **CE1.5** | -2.27% | -2.80% | -2.02% | -2.00% |
| **CE1.6** | -2.26% | -2.78% | -2.00% | -1.98% |
| **CE1.7** | -0.66% | -0.59% | -0.69% | -0.70% |
| **CE1.8** |  |  |  |  |
| **LDB** | **CE1.1** | -0.58% | -0.53% | -0.61% | -0.62% |
| **CE1.2** | -0.66% | -0.61% | -0.69% | -0.70% |
| **CE1.3** |  |  |  |  |
| **CE1.4-A** | -1.55% | -1.51% | -1.58% | -1.57% |
| **CE1.4-B** | -1.53% | -1.49% | -1.55% | -1.54% |
| **CE1.5** | -1.33% | -1.30% | -1.35% | -1.34% |
| **CE1.6** | -1.32% | -1.31% | -1.34% | -1.32% |
| **CE1.7** | -0.73% | -0.67% | -0.75% | -0.76% |
| **CE1.8** |  |  |  |  |
| **RA** | **CE1.1** | -0.51% | -0.50% | -0.50% | -0.52% |
| **CE1.2** | -0.60% | -0.61% | -0.58% | -0.60% |
| **CE1.3** |  |  |  |  |
| **CE1.4-A** | -1.45% | -1.43% | -1.46% | -1.45% |
| **CE1.4-B** | -1.45% | -1.43% | -1.47% | -1.46% |
| **CE1.5** | -1.23% | -1.24% | -1.24% | -1.23% |
| **CE1.6** | -1.23% | -1.24% | -1.22% | -1.21% |
| **CE1.7** | -0.66% | -0.67% | -0.64% | -0.66% |
| **CE1.8** |  |  |  |  |

Table 2.4. Reported run-time estimates RRC tests, CE 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% | 100% | 100% | 100% | 103% | 103% | 104% | 102% |
| **CE1.2** | 102% | 102% | 102% | 103% | 102% | 103% | 101% | 99% |
| **CE1.3** |  |  |  |  |  |  |  |  |
| **CE1.4-A** | 102% | 101% | 103% | 101% | 106% | 108% | 109% | 121% |
| **CE1.4-B** | 103% | 102% | 104% | 103% | 107% | 108% | 109% | 124% |
| **CE1.5** | 101% | 100% | 103% | 103% | 104% | 109% | 109% | 128% |
| **CE1.6** | 102% | 100% | 103% | 102% | 105% | 109% | 111% | 128% |
| **CE1.7** | 101% | 100% | 103% | 100% | 105% | 101% | 109% | 97% |
| **CE1.8** |  |  |  |  |  |  |  |  |
| **LDB** | **CE1.1** | 100% | 99% | 101% | 98% | 103% | 105% | 104% | 103% |
| **CE1.2** | 100% | 100% | 103% | 101% | 101% | 101% | 100% | 99% |
| **CE1.3** |  |  |  |  |  |  |  |  |
| **CE1.4-A** | 103% | 105% | 102% | 101% | 105% | 111% | 117% | 115% |
| **CE1.4-B** | 104% | 108% | 103% | 102% | 106% | 114% | 118% | 119% |
| **CE1.5** | 101% | 101% | 100% | 101% | 101% | 107% | 117% | 119% |
| **CE1.6** | 101% | 101% | 101% | 101% | 101% | 105% | 117% | 117% |
| **CE1.7** | 101% | 100% | 101% | 100% | 103% | 101% | 106% | 99% |
| **CE1.8** |  |  |  |  |  |  |  |  |
| **RA** | **CE1.1** | 101% | 98% | 101% | 100% | 103% | 103% | 106% | 102% |
| **CE1.2** | 104% | 103% | 102% | 101% | 102% | 103% | 102% | 99% |
| **CE1.3** |  |  |  |  |  |  |  |  |
| **CE1.4-A** | 104% | 101% | 103% | 101% | 105% | 108% | 121% | 119% |
| **CE1.4-B** | 103% | 101% | 103% | 102% | 106% | 107% | 120% | 117% |
| **CE1.5** | 101% | 100% | 101% | 101% | 100% | 106% | 114% | 117% |
| **CE1.6** | 100% | 99% | 101% | 101% | 101% | 106% | 114% | 116% |
| **CE1.7** | 101% | 100% | 101% | 100% | 103% | 101% | 106% | 99% |
| **CE1.8** |  |  |  |  |  |  |  |  |

Table 2.5. Simulation results for RRC tests, HBD/HBR/HFR CTC, NormQP test configuration.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | | | **HDR HLG** | | |
|  | DE100 | PSNR-L100 | wY | wU | wV | Y | U | V |
| **AI** | **CE1.1** | -0.07% | -0.01% | -0.02% | 0.07% | -0.07% | -0.01% | -0.04% | 0.03% |
| **CE1.2** | 0.01% | -0.03% | -0.03% | 0.09% | -0.08% | -0.01% | -0.10% | 0.05% |
| **CE1.3** |  |  |  |  |  |  |  |  |
| **CE1.4-A** | 0.02% | -0.02% | -0.03% | 0.12% | -0.06% | -0.01% | -0.02% | 0.04% |
| **CE1.4-B** | 0.00% | -0.02% | -0.02% | 0.06% | -0.07% | -0.01% | 0.00% | -0.04% |
| **CE1.5** | -0.01% | 0.00% | 0.00% | -0.05% | -0.06% | 0.00% | 0.00% | 0.10% |
| **CE1.6** | -0.01% | 0.00% | 0.00% | -0.05% | -0.06% | 0.00% | 0.00% | 0.10% |
| **CE1.7** | -0.03% | -0.02% | -0.02% | 0.11% | -0.02% | -0.01% | 0.03% | -0.03% |
| **CE1.8** |  |  |  |  |  |  |  |  |
| **RA** | **CE1.1** | -0.03% | 0.02% | 0.00% | -0.29% | -0.05% | 0.00% | -0.11% | -0.21% |
| **CE1.2** | 0.09% | 0.03% | 0.02% | -0.14% | -0.22% | -0.01% | 0.11% | -0.30% |
| **CE1.3** |  |  |  |  |  |  |  |  |
| **CE1.4-A** | 0.10% | -0.02% | -0.03% | -0.17% | -0.15% | 0.02% | -0.07% | 0.01% |
| **CE1.4-B** | -0.01% | -0.03% | -0.05% | -0.29% | -0.25% | 0.00% | -0.02% | -0.09% |
| **CE1.5** | 0.25% | -0.01% | -0.03% | 0.12% | 0.20% | 0.02% | 0.10% | 0.14% |
| **CE1.6** | 0.25% | -0.01% | -0.02% | 0.12% | 0.21% | 0.02% | 0.10% | 0.14% |
| **CE1.7** | -0.16% | -0.06% | -0.04% | -0.24% | -0.04% | 0.00% | -0.03% | -0.06% |
| **CE1.8** |  |  |  |  |  |  |  |  |

Table 2.6. Run-time estimates for RRC tests, CE CTC, NormQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | **HDR HLG** | |
|  | Enc | Dec | Enc | Dec |
| **AI** | **CE1.1** | 100% | 98% | 99% | 97% |
| **CE1.2** | 100% | 99% | 100% | 101% |
| **CE1.3** |  |  |  |  |
| **CE1.4-A** | 101% | 96% | 101% | 99% |
| **CE1.4-B** | 101% | 100% | 100% | 97% |
| **CE1.5** | 98% | 99% | 98% | 99% |
| **CE1.6** | 98% | 100% | 98% | 99% |
| **CE1.7** | 101% | 100% | 102% | 100% |
| **CE1.8** |  |  |  |  |
| **RA** | **CE1.1** | 98% | 98% | 98% | 98% |
| **CE1.2** | 100% | 101% | 100% | 100% |
| **CE1.3** |  |  |  |  |
| **CE1.4-A** | 99% | 99% | 98% | 98% |
| **CE1.4-B** | 98% | 99% | 96% | 98% |
| **CE1.5** | 101% | 101% | 102% | 102% |
| **CE1.6** | 101% | 101% | 102% | 102% |
| **CE1.7** | 100% | 100% | 100% | 100% |
| **CE1.8** |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table 2.7. Simulation results for RRC tests, CE CTC, lossless test configuration. | **Test** | bit-rate saving | |  |  |
|  |  | HDR PQ | HLG PQ | SVT12 | SVT16 |
| **AI** | **CE1.1** | -0.98% | -1.50% | -1.38% | 0.10% |
|  | **CE1.2** | -0.57% | -1.16% | -0.42% | -0.82% |
|  | **CE1.3** |  |  |  |  |
|  | **CE1.4-A** | -2.31% | -3.49% | -2.88% | -1.38% |
|  | **CE1.4-B** | -2.03% | -3.06% | -2.84% | -1.38% |
|  | **CE1.5** | -1.81% | -2.96% | -2.91% | -1.46% |
|  | **CE1.6** | -1.82% | -2.97% | -2.92% | -1.47% |
|  | **CE1.7** | -1.20% | -1.70% | -1.53% | -0.20% |
|  | **CE1.8** |  |  |  |  |
| **LDB** | **CE1.1** | -1.55% | -2.51% | -0.77% | 0.17% |
|  | **CE1.2** | -1.03% | -1.78% | -0.59% | -0.52% |
|  | **CE1.3** |  |  |  |  |
|  | **CE1.4-A** | -3.06% | -4.30% | -1.51% | -0.39% |
|  | **CE1.4-B** | -2.87% | -4.05% | -1.51% | -0.38% |
|  | **CE1.5** | -2.29% | -3.36% | -1.56% | -0.40% |
|  | **CE1.6** | -2.30% | -3.38% | -1.61% | -0.43% |
|  | **CE1.7** | -1.96% | -3.02% | -1.45% | -0.09% |
|  | **CE1.8** |  |  |  |  |
| **RA** | **CE1.1** | -1.63% | -2.58% | -0.71% | 0.18% |
|  | **CE1.2** | -0.99% | -1.74% | 0.60% | -0.44% |
|  | **CE1.3** |  |  |  |  |
|  | **CE1.4-A** | -2.98% | -4.22% | -1.42% | -0.37% |
|  | **CE1.4-B** | -2.79% | -3.98% | -1.42% | -0.36% |
|  | **CE1.5** | -2.22% | -3.29% | -1.49% | -0.39% |
|  | **CE1.6** | -2.23% | -3.31% | -1.53% | -0.42% |
|  | **CE1.7** | -2.00% | -3.05% | -1.39% | -0.08% |
|  | **CE1.8** |  |  |  |  |

“Local methods”: 1.2, 1.2, 1.7

“History based”: 1.4, 1.5, 1.6

The anchor still had losses versus HM in 16 bit coding. This would be compensated with local methods, an be converted into small gain with history based.

Would be interesting to analyse the gain directly compared to HM. Seems to become lower for higher bit depth.

Tests on Transform Skip Residual Coding (TSRC)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test** | **Proponent(s)** | **Description Document** | **Cross checker** | **Cross-chek document** |
| CE-2.2 | Sony | [JVET-V0050](https://jvet-experts.org/doc_end_user/current_document.php?id=10697) | Kwai | [JVET-V0143](https://jvet-experts.org/doc_end_user/current_document.php?id=10804) |
| CE-2.3 | Sony | [JVET-V0050](https://jvet-experts.org/doc_end_user/current_document.php?id=10697) | Kwai | [JVET-V0143](https://jvet-experts.org/doc_end_user/current_document.php?id=10804) |

Table 3.2. Simulation results for TSRC tests, 12 bits data, CE CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
|  | wY | wU | wV | Y | U | V | Aver.RGB | G | B | R |
| **AI** | **CE-2.2** | -0.02% | -0.02% | -0.03% | -0.07% | -0.09% | -0.09% | -0.14% | -0.13% | -0.14% | -0.13% |
| **CE-2.3** | -0.02% | -0.03% | -0.03% | -0.06% | -0.09% | -0.09% | -0.12% | -0.12% | -0.12% | -0.12% |
| **LDB** | **CE-2.2** | 0.01% | -0.02% | -0.03% | 0.01% | 0.00% | -0.01% | -0.09% | -0.11% | -0.07% | -0.08% |
| **CE-2.3** | 0.01% | -0.03% | -0.03% | 0.01% | 0.00% | -0.01% | -0.07% | -0.09% | -0.05% | -0.06% |
| **RA** | **CE-2.2** | 0.02% | 0.00% | -0.02% | 0.01% | 0.00% | -0.01% | -0.09% | -0.10% | -0.08% | -0.09% |
| **CE-2.3** | 0.02% | 0.00% | -0.02% | 0.02% | 0.00% | 0.00% | -0.07% | -0.08% | -0.05% | -0.07% |

Table 3.3. Simulation results for TSRC tests, 16 bits data, CE CTC, LowQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **SVT16 RGB** | | | |
|  |  | Aver.RGB | G | B | R |
| **AI** | **CE-2.2** | -0.20% | -0.26% | -0.17% | -0.16% |
| **CE-2.3** | -0.18% | -0.24% | -0.15% | -0.14% |
| **LDB** | **CE-2.2** | -0.14% | -0.14% | -0.14% | -0.12% |
| **CE-2.3** | -0.08% | -0.10% | -0.08% | -0.07% |
| **RA** | **CE-2.2** | -0.15% | -0.13% | -0.16% | -0.15% |
| **CE-2.3** | -0.10% | -0.10% | -0.10% | -0.10% |

Table 3.4. Reported run-time estimates TSRC tests, CE CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | **HDR HLG** | | **SVT12 RGB** | | **SVT16 RGB** | |
|  | Enc | Dec | Enc | Dec | Enc | Dec | Enc | Dec |
| AI | **CE-2.2** | 98% | 99% | 99% | 100% | 99% | 101% | 100% | 100% |
| **CE-2.3** | 98% | 98% | 98% | 98% | 101% | 100% | 100% | 100% |
| LDB | **CE-2.2** | 97% | 97% | 101% | 101% | 101% | 105% | 102% | 105% |
| **CE-2.3** | 98% | 96% | 99% | 99% | 101% | 104% | 101% | 103% |
| RA | **CE-2.2** | 101% | 102% | 100% | 101% | 99% | 103% | 101% | 104% |
| **CE-2.3** | 101% | 102% | 100% | 100% | 99% | 102% | 100% | 103% |

Table 3.5. Simulation results for TSRC tests, HBD/HBR/HFR CTC, NormQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | | | | | **HDR HLG** | | | | |
|  | dE100 | L100 | wY | wU | wV | Enc | Dec | Y | U | V | Enc | Dec |
| **AI** | **CE-2.2** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 97% | 98% | 0.00% | 0.00% | 0.00% | 97% | 98% |
| **CE-2.3** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 99% | 99% | 0.00% | 0.00% | 0.00% | 97% | 98% |
| **RA** | **CE-2.2** | -0.03% | -0.02% | -0.02% | -0.02% | -0.04% | 101% | 101% | -0.01% | -0.03% | 0.07% | 102% | 102% |
| **CE-2.3** | -0.03% | -0.02% | -0.02% | -0.02% | -0.03% | 101% | 101% | -0.01% | -0.03% | 0.07% | 101% | 102% |

Table 2.6. CE2.1/CE2.2/CE2.3 simulation results, 12 bits data (8 bit data increased to 12), SCC TGM, LowQP test configuration.(Compared to VTM12.0, anchor included below).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **SCC TGM** | | | | | |
|  | Aver.RGB | G | B | R | Enc | Dec |  |
| **AI** | **CEAnc (CE1.1+CE2.1)** | -3.00% | -3.03% | -2.98% | -2.99% | 100% | 98% |  |
| **CE2.1 encoder fix+CE1.1** | -3.06% | -3.11% | -3.03% | -3.04% | 101% | 96% |  |
| **CE2.2+CE1.1** | -3.26% | -3.38% | -3.20% | -3.19% | 105% | 99% |  |
| **CE2.3+CE1.1** | -3.23% | -3.37% | -3.15% | -3.17% | 105% | 99% |  |
| **LDB** | **CEAnc (CE1.1+CE2.1)** | -1.16% | -1.23% | -1.11% | -1.14% | 100% | 99% |  |
| **CE2.1 encoder fix+CE1.1** | -1.54% | -1.46% | -1.57% | -1.58% | 101% | 99% |  |
| **CE2.2+CE1.1** | -1.76% | -1.76% | -1.76% | -1.75% | 103% | 100% |  |
| **CE2.3+CE1.1** | -1.68% | -1.41% | -2.25% | -1.40% | 103% | 100% |  |
| **RA** | **CEAnc (CE1.1+CE2.1)** | -1.13% | -0.93% | -1.48% | -1.00% | 101% | 99% |  |
| **CE2.1 encoder fix+CE1.1** | -1.67% | -1.75% | -1.70% | -1.56% | 101% | 99% |  |
| **CE2.2+CE1.1** | -2.78% | -2.51% | -2.74% | -3.11% | 105% | 101% |  |
| **CE2.3+CE1.1** | -2.32% | -2.13% | -2.36% | -2.48% | 105% | 101% |  |

Gain of 2.2/2.3 compared to the anchor is roughly 0.2% for AI/LB, 1% for RA (for TGM class). No relevant gain for camera captured content. These numbers are however not for native 12 bit data.

**Transform Coefficient Clipping**

This category includes tests CE3.1 and CE3.2 that are studing methods proposed in JVET-U0052.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test** | **Proponent(s)** | **Document description** | **Cross-checker(s)** | **Cross-check document** |
| CE-3.1 | Sharp | JVET-V0047 | Sony | JVET-V0136 |
| CE-3.2 | Sharp | JVET-V0047 | Sony, Qualcomm | JVET-V0136 |

Table 4.2. CE3.1/CE3.2 simulation results, 12 bits data, CE CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
|  | wY | wU | wV | Y | U | V | Aver.RGB | G | B | R |
| **AI** | **CE-3.1** | -1.23% | -1.69% | -1.82% | -0.47% | -0.83% | -0.86% | -1.11% | -0.80% | -1.25% | -1.27% |
| **CE-3.2** | -1.12% | -1.55% | -1.67% | -0.43% | -0.76% | -0.79% | -0.92% | -0.67% | -1.04% | -1.06% |
| **LDB** | **CE-3.1** | -0.67% | -0.65% | -0.72% | -0.62% | -0.80% | -0.84% | -0.70% | -0.79% | -0.65% | -0.66% |
| **CE-3.2** | -0.64% | -0.61% | -0.68% | -0.58% | -0.74% | -0.78% | -0.61% | -0.69% | -0.57% | -0.58% |
| **RA** | **CE-3.1** | -0.60% | -0.63% | -0.69% | -0.50% | -0.73% | -0.78% | -0.63% | -0.61% | -0.64% | -0.65% |
| **CE-3.2** | -0.57% | -0.59% | -0.66% | -0.47% | -0.67% | -0.73% | -0.56% | -0.53% | -0.57% | -0.58% |

Table 4.3. CE3.1/CE3.2 simulation results, 16 bits data, CE CTC, LowQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **SVT16 RGB** | | | |
|  |  | Aver.RGB | G | B | R |
| **AI** | **Vs. VTM12** | -6.53% | -6.82% | -6.39% | -6.37% |
| **Vs. CE anchor** | -0.53% | -0.69% | -0.45% | -0.45% |
| **LDB** | **Vs. VTM12** | -2.92% | -2.98% | -2.91% | -2.88% |
| **Vs. CE anchor** | -0.22% | -0.20% | -0.23% | -0.22% |
| **RA** | **Vs. VTM12** | -2.77% | -2.60% | -2.88% | -2.83% |
| **Vs. CE anchor** | -0.22% | -0.05% | -0.32% | -0.30% |

Table 4.4. Reported run-time estimates RRC tests, CE CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | **HDR HLG** | | **SVT12 RGB** | | **SVT16 RGB** | |
|  | Enc | Dec | Enc | Dec | Enc | Dec | Enc | Dec |
| AI | **CE-3.1** | 101% | 94% | 108% | 93% | 95% | 87% | 107% | 98% |
| **CE-3.2** | 101% | 95% | 107% | 94% | 95% | 90% | 95% | 90% |
| LDB | **CE-3.1** | 100% | 98% | 102% | 95% | 94% | 93% | 103% | 100% |
| **CE-3.2** | 100% | 98% | 101% | 96% | 100% | 93% | 103% | 100% |
| RA | **CE-3.1** | 100% | 99% | 102% | 97% | 98% | 96% | 102% | 99% |
| **CE-3.2** | 100% | 94% | 101% | 90% | 99% | 83% | 102% | 89% |

Table 4.5. CE3.1/CE3.2 simulation results, 12 bits data, HBD/HBR/HFR CTC, NormQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | | | | | **HDR HLG** | | | | |
|  | dE100 | L100 | wY | wU | wV | Enc | Dec | Y | U | V | Enc | Dec |
| **AI** | **CE3.1** | 0.02% | 0.02% | 0.01% | 0.04% | -0.01% | 100% | 100% | 0.01% | -0.11% | -0.14% | 102% | 100% |
| **CE3.2** | -0.07% | -0.01% | -0.01% | 0.12% | 0.01% | 100% | 100% | 0.00% | -0.05% | 0.00% | 101% | 100% |
| **RA** | **CE3.1** | 0.00% | 0.02% | 0.02% | -0.23% | -0.30% | 101% | 100% | 0.01% | -0.06% | 0.20% | 106% | 100% |
| **CE3.2** | -0.05% | -0.02% | 0.00% | -0.40% | -0.41% | 101% | 99% | 0.03% | -0.23% | 0.15% | 106% | 102% |

To assess performance of the coding with extended precision of transform coefficients (18 bits instead of 15 bits of anchor) for 12 bit data coding, coding gain improvement reported by CE3.1 over different codebase is summarized in the Table 4.6. Test CE3.1 was selected for this summarization, since it does not change Log2TransformRange derivation for 12 bits data compared to the respective anchors (e.g. VTM12.0).

Results of CE3.1-VTM provides coding gain achieved by CE3.1 being integratated in the VTM12.0 and tested against the corresponding VTM12.0 anchor. Results of CE3.1-CE provides coding gain achieved by CE3.1 being integratated in the VTM12.0 and tested against the corresponding VTM12.0 anchor.

Table 4.6. CE3.1 simulation results, 12 bits data, HBD/HBR/HFR CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
|  | wY | wU | wV | Y | U | V | R | G | B | Aver.RGB |
| **AI** | **CE3.1-VTM** | -1.71% | -2.18% | -2.34% | -0.90% | -1.18% | -1.21% | -3.50% | -3.86% | -3.86% | -3.74% |
| **CE3.1-CE** | -1.23% | -1.69% | -1.82% | -0.47% | -0.83% | -0.86% | -0.80% | -1.25% | -1.27% | -1.11% |
| **LDB** | **CE3.1-VTM** | -0.74% | -0.71% | -0.78% | -0.60% | -0.82% | -0.90% | -1.22% | -1.15% | -1.16% | -1.18% |
| **CE3.1-CE** | -0.67% | -0.65% | -0.72% | -0.62% | -0.80% | -0.84% | -0.79% | -0.65% | -0.66% | -0.70% |
| **RA** | **CE3.1-VTM** | -0.73% | -0.76% | -0.83% | -0.51% | -0.77% | -0.84% | -1.11% | -1.15% | -1.16% | -1.14% |
| **CE3.1-CE** | -0.60% | -0.63% | -0.69% | -0.50% | -0.73% | -0.78% | -0.61% | -0.64% | -0.65% | -0.63% |

The latter table shows that the benefit of using extended precision flag also with 12 bit data is larger when used on top of VTM than on top of CE anchor (results 3.1-CE are equivalent with table 4.2)

From the discussion:

It is noted that both 3.1 and 3.2 change the binarization (by implementing the extended precision mechanism of HEVC) also for the case of 12 bit data. The anchor is VVC without enabling extended precision. In the tests CE1.x and CE2.x, the extended precision flag was only enabled for 16 bit data. Table 4.2 shows there is also some benefit for 12 bit data.

3.1 is using the HEVC method (as the CE anchor), but implements an additional clipping to 20 bit maximum precision (the anchor would have 22 in case of 16 bit data). 3.2 decreases precision relative to the HEVC method for 12 bit data, but performs the same clipping. Table 4.3 indicates that this additional clipping may have some benefit compared to the HEVC method.

The reason for the benefit of clipping might be that entropy coding uses less bits.

The runtime results from table 4.4 may not be reliable. Decoder run time decrease might not be expected in 12 bit case.

All methods from CE 1.x, 2.x, 3.x have negligible impact in normal QP range.

Subsequent results show combinations.

Table 5.2. Simulation results for combined tests, 12 bits data, CE CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | | | |
|  | wY | wU | wV | Y | U | V | Aver.GBR | G | B | R |
| **AI** | **CE4.1** | -1.42% | -1.80% | -1.93% | -0.89% | -1.01% | -1.04% | -1.81% | -1.86% | -1.77% | -1.79% |
| **CE4.2** | -1.32% | -1.66% | -1.78% | -0.85% | -0.96% | -0.99% | -1.78% | -1.85% | -1.73% | -1.75% |
| **CE4.3** | -1.31% | -1.65% | -1.77% | -0.84% | -0.95% | -0.97% | -1.75% | -1.82% | -1.71% | -1.73% |
| **CE4.4** | -1.14% | -1.57% | -1.68% | -0.47% | -0.81% | -0.83% | -0.93% | -0.68% | -1.05% | -1.07% |
| **CE4.5-CE3.1** | -1.24% | -1.63% | -1.76% | -0.49% | -0.71% | -0.74% | -1.04% | -0.87% | -1.11% | -1.13% |
| **CE4.5-CE3.2** | -1.14% | -1.51% | -1.62% | -0.45% | -0.66% | -0.69% | -0.95% | -0.83% | -1.00% | -1.03% |
| **CE4.6-CE3.1** | -1.30% | -1.72% | -1.86% | -0.54% | -0.85% | -0.88% | -1.40% | -1.16% | -1.51% | -1.54% |
| **CE4.6-CE3.2** | -1.20% | -1.59% | -1.71% | -0.51% | -0.80% | -0.82% | -1.26% | -1.07% | -1.35% | -1.37% |
| **CE4.7-A-CE3.1** | -1.52% | -1.89% | -2.02% | -0.95% | -1.07% | -1.10% | -1.93% | -1.89% | -1.94% | -1.97% |
| **CE4.7-A-CE3.2** | -1.42% | -1.75% | -1.87% | -0.91% | -1.02% | -1.05% | -1.89% | -1.88% | -1.88% | -1.90% |
| **CE4.7-B-CE3.1** | -1.52% | -1.88% | -2.02% | -0.95% | -1.07% | -1.11% | -1.92% | -1.89% | -1.93% | -1.95% |
| **CE4.7-B-CE3.2** | -1.42% | -1.74% | -1.88% | -0.91% | -1.02% | -1.05% | -1.88% | -1.88% | -1.87% | -1.89% |
| **LDB** | **CE4.1** | -0.70% | -0.69% | -0.74% | -0.65% | -0.82% | -0.87% | -1.07% | -1.20% | -1.00% | -1.01% |
| **CE4.2** | -0.66% | -0.65% | -0.72% | -0.61% | -0.77% | -0.81% | -1.04% | -1.16% | -0.98% | -0.99% |
| **CE4.3** | -0.66% | -0.64% | -0.71% | -0.61% | -0.78% | -0.80% | -1.03% | -1.15% | -0.97% | -0.98% |
| **CE4.4** | -0.63% | -0.63% | -0.70% | -0.57% | -0.75% | -0.79% | -0.62% | -0.70% | -0.58% | -0.59% |
| **CE4.5-CE3.1** | -0.69% | -0.67% | -0.73% | -0.63% | -0.81% | -0.84% | -0.79% | -0.88% | -0.74% | -0.75% |
| **CE4.5-CE3.2** | -0.65% | -0.63% | -0.69% | -0.59% | -0.76% | -0.80% | -0.70% | -0.78% | -0.65% | -0.66% |
| **CE4.6-CE3.1** | -0.64% | -0.63% | -0.68% | -0.56% | -0.73% | -0.75% | -0.83% | -0.92% | -0.77% | -0.79% |
| **CE4.6-CE3.2** | -0.62% | -0.59% | -0.65% | -0.53% | -0.69% | -0.72% | -0.76% | -0.84% | -0.71% | -0.72% |
| **CE4.7-A-CE3.1** | -0.72% | -0.70% | -0.76% | -0.63% | -0.80% | -0.83% | -1.10% | -1.19% | -1.04% | -1.06% |
| **CE4.7-A-CE3.2** | -0.69% | -0.67% | -0.74% | -0.59% | -0.74% | -0.79% | -1.06% | -1.15% | -1.01% | -1.02% |
| **CE4.7-B-CE3.1** | -0.72% | -0.70% | -0.75% | -0.63% | -0.78% | -0.83% | -1.09% | -1.18% | -1.04% | -1.05% |
| **CE4.7-B-CE3.2** | -0.69% | -0.67% | -0.74% | -0.59% | -0.74% | -0.77% | -1.06% | -1.14% | -1.01% | -1.02% |
| **RA** | **CE4.1** | -0.63% | -0.65% | -0.72% | -0.55% | -0.75% | -0.81% | -0.99% | -1.05% | -0.95% | -0.96% |
| **CE4.2** | -0.60% | -0.62% | -0.68% | -0.51% | -0.70% | -0.85% | -0.96% | -1.02% | -0.92% | -0.94% |
| **CE4.3** | -0.60% | -0.63% | -0.69% | -0.50% | -0.69% | -0.75% | -0.95% | -1.02% | -0.91% | -0.93% |
| **CE4.4** | -0.56% | -0.58% | -0.65% | -0.45% | -0.67% | -0.72% | -0.57% | -0.54% | -0.58% | -0.60% |
| **CE4.5-CE3.1** | -0.61% | -0.64% | -0.71% | -0.51% | -0.73% | -0.80% | -0.73% | -0.71% | -0.74% | -0.75% |
| **CE4.5-CE3.2** | -0.58% | -0.61% | -0.68% | -0.48% | -0.68% | -0.73% | -0.65% | -0.62% | -0.66% | -0.68% |
| **CE4.6-CE3.1** | -0.60% | -0.63% | -0.69% | -0.47% | -0.69% | -0.74% | -0.76% | -0.75% | -0.76% | -0.78% |
| **CE4.6-CE3.2** | -0.57% | -0.60% | -0.67% | -0.44% | -0.64% | -0.68% | -0.70% | -0.69% | -0.70% | -0.72% |
| **CE4.7-A-CE3.1** | -0.67% | -0.70% | -0.77% | -0.53% | -0.73% | -0.78% | -0.98% | -1.01% | -0.96% | -0.98% |
| **CE4.7-A-CE3.2** | -0.64% | -0.66% | -0.73% | -0.49% | -0.68% | -0.74% | -0.95% | -0.98% | -0.93% | -0.95% |
| **CE4.7-B-CE3.1** | -0.68% | -0.69% | -0.76% | -0.53% | -0.73% | -0.79% | -0.98% | -1.01% | -0.96% | -0.98% |
| **CE4.7-B-CE3.2** | -0.65% | -0.67% | -0.73% | -0.50% | -0.68% | -0.74% | -0.95% | -0.97% | -0.93% | -0.95% |

Table 5.3. Simulation results for combined tests, 16 bits data, CE CTC, LowQP test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Test** | **SVT16 RGB** | | | |
|  | Aver.GBR | G | B | R |
| **AI** | **CE4.1** | -2.45% | -2.96% | -2.21% | -2.18% |
| **CE4.2** | -2.45% | -2.96% | -2.21% | -2.18% |
| **CE4.3** | -2.44% | -2.94% | -2.20% | -2.17% |
| **CE4.4** | -0.63% | -0.80% | -0.54% | -0.54% |
| **CE4.5-CE3.1** | -1.21% | -1.39% | -1.12% | -1.13% |
| **CE4.5-CE3.2** | -1.21% | -1.39% | -1.12% | -1.13% |
| **CE4.6-CE3.1** | -1.24% | -1.40% | -1.16% | -1.17% |
| **CE4.6-CE3.2** | -1.24% | -1.40% | -1.16% | -1.17% |
| **CE4.7-A-CE3.1** | -2.48% | -2.90% | -2.27% | -2.26% |
| **CE4.7-A-CE3.2** | -2.48% | -2.90% | -2.27% | -2.26% |
| **CE4.7-B-CE3.1** | -2.45% | -2.88% | -2.24% | -2.22% |
| **CE4.7-B-CE3.2** | -2.45% | -2.88% | -2.24% | -2.22% |
| **LDB** | **CE4.1** | -1.40% | -1.39% | -1.41% | -1.40% |
| **CE4.2** | -1.40% | -1.39% | -1.41% | -1.40% |
| **CE4.3** | -1.39% | -1.40% | -1.40% | -1.38% |
| **CE4.4** | -0.27% | -0.26% | -0.28% | -0.27% |
| **CE4.5-CE3.1** | -0.80% | -0.78% | -0.82% | -0.81% |
| **CE4.5-CE3.2** | -0.80% | -0.78% | -0.82% | -0.81% |
| **CE4.6-CE3.1** | -0.87% | -0.86% | -0.88% | -0.87% |
| **CE4.6-CE3.2** | -0.87% | -0.86% | -0.88% | -0.87% |
| **CE4.7-A-CE3.1** | -1.62% | -1.60% | -1.64% | -1.63% |
| **CE4.7-A-CE3.2** | -1.62% | -1.60% | -1.64% | -1.63% |
| **CE4.7-B-CE3.1** | -1.60% | -1.57% | -1.62% | -1.60% |
| **CE4.7-B-CE3.2** | -1.60% | -1.57% | -1.62% | -1.60% |
| **RA** | **CE4.1** | -1.31% | -1.22% | -1.36% | -1.35% |
| **CE4.2** | -1.31% | -1.22% | -1.36% | -1.35% |
| **CE4.3** | -1.32% | -1.23% | -1.35% | -1.38% |
| **CE4.4** | -0.27% | -0.10% | -0.36% | -0.35% |
| **CE4.5-CE3.1** | -0.77% | -0.59% | -0.86% | -0.85% |
| **CE4.5-CE3.2** | -0.77% | -0.59% | -0.86% | -0.85% |
| **CE4.6-CE3.1** | -0.85% | -0.68% | -0.94% | -0.93% |
| **CE4.6-CE3.2** | -0.85% | -0.68% | -0.94% | -0.93% |
| **CE4.7-A-CE3.1** | -1.52% | -1.41% | -1.58% | -1.57% |
| **CE4.7-A-CE3.2** | -1.52% | -1.41% | -1.58% | -1.57% |
| **CE4.7-B-CE3.1** | -1.53% | -1.41% | -1.59% | -1.58% |
| **CE4.7-B-CE3.2** | -1.53% | -1.41% | -1.59% | -1.58% |

Difficult to draw conclusions from the combinations, as it would be necessary to compare various tables. This should be done offline by Excel.

Complexity analysis for CE1.x:

|  |  |  |  |
| --- | --- | --- | --- |
| **Test** | **Complexity increase vs. VVC, per TU4x4.** | **Memory requirement** | **Dependencies** |
| *CE-1.1* | 32 | 1 LUT of 4 entries  1 LUT of 5 entries |  |
| *CE-1.2* | 48 |  |  |
| *CE1.4* | 68 (Total):   * 48 in critical pass, * 20 in non-critical pass. | 3 variables (3x4 bits). | Cross-TU dependency.  CE1.4-A: No cross-CTU.  CE1.4-B: Cross-CTU. |
| *CE1.5* | 95 (Total)   * 32 in critical pass * 63 in non-critical pass | 8 variables (8 \* 10 bits)  6 \* 16 \* 2 bits LUT | Cross-subTU  Cross-TU  Cross-CTU |
| *CE1.6* | 95 (Total)   * 32 in critical pass * 63 in non-critical pass | 8 variables (8 \* 10 bits)  6 \* 16 \* 2 bits LUT | Cross-TU  Cross-CTU |
| *CE1.7* | 48 | No | No |

The numbers above are just counting operations, but not documenting what types of operations. More details can be found in the last meeting’s BoG report and in the detailed CE contributions.

Complexity analysis for CE2.x:

|  |  |  |  |
| --- | --- | --- | --- |
| **Test** | **Complexity increase vs. VVC, per TU4x4.** | **Memory requirement** | **Dependencies** |
| *CE-2.1* | 1 condition and 1 add per slice | No memory increase | No Dependency |
| *CE-2.2* | 61 (Total)   * 16 in critical pass * 45 in non-critical pass | 2 variables (2 \* 10 bits) | Cross-subTU  Cross-TU  Cross-CTU |
| *CE-2.3* | 61 (Total)   * 16 in critical pass * 45 in non-critical pass | 2 variables (2 \* 10 bits) | Cross-TU  Cross-CTU |

Complexity analysis for CE3.x/CE4.x still to be done.

Further analysis should be provided in comparison to HEVC low QP range 12 bit:

* Comparing CE anchor (available)
* Comparing anchor plus extended precision flag
* Comparing CE1.4 as a representative of better compression in entropy coding plus extended precision flag

Based on that, a reasonable compression vs. complexity tradeoff for a 12 bit profile could be discussed.

Similar analysis for 16 bit case, but invoking the 20-bit clipping from 3.1 additionally, to see how much gain is possible compared to HEVC.

Further analysis is provided in JVET-V0177.

The additional methods from CE2.x show relatively low gain (1% for RA in TGM compared to anchor), while they introduce quite some complexity. The TSRC modification from the anchor already provides most of the gain that is possible by resolving the Rice parameter problem.

[JVET-V0177](https://jvet-experts.org/doc_end_user/current_document.php?id=10840) CE-related: Additional information for CE on HBD coding [A. Browne, T. Hashimoto, H. Jhu, D. Rusanovskyy, K. Kawamura, T. Zhou] [late]

Presented and discussed in session 20, 0010-0130 UTC, Tuesday 27 April (chaired by JRO and GJS)

This contribution presents additional results on the RRC methods for high bit depth video coding being studied in CE. Results include performance comparison for selected RRC methods against the HM16.23. as well as additional results on the performance of the Method 2 presented in JVET-T0105 [6]. Complexity increase compare to VVC is provided for selected RRC methods.

Results below have been produced by comparing reported results for CE tests against the HM16.23 anchor, provided by AhG8. Complete Excel tables are provided along this contribution.

Table 1 shows coding gain achieved by selected methods against the HM16.23. Impact of EGk invoking parameters change on top of selected tests is provided for SVT16 data separately in the CE3.1 (EGk) column.

Table 1. Summary of results (BD-rate gain) for selected RRC tests, CE CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | HBD CTC | | | | | | | | CE3.1  (Egk) |
|  | **Test  vs  HM16.23** | **HDR PQ** | | | **HDR HLG** | | | **SVT12  RGB** | **SVT16  RGB** | **SVT16  RGB** |
|  | wY | wU | wV | Y | U | V | Aver.  GBR | Aver. GBR | Aver. GBR |
| **AI** | **VTM12** | -6.79% | -9.19% | -9.75% | -3.68% | -6.06% | -6.19% | 0.46% | 34.93% | 26.67% |
| **AnchorCE** | -7.24% | -9.64% | -10.23% | -4.09% | -6.39% | -6.53% | -2.17% | 0.77% | 0.24% |
| **CE1.4-A** | -7.56% | -9.85% | -10.44% | -4.57% | -6.67% | -6.82% | -3.56% | -1.55% | -1.74% |
| **LDB** | **VTM12** | -6.61% | -7.67% | -8.77% | -5.14% | -8.18% | -7.01% | -1.35% | 21.09% | 17.88% |
| **AnchorCE** | -6.68% | -7.73% | -8.82% | -5.12% | -8.20% | -7.06% | -1.79% | 0.58% | 0.35% |
| **CE1.4-A** | -6.73% | -7.79% | -8.89% | -5.14% | -8.21% | -7.05% | -2.43% | -0.99% | -1.08% |
| **RA** | **VTM12** | -6.81% | -7.91% | -8.64% | -5.72% | -8.64% | -7.32% | -1.64% | 19.55% | 16.42% |
| **AnchorCE** | -6.93% | -8.04% | -8.77% | -5.73% | -8.68% | -7.37% | -2.11% | 0.32% | 0.08% |
| **CE1.4-A** | -7.01% | -8.10% | -8.85% | -5.77% | -8.69% | -7.38% | -2.67% | -1.14% | -1.22% |

Table 2 shows coding gain achieved by enabling extended precission flag, as per CE3.1, for selected methods against the HM16.23 for 12 bits data.

Table 2. Summary of results (BD-rate gain) for selected RRC tests, CE CTC, LowQP test configuration.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | CE3.1 Extended precission flag + Test vs. HM16.23 | | | | | | |
|  | **Testvs HM** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** |
|  | wY | wU | wV | Y | U | V | Aver.GBR |
| **AI** | **VTM12+ce3.1** | -7.88% | -10.64% | -11.29% | -4.16% | -6.81% | -6.96% | -0.41% |
| **AnchorCE+ce3.1** | -8.35% | -11.13% | -11.81% | -4.54% | -7.14% | -7.30% | -3.27% |
| **CE1.4-A+ce3.1** | -8.63% | -11.31% | -11.99% | -5.00% | -7.38% | -7.54% | -4.06% |
| **LDB** | **VTM12+ce3.1** | -7.21% | -8.24% | -9.37% | -5.69% | -8.85% | -7.78% | -2.07% |
| **AnchorCE+ce3.1** | -7.27% | -8.30% | -9.44% | -5.69% | -8.88% | -7.81% | -2.48% |
| **CE1.4-A+ce3.1** | -7.31% | -8.35% | -9.49% | -5.71% | -8.88% | -7.79% | -2.86% |
| **RA** | **VTM12+ce3.1** | -7.32% | -8.44% | -9.22% | -6.17% | -9.25% | -8.01% | -2.30% |
| **AnchorCE+ce3.1** | -7.45% | -8.58% | -9.37% | -6.19% | -9.29% | -8.05% | -2.74% |
| **CE1.4-A+ce3.1** | -7.52% | -8.64% | -9.44% | -6.22% | -9.29% | -8.05% | -3.07% |

Table 3. Summary of results (BD-rate gain) for selected RRC tests, CE CTC, Lossless test configuration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | **Test** | bit-rate saving | | | |
|  | HDR PQ | HDR HLG | SVT12 | SVT16 |
| **AI** | **VTM** | 2.93% | 7.05% | 31.02% | 73.47% |
| **AnchorCE** | 0.36% | 2.82% | 3.12% | 1.46% |
| **CE1.4-A** | -1.97% | -0.77% | 0.15% | 0.05% |
| **LDB** | **VTM** | 3.67% | 2.09% | 29.88% | 80.20% |
| **AnchorCE** | 4.10% | 3.96% | 2.71% | 1.18% |
| **CE1.4-A** | 0.90% | -0.52% | 1.16% | 0.79% |
| **RA** | **VTM** | 3.30% | 1.85% | 29.30% | 80.25% |
| **AnchorCE** | 3.80% | 3.79% | 2.55% | 1.11% |
| **CE1.4-A** | 0.70% | -0.59% | 1.10% | 0.74% |

Benefit of history based vs. local based is around 0.3% for AI, 0.1% for RA in case of 12 bit low QP

Gain over HM is around 7% for 12 bit low QP in PQ, less for HLG

Additional gain of extended prec is >1% for AI, 0.5% for RA

For 16 bit data, only history based methods are able to provide (small) gain over HM

Definitely a modified entropy coding method (and extended transform precision) is necessary for professional application scenario, in particular 16 bit case.

Profile definitions have not yet been discussed, one aspect that might considered if there will be different 12 bit profiles (or tiers) for consumer and professional use cases.

Is the technology investigated in the CE mature for being adopted?

For TSRC, the method from the CE anchor (CE2.1) should be adpted, as it is simple, and the more complex solutions do not provide significant gain.

Decision: Adopt JVET-V0054, method CE2.1 as entropy coding method for high bit depth in TSRC

For RRC, two major candidates, CE1.4-A, and JVET-V0084 (which is a CE related contribution, modification of CE1.6)

In CE1.4-A, the history based approach can be disabled (then it falls back to 1.2). An alternative is proposed in JVET-V0106 (CE related), where the same history based method of CE1.4-A I built on top of CE1.1 (the latter almost identical with the CE anchor. CE1.1 is more hardware implementation friendly than CE1.2. All these have been thoroughly investigated in CE

JVET-V0084 is only history based (cannot be disabled)

Decision: Adopt JVET-V0106 as entropy coding method for high bit depth in RRC. High level flag to disable the history based part.

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

[JVET-V0046](https://jvet-experts.org/doc_end_user/current_document.php?id=10693) CE-1.7: Rice parameter derivation for high bit-depth coding [T. Hashimoto, T. Ikai (Sharp)]

[JVET-V0047](https://jvet-experts.org/doc_end_user/current_document.php?id=10694) CE-3.1 and CE-3.2: Transform coefficients range extension for high bit-depth coding [T. Zhou, T. Chujoh, T. Ikai (Sharp)]

[JVET-V0048](https://jvet-experts.org/doc_end_user/current_document.php?id=10695) CE-4.1: Combination of CE-3.1, CE-1.5 and CE-2.1 [T. Zhou, T. Chujoh, T. Ikai (Sharp)]

[JVET-V0049](https://jvet-experts.org/doc_end_user/current_document.php?id=10696) CE-4.2: Combination of CE-3.2, CE-1.5 and CE-2.1 [T. Zhou, T. Chujoh, T. Ikai (Sharp)]

[JVET-V0050](https://jvet-experts.org/doc_end_user/current_document.php?id=10697) CE-1.5, CE-1.6, CE-2.2 and CE-2.3: Rice parameter selection for high bit depths [A. Browne, S. Keating, K. Sharman (Sony)]

[JVET-V0051](https://jvet-experts.org/doc_end_user/current_document.php?id=10698) CE-4.3 and CE-4.4: Combinations of CE-1.6 and CE-2.3 with CE-3.2 [A. Browne, S. Keating, K. Sharman (Sony)]

[JVET-V0052](https://jvet-experts.org/doc_end_user/current_document.php?id=10699) CE-1.1, CE-1.2 and CE-1.4: On the Rice parameter derivation for high bit-depth coding [D. Rusanovskyy, M. Karczewicz, L. P. Van, M. Coban (Qualcomm)]

[JVET-V0053](https://jvet-experts.org/doc_end_user/current_document.php?id=10700) CE-4.5, CE-4.6 and CE-4.7: Combinations of RRC tests CE-1.1, CE-1.2 and CE-1.4 with CE-3.1/CE-3.2 [D. Rusanovskyy, M. Karczewicz, L. P. Van, M. Coban (Qualcomm)]

[JVET-V0054](https://jvet-experts.org/doc_end_user/current_document.php?id=10702) CE-2.1: Slice based Rice parameter selection for transform skip residual coding [H.-J. Jhu, X. Xiu, Y.-W. Chen, W. Chen, C.-W. Kuo, X. Wang (Kwai Inc.)]

[JVET-V0171](https://jvet-experts.org/doc_end_user/current_document.php?id=10834) CE-1.8: Results of Rice Parameter Derivation with Content Adaptation [K. Kawamura, K. Unno (KDDI)] [late]

No need for presentation according to contributors. This experiment had been withdrawn from the CE.

[JVET-V0134](https://jvet-experts.org/doc_end_user/current_document.php?id=10795) Crosscheck of CE-1.4 from JVET-V0052 (CE-1.1, CE-1.2 and CE-1.4: On the Rice parameter derivation for high bit-depth coding) [A. Browne (Sony)] [late]

[JVET-V0135](https://jvet-experts.org/doc_end_user/current_document.php?id=10796) Crosscheck of CE-4.5 and CE-4.7 from JVET-V0053 (CE-4.5, CE-4.6 and CE-4.7: Combinations of RRC tests CE-1.1, CE-1.2 and CE-1.4 with CE-3.1/CE-3.2) [A. Browne (Sony)] [late]

[JVET-V0136](https://jvet-experts.org/doc_end_user/current_document.php?id=10797) Crosscheck of JVET-V0047 (CE-3.1 and CE-3.2: Transform coefficients range extension for high bit-depth coding) [A. Browne (Sony)] [late]

[JVET-V0138](https://jvet-experts.org/doc_end_user/current_document.php?id=10799) Crosscheck of CE-1.2 from JVET-V0052 (CE-1.1, CE-1.2 and CE-1.4: On the Rice parameter derivation for high bit-depth coding) [T. Hashimoto (Sharp)] [late]

[JVET-V0139](https://jvet-experts.org/doc_end_user/current_document.php?id=10800) Cross-check report for CE1.5 and CE1.6 of JVET-V0050 and CE4.3 of JVET-V0051 [D. Rusanovskyy (Qualcomm)] [late]

[JVET-V0140](https://jvet-experts.org/doc_end_user/current_document.php?id=10801) Cross-check report for CE-1.7 of JVET-V0046 and CE-4.2 of JVET-V0049 [D. Rusanovskyy (Qualcomm)] [late]

[JVET-V0142](https://jvet-experts.org/doc_end_user/current_document.php?id=10803) Crosscheck of JVET-V0048 (CE-4.1: Combination of CE-3.1, CE-1.5 and CE-2.1) [H.-J. Jhu (Kwai Inc.)] [late]

[JVET-V0143](https://jvet-experts.org/doc_end_user/current_document.php?id=10804) Crosscheck of CE-2.2 and CE-2.3 from JVET-V0050 (CE-1.5, CE-1.6, CE-2.2 and CE-2.3: Rice parameter selection for high bit depths) [H.-J. Jhu (Kwai Inc.)] [late]

[JVET-V0144](https://jvet-experts.org/doc_end_user/current_document.php?id=10805) Crosscheck of CE-4.4 from JVET-V0051 (CE-4.3 and CE-4.4: Combinations of CE-1.6 and CE-2.3 with CE-3.2) [H.-J. Jhu (Kwai Inc.)] [late]

[JVET-V0145](https://jvet-experts.org/doc_end_user/current_document.php?id=10806) Crosscheck of CE-4.6 from JVET-V0053 (CE-4.5, CE-4.6 and CE-4.7: Combinations of RRC tests CE-1.1, CE-1.2 and CE-1.4 with CE-3.1/CE-3.2) [T. Zhou (Sharp)] [late]

[JVET-V0155](https://jvet-experts.org/doc_end_user/current_document.php?id=10816) Crosscheck of CE-1.1 from JVET-V0052 (CE-1.1, CE-1.2 and CE-1.4: On the Rice parameter derivation for high bit-depth coding) [M.G. Sarwer (Alibaba)] [late]

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

Contributions in this area were discussed in session 4 at 2330–0105 UTC on Tuesday/Wednesday 20/21 April 2021 (chaired by JRO and GJS).

[JVET-V0084](https://jvet-experts.org/doc_end_user/current_document.php?id=10732) CE-related: On Rice parameter selection for regular residual coding (RRC) at high bit depths [A. Browne, S. Keating, K. Sharman (Sony)]

This contribution describes modifications to the method for the selection of Rice parameters for regular residual coding, for consideration in VVC version 2. The modifications offer BD-rate gains when residual coefficient values become larger but do not affect version 1 operating points. It is suggested that the modification extends the operating point of the existing residual coding method to be applicable at all currently considered bit depths. A further modification is described which alters the template function used to calculate a clipped sum of neighbouring coefficients used during Rice parameter evaluation. Results are provided both with and without this further modification.

Using the high bit depth, high bit rate CTC the following AI results were obtained for the low QP range against the CE anchor:

With template modification:

PQ: -0.29%/-0.18%/-0.18%

HLG: -0.56%/-0.34%/0.35%

SVT-12: -1.67%/-1.26%/-1.27%

SVT-16: -2.94%/-2.09%/-2.07%

Without template modification

PQ: -0.26%/-0.16%/-0.16%

HLG: -0.53%/-0.33%/-0.34%

SVT-12: -1.65%/-1.26%/-1.26%

SVT-16: -2.94%/-2.10%/-2.08%

Based on CE-1.6, history based method using counters, but is claimed to have various simplifications. Operates on sub-TU basis. Improved results compared to CE-1.6. Compared to CE-1.4 (which is better than CE1.6 generally), it is sometimes better, sometimes worse.

Would require more analysis on complexity in comparison to other proposals and the CE anchor.

Why is decoder runtime increased significantly for SVT class? Analysis shows that more 4x4 blocks are used than with other methods, which is closer to the worst case scenario.

[JVET-V0157](https://jvet-experts.org/doc_end_user/current_document.php?id=10818) Crosscheck of JVET-V0084 (CE-related: On Rice parameter selection for regular residual coding (RRC) at high bit depths) [M. G. Sarwer (Alibaba)] [late]

[JVET-V0085](https://jvet-experts.org/doc_end_user/current_document.php?id=10733) CE-related: On Rice parameter selection for transform skip residual coding (TSRC) at high bit depths [A. Browne, S. Keating, K. Sharman (Sony)]

This contribution describes modifications to the method for the selection of Rice parameters for transform skip residual coding, for consideration in VVC version 2. The modifications offer BD-rate gains, particularly for screen content, when residual coefficient values become larger but do not affect version 1 operating points. It is suggested that the modification extends the operating point of the existing residual coding method to be applicable at all currently considered bit depths.

Using the high bit depth, high bit rate CTC the following BD-rate results were obtained for screen content coding:

AI: -0.41%/-0.14%/-0.25%

RA: -0.86%/-0.85%/-0.86%

LDB: -0.59%/-1.08%/-0.32%

Similar to method CE2.3, simplification conceptually similar to V0084 with single threshold. Gain too low to be considered.

[JVET-V0169](https://jvet-experts.org/doc_end_user/current_document.php?id=10832) Crosscheck of JVET-V0085 (CE-related: On Rice parameter selection for transform skip residual coding (TSRC) at high bit depths) [?? (Sharp)] [late]

[JVET-V0106](https://jvet-experts.org/doc_end_user/current_document.php?id=10754) CE-related: On history-enhanced method of Rice parameter derivation for regular residual coding (RRC) at high bit depths [D. Rusanovskyy, L. Pham Van, M. Coban, M. Karczewicz (Qualcomm)]

This contribution describes a history-based extension of Rice parameters derivation for regular residual coding (RRC), for VVCv2 development. It is advocated that a local (template based) methods of Rice parameter derivation in RRC, can be further improved by enabling history (across TUs) utilization during the template computation. One example of such history-based extension is currently being studied in CE1.4, where local method of CE1.2 is being extended with a history usage. This proposal present additional example of history-based extension for local CE1.1 derivation and present improvement provided by the said extension compare to the local methods.

It is asserted by the proponent, that proposed history-enhancement for local RRC derivation provides a complexity and dependency scalable RRC parameter derivation method with each component (of the method) providing a gain in all tested class. Method is proposed for adoption to VVCv2.

The following results (against the CE anchor) are reported for the described inhere local CE1.1 and CE1.2 RRC methods and their enhanced versions, respectively:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Test** | **HDR PQ** | | | **HDR HLG** | | | **SVT12 RGB** | **SVT16 RGB** |
|  | wY | wU | wV | Y | U | V | Aver.GBR | Aver.GBR |
| **AI** | **CE1.1** | -0.08% | -0.07% | -0.08% | -0.07% | -0.02% | -0.02% | -0.27% | -0.71% |
| **CE1.2** | -0.09% | -0.06% | -0.07% | -0.10% | -0.03% | -0.04% | -0.41% | -0.79% |
| **CE1.1+Hist** | -0.31% | -0.21% | -0.22% | -0.45% | -0.25% | -0.26% | -1.33% | -2.36% |
| **CE1.2+Hist** | -0.33% | -0.23% | -0.25% | -0.50% | -0.30% | -0.32% | -1.43% | -2.27% |
| **LDB** | **CE1.1** | -0.02% | -0.02% | -0.04% | -0.01% | 0.00% | 0.01% | -0.13% | -0.58% |
| **CE1.2** | 0.03% | 0.03% | 0.02% | 0.02% | 0.05% | 0.06% | -0.19% | -0.66% |
| **CE1.1+Hist** | -0.09% | -0.10% | -0.12% | -0.05% | -0.04% | -0.05% | -0.67% | -1.71% |
| **CE1.2+Hist** | -0.06% | -0.08% | -0.09% | -0.06% | -0.01% | 0.00% | -0.66% | -1.53% |
| **RA** | **CE1.1** | -0.02% | -0.03% | -0.03% | -0.01% | 0.00% | 0.00% | -0.12% | -0.51% |
| **CE1.2** | -0.01% | -0.01% | -0.02% | 0.01% | 0.03% | 0.03% | -0.17% | -0.60% |
| **CE1.1+Hist** | -0.11% | -0.10% | -0.12% | -0.07% | -0.05% | -0.06% | -0.60% | -1.61% |
| **CE1.2+Hist** | -0.08% | -0.08% | -0.09% | -0.04% | -0.02% | -0.02% | -0.57% | -1.45% |

Proponents claim that the history based extension mechanism might be useful in a potential 16 bit profile, when a 12 bit profile uses a local based method such as CE1.1 (complexity scalable solution).

Question: Would this be a benefit for implementation? Would the same devices be used for 12 and 16 bit profiles? It would nevertheless simplify a unified specification.

CE1.1 is claimed to be similarly complex as the anchor. Would it be possible to extend the CE anchor by history based extension? In principle yes.

The anchor should be included in the complexity analysis of CE1. It is mentioned that the anchor had been specified in JVET-U0064 which was decided to become the anchor in the last meeting. One expert mentioned that a possibly simpler method had earlier been proposed in JVET-T0105 which had similar performance for 12 bit data.

[JVET-V0156](https://jvet-experts.org/doc_end_user/current_document.php?id=10817) Crosscheck of JVET-V0106 (CE-related: On history-enhanced method of Rice parameter derivation for regular residual coding (RRC) at high bit depths) [M. G. Sarwer (Alibaba)] [late]

[JVET-V0123](https://jvet-experts.org/doc_end_user/current_document.php?id=10772) CE-related: On prefix code length of remaining level coding for high bit depth and high bit rate coding [Y. Yu, Z. Xie, F. Wang, H. Yu, D. Wang (OPPO)]

This contribution proposes a modification to the prefix code length of remaining level coding for high bit-depth and high bit-rate coding. A smaller fixed value equal to 1 or 4 is proposed to be used in the VTM12 or on the top of CE anchor. Simulations results show that there are gain for SVT 16 bit video while there are loss for 12 bit video.

Over VTM12, SVT 16:

AI: -19.45% (Y), -17.97%(U), -17.97%(V); LDB:-11.39%(Y),-10.79%(U),-10.80%(V); RA: -10.38%(Y), -10.27%(U), -10.24%(V)

Over CE anchor, SVT 16:

AI:-0.3% (Y),-0.25%(U), -0.25%(V); LDB:-0.06%(Y),-0.07%(U),-0.07%(V); RA: -0.02%(Y), -0.06%(U), -0.05%(V)

It is suggested that a smaller fixed number of prefix code length could be used for 16-bit video while retaining the current prefix code length unchanged for 12-bit input video.

Compared to CE anchor, still loss is observed for 12 bit data. Also for 16 bit data, some CE proposals showed better performance.

Benefit of the method compared to other proposals not obvious.

[JVET-V0161](https://jvet-experts.org/doc_end_user/current_document.php?id=10822) Crosscheck of JVET-V0123 (CE-related: On prefix code length of remaining level coding for high bit depth and high bit rate coding) [H.-J. Jhu (Kwai Inc.)] [late]

### Adaptation of other tools for high bit rate and high bit depth (10)

Contributions in this area were discussed in session 8 at 0105–0120 UTC, in sessions 9/10 at 1500-1720 UTC on Wednesday 21 April 2021 (chaired by JRO and GJS) , and in session 13 at 1405-1435 on Friday 23 April 2021 (chaired by JRO).

[JVET-V0059](https://jvet-experts.org/doc_end_user/current_document.php?id=10707) AHG8: CABAC-bypass alignment for high bit-depth coding [M. G. Sarwer, J. Chen, Y. Ye, R. -L. Liao (Alibaba)]

In the last JVET meeting, a CABAC bypass alignment flag (similar to HEVC range extension) was proposed in JVET-U0069 to improve the throughput of the CABAC engine in high bit-rate applications. It was generally agreed that CABAC bypass alignment method in JVET-U0069 could improve the throughput in very high bit-rate professional applications, however, further study was recommended to investigate at which level (i.e. profile level, sequence level etc) this functionality is desirable. This contribution proposed to enable CABAC bypass alignment method in VVC High Throughput 4:4:4 16 Intra Profile.

Same as JVET-U0069, this contribution proposes two alignment options. In the first option, the bypass alignment is applied only to coefficient coding within a coefficient group (CG), without affecting the coding of sb\_coded\_flag of a transform block (TB). In the second option, in addition to CABAC bypass alignment, it is also proposed to switch to bypass coding of sb\_coded\_flag after the limit of context coded bins has been reached for the current TB. It is asserted that with the second option, alignment is needed only once after the limit of the context coded bins has been reached for the TB.

Following results are reported. Note that only SVT 16-bit sequences will be affected by the proposed changes as all other sequences are 12-bit sequences.

As compared to CE-1 anchor (low QP):

* Option1: SVT-16 bit sequences:
  + All Intra: 0.47%(G), 0.49% (B), 0.49 % (R), 100% (EncT), 93%(DecT)
  + LDB: 0.55%(G), 0.55 % (B), 0.55 % (R), 100% (EncT), 94%(DecT)
  + Random access: 0.57%(G), 0.56 % (B), 0.56 % (R), 100% (EncT), 93%(DecT)
* Option2: SVT-16 bit sequences:
  + All Intra: 0.47%(G), 0.49 % (B), 0.49 % (R), 99% (EncT), 94%(DecT)
  + LDB: 0.55%(G), 0.54 % (B), 0.54 % (R), 100%(EncT), 94%%(DecT)
  + Random access: 0.57%(G), 0.56 % (B), 0.56 % (R), 100%(EncT), 94%(DecT)

From the discussion:

* Should there be two 16 bit profiles?
* If yes, should they be “onion shell” structured, i.e. high throughput below the non-restricted?
* How does definition of a profile relate to the benefit compared to HEVC?
* The lack of gain may also be related to the current test data, and the versatility of VVC should include that it is able to support 16 bit

See further notes under JVET-V0178.

[JVET-V0141](https://jvet-experts.org/doc_end_user/current_document.php?id=10802) Cross-check report of AHG8: CABAC-bypass alignment for high bit-depth coding (JVET-V0059) [D. Rusanovskyy (Qualcomm)] [late]

[JVET-V0066](https://jvet-experts.org/doc_end_user/current_document.php?id=10714) AHG8: Encoder improvements to palette coding for high bit depth [T. Tsukuba, M. Ikeda, T. Suzuki (Sony)]

This contribution proposes two encoder modifications to palette coding for high bit depth: Proposal#1 uses on-the-fly bit calculation functions instead of pre-defined lookup tables to reduce memory usage and Proposal#2 fixes QP offset derivation used for an error limit to improve coding performance of palette coding. It is reported that Proposal#1 has no impact to coding performance and can reduce memory size. Regarding Proposal#2, it is reported that average bd-rate changes for YUV at normal QP are (-8.70%, -10.86%, -11.66%) for AI and (-6.09%, -8.43%, -9.16%) for RA, and average bd-rate changes for RGB are (-8.51%, -8.65%, -8.32%) for AI and (-5.64%, -5.86%, -5.73%) for RA; average bd-rate changes for YUV at low QP are (-1.80%, -2.08%, -2.09%) for AI and (-1.34%, -1.27%, -1.31%) for RA, and average bd-rate changes for RGB are (-1.87%, -1.93%, -1.96%) for AI and (-1.19%, -1.33%, -1.27%) for RA, respectively.

Proposal2 is obvious bug. It is asked if Proposal1 would have impact on CTC conditions in terms of run time. It would be desirable to replace the lookup table by on the fly computation for all bit depth cases. Proponents were asked to investigate on this aspect. It was later confirmed that there is no impact when the on-the-fly computation is exercised in CTC 4:4:4 condition, neither on runtime nor on performance.

Decision (SW/BF): Adopt JVET-V0066 both aspects. The method should replace the lookup table approach (which also removes some macros) and be uniquely applied for both “normal” CTC and high bit depth.

[JVET-V0162](https://jvet-experts.org/doc_end_user/current_document.php?id=10823) Crosscheck of JVET-V0066 (AHG8: Encoder improvements to palette coding for high bit depth) [H.-J. Jhu (Kwai Inc.)] [late]

[JVET-V0067](https://jvet-experts.org/doc_end_user/current_document.php?id=10715) AHG8: A constraint of max transform size for high bit depth and high bit rate coding [K. Kondo, M. Ikeda (Sony)]

This contribution proposes to restrict max transform size for high bit depth. This constraint helps to reduce memory size in hardware architecture. To investigate the impact of coding efficiency, experiments to restrict max transform size to 32 was carried out. 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.

* Low QP range
  + {0.01%, 0.02%, 0.01%} for wPSNR and PQ contents
  + {0.00%, 0.02%, 0.01%} for PSNR and PQ contents
  + {0.00%, 0.01%, 0.00%} for HLG contents
  + {0.00%, 0.00%, 0.00%} for SVT contents.
* Normal QP range
  + {0.35% , 4.45%, 4.64%} overall in RA case
  + {0.13%, 6.08%, 6.43%} overall in AI case

Proponents propose the constraint for 16 bit and high bit rate (low QP).

It is commented that it would only be useful as a profile constraint, as a decoder would still need to reserve the memory when the QP would be QP dependent.

It is asked what the overall saving in memory would be. With higher bit depth, the internal memory is increased not only in the transform stage, so the overall benefit would be limited.

It is also asked if a 16 bit profile decoder would be able to decode a Main 10 stream? In that case, the memory and logic for a 64-size transform would need to be provided anyway.

Further study in the overall context of a possible definition of a 16 bit profile.

[JVET-V0159](https://jvet-experts.org/doc_end_user/current_document.php?id=10820) Crosscheck of JVET-V0067 (AHG8: A constraint of max transform size for high bit depth and high bit rate coding) [T. Hashimoto (Sharp)] [late]

[JVET-V0068](https://jvet-experts.org/doc_end_user/current_document.php?id=10716) AHG8: A constraint of max CTU size for high bit depth and 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. To investigate the impact of coding efficiency, max CTU size 64 was 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.

* Low QP range
  + {-0.05%, -0.20%, -0.26%} for wPSNR and PQ contents
  + {-0.04%, -0.17%, -0.23%} for PSNR and PQ contents
  + {0.01%, -0.07%, -0.13%} for HLG contents
  + {0.02%, 0.00%, 0.00%} for SVT contents.
* Normal QP range
  + {1.74%, 4.53%, 4.29%} overall in RA case
  + {0.35%, 1.49%, 0.46%} overall in AI case

In case of 4K video, 16 threads could be used at decoder with 128 CTU size. Is this not sufficient for the high bit rate range? Not known, further analysis would be necessary.

It is further pointed out that a decoder would not be able to do parsing in wavefront mode if it is not enabled in the bitstream. A profile intended for extreme high bit rate might need to enforce that.

Further study in the overall context of a possible definition of a 16 bit profile.

[JVET-V0160](https://jvet-experts.org/doc_end_user/current_document.php?id=10821) Crosscheck of JVET-V0068 (AHG8: A constraint of max CTU size for high bit depth and high bit rate coding) [T. Hashimoto (Sharp)] [late]

[JVET-V0121](https://jvet-experts.org/doc_end_user/current_document.php?id=10770) AHG8: On 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)]

This contribution proposes a modification to the derivation method of last significant coefficient position in regular residual coding (RRC), for consideration in VVC version 2. In the current VVC specification, the coordinate of last significant coefficient (LastSignificantCoeffX, LastSignificantCoeffY) is measured in reference to the top-left corner of a transform block (TB) and coded with syntax elements last\_sig\_coeff\_x\_prefix, last\_sig\_coeff\_y\_prefix, last\_sig\_coeff\_x\_suffix, and last\_sig\_coeff\_y\_suffix. It is observed that, when coding for high quality (or high bit-rate) with high bit-depth video, the number of significant coefficients become larger and the last significant coefficient often occur in the bottom-right portion of a TB. To accommodate this new distribution aspect, this contribution proposes to code the relative coordinates of last significant coefficient in reference to the bottom-right corner of a zero-out transform block. The simulation results are summarized below, where the anchor was produced by the CE base software.

Low QP:

* PQ AI -0.01%, -0.04%, -0.04% LD 0.00%, -0.01%, 0.03% RA 0.01%, -0.02%, 0.01%
* HLG AI -0.20%, -0.26%, -0.33% LD -0.22%, -0.27%, -0.35% RA -0.22%, -0.28%, -0.38%
* SVT AI -0.21%, -0.20%, -0.21% LD -0.20%, -0.19%, -0.20% RA -0.14%, -0.12%, -0.12%

Lossless:

* PQ AI -0.27% LD -0.36% RA -0.35%
* HLG AI -0.26% LD -0.29% RA -0.27%
* SVT AI -0.11% LD -0.10% RA -0.09%

One explanation for the gain is that small blocks are more frequently used, such that saving rate for last coefficient position (which could occur in every block) could be relevant

The proposal is to introduce a flag that signals if the last position is coded with reference to the top left or bottom right corner. It is then an encoder decision how to use it.

Implementation-wise this appears simple.

Study in CE.

[JVET-V0158](https://jvet-experts.org/doc_end_user/current_document.php?id=10819) Crosscheck of JVET-V0121 (AHG8: on coding of last significant coefficient position for high bit depth and high bit rate extensions) [M. G. Sarwer (Alibaba)] [late]

[JVET-V0122](https://jvet-experts.org/doc_end_user/current_document.php?id=10771) AHG8: A full-bypass mode in residual coding for high bit depth and high bit rate extensions [F. Wang, Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO)]

This contribution proposes a full-bypass mode in residual coding for VVC version 2. It is proposed to add this new mode in both regular residual coding (RRC) and transform skip residual coding (TSRC). When enabled, this new mode will code coefficient or residual levels only through the bypass mode of the arithmetic coding engine without using any context-coded bin. This mode is primarily designed to improve the throughput of arithmetic coding for high quality applications with high bit-depth video signals, such as content capture, creation, and editing. It is observed that the simulation results showed over 15% reduction in average decoding time, even with a slight coding efficiency gain, from the 16-bit SVT sequences. The tests were conducted by using the CE anchor software and running the latest CTC in JVET-U2018.

For the HLG and PQ sequences, significant loss (10% BR increase in RA/LB, 5-6% in AI) is observed.

It is mentioned that modification only for transform coefficient coding may not be the only problem. Percentage of bits spent on other information such as MV may also be relevant for small block sizes.

It is also pointed out that in a true bypass mode the entire entropy coding engine might need to be bypassed.

Further study in the context of a possible definition of a high-throughput mode (see also relation with JVET-V0059 and JVET-V0150).

The problem needs to be better understood and analysed e.g. the ratio of context coded and bypass coded bins, in particular for 16 bit data in high rate.

[JVET-V0163](https://jvet-experts.org/doc_end_user/current_document.php?id=10824) Crosscheck of JVET-V0122 (AHG8: a full-bypass mode in residual coding for high bit depth and high bit rate extensions) [A. Browne (Sony)] [late]

[JVET-V0124](https://jvet-experts.org/doc_end_user/current_document.php?id=10773) AHG8: Weighted Prediction for VVC High Bit Depth Extension [Y. Yu, H. Yu, Z. Xie, F. Wang, D. Wang (OPPO)]

VVC high bit-depth extension is expected to support video formats with bit-depth higher than 10 bits. The weighted prediction in the current VVC specification is limited with less precision cannot handle video well if the bit depth is more than 10. This contribution proposes a weighted prediction method. If the bit depth of input video is not more than 10, weighted prediction will follow current VVC. If the bit depth of input video is more than 10, weighted prediction will follow the way defined at HEVC range extension.

The proposal proposes a similar method as in HEVC, increasing the precision of weighted prediction parameters for higher bit depth. Currently in VVC, they are always coded as 8 bits, and then aligned with the video’s bit depth.

Evidence should be brought that the increased precision provides benefit in the context of VVC. When it was introduced in HEVC, some tool had been used to generate fading sequences for such a purpose.

Further study.

[JVET-V0131](https://jvet-experts.org/doc_end_user/current_document.php?id=10792) AHG8: Fixing the forward transform matrices for high bit depth coding [K. Naser, F. Galpin, F. F. Le Léannec, P. De Lagrange (InterDigital)]

In High Bit-depth coding of VVC, the forward transform matrices are represented with higher precision than the standard coding. This modification is reported to alter the properties of DST-VII and DCT-VIII 4x4 transforms that leads to difference between matrix multiplication- and fast algorithm-based computation. This contribution proposes modifying these transform matrices such that both matrix multiplication and fast algorithm are identical and therefore restore the dual-implementation property.

The following experimental results of bdrate gain are reported:

Low QP

* 0.00 %, 0.00 % and -0.01 % in bdrateY in PQ, HLG and SVT in AI
* -0.01 %, -0.01 % and 0.00 % in bdrateY in PQ, HLG and SVT in LDB
* X.XX %, X.XX % and X.XX % in bdrateY in PQ, HLG and SVT in RA

Normal QP

* -0.07 % and 0.00 % in bdrateY in H1 and H2 in AI
* - 0.04 % and X.XX % in bdrateY in H1 and H2 in RA

Presented in session 13 at 1405 UTC (chaired by JRO)

This does not have any impact on the spec, as the inverse transform does not require increased precision. It is however useful to have it in the software, as a reference for implementers.

Decision (SW): Adopt JVET-V0131

[JVET-V0170](https://jvet-experts.org/doc_end_user/current_document.php?id=10833) Crosscheck of JVET-V0131 (AHG8: Fixing the forward transform matrices for high bit depth coding) [H.-J. Jhu (Kwai Inc.)] [late]

[JVET-V0133](https://jvet-experts.org/doc_end_user/current_document.php?id=10794) AHG8: Content Adaptive Transform Precision for High Bit Depth Coding [[K. Naser](mailto:karam.naser@interdigital.com), F. Galpin, F. Le Léannec, P. De Lagrange (InterDigital)]

The intermediate transform coefficient precision is fixed to 16 bit signed representation in VVC. Therefore, a pre-computed down-shift is performed after inverse quantization and inverse transform in order to avoid exceeding the 16 bit buffer. It is argued that this down-shift is too conservative as the prediction mechanism results in low magnitude residuals. For High Bit-Depth coding, the down-shift is increased to ensure 16 bit representation, which further decreases the quality of the inverse transformed coefficients. This contribution proposes to adaptively assign the down-shift according the contents.

The following experimental results of bdrate gain are reported, for CATP0, Low QP:

* 0.00 %, -0.12 % and -0.06 % in bdrateY in PQ, HLG and SVT in AI
* -0.22 %, -0.18 % and -0.08 % in bdrateY in PQ, HLG and SVT in LDB
* X.XX %, X.XX % and X.XX % in bdrateY in PQ, HLG and SVT in RA

Normal QP

* -0.04 % and 0.00 % in bdrateY in H1 and H2 in AI
* - 0.17 % and X.XX % in bdrateY in H1 and H2 in RA

Additional results Low QP, SVT 16 with *TrDynamicRange = bitDepth*

* -0.41 % for SVT16 in AI

CATP1: Low QP

* -0.50 %, -0.20 % and -0.11 % in bdrateY in PQ, HLG and SVT in AI
* -0.39 %, -0.28 % and -0.13 % in bdrateY in PQ, HLG and SVT in LDB
* X.XX %, X.XX % and X.XX % in bdrateY in PQ, HLG and SVT in RA

Normal QP

* -0.05 % and 0.00 % in bdrateY in H1 and H2 in AI
* - 0.15 % and X.XX % in bdrateY in H1 and H2 in RA

Additional results Low QP, SVT 16 with *TrDynamicRange = bitDepth*

* -1.50 % for SVT16 in AI

Presented in session 13 1415-1435 UTC (chaired by JRO)

The approach is to analyse the maximum value of transform coefficient in current block, and determine the amount of shift. This keeps the buffer size lower, and also less precision in multiplier.

Questions:

* Would the gain still appear with precision extension flag enabled? Likely not.
* Was it tested with the CE base anchor? No.

Implementing precision extension flag increases complexity and requires some re-design. The same is true for this proposal, but more investigation would be necessary to assess if it is more implementation friendly.

Study in CE.

[JVET-V0166](https://jvet-experts.org/doc_end_user/current_document.php?id=10828) Cross-check report on AHG8: Content Adaptive Transform Precision for High Bit Depth Coding (JVET-V0133) [D. Rusanovskyy (Qualcomm)] [late] [miss]

[JVET-V0178](https://jvet-experts.org/doc_end_user/current_document.php?id=10841) AHG8: a combination of JVET-V0059 option 2 and JVET-V0122 for high bit depth and high bit rate extensions [F. Wang, Z. Xie, Y. Yu, H. Yu, D. Wang (OPPO)] [late]

In JVET-V0059, a method of using CABAC bypass alignment and a bypass coding of sb\_coded\_flag is proposed as option 2 to improve the throughput. It is asserted that with the proposed method, alignment is needed only once after the limit of the context coded bins has been reached for the TB. In JVET-V0122, a full-bypass mode in residual coding is proposed to improve the throughput and software encoding/decoding speed. It is asserted that with the proposed method, the coefficient or residual levels are only coded through bypass mode of the arithmetic coding engine without using any context-coded bin. It is observed that the simulation results showed over 15% reduction in average encoding and decoding time, even with a slight coding efficiency gain, from the 16-bit SVT sequences.

This contribution proposes a combination of JVET-V0059 option 2 and JVET-V0122. It is asserted that with the proposed method, all the syntax elements in residual coding, except for last significant coefficient position in RRC, are coded through bypass mode and alignment is needed only once after the last significant coefficient position in RRC and at the very beginning of a TB in TSRC. It is suggested to apply the proposed method in 16 bit high throughput configuration.

Following results for SVT 16-bit sequences are reported. It is believed that the proposed method in this contribution can truly enable decoding of all the coefficients in a TB simply by using a shift register and multiple coded bins can be decoded in parallel.

As compared to CE anchor (low QP):

* AI: 0.49 % (G), 0.45 % (B), 0.43 % (R), 86 % (EncT), 77 % (DecT)
* LDB: x.xx % (G), x.xx % (B), x.xx % (R), xx % (EncT), xx % (DecT)
* RA: x.xx % (G), x.xx % (B), x.xx % (R), xx % (EncT), xx % (DecT)

For reference, below are the results reported in JVET-V0059 option 2:

* AI: 0.47 % (G), 0.49 % (B), 0.49 % (R), 99 % (EncT), 94 % (DecT)
* LDB: 0.55 % (G), 0.54 % (B), 0.54 % (R), 100 % (EncT), 94 % (DecT)
* RA: 0.57 % (G), 0.56 % (B), 0.56 % (R), 100 % (EncT), 94 % (DecT)

The full simulation results are included in the following of this document.

Was presented Tuesday 27 April 2100-2130 UTC (chaired by JRO and GJS)

For AI coding of HDR in 12 bit, there is roughly 6.8% and 4.5% loss for PQ and HLG, respectively.

It is commented that SVT16 is not representative for typical high bit depth sequences, as these sequences are rather noisy (as an example, it is pointed to the analysis in JVET-U0069). As a consequence, the number of bypass coded bins is significantly higher than the number of context coded bins.

Further study is recommended. However, the problem still needs to be better understood. Throughput problems could also be caused by a very high number of bypass coded bins.

Study in CE. Also use HEVC high throughput profile as an additional reference point, and also analyse the throughput compared to that.

## AHG11: Neural network-based video coding (19)

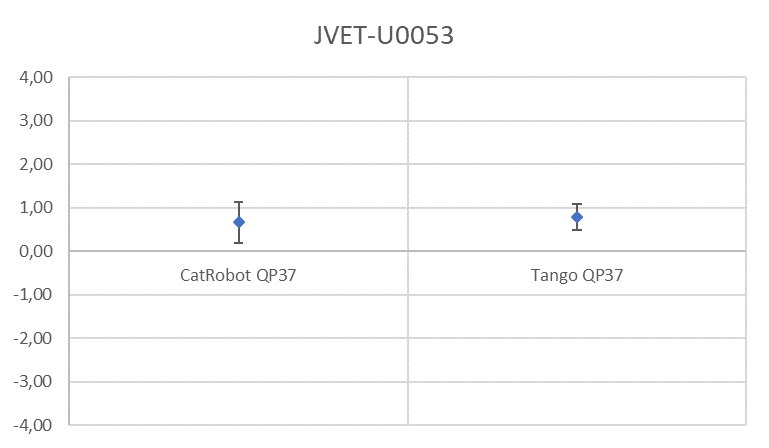
### General (1)

Contributions in this area were discussed in session 20 at 2330–0010 UTC on Monday 26 April 2021 (chaired by JRO and GJS).

[JVET-V0173](https://jvet-experts.org/doc_end_user/current_document.php?id=10836) EE1-related: Report on results of remote viewing session [M. Wien (RWTH Aachen University)]

Remote expert viewing tests were conducted during the 22nd 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, 19 experts volunteered in participating in the tests. The testing procedure follows an online variant of the method used for visual testing during the 13th and 14th JVET meetings. The test revealed that a revision of the test configurations might be indicated in order to align the duration of the video sequences in the visual assessment. A duration of at least 5 s or more was suggested. Furthermore, the presentation of HD video sequences was recommended in order to enlarge the number of JVET experts qualified to participate in the viewing sessions.

The results of all but one participant were included in the evaluation, no further processing of the data was applied. Some participants reported problems with the presentation of the sequences for JVET-U0061. Therefore, the presented results for this proposal may considered not to be conclusive.



QP should be 37 here

The results reveal performance improvements for the proposals compared to the VTM anchor. For the superresolution approach, a more significant improvement was observed. Participants found the very short LD sequences difficult to watch. They also found it difficult to cope with the different sequence lengths and the fact that one sequence was of full length and shown only once. It is noted that the EE configuration has been previously used for objective evaluation only and that the suitability for subjective evaluation was not taken into account.

In general, it is suggested to align the length of the sequences to be assessed visually. A sequence length of 5 – 10 s would be more suitable for subjective evaluation. It is further suggested to consider a lower resolution of the test sequences used in the subjective evaluation since the number of remote participants with access to HD displays may be higher than the number of available UHD displays. Therefore, HD sequences might be considered for this task. If UHD characteristics should be maintained, cropping an appropriate region from the full resolution could be used.

U0053, U0096, U0099 are NN based superresolution approaches, plus VTM11.RPR2 as conventional Down/upsampling, From the current results, conventional method was judged to be better.

U0061, U0104, U0115 are NN based loop filtering.

It can be concluded that the test methodology points in the correct direction, and initially shows some visual benefit. Needs to be extended to a wider range of QPs, and longer sequences, and better matching of the rates.

To give more experts a chance to participate, cropping into HD resolution might be considered, or viewing with class B.

To be further refined in preparation of EE1 description during session 22 Tue 1520

* Define better testing condition for SR
* Structure of sub-EEs (see meeting notes which contributions)
* Preparation of next viewing

### EE1 contributions: Neural network-based video coding (7)

Contributions in this area were discussed in sessions 7 and 8 at 2100–0040 UTC on Wednesday/Thursday 21/22 April 2021 (chaired by JRO and GJS).

[JVET-V0023](https://jvet-experts.org/doc_end_user/current_document.php?id=10829) EE Summary Report: Neural Network-based Video Coding [E. Alshina, S. Liu, W. Chen, Y. Li, R.-L. Liao, Z. Ma, H. Wang]

This contribution provides a summary report for the Exploration Experiments on Neural Network-based Video Coding.In total 6 tests have been completed within this EE between the JVET U and V meetings to study and evaluate NNVC technologies, analyze their performance and complexity aspects.

NN CTC anchor changed at JVET-U meeting (GOP size enlarged and MCTF enabled). It was anticipated that gain on NN-based filters and MCTF can overlap. But EE results how that this is not the case. Same EE proponents report even higher compression gain.

There are 2 categories of tests in this EE: in-loop filetrs and superresolution-like technologies. Algorithms with different level of complexity were demonstrated in both categories.

First attempts to cross-check test presented in EE, by not just running code and checking performance, but also verifying complexity assessment were taken. Several aspects which are not absolutely clear in NN CTC (for example, MAC/pixel computation) have been clarified trhough cross-check activity.

As part of EE study prior to JVET-V meeting viewing was prepared in coordination with AG05. It was noticed that the length of 4K sequences in NN CTC is 5 sec in RA and only 3 sec in LD-B/P tests, and it was asserted this is too short for visual tests.

All proponents tested their proposals under NN CTC, unfortunately not all use official results reporting template.

**Exploration Experiment on NN-based in-loop filters**

Proposals in this category were tested compared to VTM11.0, QP=22,..., 42. High level summary of results is shown in Table 1.

Table 1. Test results in NN-based in-loop filters category.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **#** | **Performance** | **Dec.T.** | **Features** | **#params** | **kMAC /pxl** |
| [JVET-V0114](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10763) | RA:**7.2**%(Y); 13% (Ch) RA:**7.5**%(Y); 14% (Ch) | ×137 ×146  (CPU) | Input 10 planes: 4\*Y+U+V+QP+3\*Bsinfo, 72×72, #channels K=96 Residual scalling (per color, per picture); Supplement: Refining training improves performance | 4×1M | 319 |
| [JVET-V0115](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10764) | RA:**8.3**%(Y); 20% (Ch) RA:**8.5**%(Y); 22% (Ch) | ×337 ×344  (CPU) | Same, as JVET-V0114, but with LongActivation = CONV1×1,K×M + LeakyReLu+CONV1×1,M×K ; M=216> K=72 Supplement: Refining training improves performance | 8×1M | 624 |
| [JVET-V0137](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10798) | RA:**1.2**%(Y); 4% (Ch)  AI:**1.6**%(Y); 4%(Ch) | ×30  ×34  (CPU) | Input YUV444, 4 planes + QP map, 64×64, DepthwiseSeparableConvolutions (DSC) in 4 Residual Blocks | 23K | 23 |

For the proposal which reports BD-rate computed based on those 2 quality metrics (both PSNR and MS-SSIM) performance difference was close. Interesting model refining hint, which provides ~0.3% compression performance improvement (likely to be useful for other proponents), was demonstrated by Qualcomm.

For the simplicity of comparison NN structures for all 3 tests in this category are shown on Fig. 1-3.



Fig. 1. "NN-based de-block" from [JVET-V0114](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10763). 

Fig. 2. "NN-based de-block" with long activation function from JVET-V0115.

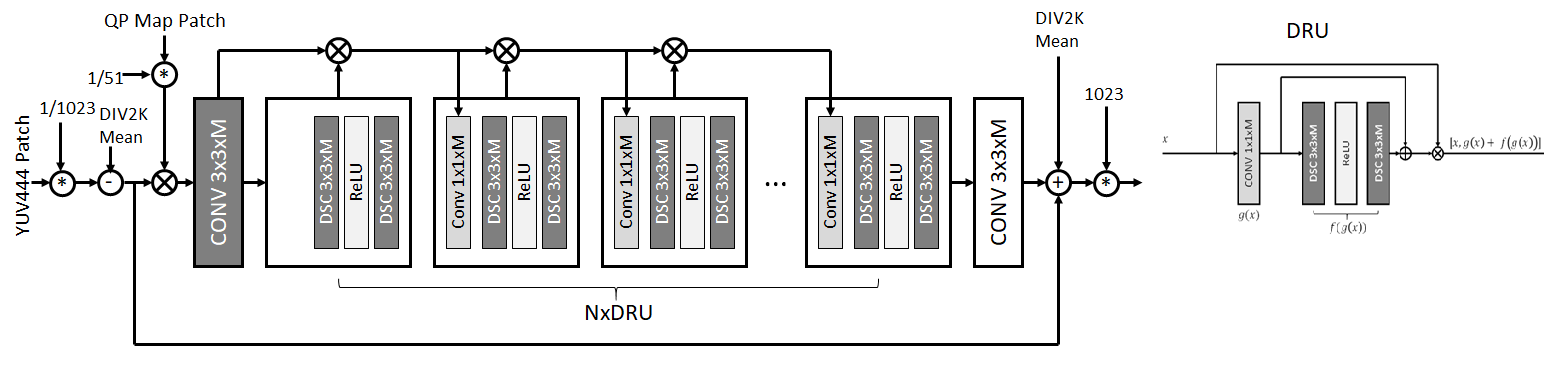


Fig. 3. Low-complexity NN-based in-loop filter with DSC from [JVET-V0137](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10798).

For the training proponents used DIV2K [2] and BVI-DVC [3].

**Exploration Experiment on superresolution-like techniques**

Proposals in this category were tested compared to VTM11.0, QP=27,..., 47. High level summary of results is shown in Table 2.

Table 2. Test results of superresolution-like technologies in EE1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **#** | **Performance** | **Dec.T.** | **Features** | **#params** | **kMAC /pxl** |
| [JVET-V0073](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10721) | RA(4K):**2.5**%(Y), ~ 30% drop (Ch) RA(4K):**5.8**%(Y), ~ 20% drop (Ch) RA(4K):**6.5**%(Y), ~ 10% drop (Ch) | ×0.3 ×1.5 ×1.5  (×90 CPU) | Test0: RPR2 (↓VVC↑­)  Test1: RPR2 (↓VVC↑­) CNN with const.param Test2: RPR2 (↓VVC↑­) CNN **signalled** param NCTM5.0 used for CNN param compression; Simplified [ESRGAN](https://arxiv.org/pdf/1809.00219.pdf) | no info (483.6K) | no info (122) |
| [JVET-V0096](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10744) | RA(4K):**8.1**%(Y), ~20% drop (Ch) AI(4K): **11**% (Y), ~24% drop (Ch) | ×0.3 ×0.4  (×90 CPU) | RPR2 (↓VVC­↑) YUV444 CNN with const.param Simplified [EDSR](https://openaccess.thecvf.com/content_cvpr_2017_workshops/w12/papers/Lim_Enhanced_Deep_Residual_CVPR_2017_paper.pdf) | 2.7M (2×1.3M) | 344 |
| [JVET-V0149](https://jvet-experts.org/doc_end_user/current_document.php?id=10810) | RA(4K):**9.6**%(Y), ~3% drop (Ch) LD(4K):**18**%(Y), ~25% drop (Ch) | no info | DCS-based Video Coding:  Intra = spatial texture frames are coded in original resolution, Inter = temporal motion frames frames are coded ×2 ↓, synthesis with Motion Compensation & Texture Transfer Networks | 8.5M  (2×4.25M) | no info |

“Test 0” above is VVC using RPR, no NN approach.

“drop” means loss (bit rate increase or PSNR decrease) in chroma.

The first two proposals only tested 4K, the third also includes classes B and E, where also gain was achieved.

Comments in brackets () show estimation by analysis performed by Franck Galpin (kMAC/pxl) and Elena Alshina (Dec. Run Time). Franck Galpin used automatic flops (~MAC) estimation procedure [1], scripts were shared to porponents with explanation on usage. CPU Decoding Run time was estimated, using logic described in APPENDIX.

Technologies in this category typically provide compression gain (measured in Y-PSNR) for 4K videos in NN CTC (classes A1, A2, H2). Authors of [JVET-V0149](https://jvet-experts.org/doc_end_user/current_document.php?id=10810) have managed to achieve gain [for](https://jvet-experts.org/doc_end_user/current_document.php?id=10810) classes B and E as well. For most of the cases Luma gain is accomplished by loss in Chroma.

By just encoding 4K content from NN CTC “in quarter resolution” (donsampling ×2, coding with VVC, upsamling with VVC filetrs designed for RPR2) one gets 2.5% gain. By changing re-sampling RPR2 filters to NN-based off-line trained ones (which also operate as enhancement filters), compression gain in a range 6% ... 8% ([JVET-V0073](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10721), [JVET-V0096](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10744)) can be achieved. If additionally NN-filter parameters are tunned for concrete video sequences and included to the bit-stream ([JVET-V0073](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10721)) additionally ~1% compression gain can be achieved.

If Intra frames are coded in full resolution, Inter-frames are encoded in quarter resolution and Motion Compensaiton Network with Texture Transfer Network are used for full resolution reconstruction synthesis ([JVET-V0149](https://jvet-experts.org/doc_end_user/current_document.php?id=10810)) gain of almost 10% (in average for 4K content in RA) can be achieved.

One cross-check was performed for proposals in this category([JVET-V0148](https://jvet-experts.org/doc_end_user/documents/22_Teleconference/wg11/JVET-V0148-v1.zip)).

Proponents of [JVET-V0096](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10744) reported that they used BVI-DVC [2] to fine tune [EDSR](https://openaccess.thecvf.com/content_cvpr_2017_workshops/w12/papers/Lim_Enhanced_Deep_Residual_CVPR_2017_paper.pdf) which was originaly trained on DVI2K [3].

This category can be classified as post processing.

**Viewing for test materials in EE1**

Candidates for viewing were suggested by EE proponents, discussed during AhG11 telco and made available via JVET ftp. All proponents selected 4K sequences to illustrate benefits of their proposals.

It was noticed that 4K sequences in NN CTC are coded as 5 sec video in RA configuration (Tango is even shorter). Low-delay configuration in NN CTC is optional for 4K sequences and only 3 sec of 4K video are encoded. Short length of video sequences definitely not sufficient for the comprehensive viewing. 5 sec can be used for viewing, 3 sec as asserted by AG05 is too short. It might be reasonable to equalize number of encoded frames in RA and LD configurations for 4K videos in NN CTC.

Bit-rate difference between anchor (VTM) and proposal is shown in Table 3. In NN-based in-loop filter category the difference is within 2%. In super-resolution cathegory difference is up to 10%. Note that for files name on ftp use JVET-U documents numbers were used.

Table. 3. Sequences selected for viewing.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| VTM11.0 | QP | kbps | Proposal # | QP | kbps | diff. % |
| Campfire RA QP37 | 37 | **3492.338** | **JVET-U0061 (Tencent)** | 37 | **3509.302** | 0% |
| Campfire RA QP42 | 42 | **1643.751** | [JVET-V0137](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10798) | 42 | **1652.949** | 1% |
|  |  |  |  |  |  |  |
| Campfire RA | 42 | **1643.751** | **JVET-U0104(Qualcomm)** | 42 | **1645.79** | 0% |
| DayLightRoad\_LB | 42 | **1092.04** | [JVET-V0115](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10764) | 42 | **1086.296** | -1% |
| FoodMarket\_LB | 42 | **1009.283** |  | 42 | **985.357** | -2% |
|  |  |  |  |  |  |  |
| Campfire RA | 42 | **1643.751** | **JVET-U0115 (Qualcomm)** | 42 | **1640.715** | 0% |
| DayLightRoad\_LB | 42 | **1092.04** |  | 42 | **1081.968** | -1% |
| FoodMarket\_LB | 42 | **1009.283** | [JVET-V0114](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10763) | 42 | **984.477** | -2% |
|  |  |  |  |  |  |  |
| Tango RA | 44 | **898.2073** | **JVET-U0053 (Sharp)** | 37 | **924.7612** | 3% |
| CatRobot RA | 44 | **922.8992** | [JVET-V0073](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10721) | 37 | **986.4184** | 7% |
|  |  |  |  |  |  |  |
| Campfire RA | 44 | **1210.127** | **JVET-U0099 (Qualcomm)** | 42 | **1331.5** | 10% |
| FoodMarket4 RA | 44 | **851.0048** | [JVET-V0096](file:///C:\Users\e00443164\Downloads\current_document.php%3fid=10744) | 42 | **921.43** | 8% |
| Tango RA | 44 | **898.2073** | 42 | **925.31** | 3% |
|  |  |  |  |  |  |  |
| Campfire RA | 44 | **1210.127** | **VTM11.0 RPR2** | 37 | **1323.618** | 9% |
| FoodMarket4 RA | 44 | **851.0048** |  | 37 | **915.4704** | 8% |
| Tango RA | 44 | **898.2073** | 37 | **922.3788** | 3% |
| CatRobot RA | 44 | **922.8992** | 37 | **983.9776** | 7% |
|  | | | | | | |
| Tango2 RA | 44 | **898.2073** | **JVET-U0096 (NJU)** | 37 | **970.75** | 8% |
| Campfire RA | 43 | **1417.397** | [JVET-V0149](https://jvet-experts.org/doc_end_user/current_document.php?id=10810) | 37 | **1378.44** | -3% |
| DaylightRoad2 RA | 41 | **1240.958** | 47 | **1123.56** | -9% |

Viewing is supposed to be conducted during JVET-V meeting.

Sequences prepared for the viewing are available on <ftp://jvet-ul1@ftp.ient.rwth-aachen.de/JVET-V_EE-DNN/>.

Mathias Wien proposes to perform viewing also for the short sequences (3 sec) with an appropriate arrangement (e.g. looping). He also points out that for some of the sequences the first seconds are not the most suitable ones for viewing.

Viewing could be perfomed Thu during session 10 (1520 UTC).

From the discussion:

* One expert points out that it would be interesting to collect information what are reasonable sizes of networks, including the capability of mobile devices.
* It is also pointed out that the requirements of single images and video are quite different in terms of realtime processing capability and power consumption.
* One expert mentions that a processing power of 1 kMAC per pixel is currently possible with high-end mobile devices (8 bit integer precision).

[JVET-V0073](https://jvet-experts.org/doc_end_user/current_document.php?id=10721) EE1.2: Additional experimental results of NN-based super resolution (JVET-U0053) [T. Chujoh, T. Ikai (Sharp)]

In this contribution, several additional experimental results of EE1.2 NN-based super-resolution (JVET-U0053) are reported. JVET-U0053 was based on a framework for introducing a super-resolution post-filter using CNN (Neural Network). 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, sending partial network parameters as network updated by each sequence and QP are tested. As experimental results, by sending partial network parameters, the coding efficiency has been improved obviously compared to the fixed network.

Presentation deck to be provided.The gains are not homogeneous over all test sequences, some sequences may not be so suitable for superresolution upsampling. However, for the NN approaches, it is more homogeneous than for RPR.

In test 2, only the last layer is updated (once per sequence). Basically, this introduces a delay of the whole sequence. Might better be done at least once per RA point. MPEG NN compression is used for parameters.

Large losses in chroma (also in RPR) – would it be possible to make it more balanced?

It is pointed out that the superresolution method may not be suitable for all QP points. Actually it may have losses at lower QP points. If RD plots are crossing each other, this might make the computation of BD numbers meaningless.

[JVET-V0096](https://jvet-experts.org/doc_end_user/current_document.php?id=10744) EE1-2.3: Neural Network-based Super Resolution [A. M. Kotra, K. Reuzé, J. Chen, H. Wang, M. Karczewicz, J. Li (Qualcomm)]

This contribution reports the EE test results of JVET-U0099. JVET-U0099 studied the performance of applying a Neural-Network based super-resolution used as upsampling filter in the context of VVC RPR. Prior to encoding, a given picture is downsampled by a factor of 2x using the inbuilt RPR mechanism of VTM11. PSNR of the coded frame is computed by calculating the MSE between the original picture and the upsampled version of the decoded picture. The upsampled picture is generated by the Neural Network-based upsampling filter instead of the existing VTM up-sampling filter.

The objective results over VTM-11 with Anchor QPs as 27, 32, 37, 42, 47 (as suggested by NN-based super-resolution category of JVET-U2023) are as follows:

On average, for Class A1 sequences, Luma BD-Rate gains of 12.47% and 11.96% for RA and AI configurations respectively were observed.

It is pointed out that the chroma loss might be caused by the multiple down- and upsampling (for the network input, chroma is upsampled to luma solution). A possible approach might be to not downsample chroma, but code as 444 instead (which is however not supported by VVC’s RPR).

Could the performance be improved when using three models instead of two? The proponent says this might likely be the case.

Similar to the comments made for the JVET-V0073, the test conditions might be re-considered, e.g. only applying super resolution where it is meaningful (both for conventional and NN based method). RPR should be used as anchor.

Reporting template needs to be fixed to produce RD plots.

[JVET-V0148](https://jvet-experts.org/doc_end_user/current_document.php?id=10809) Crosscheck of JVET-V0096 (EE1-2.3: Neural Network-based Super Resolution) [T. Ikai, T. Chujoh (Sharp)] [late]

[JVET-V0114](https://jvet-experts.org/doc_end_user/current_document.php?id=10763) EE1-1.3: Test on Neural Network-based In-Loop Filter with No Deblocking Filtering stage [H. Wang, J. Chen, A. M. Kotra, K. Reuzé, M. Karczewicz (Qualcomm)]

This contribution reports the EE test results of JVET-U0115. In the proposed Neural Network-based method, 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 information to the NN filter but the actual deblocking filtering process is bypassed. Simulation results reportedly show 7.24 %, 13.54 % and 13.13 % BD rate saving for RA and 6.64 %, 11.67 %, 12.84 % BD rate saving for AI, for Y, Cb and Cr components respectively.

QP is also fed into the network, so the operation would be different for different temporal layers.

4 different models are used (two for intra and two for inter), model selected at picture level, on/off at CTU level.

Processing per CTU, input is a 144x144 block (overlapping 8 luma samples at each side into neighbors), luma is split into four planes of 72x72 each, output is cropped to 64x64.

Compared to the last meeting’s proposal, training was refined with larger training set.

Compared to an approach with separate deblocking, the complexity is increased due to the extended block size.

Visual performance would also need to be investigated to assess if the combination with deblocking makes sense, or if it is better to keep a separate deblocking (where the latter is less complex).

[JVET-V0115](https://jvet-experts.org/doc_end_user/current_document.php?id=10764) EE1-1.4: Test on Neural Network-based In-Loop Filter with Large Activation Layer [H. Wang, J. Chen, A. M. Kotra, K. Reuzé, M. Karczewicz (Qualcomm)]

This contribution reports the EE test results of combining JVET-U0104 and JVET-U0115. In the proposed Neural Network-based method, one NN-based filter can be selected out of four candidates as an in-loop filter for each color channel type (luma/chroma) of a 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. Simulation results reportedly show 8.23 %, 19.88 % and 20.41 % BD rate saving for RA and 7.46 %, 17.94 %, 19.72 % BD rate saving for AI, for Y, Cb and Cr components respectively.

[JVET-V0137](https://jvet-experts.org/doc_end_user/current_document.php?id=10798) EE1-1.1: neural network based in-loop filter using depthwise separable convolution and regular convolution (JVET-U0061) [Z. Li, C. Auyeung, X. Xu, W. Wang, X. Li, S. Liu (Tencent)] [late]

This contribution presents the updated result of the neural-network-based in loop filter using depthwise separable convolution (DSC) in EE1.1 with new test condition that GOP=32, and QP= {22,27,32,37,42}. With VTM-11.0 as anchor, on average results from RA/AI configurations, report gains for luma BD-Rate are 1.15% and 1.61% respectively.

Presentation deck to be provided.

Filter is a new stage between deblocking and SAO.

QP information is input to the network.

Only one model

No local (e.g. CTU based) control.

[JVET-V0149](https://jvet-experts.org/doc_end_user/current_document.php?id=10810) EE1: Tests on Decomposition, Compression and Synthesis (DCS)-based Technology (JVET-U0096) [M. Lu, Z. Ma (NJU), Z. Dai, D. Wang (OPPO)] [late]

This contribution reports the EE test results of JVET-V0149 with Anchor as VVC. The pipeline of the proposal is shown in Figure 1 below.

图示

描述已自动生成

Figure 1. Pipeline of the proposed DCS-based Video Coding: ↓ (in red) is for down-sampling, and ↑ (in red) is for up-sampling; E, D and DPB represent video encoder, decoder and decoded picture buffer respectively.

**Decomposition**: A raw video can be decomposed as a combined representation of spatial texture and temporal motion, having the spatial texture frames (STFs) and temporal motion frames (TMFs) respectively. We downscale the TMFs to a lower spatial resolution, while keeping the STFs the same native frame size. Ideally, we wish to use STFs to preserve the spatial texture details, and have TMFs to well capture the temporal motion smoothness.

**Compression**: To support the general compatibility with existing video standards, the STF is enforced as the intra frame of a specific group of pictures (GoP), while the rest frames in this GoP are TMFs. We choose to code STFs and low-resolution TMFs using the mainstream VVC model in this work. Other standards, e.g., HEVC, or emerging DNN-based video coding can be applied as well. The STFs are VVC intra coded, reconstructed, and downscaled to the same size as the TMF to input into the DPB for prediction reference in Fig. 1.

**Synthesis**: In a GoP, decoded STF and TMFs are used to produce high-fidelity video at its native spatiotemporal resolution. This is a super resolution based synthesis problem, aiming for the high-quality restoration to alleviate both compression and re-sampling noises for preserving both spatial details and temporal smoothness. Towards this purpose, a motion compensation network (MCN) is first utilized to generate temporally smooth and spatially fine-grained motion representation of current TMF by aggregating the information across neighboring TMFs. This MCN is comprised of a deformable convolution network-based multi-frame motion feature alignment, and a separable temporal-spatial attention-based motion feature aggregation. Then, temporal motion features generated by MCN, together with decoded TMFs, decoded STF as well as its re-sampled version, are fed into a texture transfer network (TTN) to learn and transfer cross-resolution information for high-fidelity frame restoration with both spatial details and temporal smoothness, where the TTN has utilized features to guide accurate refinement for final reconstruction.

Presentation deck to be provided.

One model for each QP value 22,27,32,37,42, and separate luma/chroma (10 models in total).

Is it necessary to have this high number of models? Better train for certain QP ranges, or feed QP into the network?

Generally, more gain is observed at higher QPs.

Similar comments apply as for other SR proposals.

### EE1 related contributions: Neural network-based video coding (0)

Contributions in this area were discussed in session X at XXXX–XXXX UTC on XXday 2X April 2021 (chaired by XXX).

### Tools in “hybrid” architectures (9)

Contributions in this area were discussed in session 8 at 0045–0125 UTC on Thursday 22 April 2021 (chaired by JRO and GJS), and in sessions 13/14 at 1300-1400, 1440-1510 and 1530-1620 UTC on Friday 23 April 2021 (chaired by JRO).

[JVET-V0074](https://jvet-experts.org/doc_end_user/current_document.php?id=10722) AHG11: Separate density attention network for loop filtering [[Z. Wang](mailto:baixiu.wz@alibaba-inc.com), C. Ma, R.-L. Liao, Y. Ye (Alibaba)]

To achieve better trade-off between the model performance and the complexity, this contribution presents a separate density attention network (SDAN) for loop filtering. The SDAN has less than 1M parameters, a reduction of 70% compared to the multi-density network (MDN) presented in JVET-U0054. The experimental results show that our method can achieve 5.52%, 14.96% and 14.67% coding gain over VVC for Y, U and V components under RA configuration.

Presentation deck to be uploaded.

Network located between deblocking and SAO

Block level on/off

5 models for different QP ranges

Scaling at frame level, provides approx. 0.5% gain.

Decoding time (CPU) roughly 100x on 16-CPU

[JVET-V0075](https://jvet-experts.org/doc_end_user/current_document.php?id=10723) AHG11: Content-adaptive neural network post-processing filter [Y. Lam, M. Santamaria, J. Lainema, F. Cricri, R. Ghaznavi-Youvalari, A. Zare, H. Zhang, H. R. Tavakoli, M. M. Hannuksela (Nokia)]

This document presents a content-adaptive post-processing filter based on Convolutional Neural Networks (CNNs). The filter is first trained offline on a large dataset, and then finetuned on each test video sequence. The filter is applied at CTU level and two flags are used to signal its usage for luma and chroma. The proposed filter was implemented on top of VTM 11.0 and evaluated under Random Access (RA) configuration. The results showed that our proposed content-adaptive method outperforms both VTM 11.0 and VTM 11.0 augmented with a pretrained post-processing neural network (NN) filter.

Presentation deck to be uploaded.

Pretrained network with finetuning adaptation by updating the bias term of the network nodes

* At which level? Sequence level
* How many parameters? 323 bias terms

Switch (separate for luma and chroma) at CTU level

BD rate gain is 2-3% for classes C and D. Out-of-band signalling, i.e. rate for the bias term is not considered

It was suggested that it might be interesting to consider updates of parameters at finer granularity (e.g. picture, GOP). Would the gain be retained?

[JVET-V0076](https://jvet-experts.org/doc_end_user/current_document.php?id=10724) AHG11: Deep-learning based inter prediction blending [F. Galpin, P. Bordes, T. Dumas, A. Robert, P. Nikitin, F. Le Léannec (InterDigital)]

This contribution proposes a Neural Network-based Inter prediction blending. The proposed NN replaces the regular blending of bi-prediction of motion compensated blocks. Several networks with different complexity are proposed. All networks use a number of MACs per pixel between 11 and 17 kMACs.

It is reported that averagely 0.99%, 0.34% and 0.31% BD rate saving for RA for Y, Cb, and Cr components respectively, are achieved for the most complex network.

Networks with 5 or 6 layers. One single network used for all block sizes (8x8 … 128x128)

Concatenation: input and output of network are mixed for final prediction

“Fast version” does not apply to prediction from certain modes such as affine, CIIP, BCW

BDOF is not needed (as it intends to perform a similar refinement) and therefore disabled

SATD as cost function in training (gave 0.2%-0.3% better performance)

More gain for lower resolution classes such as C,D

Approx 17 kMAC/pixel

Network parameters quantized to 16 bit integer (important for avoiding mismatch in prediction loop)

It is pointed out that it might be useful to specifically optimize for different POC distances.

It was suggested to study in EE. The proponents later expressed that they would not be able to contribute to the EE.

[JVET-V0090](https://jvet-experts.org/doc_end_user/current_document.php?id=10738) AHG11: Neural network based temporal processing [B. Choi, Z. Li, W. Wang, W. Jiang, X. Xu, S. Liu (Tencent)]

This informative proposal reports the test results of a neural network (NN) based temporal filtering for both detail enhancement and inter-prediction. The proposed NN processing is composed of two steps, NN-based reconstruction process and NN-based prediction process. Both processes has the same NN model utilizing temporal features, but the parameters of each model was individually trained for improving the quality of the reconstruction or minimizing error of the motion compensated prediction. The test results show coding gains 3.14%, 4.69% and 6.87% respectively, for classes B, C & D in luma for RA.

No class A results yet.

Two stages of NN, each with temporal processing invoking additional reference pictures (using past and future pictures):

* Reconstruction/loop filter (before DPB)
* Prediction filter (after DPB, before MC)

The proponent asserts that this is mainly effective in case of small motion, but might have problems with large motion.

Would it make sense to use temporal processing only for the second stage, and employ another filter as loop filter? What are the gains of stage 1 and stage 2 individually? Has only roughly been investigated so far. Is there overlap in the gain of the two filters? Yes.

The current approach replaces one of the reference pictures used for prediction by the temporally processed version. It is pointed out that this might not be optimum.

The benefit of stage 2 is probably more interesting.

Further study encouraged, but would be premature investigating in an EE.

[JVET-V0092](https://jvet-experts.org/doc_end_user/current_document.php?id=10740) AHG11: Replacing SAO in-loop filter with Neural Networks [P. Bordes, F. Galpin, T. Dumas, P. Nikitin (InterDigital)]

This contribution proposes some variants to test performance of the Neural Network-based In-loop filter (NN filter) proposed in [1]. The proposed NN filter replaces the regular SAO filter while re-using some of its features.

It is reported that in average 3.50%, 7.23%, and 6.32% BD rate saving for RA, 3.61%, 12.14%, and 9.35% BD rate saving for LDB and 3.50%, 6.43%, and 6.49% BD rate saving for AI, for Y, Cb, and Cr components respectively, are achieved compared to VTM-11 anchors.

5 models for AI, 9 models for RA/LB. Models depend on QP. Any two Models can be combined in a weighted linear fashion. This is signaled at CTU level. The decision on the combination causes a high increase in encoder run time. 3 bits for the scaling factors.

Using different models for intra and inter gave approx. 1% additional gain.

Parameters are quantized to 16 bit integer. Almost no difference due to quantization.

Replaces SAO, before ALF.

Combines the luma and chroma channels.

Would the gain be additive with other NN based loop filters?

It was suggested to study in EE. The proponents later expressed that they would not be able to contribute to the EE.

[JVET-V0100](https://jvet-experts.org/doc_end_user/current_document.php?id=10748) AHG11: Deep In-Loop Filter with Adaptive Model Selection [Y. Li, L. Zhang, K. 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-U0068. Several changes are further introduced: migration of the algorithm to VTM-11.0, improved network structure, refined model candidates, adaptive inference size. Compared with VTM-11.0, the proposed method reportedly shows on average {8.79%, 22.54%, 22.90%}, {11.79%, 27.62%, and 27.26%}, and {10.85%, 27.60%, and 27.79%} BD-rate reductions for {Y, Cb, Cr}, under AI, RA, and LDB configurations, respectively.

Model selection at CTU level.

Models trained depending on QP, but a model trained for another QP can be used for current block. In AI only on/off.

Training first with L1 loss, after convergence refined with L2 loss.

DIV2K set was extended by variants, such as flipping etc. 128x128 patches in training

Would the gain be retained when parameters are quantized? Would likely go down. Information on this would be highly desirable, as the floating point precision might lead to encoder/decoder mismatch.

The question is raised if (generally, not just for this proposal) the gain would go down for a not fully optimized encoder?

Further study recommended, proponents think it is too early including in EE

[JVET-V0101](https://jvet-experts.org/doc_end_user/current_document.php?id=10749) AHG11: Conditional In-Loop Filter with Parameter Selection [Y. Li, L. Zhang, K. Zhang (Bytedance)]

This contribution presents a conditionalin-loop filtering method. It is developed from the prior contribution JVET-U0068. The proposed CNN-architecture is conditioned on an auxiliary parameter based on QP, leading to a unified model to handle different quality levels. To better capture local characteristics of an image, the proposed technique enables an adaptive selection of the auxiliary parameter at block level and slice level. Compared with VTM-11.0, the proposed method reportedly shows on average {8.65%, 21.33%, 22.09%}, {11.40%, 23.66%, 22.60%}, and {10.85%, 22.42%, 21.23%} BD-rate reductions for {Y, Cb, Cr}, under AI, RA, and LDB configurations, respectively.

In comparison to the contribution V0100, local QP is used in the network. Only one model for different QPs.

Further study recommended, proponents think it is too early including in EE

[JVET-V0105](https://jvet-experts.org/doc_end_user/current_document.php?id=10753) AHG11: neural network-based intra prediction: updated signaling [T. Dumas, F. Galpin, P. Bordes, P. Nikitin (InterDigital)]

This contribution suggests several updates of the signaling of the single additional neural network-based intra prediction mode in VVC introduced in “AHG11: neural network-based intra prediction with transform selection in VVC” [1]. Moreover, it compares the trade-off between rate-distortion performance and encoder/decoder running times variations when the neural networks run inside VTM-11.0 using two different C++ frameworks and with/without neural network quantization.

For VTM-11.0 including the single additional neural network-based intra prediction mode, the parameters of its neural networks being quantized on 16-bit, -3.46%, -3.21%, -3.29% and -1.69%, -0.71%, -1.01% of mean BD-rate reduction are reported in AI and RA for Y, Cb, Cr respectively. The average encoder and decoder running times for AI and RA are 390%, 3560% and 168%, 551% respectively.

By changing position in MPM list, approx. 0.2% gain is achieved over the previous proposal.

Quantization of parameters does not change the compression performance (very small gain in chroma). Run time is decreased due to quantization (on CPU). Beyond run time and implementation friendliness, the avoidance of encoder/decoder mismatch is an important aspect of integerization.

Further study recommended, proponents think it is too early including in EE, perhaps next meeting.

[JVET-V0102](https://jvet-experts.org/doc_end_user/current_document.php?id=10750) AHG11 & AHG12: Deep In-Loop Filter with Adaptive Model Selection for Enhanced Compression Beyond VVC Capability [Y. Li, L. Zhang, K. Zhang (Bytedance)]

See under 5.3.4.

### “End to end” architecture concepts (1)

Contributions in this area were discussed in session 14 at 1630–1705 UTC on Friday 23 April 2021 (chaired by JRO).

[JVET-V0055](https://jvet-experts.org/doc_end_user/current_document.php?id=10703) Improved end to end deep intra frame compression with cross channel context model and loop filter [C. Ma, Z. Wang, R.-L. Liao, Y. Ye (Alibaba)]

Currently, joint autoregressive and hierarchical prior entropy models are used in most end-to-end deep intra frame compression frameworks, where the autoregressive models are used to capture the local contexts from the quantized latent elements and hierarchical models are used to capture the global contexts from the hyper latents. For the local context, 2D mask convolution is widely used to capture the spatial context. However, it is observed that there are strong correlations between different channels in the latents. To utilize the cross channel correlations, a cross channel context model complement with the spatial context model is applied. In addition, considering the important role of loop filter in traditional video codec, loop filter is integrated with the deep intra frame compression framework to further improve performance. The experimental results demonstrate that when combined, the cross channel context model and loop filter methods outperform VTM-11.0 by 1.44% for Y with a coding performance loss of 24.74% and 11.91% for the U and V components, respectively. Over the baseline deep compression framework, the combined method provides 9.10%, 12.27%, and 12.68% performance improvement for Y, U and V, respectively.

Build on top of “Cheng 2020” baseline model (ref. 2 in contribution). Was re-trained with the additional elements.

It is noted that the “loop filter” is actually a post filter here. It is not adaptive, no parameters, no switching.

It is interesting to see an end-to-end approach which seems to be close to VVC intra in terms of performance. Would further be interesting (if possible) to get a better balance between luma and chroma compression. Currently, chroma seems to be worse in PSNR.

Further study recommended. No specific action suggested at this point.

### NN related HLS signalling (1)

Contributions in this area were discussed in session 14 at 1705–1720 UTC on Friday 23 April 2021 (chaired by JRO).

[JVET-V0091](https://jvet-experts.org/doc_end_user/current_document.php?id=10739) AHG9/AHG11: SEI messages for carriage of neural network information for post-filtering [B. Choi, Z. Li, W. Wang, W. Jiang, X. Xu, S. Wenger, S. Liu (Tencent)]

This contribution proposes an updated SEI message design for carriage of a neural network (NN) topology and parameters that are utilized for post filtering with neural network models. In the update design, two SEI messages are proposed; the first SEI specifies the internal or external carriage of a NN topology and its parameters, and the second one specifies the organization of multiple NN models, which are carried by the first SEI messages, for the post-processing. The second SEI message contains the information on which model is applied to each picture or block among multiple candidate models. The proposed high-level syntax structure is applicable to SEI messages for post-filtering as well as other NN-based EE proposals.

Further study recommended to understand

* What is required in terms of information (technically), i.e. information that cannot be generated at the decoder side, but requires knowledge about the original signal
* If there is a chance to be adopted by industry, or if e.g. display manufacturers would develop their own solutions
* Is an SEI message the right place, in particular if the number of parameters is large?
* Relation with MPEG NNR?

## AHG12: Enhanced compression beyond VVC capability (24)

### General (0)

See also JVET-V0108 colour transform information SEI message proposal (with benefit specific to PQ content).

Contributions in this area were discussed in session X at XXXX–XXXX UTC on XXday 2X April 2021 (chaired by XXX).

### EE2 contributions: Enhanced compression beyond VVC capability (4)

Contributions in this area were discussed in session 11 at 2100–2210 UTC on Thursday 22 April 2021 (chaired by JRO and GJS).

[JVET-V0024](https://jvet-experts.org/doc_end_user/current_document.php?id=10825) 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 a base/general modification, intra prediction, inter prediction, transform and coefficient coding, in-loop filtering, entropy coding, and combination tests.

The software basis for this EE was JVET-U0100 based on VMT-10.0, which can be particularly configured to perform tool-on (on top VTM) and overall tests. Configuration files were distributed with the EE software and follow VTM CTC. VTM-11.0 and EE2 U0100 software are used as tool-on and tool-off anchors, respectively.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Tests** | **Tester** | **Cross-checker** |
| **1 Base** | |  |  |
| 1.1 | Non-normative changes introduced by macros and cfg file parameter changes | Qualcomm  JVET-V0120 |  |
| 1.2 | Normative changes introduced by macros | Qualcomm  JVET-V0120 |  |
| 1.3 | Non-normative and normative changes (1.1+1.2) | Qualcomm  JVET-V0120 |  |
| **2 Intra prediction** | |  |  |
| 2.1 | Intra template matching (JVET-U0048) | InterDigital  JVET-V0130 |  |
| 2.2 | Multi-model LM (MMLM) | Qualcomm  JVET-V0120 |  |
| 2.3 | Gradient PDPC | Qualcomm  JVET-V0120 |  |
| 2.4 | Secondary MPM | Qualcomm  JVET-V0120 |  |
| 2.5 | Reference sample interpolation and smoothing for intra-prediction | Qualcomm  JVET-V0120 |  |
| 2.6 | Decoder side intra mode derivation (DIMD) | Qualcomm  JVET-V0120 |  |
| **3 Inter prediction** | |  |  |
| 3.1 | Local illumination compensation (LIC) | Qualcomm  JVET-V0120 |  |
| 3.2 | Non-adjacent spatial candidates | Qualcomm  JVET-V0120 |  |
| 3.3 | Template matching (TM) | Qualcomm  JVET-V0120 | Huawei  JVET-V0146 |
| 3.4 | Multi-pass decoder-side motion vector refinement | Qualcomm  JVET-V0120 | Huawei  JVET-V0146 |
| 3.5 | Multi-pass decoder-side motion vector refinement with sample-based BDOF | Qualcomm  JVET-V0120 | Huawei  JVET-V0146 |
| 3.6 | OBMC | Qualcomm  JVET-V0120 |  |
| 3.7 | 12-tap interpolation | Qualcomm  JVET-V0120 |  |
| 3.8 | Multi-hypothesis prediction (MHP) | HHI  JVET-V0077  Qualcomm  JVET-V0120 |  |
| ~~3.8.b~~ | ~~MHP without the restriction on non-equal weight in BCW~~ | withdrawn |  |
| 3.9 | Affine MMVD | Qualcomm  JVET-V0120 |  |
| 3.10 | CIIP with PDPC blending | Qualcomm  JVET-V0120 |  |
| **4 Transform and coefficient coding** | |  |  |
| 4.1 | 8-states DQ | HHI  JVET-V0082  Qualcomm  JVET-V0120 |  |
| 4.2 | Extended LFNST | Qualcomm  JVET-V0120 |  |
| 4.3 | Transform sign prediction | Qualcomm  JVET-V0120 |  |
| **5 In-loop filtering** | |  |  |
| 5.1 | ALF with multiple fixed filters | Qualcomm  JVET-V0120 |  |
| 5.2 | Bilateral filter | Ericsson  JVET-V0094 | HHI  JVET-V0147 |
| **6 Entropy coding** | |  |  |
| 6.1 | Extended precision | Qualcomm  JVET-V0120 |  |
| 6.2 | Extended precision with slice type based CABAC window | Qualcomm  JVET-V0120 |  |
| **7 Combinations** | |  |  |
| 7.1 | Intra prediction (2.2+…+2.6) | Qualcomm  JVET-V0120 |  |
| 7.2 | Template + bilateral matching (3.3 + 3.5) | Qualcomm  JVET-V0120 | Huawei  JVET-V0146 |
| 7.4 | JVET-U0100 (EE common software) | Qualcomm  JVET-V0120 | It is used in multiple EE and EE related contributions, where the results match |

2 Description of tests

2.1 Base modification

2.1.1 Test 1.1 - Non-normative base

The following non-normative changes are introduced in the EE base software:

• Lossy encoder speedups for affine, AMVR, and merge

• Use original samples for SAO and ALF (separately reported in JVET-V0095)

• More SIMD

2.1.2 Test 1.2 - Normative base

The following normative changes are introduced in the EE base software. They either extend VVC capability or remove restrictions (no individual results on each of these):

• Increase CTU and TU sizes to 256

• Remove MTS zero out

• Remove VPDU

• Remove constraints on small intra (SCIPU), inter, and CIIP blocks; CCLM latency restriction, use LMS linear model for CCLM, affine and BDOF constraints.

• Adjust deblocking parameters based on temporal ID

2.1.3 Test 1.3 - Normative and non-normative base changes

This is a combinational test of test 1.1 and test 1.2

2.2 Intra prediction

2.2.1 Test 2.1 - Intra template matching

The prediction signal is generated at the decoder side by matching the L-shaped causal neighbor of the current block with another block in a predefined search area consisting of:

R1: current CTU

R2: top-left CTU

R3: above CTU

R4: left CTU

SAD is used as a cost function.

2.2.2 Test 2.2 - Multi-model LM

CCLM included in VVC is extended by adding three Multi-model LM (MMLM) modes, left and top, top-only, and left-only, as proposed in JVET-D0110. In each MMLM mode, the reconstructed neighboring samples are classified into two classes using a threshold which is the average of the luma reconstructed neighboring samples. The linear model of each class is derived by using the Least-Mean-Square (LMS) method, if enabled, or min/max method of VVC.

MMLM modes are added to the CCLM modes and are signaled with an index using truncated unary binarization.

2.2.3 Test 2.3 - Gradient PDPC

In VVC, for a few scenarios, PDPC may not be applied due to the unavailability of the secondary reference samples. As proposed in JVET-Q0391, in these cases, a gradient based PDPC, extended from horizontal/vertical mode, is applied. The PDPC weights (wT / wL) and nScale parameter for determining the decay in PDPC weights with respect to the distance from left/top boundary are set equal to corresponding parameters in horizontal/vertical mode, respectively. When the secondary reference sample is at a fractional sample position, bilinear interpolation is applied.

2.2.4 Test 2.4 - Secondary MPM

Secondary MPM list is introduced as described in JVET-D0114.The existing primary MPM (PMPM) list consists of 6 entries and the secondary MPM (SMPM) list includes 16 entries. A general MPM list with 22 entries is constructed first, and then the first 6 entries in this general MPM list are included into the PMPM list, and the rest of entries form the SMPM list. The first entry in the general MPM list is the Planar mode. The remaining entries are composed of the intra modes of the left (L), above (A), below-left (BL), above-right (AR), and above-left (AL) neighbouring blocks, the directional modes with added offset from the first two available directional modes of neighbouring blocks, and the default modes.

If a CU block is vertically oriented, the order of neighbouring blocks is A, L, BL, AR, AL; otherwise, it is L, A, BL, AR, AL.

A PMPM flag is parsed first, if equal to 1 then a PMPM index is parsed to determine which entry of the PMPM list is selected, otherwise an SMPM flag is parsed to determine whether to parse the SMPM index or the remaining modes.

2.2.5 Test 2.5 - Reference sample interpolation and smoothing for intra-prediction

Three modifications are proposed in this test. First, the 4-tap cubic interpolation is replaced with a 6-tap cubic interpolation filter, as described in JVET-D0119, to derive fractional samples from the reference samples. Second, for the coding blocks where a gaussian interpolation filter is applied to filter the reference samples, a 6-tap gaussian filter is applied for larger blocks (W >= 32 and H >=32) and the existing VVC 4-tap gaussian interpolation filter is used otherwise. Third, to derive the extended intra reference samples, the 4-tap interpolation filter is used instead of the nearest neighbor rounding.

2.2.6 Test 2.6 - Decoder side intra mode derivation

When decoder side intra mode derivation (DIMD) is applied, two intra modes are derived from the reconstructed neighbor samples, and those two predictors are combined with the planar mode predictor with the weights derived from the gradients as described in JVET-O0449.

Derived intra modes are included into the primary list of the intra most probable modes (MPM), so the DIMD process is performed before the MPM list is constructed. The primary derived intra mode of a DIMD block is stored with a block and is used for MPM list construction of the neighboring blocks.

2.3 Inter prediction

2.3.1 Test 3.1 - Local illumination compensation

Local illumination compensation (LIC) previously proposed in JVET-O0066 is an inter prediction technique to model local illumination variation between a current block and its prediction block with a linear function α\*p[x]+β, where p[x] is a reference sample pointed to by MV at a location x in the reference picture, α and β are derived based on the current block and the reference block templates. LIC flag is inherited for merge mode and is is signaled for AMVP.

2.3.2 Test 3.2 - Non-adjacent spatial candidate

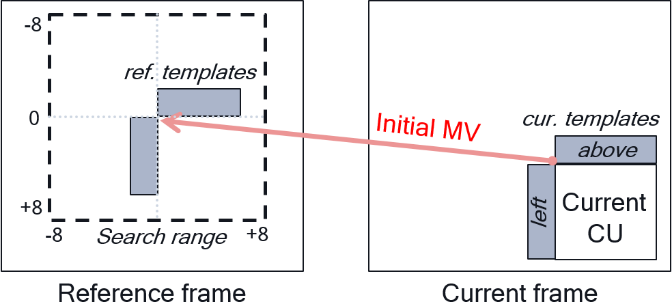
The non-adjacent spatial merge candidates as described in JVET-L0399 are inserted after the TMVP in the regular merge candidate list. The pattern of the spatial merge candidates is shown on Figure 1. The distances between the non-adjacent spatial candidates and the current coding block are based on the width and height of the current coding block.



**Figure 1. Spatial neighboring blocks used to derive the spatial merge candidates**

2.3.3 Test 3.3 - Template matching

Template matching (TM) previously proposed in JVET-J0021 is a decoder-side MV derivation method to refine the motion information of the current CU by finding the closest match between a template (i.e., top and/or left neighbouring blocks of the current CU) in the current picture and a block in a reference picture as illustrated in Figure 2. A better MV is to be searched around the initial motion of the current CU within a [– 8, +8]-pel search range.



**Figure 2. Template matching performs on a search area around initial MV.**

In AMVP mode, an MVP candidate is determined based on the template matching error to pick up the one which reaches the minimum difference between the current block and the reference block templates, and then TM performs only for this particular MVP candidate for MV refinement. TM refines this MVP candidate, starting from full-pel MVD precision (or 4-pel for 4-pel AMVR mode) within a [–8, +8]-pel search range by using iterative diamond search. The AMVP candidate may be further refined by using cross search with full-pel MVD precision (or 4-pel for 4-pel AMVR mode), followed sequentially by half-pel and quarter-pel ones depending on AMVR mode as specified in Table 1. This search process ensures that the MVP candidate still keeps the same MV precision as indicated by AMVR mode after TM process.

**Table 1. Search patterns of AMVR and merge mode with AMVR.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Search pattern** | **AMVR mode** | | | | **Merge mode** | |
| **4-pel** | **Full-pel** | **Half-pel** | **Quarter-pel** | **AltIF=0** | **AltIF=1** |
| 4-pel diamond | v |  |  |  |  |  |
| 4-pel cross | v |  |  |  |  |  |
| Full-pel diamond |  | v | v | v | v | v |
| Full-pel cross |  | v | v | v | v | v |
| Half-pel cross |  |  | v | v | v | v |
| Quarter-pel cross |  |  |  | v | v |  |
| 1/8-pel cross |  |  |  |  | v |  |

In the merge mode, similar search method is applied to the merge candidate indicated by the merge index. As Table 1 shows, TM may perform all the way down to 1/8-pel MVD precision or skipping those beyond half-pel MVD precision, depending on whether the alternative interpolation filter (that is used when AMVR is of half-pel mode) is used according to merged motion information. Besides, when TM mode is enabled, template matching may work as an independent process or an extra MV refinement process between block-based and subblock-based bilateral matching (BM) methods, depending on whether BM can be enabled or not according to its enabling condition check. When BM and TM are both enabled for a CU, the search process of TM stops at half-pel MVD precision and the resulted MVs are further refined by using the same model-based MVD derivation method as in DMVR.

2.3.4 Test 3.4 - Multi-pass decoder-side motion vector refinement

Multi-pass decoder-side motion vector refinement is applied instead of DMVR. In the first pass, bilateral matching (BM) is applied to a coding block. In the second pass, BM is applied to each 16x16 subblock within the coding block. In the third pass, MV in each 8x8 subblock is refined by applying bi-directional optical flow (BDOF). The refined MVs are stored for both spatial and temporal motion vector prediction.

First pass – Block based bilateral matching MV refinement

In the first pass, a refined MV is derived by applying BM to a coding block. Similar to decoder-side motion vector refinement (DMVR), the refined MV is searched around the two initial MVs (MV0 and MV1) in the reference picture lists L0 and L1. The refined MVs (MV0\_pass1 and MV1\_pass1) are derived around the initiate MVs based on the minimum bilateral matching cost between the two reference blocks in L0 and L1.

BM performs local search to derive integer sample precision intDeltaMV and half-pel sample precision halfDeltaMv. The local search applies a 3×3 square search pattern to loop through the search range [–sHor, sHor] in a horizontal direction and [–sVer, sVer] in a vertical direction, wherein, the values of sHor and sVer are determined by the block dimension, and the maximum value of sHor and sVer is 8.

The bilateral matching cost is calculated as: bilCost = mvDistanceCost + sadCost. When the block size cbW \* cbH is greater than 64, MRSAD cost function is applied to remove the DC effect of the distortion between the reference blocks. When the bilCost at the center point of the 3×3 search pattern has the minimum cost, the intDeltaMV or halfDeltaMV local search is terminated. Otherwise, the current minimum cost search point becomes the new center point of the 3×3 search pattern and the search for the minimum cost continues, until it reaches the end of the search range.

The existing fractional sample refinement is further applied to derive the final deltaMV. The refined MVs after the first pass are then derived as:

· MV0\_pass1 = MV0 + deltaMV

· MV1\_pass1 = MV1 – deltaMV

Second pass – Subblock based bilateral matching MV refinement

In the second pass, a refined MV is derived by applying BM to a 16×16 grid subblock. For each subblock, the refined MV is searched around the two MVs (MV0\_pass1 and MV1\_pass1), obtained on the first pass for the reference picture list L0 and L1. The refined MVs (MV0\_pass2(sbIdx2) and MV1\_pass2(sbIdx2)) are derived based on the minimum bilateral matching cost between the two reference subblocks in L0 and L1.

For each subblock, BM performs full search to derive integer sample precision intDeltaMV. The full search has a search range [–sHor, sHor] in a horizontal direction and [–sVer, sVer] in a vertical direction, wherein, the values of sHor and sVer are determined by the block dimension, and the maximum value of sHor and sVer is 8.

The bilateral matching cost is calculated by applying a cost factor to the SATD cost between the two reference subblocks, as: bilCost = satdCost \* costFactor. The search area (2\*sHor + 1) \* (2\*sVer + 1) is divided up to 5 diamond shape search regions shown on Figure 4. Each search region is assigned a costFactor, which is determined by the distance (intDeltaMV) between each search point and the starting MV, and each diamond region is processed in the order starting from the center of the search area. In each region, the search points are processed in the raster scan order starting from the top left going to the bottom right corner of the region. When the minimum bilCost within the current search region is less than a threshold equal to sbW \* sbH, the int-pel full search is terminated, otherwise, the int-pel full search continues to the next search region until all search points are examined.



**Figure 3. Diamond regions in the search area.**

BM performs local search to derive half sample precision halfDeltaMv. The search pattern and cost function are the same.

The existing VVC DMVR fractional sample refinement is further applied to derive the final deltaMV(sbIdx2). The refined MVs at second pass is then derived as:

· MV0\_pass2(sbIdx2) = MV0\_pass1 + deltaMV(sbIdx2)

· MV1\_pass2(sbIdx2) = MV1\_pass1 – deltaMV(sbIdx2)

Third pass – Subblock based bi-directional optical flow MV refinement

In the third pass, a refined MV is derived by applying BDOF to an 8×8 grid subblock. For each 8×8 subblock, BDOF refinement is applied to derive scaled Vx and Vy without clipping starting from the refined MV of the parent subblock of the second pass. The derived bioMv(Vx, Vy) is rounded to 1/16 sample precision and clipped between -32 and 32.

The refined MVs (MV0\_pass3(sbIdx3) and MV1\_pass3(sbIdx3)) at third pass are derived as:

· MV0\_pass3(sbIdx3) = MV0\_pass2(sbIdx2) + bioMv

· MV1\_pass3(sbIdx3) = MV0\_pass2(sbIdx2) – bioMv

2.3.5 Test 3.5 - Sample-based BDOF

In the sample-based BDOF, instead of deriving motion refinement (Vx, Vy) on a block basis, it is performed per sample.

The coding block is divided into 8×8 subblocks. For each subblock, whether to apply BDOF or not is determined by checking the SAD between the two reference subblocks against a threshold. If decided to apply BDOF to a subblock, for every sample in the subblock, a sliding 5×5 window is used and the existing BDOF process is applied for every sliding window to derive Vx and Vy. The derived motion refinement (Vx, Vy) is applied to adjust the bi-predicted sample value for the center sample of the window.

This test is performed together with multi-pass DMVR (test 3.4).

2.3.6 Test 3.6 - Overlapped block motion compensation

When overlapped block motion compensation (OBMC) is applied, top and left boundary pixels of a CU are refined using neighboring block’s motion information with a weighted prediction as described in JVET-L0101. An OBMC flag is inherited to be true for merge mode and is signaled to indicate the mode otherwise.

Conditions of not applying OBMC are as follows:

• When OBMC is disabled at SPS level

• When current block has intra mode or IBC mode

• When current block applies LIC

• When current luma block area is smaller or equal to 32

• OBMC flag is false

A subblock-boundary OBMC is further refined by applying the same blending to the top, left, bottom, and right subblock boundary pixels using neighboring subblock’s motion information. It is enabled for the subblock based coding tools:

• Affine AMVP modes;

• Affine merge modes and subblock-based temporal motion vector prediction (SbTMVP);

• Subblock-based bilateral matching.

2.3.7 Test 3.7 - 12-tap interpolation

The 8-tap interpolation filter used in VVC is replaced with a 12-tap filter in this test. The proposed interpolation filter is derived from the sinc function of which the frequency response is cut off at Nyquist frequency, and cropped by a cosine window function.

2.3.8 Test 3.8 - Multi-hypothesis prediction

The multi-hypothesis prediction (MHP) previously was proposed in JVET-M0425. Up to two additional predictors are signalled on top of inter AMVP mode, regular merge mode, affine merge and MMVD mode. The resulting overall prediction signal is accumulated iteratively with each additional prediction signal.

The weighting factor *α* is specified according to the following table:

|  |  |
| --- | --- |
| **add\_hyp\_weight\_idx** | ***α*** |
| 0 | 1/4 |
| 1 | -1/8 |

For inter AMVP mode, MHP is only applied if non-equal weight in BCW is selected in bi-prediction mode.

The additional hypothesis can be either merge or AMVP mode. In the case of merge mode, the motion information is indicated by a merge index, and the merge candidate list is the same as in the Geometric Partition Mode. In the case of AMVP mode, the reference index, MVP index, and MVD are signaled.

2.3.9 Test 3.9 - Affine MMVD

The affine MMVD mode that was previously proposed in JVET-N0378. In Affine MMVD mode, there is only one base vector candidate which is the first affine candidate in the subblock merge candidate list, and thus there is no base vector candidate flag to be signaled. There are three sets of distance offsets to be selected based on the sequence resolution, as follows:

• {1/16, 1/8, 1/4, 1/2, 1} for sequences smaller than 1280x720;

• {1/4, 1/2, 1, 2, 4} for sequences smaller than 2560x1600;

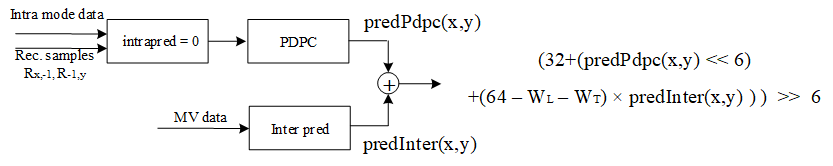
• {1/2, 1, 2, 4, 8} for sequences larger than or equal to 2560x1600.

Same as MMVD, affine MMVD also supports 4 directions (i.e., (+,0), (-,0), (0,+), (0,-)) for each of its distance offsets. When the base candidate is of uni-prediction, the offset vector (i.e., distance offset times direction) is added up equally to each CPMV. For the bi-prediction candidate, the offset vector is added up equally to L0 CPMVs, while the offset vector is firstly mirrored based on POC distance and is then added up to L1 CPMVs.

2.3.10 Test 3.10 - CIIP with PDPC blending

The CIIP mode is extended as described in JVET-O0537. In this extended mode (CIIP\_PDPC), the prediction of the regular merge mode is refined using the above (Rx, -1) and left (R-1, y) reconstructed samples. This refinement inherits the position dependent prediction combination (PDPC) scheme. The flowchart of the prediction of the CIIP\_PDPC mode can be depicted as in Figure 4, where WT and WL are the weighted values which depend on the sample position in the block as defined in PDPC.

The CIIP\_PDPC mode is signaled together with CIIP mode. When CIIP flag is true, another flag, namely CIIP\_PDPC flag, is further signaled to indicate whether to use CIIP\_PDPC.



**Figure 4. CIIP\_PDPC flowchart of the extended CIIP mode using PDPC**

2.4 Transform and coefficient coding

2.4.1 Test 4.1 - 8-states DQ

The 8-state dependent quantization proposed in JVET-Q0243 is tested. This is implemented in the unified manner, i.e. it is configurable to have 4 or 8 states. A parameter is signaled is a slice header to indicate it.

2.4.2 Test 4.2 - Extended LFNST

The LFNST design in VVC is extended as follows:

• The number of LFNST sets (S) and candidates (C) are extended to S=35 and C=3, and the LFNST set (lfnstTrSetIdx) for a given intra mode (predModeIntra) is derived according to the following formula:

o For predModeIntra < 2, lfnstTrSetIdx is equal to 2

o lfnstTrSetIdx = predModeIntra, for predModeIntra in [0,34]

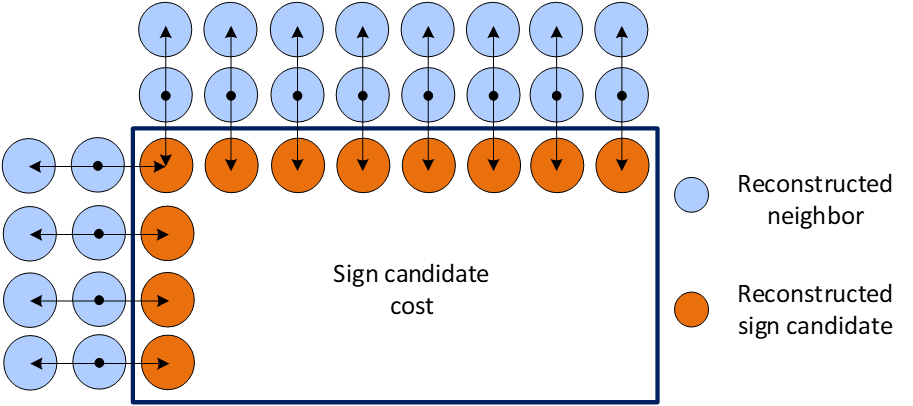
o lfnstTrSetIdx = 68 – predModeIntra, for predModeIntra in [35,66]

• The 8x8 LFNST matrix size (i.e., the transform support) in VVC is extended to full 64x64, such that it is not reduced as no zeroing-out process is applied to secondary transform coefficients.

2.4.3 Test 4.3 - Transform sign prediction

Following JVET-D0031 and JVET-J0021, basic idea of the coefficient sign prediction method is for applicable transform coefficients to calculate reconstructed residual for both negative and positive sign combinations and select the hypothesis that minimizes a cost function.

To derive the best sign, the cost function is defined as discontinuity measure across block boundary shown on Figure 5. It is measured for all hypotheses, and the one with the smallest cost is selected as a predictor for coefficient signs.

****

**Figure 5. Discontinuity measure.**

The cost function is defined as a sum of absolute second derivatives in the residual domain for the above row and left column as follows

where *R* is reconstructed neighbors, *P* is prediction of the current block, and *r* is the residual hypothesis. The term can be calculated only once per block and only residual hypothesis is subtracted.

Up to 8 signs are predicted and for those coefficients, instead of the coefficient sign, a bin indicating whether the sign is equal to the selected hypothesis is signaled. This sing prediction bin is context coded, where context is derived whether a coefficient is DC or not, and contexts are separated for intra and inter blocks, for luma and chroma components.

2.5 In-loop filtering

2.5.1 Test 5.1 - Adaptive loop filter

**ALF simplification removal**

In this test, ALF gradient subsampling and ALF virtual boundary processing are removed. Block size for classification is reduced from 4x4 to 2x2. Filter size for both luma and chroma, for which ALF coefficients are signalled, is increased to 9x9.

**ALF with fixed filters**

To filter a luma sample, three different classifiers (C0, C1, and C2) and three different sets of filters (F0, F1, and F2) are used. Sets F0 and F1 contain fixed filters, with coefficients trained for classifiers C0 and C1. Coefficients of filters in F2 are signalled. Which filter from a set Fi is used for a given sample is decided by a class assigned to this sample using classifier Ci

**Filtering**

At first, two 13x13 diamond shape fixed filters F0 and F1 are applied to derive two intermediate samples and . After that, F2 is applied to , , and neighboring samples to derive a filtered sample as

where is the clipped difference between a neighboring sample and current sample and is the clipped difference between and the current sample. The filter coefficients are signalled.



**Figure 6. ALF process.**

**Classification**

Based on directionality and activity , a class is assigned to each 2x2 block:

where represents the total number of directionalities .

As in VVC, values of the horizontal, vertical, and two diagonal gradients are calculated for each sample using 1-D Laplacian. The sum of the sample gradients within a 4×4 window that covers the target 2×2 block is used for classifier C0 and the sum of sample gradients within a 12×12 window is used for classifiers C1 and C2. The sums of horizontal, vertical and two diagonal gradients are denoted, respectively, as , , and . The directionality is determined by comparing

with a set of thresholds.

To obtain activity , the sum of vertical and horizontal gradients is mapped to the range of 0 to , where is equal to 4 for and 15 for and .

In an ALF\_APS, up to 4 luma filter sets are signalled, each set may have up to 25 filters.

**Test 5.2 - Bilateral filter**

The proposed filter is carried out in the sample adaptive offset (SAO) loop-filter stage, as shown on Figure 7. Both the proposed bilateral filter (BIF) and SAO are using samples from deblocking as input. Each filter creates an offset per sample, and these are added to the input sample and then clipped, before proceeding to ALF.

Diagram

Description automatically generated

**Figure 7 Both the proposed filter (BIF) and SAO uses samples from the deblocking stage as input. Both create an offset, and these are added to the input sample and clipped**

In detail, the output sample is obtained as

where is the input sample from deblocking, is the offset from the bilateral filter and is the offset from SAO.

The proposed implementation provides the possibility for the encoder to enable or disable filtering at the CTU and slice level. The encoder takes a decision by evaluating the RDO cost.

The samples surrounding the center sample are denoted according to Figure 8, where A, B, L and R stands for above, below, left and right and where NW, NE, SW, SE stands for north-west etc. Likewise, AA stands for above-above, BB for below-below etc.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Figure 8 Naming convention for samples surrounding the center sample, I\_C.**

Each surrounding sample , etc will contribute with a corresponding modifier value computed using LUT and intensity difference between the center sample and the surrounding samples.

Entropy ***coding***

**Test 6.1 - Extended precision**

The intermediate precision used in the arithmetic coding engine is increased, including three elements. First, the precisions for two probability states are both increased to 15 bits, in comparison to 10 bits and 14 bits in VVC. Second, the LPS range update process is modified as below,

if q >= 16384

q = 215 – 1 – q

RLPS = ((range \* (q>>6)) >>9) + 1,

where range is a 9-bit variable representing the width of the current interval, q is a 15-bit variable representing the probability state of the current context model, and RLPS is the updated range for LPS. This operation can also be realized by looking up a 512×256-entry in 9-bit look-up table. Third, at the encoder side, the 256-entry look-up table used for bits estimation in VTM is extended to 512 entries.

**Test 6.2 - Slice-type-based window size**

Since statistics are different with different slice types, in this test for each context model, three window sizes are pre-defined for I-, B-, and P-slices, respectively, like the initialization parameters to provide a context’s probability state updated at a rate that is optimal under the given slice type.

The context initialization parameters and window sizes are retrained.

Results

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| # | Test | AI | | | | | RA | | | | | LB | | | | |
|  |  | Y | U | V | Enc | Dec | Y | U | V | Enc | Dec | Y | U | V | Enc | Dec |
| 1,1 | Non-normative base | 0,00% | 0,00% | 0,00% | 101% | 103% | -0,07% | -0,05% | -0,03% | 93% | 100% | 0,00% | 0,04% | -0,15% | 92% | 99% |
| 1,2 | Normative base | -1,23% | -3,39% | -2,23% | 107% | 130% | -2,01% | -3,66% | -2,59% | 125% | 136% | -1,93% | -2,59% | -2,50% | 111% | 135% |
| 1,3 | Base (1.1 + 1.2) | -1,23% | -3,39% | -2,23% | 109% | 113% | -2,17% | -3,60% | -2,69% | 108% | 111% | -1,84% | -2,54% | -2,65% | 99% | 113% |
| 2,1 | Intra template matching | -0,40% | -0,47% | -0,49% | 178% | 168% | -0,09% | -0,08% | 0,01% | 118% | 101% |  |  |  |  |  |
| Classes F+TGM variant 1 | -4,31% | -4,08% | -4,11% | 133% | 373% |  |  |  |  |  |  |  |  |  |  |
| Classes F+TGM default | **-3,54%** | **-3,35%** | **-3,37%** | **124%** | **286%** | **-1,91%** | **-1,93%** | **-1,82%** | **117%** | **128%** |  |  |  |  |  |
| Classes F+TGM variant 2 | -3,82% | -3,60% | -3,66% | 127% | 245% |  |  |  |  |  |  |  |  |  |  |
| Classes F+TGM variant 3 | -1,16% | -1,10% | -1,11% | 106% | 125% |  |  |  |  |  |  |  |  |  |  |
| tool-off | 0,34% | 0,34% | 0,32% | 75% | 88% | 0,06% | 0,08% | 0,12% | 94% | 99% |  |  |  |  |  |
| Classes F+TGM default | **2,84%** | **2,58%** | **2,60%** | **88%** | **58%** | **1,66%** | **1,63%** | **1,64%** | **94%** | **92%** |  |  |  |  |  |
| 2,2 | Multi-model LM | -0,21% | -1,57% | -2,05% | 101% | 99% | -0,13% | -1,24% | -1,52% | 101% | 100% |  |  |  |  |  |
| 2,3 | Gradient PDPC | -0,08% | -0,10% | -0,12% | 101% | 102% | -0,04% | -0,08% | 0,05% | 100% | 100% |  |  |  |  |  |
| 2,4 | Secondary MPM | -0,26% | -0,24% | -0,27% | 101% | 101% | -0,14% | -0,04% | -0,02% | 101% | 101% |  |  |  |  |  |
| 2,5 | Reference interpolation and smoothing | -0,14% | -0,17% | -0,19% | 104% | 102% | -0,08% | -0,04% | 0,05% | 101% | 102% |  |  |  |  |  |
| 2,6 | Decoder side intra mode derivation | -0,45% | -0,26% | -0,30% | 113% | 110% | -0,26% | -0,21% | -0,19% | 102% | 106% |  |  |  |  |  |
| 3,1 | LIC |  |  |  |  |  | -0,94% | -0,51% | -0,56% | 122% | 100% | -0,69% | -0,40% | -0,50% | 121% | 101% |
| tool-off |  |  |  |  |  | 0,72% | 0,33% | 0,37% | 91% | 102% | 0,59% | 0,27% | 0,51% | 87% | 104% |
| 3,2 | Non-adjacent candidates |  |  |  |  |  | -0,41% | -0,44% | -0,40% | 102% | 102% | -0,58% | -0,69% | -0,65% | 101% | 95% |
| tool-off |  |  |  |  |  | 0,44% | 0,39% | 0,42% | 101% | 104% |  |  |  |  |  |
| 3,3 | Template matching |  |  |  |  |  | -2,26% | -2,45% | -2,58% | 116% | 115% | -1,66% | -2,02% | -2,08% | 123% | 117% |
| tool-off |  |  |  |  |  | 1,85% | 2,07% | 2,21% | 89% | 94% | 1,46% | 1,94% | 2,37% | 86% | 96% |
| 3,4 | Muti-pass DMVR |  |  |  |  |  | -1,79% | -1,69% | -1,69% | 112% | 202% |  |  |  |  |  |
| 3,5 | Muti-pass DMVR with sample based BDOF |  |  |  |  |  | -1,85% | -1,71% | -1,74% | 113% | 204% |  |  |  |  |  |
| tool-off |  |  |  |  |  | 1,97% | 1,79% | 1,81% | 84% | 59% |  |  |  |  |  |
| 3,6 | OBMC |  |  |  |  |  | -0,47% | -1,28% | -1,22% | 110% | 159% | -0,96% | -1,54% | -1,62% | 113% | 160% |
| tool-off |  |  |  |  |  | 0,61% | 1,30% | 1,33% | 94% | 75% |  |  |  |  |  |
| 3,7 | 12-tap interpolation |  |  |  |  |  | -0,71% | -0,31% | -0,21% | 104% | 103% | -1,04% | -0,78% | -0,52% | 104% | 104% |
| tool-off |  |  |  |  |  | 0,51% | 0,19% | 0,17% | 96% | 98% |  |  |  |  |  |
| 3,8 | MHP |  |  |  |  |  | -0,31% | -0,30% | -0,21% | 106% | 107% | -0,49% | -0,15% | 0,04% | 112% | 109% |
| tool-off |  |  |  |  |  | 0,27% | 0,20% | 0,23% | 99% | 101% |  |  |  |  |  |
| 3,9 | Afiine MMVD |  |  |  |  |  | -0,17% | -0,13% | -0,02% | 110% | 101% | -0,13% | -0,05% | -0,16% | 112% | 102% |
| 3,10 | CIIP with PDPC |  |  |  |  |  | -0,11% | -0,09% | 0,02% | 101% | 102% | -0,18% | -0,30% | -0,35% | 99% | 97% |
| 4,1 | 8-states DQ | -0,44% | 0,15% | 0,16% | 115% | 97% | -0,33% | 0,02% | 0,18% | 106% | 99% | -0,41% | -0,16% | 0,29% | 106% | 97% |
| tool-off | 0,40% | -0,14% | -0,11% | 87% | 101% | 0,32% | 0,05% | -0,08% | 97% | 101% | 0,37% | -0,09% | -0,29% | 99% | 104% |
| 4,2 | Extended LFNST | -0,82% | -0,63% | -0,59% | 122% | 102% | -0,47% | -0,41% | -0,28% | 106% | 101% |  |  |  |  |  |
| tool-off | 0,70% | 0,56% | 0,47% | 84% | 101% | 0,40% | 0,26% | 0,19% | 98% | 101% |  |  |  |  |  |
| 4,3 | Sign prediction | -0,29% | -0,50% | -0,55% | 118% | 122% | -0,85% | -0,85% | -0,90% | 109% | 109% | -0,78% | -0,71% | -0,71% | 107% | 109% |
| tool-off | 0,28% | 0,43% | 0,53% | 91% | 93% | 0,67% | 0,64% | 0,73% | 98% | 100% |  |  |  |  |  |
| 5,1 | ALF with fixed filters | -1,16% | -1,49% | -1,65% | 112% | 157% | -1,48% | -1,43% | -1,64% | 111% | 165% | -1,55% | -2,66% | -2,34% | 117% | 163% |
| tool-off | 1,12% | 1,50% | 1,64% | 94% | 67% | 1,42% | 1,42% | 1,57% | 96% | 85% | 1,44% | 2,83% | 2,39% | 91% | 74% |
| 5,2 | Bilateral filter | -0,48% | 0,22% | 0,23% | 107% | 106% | -0,45% | 0,14% | 0,21% | 104% | 103% |  |  |  |  |  |
| tool-off | 0,46% | -0,28% | -0,24% | 93% | 97% | 0,43% | -0,19% | -0,14% | 95% | 99% | 0,43% | -0,55% | -0,41% | 94% | 99% |
| 6,1 | Extended CABAC precision | -0,19% | -0,39% | -0,41% | 100% | 100% | -0,18% | -0,02% | 0,06% | 100% | 102% | -0,22% | 0,10% | 0,44% | 99% | 100% |
| tool-off | 0,24% | 0,43% | 0,47% | 100% | 101% | 0,21% | -0,02% | 0,06% | 101% | 101% | 0,25% | -0,10% | -0,52% | 101% | 101% |
| 6,2 | Extended precision with slice-type window | -0,23% | -0,48% | -0,56% | 101% | 101% | -0,22% | -0,15% | 0,00% | 101% | 102% | -0,23% | 0,12% | 0,26% | 100% | 101% |
| tool-off | 0,24% | 0,52% | 0,60% | 100% | 102% | 0,28% | 0,11% | 0,10% | 101% | 101% | 0,27% | -0,23% | -0,55% | 102% | 102% |
| 7,1 | All intra tools | -1,22% | -2,38% | -2,89% | 117% | 110% | -0,71% | -1,55% | -1,89% | 104% | 103% | -0,20% | -0,49% | -0,68% | 102% | 100% |
| 7,2 | Template and bilateral matching |  |  |  |  |  | -3,88% | -3,91% | -4,10% | 141% | 220% |  |  |  |  |  |
| 7,3 | EE2 base (U0100) | -5,15% | -8,15% | -7,97% | 218% | 196% | -12,14% | -13,31% | -13,62% | 265% | 459% | -9,76% | -10,58% | -10,37% | 229% | 266% |

Bilateral filter and template matching were tested additional to the base codec, the tool-off test shows the loss if disabled from the combination EE2base+addtool. The tool-on test is relative to VTM11

[JVET-V0077](https://jvet-experts.org/doc_end_user/current_document.php?id=10725) EE2-3.8: Multiple Hypothesis Prediction [M. Winken, J. Pfaff, B. Bross, H. Schwarz (HHI)]

This is rather to be interpretd as a crosscheck of JVET-V0120.

[JVET-V0082](https://jvet-experts.org/doc_end_user/current_document.php?id=10730) EE2-4.1: Results for dependent quantization with 8 states [H. Schwarz, P. Haase, T. Nguyen, J. Pfaff, D. Marpe, T. Wiegand (HHI)]

This is rather to be interpreted as a crosscheck of JVET-V0120.

[JVET-V0094](https://jvet-experts.org/doc_end_user/current_document.php?id=10742) EE2: Bilateral filter in VTM, EE2 and VVenC [J. Ström, P. Wennersten, K. Andersson, R. Sjöberg (Ericsson), S. Ikonin, E. Alshina (Huawei)]

[JVET-V0147](https://jvet-experts.org/doc_end_user/current_document.php?id=10808) Cross-check of JVET-V0094: EE2: Bilateral filter in VTM, EE2 and VVenC [A. Henkel (HHI)] [late]

[JVET-V0120](https://jvet-experts.org/doc_end_user/current_document.php?id=10769) EE2: Tests of compression efficiency methods beyond VVC [Y.-J. Chang, C.-C. Chen, J. Dong, N. Hu, H. Huang, M. Karczewicz, B. Ray, V. Seregin, Y. Zhang, Z. Zhang (Qualcomm)]

Reports everything but bilateral filter andTM

[JVET-V0146](https://jvet-experts.org/doc_end_user/current_document.php?id=10807) Crosscheck of Decoder Side Refinement tools in JVET-V0120 and Additional Results [S. Esenlik, E. Alshina (Huawei)] [late]

Additional results: Disable combination of template matching and bilateral matching (1% loss), and enable MV propagation in DMVR (0.5% gain over VTM12)

[JVET-V0130](https://jvet-experts.org/doc_end_user/current_document.php?id=10791) EE2: Intra Template Matching [K. Naser, T. Poirier, F. Le Léannec, G. Martin-Cocher (InterDigital)]

Conclusion about EE:

As a general finding, the difference between tool-on gain and tool-off loss is not very large in almost all cases, which would indicate that the benefit of the tools in the base set is somewhat additive (which would also apply to intra template matching (for screen content) and bilateral filter. However, tool-off tests were not run for all tools.

It is agreed that at this point of exploration, complexity is not of primary concern, as we are targeting to explore the potential to further increase compression performance. Nevertheless, feedback on possible implementation problems of certain tools is more than welcome.

Currently, encoder runtime is not asserted to be a problem, either. The maximum duration of a full set is asserted to run for 10 days.

One expert expressed opinion that the memory consumption of the software might be a problem on some computing platforms. However, it is reported that the current base software does not consume more than 10 GB for class A.

Decision: Adopt bilateral filter (JVET-V0094) and intra template matching (JVET-V030), the latter for screen content, into the next version of “EE base software”.

### EE2 related contributions: Enhanced compression beyond VVC capability (8)

Contributions in this area were discussed in sessions 11 and 12 at 2215–0125 UTC on Thursday/Friday 22/23 April 2021 (chaired by JRO and GJS), and in session 15a at 2100-2124 UTC on Friday 23 April 2021 (chaired by JRO).

[JVET-V0083](https://jvet-experts.org/doc_end_user/current_document.php?id=10731) EE2 related: asymmetric binary tree splitting on top of VVC [F. Le Léannec, K. Naser, T. Dumas, A. Robert, F. Galpin, E. François (InterDigital)]

Asymmetric Binary Tree (ABT) splitting was proposed in JVET-D0064, JVET-J0022 and JVET-K0197. It basically consists in a split mode that divides a coding unit into 2 sub-CUs with split ratio (1/4,3/4) or (3/4,1/4), in the horizontal or vertical direction.

This contribution re-introduces ABT on top of VTM-11. In addition to the technical aspects described in JVET-K0197, some adaptations to VVC coding tools are needed, like affine motion compensation, BDOF, DMVR, ISP and MIP.

Several obtained trade-offs between luma coding gain and encoding time over VTM-11 are reported in random access configuration, which are -1.1%, -1.5%, -1.7%, -1.83% and -2.1% luma gain with respectively 116%, 133%, 146%, 155% and 186% encoding times.

It is also shown that higher coding gains are obtained with the proposed partitioning scheme than with the QTBTT at equivalent encoder complexity, when varying the MaxMTTHierachyDepth coding parameters of VTM-11.

It is proposed to further explore the ABT partitioning in the EE2 on Enhanced Compression beyond VVC capability.

“redundancy” of splits is avoided, i.e. same block layout cannot be reached by different splits.

Additional transform sizes: 6, 12, 24

Further modifications are necessary for MIP, ISP

Comments:

* Could be substantial amount of work integrating into EE2 SW
* Transforms and other partitions with non-2^x sizes might be undesirable
* Largest block size where the new block sizes are applied? 64
* Smallest size? 6
* It is noticed that the gain is largest in the classes A1/2
* Why loss in chroma? A shift of bits was applied from chroma to luma.
* Could the prevention of redundant splits be implemented as encoder only? This would reduce some gain. Is the parsing more complicated by that? Yes, somewhat.

It is also pointed out that the slide deck has a “copyright” note with “all rights reserved”. Thi should be updated.

Study in EE. If time allows, it would be desirable also testing a version that only has power-of-two splits

[JVET-V0086](https://jvet-experts.org/doc_end_user/current_document.php?id=10734) EE2-related: OBMC fixes and updates [A. Robert, F. Galpin, F. Le Léannec, T. Poirier (InterDigital)]

This contribution proposes some OBMC fixes and updates on top of JVET-U0100. This includes:

* Fixing the AMVR index for OBMC processing,
* Limiting the number of lines/columns processed at top and left borders in case of subblock mode,
* Considering as non subblock translational affine CUs,
* Using SIF and BCW in the OBMC process,
* Subtracting OBMC of the original signal in the AMVP, MMVD and CIIP modes.

It is reported that tool-on test achieves averagely 0.05% BD rate saving in RA while saving 1% encoding and 5% decoding time, and 0.06% BD rate saving in LDB while saving 1% encoding and 3% decoding time. Tool-off test is achieving averagely 0.06% BD rate saving in RA while encoding time remains the same and saving 1% decoding time, and 0.05% BD rate saving in LDB while keeping encoding time and saving 2% decoding time

JVET-V0126 proposes also fixing the AMVR index issue. Further discuss there, if there is really a problem

In general, the aspects introduced in this contribution seem to be minor improvements.

No action.

[JVET-V0087](https://jvet-experts.org/doc_end_user/current_document.php?id=10735) EE2-Related: Improvements of Decoder-Side Intra Mode Derivation [J. Zhao, S. Paluri, S. Kim (LGE)]

This contribution presents methods to improve Decoder-side Intra Mode Derivation (DIMD) on top of EE2 common software/JVET-U0100.

Aspect 1), Combination of DIMD and ISP is not allowed.

Aspect 2), DIMD with multiple blend modes is proposed. Instead of one DIMD blend mode in JVET-U0100, in the proposed method, 3 blend modes are used. These 3 blend modes have different weights when blending the predictions from decoder derived intra modes and planar mode. An index is signalled to indicate which DIMD blend mode is to be used.

Experiments are performed using EE2 Tool-on test Base setting. Comparing to EE2 Base with only DIMD on, aspect 1 gives gain of -0.07%, -0.13%, -0.06% in AI configuration, and -0.04%, -0.09%, -0.04% in RA configuration for Y,U,V components respectively. There is no change of encoding and decoding time. Aspect 1 and 2 together gives gain of -0.23%, -0.17%, -0.16% in AI configuration, with 115% encoding time; and gives gain of -0.12%, -0.09%, -0.03% in RA configuration with 103% encoding time. There is no decoding time increase.

Gain on top of VTM is 0.67% for AI and 0.38% for RA (larger gain than on top of EE2 software)

The gain of aspect 1 comes from avoiding signalling a syntax element.

Aspect 2 introduces some new mode differentiation in addition.

It is mentioned that JVET-V0098 also investigates another modification of DIMD (including information from MPM list) which is claimed that it might work better also in the combination with ISP.

Investigate aspect 2 in EE.

Decision: Adopt aspect 1 in EE2 base software.

[JVET-V0165](https://jvet-experts.org/doc_end_user/current_document.php?id=10827) Crosscheck of JVET-V0087 (EE2-Related: Improvements of Decoder-Side Intra Mode Derivation) [Y.-J. Chang (Qualcomm)] [late]

[JVET-V0089](https://jvet-experts.org/doc_end_user/current_document.php?id=10737) EE2-related: Inter coding modes modifications [A. Robert, F. Le Léannec, F. Galpin, T. Poirier (InterDigital)]

[Ed. (GJS): Rephrase the unclear use of “losing” in the summary.]

This contribution proposes some Inter coding modes modifications on top of JVET-U0100. This includes:

* Adding a full pruning considering BCW index, SIF and LIC flags in all Inter coding modes,
* Adding a STMVP candidate in merge list,
* Modifying the pairwise merge candidate,
* Adding pairwise affine merge candidates,
* Using the SIF flag in GEO mode.

It is reported that tool-on test achieves averagely 0.24% BD rate saving in RA while losing 2% encoding and 3% decoding time, and 0.36% BD rate saving in LDB while keeping the encoding and decoding time. Tool-off test is achieving averagely 0.09% BD rate saving in RA while keeping the encoding and decoding time, and 0.xx% BD rate saving in LDB while losing x% encoding and x% decoding time.

It is reported by the proponents that the gain was larger before the last modification of the EE2 base SW.

It is claimed that the gain in LB is expected to be larger than for RA, but results are not complete yet.

5 different modifications which together give relatively small gain, and some complexity in the list construction etc may be added. Is each of them justified? Would be desirable to have separate analysis of them.

Not obvious that there is a problem with current SW.

With regard to RA, gain is not substantial enough to take action at this point. May be re-considered if LB results are complete, and if more evidence about the benefit of different modifications is brought.

[JVET-V0098](https://jvet-experts.org/doc_end_user/current_document.php?id=10746) EE2-related: Template-based intra mode derivation using MPMs [Y. Wang, L. Zhang, K. Zhang, Z. Deng, N. Zhang (Bytedance)]

This contribution proposes a template-based intra mode derivation (TIMD) method using most probable modes (MPMs). For each intra prediction mode in MPMs, the sum of absolute transformed differences (SATD) between the prediction and reconstruction samples of a template is calculated. The intra prediction mode with the minimum SATD is selected as the TIMD mode and used for intra prediction of current block. On top of the EE2 code of JVET-U0100, simulation results of the proposed method are reported as below:

Tool-on test (Anchor: VTM-11.0): AI: {-0.41%, 117%, 111%}; RA: {-0.23%, 109%, 107%}.

Tool-off test (Anchor: EE2 code): AI: {-0.27%, 118%, 109%}; RA: {-0.11%, 104%, 101%}.

Question: Was the MPM list derivation changed? Yes, but it is not using TM to construct the MPM list.

This is different from the current DIMD mode, and if tested in EE, it could also be tested in combination with ISP. There is no conflict with the aspect 1 from JVET-V0087.

It is clarified that the shape of the template depends on CU size.

Investigate in EE.

[JVET-V0104](https://jvet-experts.org/doc_end_user/current_document.php?id=10752) EE2-related: TU-level adaptive self-guided filter [W. Yin, K. Zhang, L. Zhang, Y. Wang, H. Liu (Bytedance)]

This contribution presents a TU-level adaptive self-guided filter (TU-ASGF) with five taps and a cross-shape. TU-ASGF is applied on the reconstruction samples of a TU. The filtering strength is based on the TU size, QP, and statistical information adaptively.

On top of the EE code of JVET-U0100, simulation results of the proposed method are reported as below:

Tool-on test (VTM-11 as the anchor):

AI: -0.42%, 122%, 123%; RA: -0.32%, 112%, 111%; LB: -0.41%, 108%, 113%.

Tool-off test (EE code as the anchor):

AI: -0.43%, 115%, 110%; RA: -0.24%, 104%, 101%; LB: -0.28%, 104%, 102%.

On top of the EE code of JVET-V0094 (bilateral filter), simulation results of the proposed method are reported as below:

Tool-on test (VTM-11 as the anchor):

AI: -0.71%, 126%, 122%; RA: -0.55%, 110%, 108%; LB: -0.64%, 107%, 110%.

Tool-off test (EE-bilateral filter code as the anchor):

AI: -0.22%, 115%, 114%; RA: -0.09%, 106%, 104%; LB: -0.11%, 103%, 103%.

The results indicate that there is an overlap in gain with the bilateral filter (concepts are similar, anyway)

The filter is however in the intra prediction loop whereas the bilateral filter is (currently) a loop filter.

Which TU sizes are filtered? Minimum 64 samples.

Are intra blocks filtered? Yes.

In case of neighbored blocks being intra blocks, samples from these are filtered.

When the bilateral filter was originally introduced, it was also operating in the prediction loop and had higher than currently as a post filter. Then, restrictions were imposed, such as not using it for intra coded blocks, not using samples from neighboring blocks, and the gain went down.

A question would be if the self guided filter would still give additional gain when similar restrictions are imposed as for the bilateral filter. The most consequential way would be operating it as a loop filter

Proponents are asked to perform further study if the self-guided filter could be beneficial in addition to the bilateral filter. Proponents were initially asked to investigate if some action regarding EE could be taken from this meeting.

In version 3, results with a variant which implements the filter as loop filter are reported (only AI and for some classes). The gain is substantially decreased.

Further study (not CE) is recommended, if there is potential for more gain.

[JVET-V0117](https://jvet-experts.org/doc_end_user/current_document.php?id=10766) EE2-related: Combination of GPM and template matching [R.-L. Liao, Y. Ye, X. Li, J. Chen (Alibaba)]

In this contribution, geometric partition mode with template matching is proposed. For a CU coded in GPM, each partition can decide whether the motion is refined using template matching or not. When TM is applied, motion is refined in the same way as TM applied to merge mode. It is reported that on top of EE2 common software, the overall coding performance impact for {Y, U, V, EncT, DecT} is {-0.16%, -0.30%, -0.38%, 110%, 101%} for RA and {xx%, xx%, xx%, xx%, xx%} for LDB.

Presented in session 15a.Question: Is the decoding time correct? Usually template matching increases more. However, as GPM is not as frequently used, it may be reasonable.

Encoding time increase: RD decision in encoder not optimized – this should be investigated.

Both partitions are refined independently by TM, then they are blended.

Study in EE.

[JVET-V0164](https://jvet-experts.org/doc_end_user/current_document.php?id=10826) Crosscheck of JVET-V0117 (EE2-related: Combination of GPM and template matching) [C.-C. Chen, H. Huang (Qualcomm)] [late]

[JVET-V0118](https://jvet-experts.org/doc_end_user/current_document.php?id=10767) EE2-related: Extension of template matching to Affine, CIIP, GPM merge modes, and boundary sub-blocks [[C.-C. Chen](mailto:chunchic@qti.qualcomm.com), [V. Seregin](mailto:vseregin@qti.qualcomm.com), [Y.-J. Chang](mailto:yjchang@qti.qualcomm.com), [Z. Zhang](mailto:zhizhang@qti.qualcomm.com), [H. Huang](mailto:hanhuang@qti.qualcomm.com), Y. Zhang, M. Karczewicz (Qualcomm)]

Template matching refinement applied to regular merge mode as being studied in Exploration Experiments 2 (EE2, JVET-U2024) is extended to other merge modes, namely CIIP, GPM, boundary sub-blocks, and Affine. This extension was implemented on top of the EE2 software and was tested according to the procedure described in the EE2 document. As compared with the EE2 tool-off anchor, the experimental results demonstrate 0.33% Y BD-rate reduction, with 118% encoder and 102% decoder runtime, for random access configuration. It is also reported that the Y BD-rate reduction of Class C/D/E in Low Delay B is 0.48%/0.55%/0.91%, respectively.

Presented in session 15a.GPM similar to V0117, but here only one flag is used to enable TM for both partitions.

In affine, only CPMVs are refined, and applied to all subblocks

How much gain for the different elements, affine, CIIP, etc.? Should be further studied if feasible.

Study in EE.

### Technology elements beyond EE2 (12)

Contributions in this area were discussed in sessions 15a and 16a at 2125–0115 UTC on Friday/Saturday 23/24 April 2021 (chaired by JRO except noted otherwise).

[JVET-V0080](https://jvet-experts.org/doc_end_user/current_document.php?id=10728) AHG12: Extensions to CCALF [S. Keating, A. Browne, K. Sharman (Sony)]

This contribution proposes extensions to CCALF in both the size of the filters and number of filters available. Experiments have shown that using a different filter shape for Type 0 and Type 2 chroma can give some further improvement in coding efficiency. BD-rate of the proposed Type 0 filter is 0.09%, -1.65%, -1.75% for AI and 0.01%, -1.10%, -1.15% for RA when applied over JVET-U0100 (JVET-M1010 CTC).

Gain is relatively small compared to the one that CCALF had when it was adopted. For AI, with a luma loss of 0.1%, the gain could be close to zero. What is the gain with a combined YUV metric?

As CCALF could introduce chroma artifacts at higher QPs and mechanisms are built in to prevent that, it should be checked if such an extension might not re-introduce them. This should be further studied.

[JVET-V0097](https://jvet-experts.org/doc_end_user/current_document.php?id=10745) AHG12: Unsymmetric partitioning methods in video coding [K. Zhang, L. Zhang, Z. Deng, N. Zhang, Y. Wang (Bytedance)]

This contribution presents two unsymmetric partitioning methods denoted as unsymmetric binary-tree (UBT) partitioning and unsymmetric quad-tree (UQT) partitioning. UBT and UQT can be applied for blocks with dimensions no greater than 64×64. On top of the EE code of JVET-U0100, simulation results of the proposed method are reported as below:

Tool-on test (VTM-11 as the anchor):

AI: -0.68%, 190%, 104%; RA: -1.48%, 204%, 102%; LB: -1.46%, 206%, 108%.

Tool-off test (EE code as the anchor):

AI: -0.61%, 187%, 103%; RA: -1.23%, 201%, 101%; LB: x%, x%, x%.

The ¼ vs. ¾ split (UBT) requires four new transforms: 6, 12, 24, 48 to be introduced

Also has a ternary split (UQT) 1/8 3/4 1/8.

The ¾ part must be further split into 1/3 2/3, therefore no additional transform.

Similar to JVET-V0083 (UBT is equivalent to ABT), but gain is lower.

It is suggested to study lower-complex encoder configurations.

Study in EE

[JVET-V0099](https://jvet-experts.org/doc_end_user/current_document.php?id=10747) AHG12: Adaptive Reordering of Merge Candidates with Template Matching [N. Zhang, K. Zhang, L. Zhang, H. Liu, Z. Deng, Y. Wang (Bytedance)]

In this contribution, an adaptive reordering of merge candidates with template matching (ARMC) method is proposed. With the proposed method, merge candidates are divided into several subgroups, and merge candidates in a subgroup are reordered according to cost values based on template matching. The reordering method is applied to regular merge mode, template matching merge mode, and affine merge mode.

On top of the EE code of JVET-U0100, simulation results of the proposed method are reported as below:

Tool-on test (VTM-11 as the anchor): RA: -0.45%,103%, 111%; LB: -0.52%, 101%,114%

Tool-off test (EE code as the anchor): RA: -0.31%, 100%, 103%; LB: -0.36%, 102%, 105%

Template matching uses one line and one column from neighbored block. No signalling

Is LIC considered in TM? Yes

The gain is interesting, but the design introduces the problem that the independency of MV decoding and reconstruction is given up.

Could this better be implemented in a DMVR-like approach?

Further study in EE, but consider the aspects mentioned above.

[JVET-V0102](https://jvet-experts.org/doc_end_user/current_document.php?id=10750) AHG11 & AHG12: Deep In-Loop Filter with Adaptive Model Selection for Enhanced Compression Beyond VVC Capability [Y. Li, L. Zhang, K. Zhang (Bytedance)]

This contribution evaluates the deep in-loop filter presented in JVET-V0100 on top of JVET-U0100. Compared with JVET-U0100, the proposed method reportedly shows on average {6.70%, 18.52%, 19.24%}, {xx%, xx%, xx%}, and {xx%, xx%, xx%} BD-rate reductions for {Y, Cb, Cr}, under AI, RA, and LDB configurations, respectively.

The gain when combined with EE base SW is somewhat lower than the gain over VTM (same by tendency for RA and LB, where results are not complete yet). However, still substantial gain.

Encoding time is 130% for AI, less than 200% for RA. Decoder time is 70x…80x increased.

One expert suggests it would be beneficial retaining a common code base for EE1 and EE2.

Further study is strongly encouraged.

[JVET-V0103](https://jvet-experts.org/doc_end_user/current_document.php?id=10751) AHG12: Geometric prediction mode with motion vector differences [Z. Deng, K. Zhang, L. Zhang, N. Zhang, Y. Wang (Bytedance)]

In this contribution, Geometric prediction mode with Motion Vector Difference (GMVD) is proposed. With GMVD, the motion information of geometric partitions of a GPM CU are further refined by adding up signalled MVDs to GPM merge candidates. The MVD of a GPM partition is signalled in a similar way as MMVD in VVC, wherein indices of a direction and a distance are coded. Both tool-on (Test 1) and tool-off (Test 2) tests were conducted. Simulation results are reported as below:

Test 1 (VTM-11.0 as the anchor): RA -0.22%, 104%, 99%, and LB -0.30%, 105%, 101%

Test 2 (EE common base as the anchor): RA -0.13%, 122%, 100%, and LB -0.18%, 128%, 99%

The peak BD-rate impacts for tool-on test in CTC are reported to be -0.66% and -0.79% under RA and LB configurations respectively.

For test 2, a full search was used.

Study in EE. Reduction encoder runtime should be studied as well, as this is not a good tradeoff with compression benefit.

[JVET-V0110](https://jvet-experts.org/doc_end_user/current_document.php?id=10758) AHG12: Two additional reference sample sets for CCLM [C.-Y. Teng, Y.-C. Yang (FG Innovation)]

This contribution reports the test results of adding two additional linear model (LM) modes to the EE common software. The additional linear model modes are derived based on two reference sample position sets presented in JVET-P0500, and the reference sample position is different from the LM modes in the current VVC.

The simulation results are shown as follows:

Proposed method implemented on top of EE common software (Anchor: VTM11):

-5.13/-8.63/-8.75 BD performance on Y, Cb and Cr components for AI configuration.

-11.60/-12.72/-13.26 BD performance on Y, Cb and Cr components for RA configuration.

Proposed method implemented on top of EE common software (Anchor: EE common software):

-0.04/-0.61/-0.80 BD performance on Y, Cb and Cr components for AI configuration.

-0.06/-0.23/-0.53 BD performance on Y, Cb and Cr components for RA configuration.

Proposed method implemented on top of VTM11 (Anchor: VTM11):

0.01/-1.01/-0.98 BD performance on Y, Cb and Cr components for AI configuration.

0.01/-0.79/-0.63 BD performance on Y, Cb and Cr components for RA configuration.

The gain is not substantial enough to be considered, it is extending an existing tool (making it more complex). No action at this point.

[JVET-V0116](https://jvet-experts.org/doc_end_user/current_document.php?id=10765) Enhanced Intra MTS and LFNST for compression beyond VVC [B. Ray, L. Kerofsky, M. Coban, M. Karczewicz, H. Egilmez, V. Seregin (Qualcomm)]

This contribution investigates potential improvement of MTS for intra coding and LFNST. Simulation results based on JVET-U0100 show gain for tool on configuration (BASE\_ENCODER = BASE\_NORMATIVE = TOOLS = 0). For only MTS changes, in AI, RA and LB configuration, the reported luma BD rates are -0.58%, -0.25% and -0.08%, with an encoder runtime of 135%, 113% and 115%, respectively. For only LFNST changes (method 2), the reported luma BD rates for AI and RA configuration are -1.08% and -0.51%, with an encoder runtime of 121% and 106%, respectively.

For JVET-U0100 EE2 (BASE\_ENCODER = BASE\_NORMATIVE = TOOLS = 1) configuration, for only MTS changes, in AI, RA and LDB configuration, the reported luma BD rates are -0.50%, -0.22%, and -0.xx%, with an encoder runtime of 135%, 112%, and 1xx%, respectively. For only LFNST change (method 1), for AI and RA configuration, the reported luma BD rates are -0.20% and -0.xx%, with an encoder runtime of 99% and 1xx%, respectively.

For LFNST, an alternative method (method 2) is included. Also, the results of encoder and signaling only changes for MTS, and a speed-up version of MTS are presented.

Additional primary transfoms: DCT5, DST4, DST1. Gains are retained on top of EE base SW

LFNST modified based on TU shape (method 1), 7 classes of LFNST kernels (method)

For training the same set was used as before.

For LFNST, only tool on test. Probably, the gain would only be around 0.2% on top of EE base SW, as the latter already includes an improved LFNST

Generally, the additional benefit of the LFNST extensions seems to be low, but it does not increase the encoder runtime, and the additional methods are options to explore further improvements of LFNST. It should also be studied if the combination of the two methods gives additive gain.

Reducing encoder runtime for MTS should also be studied.

Study in EE (both MTS and LFNST extensions).

[JVET-V0119](https://jvet-experts.org/doc_end_user/current_document.php?id=10768) AHG12: LFNST extension with large kernel [M. Koo, J. Zhao, J. Lim, S. Kim (LGE)]

In the VVC standard, two types of LFNST kernels, which are commonly called LFNST4 and LFNST8, are incorporated. LFNST4 has a form of 16x16 matrix and is applied to 4xN/Nx4 blocks. LFNST8 has a form of 16x48 matrix and applied to the larger block size which is not coved by LFNST4. In this contribution, a new LFNST16 kernel is proposed to deal with larger blocks whose width and height are greater than or equal to 16. LFNST16 has a form of 32x96 matrix. In addition, a new LFNST8 architecture with 32x64 matrix is introduced. In Test A, the LFNST16 is tested, and both the LFNST16 and new LFNST8 (32x64) are applied in Test B as follows;

Test A: (LFNST4, LFNST8, LFNST16\*) = (16x16, 16x48, 32x96)

Test B: (LFSNT4, LFNST8\*, LFNST16\*) = (16x16, 32x64, 32x96)

VVC: (LFNST4, LFNST8) = (16x16, 16x48)

In the above tests, the following three aspects are configured in common: 1) the total number of LFNST sets is 36, one of which is dedicated for WAIP extended modes, 2) each LFNST set consists of three kernels, and 3) the worst case multiplication count per sample is 16.

The AI results are reportedly observed that the BD-rates are reduced by 0.86% and 1.04% with 117% and 120% encoding time changes for Test A and B, respectively relative to VTM-12.0 anchor.

The RA results are reportedly observed that the BD-rates are reduced by 0.48% and 0.60% with 104% and 105% encoding time changes for Test A and B, respectively relative to VTM-12.0 anchor.

Unlike the extended LFNST of U0100, the “zero out” approach is still retained her to limit the worst case number of multiplications.

It is not known how much of the gain is retained when the method would be applied on top of the EE base SW, but the introduction of new matrices for LFNST16 seems to give most gain for the large resolutions such as class A, so it could be expected that a substantial amount of the gain would be retained.

Study in EE. It should also be tested against a version where the extended LFNST approach of the EE code base is replaced by this proposal.

[JVET-V0125](https://jvet-experts.org/doc_end_user/current_document.php?id=10774) AHG12: Evaluation of GPM with MMVD for coding efficiency improvement over VVC [X. Xiu, H.-J. Jhu, C.-W. Kuo, W. Chen, Y.-W. Chen, X. Wang (Kwai)]

In this contribution, the coding performance of enhanced geometry partition mode (GPM) by further applying motion vector (MV) refinement to the unidirectional MV of each GPM partition is studied. To reduce the signaling overhead, the MV refinements are signaled based on the MV difference (MVD) signaling method of merge mode with MVD (MMVD). Compared to VTM-10.0 anchors, simulation results reportedly show that the method provides average {Y, U, V} BD-rate savings of {0.36%, 0.58%, 0.63%} and {0.66%, 1.12%, 0.78%} for RA and LD configurations, respectively. The corresponding encoding and decoding run-time are 103% and 97% for RA and 103% and 98% for LD, respectively.

The same signalling of offsets, and context modelling is used as in VVC’s MMVD. This is different from the method of JVET-V0103, where different offset values were used.

Two separate flags for the two partitions (as in the other proposal).

On top of EE base SW, the gain is 0.08% for RA and 0.21 for LB.

Study in EE along with JVET-V0103, try identifying the best combination of the elements from the two proposals.

[JVET-V0126](https://jvet-experts.org/doc_end_user/current_document.php?id=10775) EE-related: Modified OBMC [Y. Kidani, K. Kawamura, K. Unno (KDDI)]

This contribution proposes modified OBMC based on JVET-U0100 to improve coding efficiency, which includes fixing the RDO of OBMC for AMVR. Simulation results reportedly show proposed method achieves -0.04% BD-rate saving and 10% decoding time saving with 1% increasing encoding time increase compared to the EE common software base where OBMC is only enabled.

This is an encoder-only change. BD rate saving almost negligible.

The reference for testing is not the full set from EE base software, but a configuration where only OBMC and/or the 12 tap filter is turned on. When compared against the full set, the decoding time saving would be less.

There does not seem to be a substantial problem. No urgency for action at this moment.

[JVET-V0129](https://jvet-experts.org/doc_end_user/current_document.php?id=10778) AHG12: 3D Geometry for Global Motion Compensation [H. Golestani, C. Rohlfing, M. Wien (RWTH Aachen)]

This contribution reports a research project on developing 3D-geometry-based global motion compensation for video sequences captured by monocular moving cameras. DayLightRoad, ParkRuning, and MarketPlace are examples of the target sequences. The main idea is to estimate 3D scene geometry and camera motion data from 2D video sequences and use them for global motion compensation. Since this 3D data provides a better scene understanding, integrating it into video coding can be beneficial to enhance e.g. motion compensation. The proposed method applies Structure-from-Motion (SfM) to the target 2D video sequences to estimate both their camera parameters and 3D scene geometry (in the form of either point clouds or 3D meshes). The extracted 3D data guides a rendering engine to synthesize 3D-based virtual Reference Pictures (RPs). These RPs are finally offered to the VVC encoder by adding them into Reference Picture Lists (RPLs). This contribution states that the 3D-based rendered RP can provide better motion compensation performance compared to regular RPs.

The proposed method can be applied directly to video sequences captured by an unknown monocular moving camera. Any pre-knowledge about the camera (like focal length), having stereo/RGBD cameras, using additional sensors like GPS, Inertial Measurement Unit (IMU), or LiDAR sensors (in self-driving cars) might accelerate the procedure and improve the coding gain. In this document, the proposed method is applied to seven video sequences with camera motion (refer to table 1). It is reported that the proposed method achieves around 3% bit-rate reduction over VTM10.0 on average on the used test sequences.

This contribution was discussed in session 16a at 0015 UTC on Friday 24 April 2021 (chaired & notes by Y. Ye).

VTM-10.0 was configured to operate in 8-bit domain.

2.91% and 3.33% BD rate reduction (averaged over a selected set of test sequences) was achieved for replaced reference (RR) and additional reference (AR) mode, respectively.

Encoding time is comparable with the anchor (95% and 103% for RR and AR mode respectively).

Decoding time is significantly higher than the anchor (2668% and 2674% for RR and AR mode respectively). One expert commented that the decoding time of NN methods seemed to be even higher.

The 3D-Ref was always placed at the end of the RPL in both RR and AR modes.

The POC of the 3D-Ref was set the same as the current picture.

Currently, camera parameters and mesh parameters are both sent once at the sequence level. This causes a latency at the encoder side, which can be fixed in the future by sending the parameters more frequently, e.g. at the picture level. Sending parameters more frequently can improve systhesis quality but also incur higher signalling overhead. This could be investigated in the future.

How does the fidelity of the parameters affect synthesis quality? A PCS paper by the proponents contains more detailed information on the tradeoff between signaling overhead and synthesis quality.

Multiple experts commented that the decoding complexity seemed to be manageable.

Further study encouraged.

[JVET-V0153](https://jvet-experts.org/doc_end_user/current_document.php?id=10814) AHG12: Cross-component Sample Adaptive Offset [C.-W. Kuo, X. Xiu, Y.-W. Chen, H.-J. Jhu, W. Chen, X. Wang (Kwai)] [late]

In this contribution, one coding tool named Cross-component Sample Adaptive Offset (CCSAO) is proposed for compression efficiency improvement beyond VVC. CCSAO operates as part of SAO process and utilizes the correlation between three components as guidance to further enhance the reconstructive quality of the current sample. Compared to VTM-12.0 anchors with the encoder fix in JVET-V0095, the BD-rate impact of the proposed method is summarized as below.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **All Intra** | | | **Random Access** | | | **Low Delay B** | | |
| **Y** | **U** | **V** | **Y** | **U** | **V** | **Y** | **U** | **V** |
| CCSAO | 0.13% | -2.43% | -2.39% | 0.04% | -3.62% | -3.61% | -0.15% | -8.45% | -8.64% |

The lower gain for intra is asserted to be due to the higher number of parameters that are sent. In RA, it is hardly used for higher temporal IDs.

Gains for RA and LB look interesting

Question: Could it have impact on visual quality of chroma? Not known, should be investigated.

Study in EE. It is suggested to also perform a visual investigation with ALF off.

# High-level syntax (HLS) proposals (15)

## AHG9: SEI message studies and proposals (10)

See also the notes for JVET-V0072 (on RPL constraints) and JVET-V0111 (on decoding unit information), reported as issues for VVC v1.

See also the notes for JVET-V0093 with software for encoder and decoder functionality in section 4.11, which also suggest potential standardization action. JVET-V0093 was initially discussed as an SEI message contribution and the notes were then moved to reflect its potential as an encoder and decoder software functionality enhancement.

Contributions in this area were initially discussed in session 9 at 1300–1500 UTC and session 10 at 1520–1720 UTC on Thursday 22 April 2021 (chaired by GJS and Y.-K. Wang as noted). Further discussion was held in sesssion 15b at 2100-2300 and 2330 on Friday 23 April 2021 – 0020 on Saturday 24 April 2021 (chaired by GJS) and in sesssion 18b at 1650 on Monday 26 April 2021 (chaired by GJS).

[JVET-V0058](https://jvet-experts.org/doc_end_user/current_document.php?id=10706) AHG9: On post-filter SEI [M. M. Hannuksela, E. B. Aksu, F. Cricri, H. R. Tavakoli (Nokia)]

This contribution was discussed in sesssion 9 at 1400 on Thursday 22 April 2021 (chaired by Y.-K. Wang).

The contribution proposes a post-filter nesting SEI message, which contains inter-related SEI messages jointly specifying a post-filtering operation.

It is envisioned that the nested SEI messages could include the following:

* An SEI message specifying the purpose for filtering, such as quality improvement or super-resolution.
  + This SEI message could also contain parameters specific to the purpose, such as an applicable scaling ratio range for a super-resolution filter.
* An SEI message specifying the filter itself. Such an SEI message could for example contain or reference (through a URI) an MPEG Neural Network Representation (NNR) bitstream.

This contribution leaves the design of the nested SEI messages to other contributions.

Question: What's the usefulness of the pfn\_id syntax element?

Proponent's answer: For supporting of multiple instances of the same post filter, e.g., to have different persistency scopes.

Question: Which SEI messages could be contained in the post-filter nesting SEI message?

Proponent's answer: It seems better to define this in an open-ended manner, i.e., not to preclude the inclusion of SEI messages to be specified in the future.

Question: Is it possible to nest multiple NN topology SEI messages for the same set of NN parameters?

Proponent's answer: Yes.

Question: Is it mandatory to nest an SEI message or is it allowed to be present non-nested?

Proponent's answer: Yes.

Comment: Different pfn\_id values should not be associated with the same set of SEI messages with the same content.

Comment: Nesting is not necessarily the solution for avoiding extending SEI messages.

Proponent: If extending SEI messages is allowed, some problems this contribution tries to address will go. However, it seems good to split purpose and other aspects into separate SEI messages.

Comment: Nesting can add some additional burden for extracting the content of a nested SEI message or remove some of the nested SEI messages.

Proponent: It is allowed for such SEI messages non-nested.

Comment: In the past years, the trend has been making filters in-loop, which leaves less room for post filters. Then why do we need such a generic mechanism for various post filters?

Comment: It can be useful to have NN filters as post filters.

Comment: Study is needed even for NN filters, whether it's better to have them as in-loop or post-processing, including implementation aspects, etc.

Comment: This nesting mechanism is not for associating something to multiple applicable layers, OLSs etc., but just to group different things that work together.

Comment: It seems that we should decide whether to have such a generic nesting mechanism only when studies show that there is such a need.

Further study was encouraged.

This contribution was further discussed in sesssion 18b at 1650 on Monday 26 April 2021 (chaired by GJS).

Wanting to extend the existing SEI message is part of the motivation for this.

This is more than just adding modes to the existing SEI message. Even if we choose not to extend the syntax of existing SEI messages, this contribution proposes additional action. Two primary issues are:

* Would we extend SEI messages? This may depend on whether we’re talking about a future SEI message or an existing one.
* We would need clear evidence of a need of a generic nesting mechanism (e.g., informed by NN-based filtering studies)

There was discussion on this aspect and it was intended to have a clear decision a group: Do we continue to stick to the design principle of not extending SEI messages?

Discussed in joint meeting on Tuesday. It was suggested to continue with the intent to not significantly extend existing SEI messages (at least if there was not a clear plan to do so that was established in advance, and that the existence of a reserved value is not sufficient to indicate such a plan). A NOTE in the text to describe future plans could be an approach to consider for new SEI messages.

Further study on the technical aspects of the proposal

[JVET-V0061](https://jvet-experts.org/doc_end_user/current_document.php?id=10709) AHG9: Display orientation information SEI message [Y. He, M. Coban, M. Karczewicz (Qualcomm), J. Boyce (Intel)]

This contribution was discussed in sesssion 9 at 1600 on Thursday 22 April 2021 (chaired by GJS).

This contribution proposes the inclusion of display orientation information in the VSEI specification. The display orientation SEI message informs the decoder of a transformation that is recommended to be applied to the cropped decoded picture prior to display.

Experiment results were reported to show coding efficiency impact of rotation, although was not emphasized as the rationale for the message.

The proponent said cell phones often use the same encoding regardless of camera orientation.

This does not propose fine granularity rotation angle support, although those were supported in HEVC & AVC. Its transformations are aligned with the region-wise packing SEI message.

It also proposes a dor\_constituent\_picture\_matching\_flag for combination with a frame packing arrangement SEI message. This is specified similarly in the region-wise packing SEI messages.

The combination with frame packing was questioned. A participant suggested not to put a special provision for frame packing into this SEI message.

It was asked what payload type should be used. Using the same value as in AVC and HEVC was preferred.

It was asked whether thought is needed for interactions with other features such as GCMP and sphere rotation. This is for futher study. For now, we would disallow the combination with 360° SEI messages (sphere rotation, GCMP, ERP).

Decision (additional SEI functionality): Adopt (into CD output of this meeting) without the frame packing related flag.

Further study was encouraged for interactions, restrictions and refinements.

[JVET-V0062](https://jvet-experts.org/doc_end_user/current_document.php?id=10710) AHG9: Picture quality metrics SEI message [Y. He, M. Coban, M. Karczewicz (Qualcomm)]

This contribution was discussed in sesssion 9 at 1640 on Thursday 22 April 2021 (chaired by GJS).

This contribution proposes the inclusion of picture quality metrics in the VSEI specification. The message provides a quality metric for each colour component of the current decoded picture.

Picture quality metrics are used to measure a vareity of picture distortions during the acquisition, processing, compression, storage, transmission and reproduction.

ISO/IEC 23001-10, carriage of timed metadata metrics of media in ISOBMFF, specifies the quality metrics of media such as PSNR, SSIM, MS-SSIM, VQM, PEVQ and MOS in the file format.

The quality relevant quality ranking is specified in ISO/IEC 23090-2 OMAF to indicate a relative quality order of quality ranking region. The value of the ranking may be derived from specific quality metrics, but the standard does not indicate whether this is the case and does not indicate the use of a particular metric.

ISO/IEC 23090-6, Immersive Media Metrics, also uses quality ranking to measure the viewport switching latency.

It was commented that the “green MPEG” standard also contains some quality indicators, saying “xsd\_metric\_type - indicates the type of the objective quality metric as shown in the table below. PSNR, as defined in ISO/IEC 23001 10, is the only type currently supported”.

Some objective quality metrics such as PSNR and SSIM can only be obtained at the encoder side, a SEI message to carry such information is able to provide the relevant information to the application. In addition, such SEI message may also be used for the picture coding performance indication and verification, similar as picture HASH SEI message.

The proposal has per colour component aspects that differ somewhat from ISO/IEC 23001-10.

The proposal is to support PSNR, SSIM and MS-SSIM.

The proposal did not contain precise definition of these metrics.

There would need to be a selection of exactly which metrics, and which variations of such metrics, should be supported. The list of potential metrics to consider might grow over time, and fully specifying the metrics could be a challenge.

There was some questioning of detail, such as the precision that was selected and the naming of syntax elements.

It was noted that the correctness of the number cannot, in general, be verified by the decoder. It was asked whether there could (or should) be a constraint that there needs to be some possible “original” picture that would produce the value.

It was commented that the true reference might be at a different resolution, such that the represented quality value might be something measured at that other resolution. For example, one might want to compare the fidelity of a low resolution encoding relative to a high resolution original picture.

It was commented that the specification in ISO/IEC 23001-10 may be somewhat underspecified or undesirable for some aspects.

The proposal is per-frame information. Having metrics that span multiple pictures could potentially have value, and possibly area-specific measurements. Another thing which ISO/IEC 23001-10 allows you to do is to group multiple metrics in the same sample, potentially this SEI message could also be designed to group multiple metrics together. The file format “groups” all metrics together (using a loop). This proposal only supports a single metric per SEI message.

Further study of exactly how this would be used was suggested.

This contribution was further discussed in sesssion 18b at 1715 on Monday 26 April 2021 (chaired by GJS).

It was commented that there is a related incoming liaison statement to AG5 from ITU-T SG 12.

This was further discussed in a joint meeting with AG5 (at 1400 on Tuesday 27 April).

The “reduced reference” quality metric type was also mentioned.

It was noted that ITU-T SG 12 (Q14) defines video quality metrics, and it recently had sent a liaison letter to AG5. However, it was commented that we may be at too early a stage to communicate any conclusions on this matter.

Further study was encouraged for candidate metrics and the various identified issues.

[JVET-V0063](https://jvet-experts.org/doc_end_user/current_document.php?id=10711) AHG9: On the scalability dimension information SEI message [Y. Wang, Y.-K. Wang, L. Zhang (Bytedance)]

This contribution was discussed in sesssion 9 at 1300 on Thursday 22 April (chaired by GJS).

This SEI message is under development for v2 of VVC+VSEI.

This contribution proposes the following changes to the scalability dimension information SEI message:

1. Change sdi\_view\_id\_len to sdi\_view\_id\_len\_minus1.
2. Change the condition for inference of sdi\_view\_id\_val[ i ] equal to 0 from "when not present" to "when sdi\_multiview\_info\_flag is equal to 0". Offline study of this aspect was conducted. This was further discussed in sesssion 18b at 1720 on Monday 26 April 2021 (chaired by GJS) and it was reported that this ID value is not used when it is inferred, so it was agreed to just remove the inference rule.
3. Add inference of sdi\_aux\_id[ i ] to be equal to 0 when sdi\_auxiliary\_info\_flag is equal to 0.
4. Add the following constraints to the semantics of scalability dimension information SEI message:
   1. When an SDI SEI message is present in any AU of a CVS, an SDI SEI message shall be present for the first AU of the CVS.
   2. All SDI SEI messages that apply to the same CVS shall have the same content. It was agreed that this should refer to presence in a CVS, not “applying to the” CVS.
   3. An SDI SEI message shall not be contained in a scalable nesting SEI message.
   4. It is a requirement of bitstream conformance that sdi\_max\_layers\_minus1 shall be equal to vps\_max\_layers\_minus1. It was asked what could happen if a version 1 bitstream editor processes a version 2 bitstream, removes layers from it, and changes the value of vps\_max\_layers\_minus1. This is a similar issue as for when we added bitstream manifest and prefix SEI messages to HEVC. This was further discussed in sesssion 18b at 1725 on Monday 26 April 2021 (chaired by GJS) and it was suggested by a participant to include the layer ID in the loop and not specify a requirement of matching the VPS syntax element. This suggestion was agreed. It was commented that in the VVC context there is no good reason to change layer IDs; we only need to worry about removal of layers, and changes of layer ID would have other undesired side effects.

The text changes for the proposals in JVET-V0063 (this contribution), JVET-V0064, and JVET-V0065 were included in the attachment to JVET-V0063.

It was remarked that any persistence beyond the current CVS could cause a problem with bitstream splicing. It was thus agreed that there should be no such persistence.

Decision (cleanup): Adopted as modified by the four “agreed” aspects noted above.

[JVET-V0064](https://jvet-experts.org/doc_end_user/current_document.php?id=10712) AHG9: On the MAI, DRI, and ACI SEI messages and their interactions with the SDI SEI message [Y.-K. Wang, Y. Wang, L. Zhang (Bytedance)]

This contribution was discussed in sesssion 9 at 1335 on Thursday 22 April (chaired by GJS).

This contribution proposes the following changes on the multiview acquisition information (MAI), depth representation information (DRI), and alpha channel information (ACI) SEI messages and their interactions with the scalability dimension information (SDI) SEI message:

1. It is specified that the MAI SEI message persists in decoding order from the current AU until the next AU containing an MAI SEI message for which the content differs from the current MAI SEI message or the end of the bitstream.

The following constraints are further specified:

* 1. When an MAI SEI message is present in any AU of a CVS, an MAI SEI message shall be present for the first AU of the CVS.
  2. All MAI SEI messages that apply to the same CVS shall have the same content.
  3. An MAI SEI message shall not be scalable-nested.

Similar as with JVET-V0063, it was agreed that there should be no persistence beyond the CVS.

It was remarked that HEVC already has these SEI messages, and should be checked for persistence issues. This was further discussed in sesssion 18b at 1730 on Monday 26 April 2021 (chaired by GJS) and it was concluded that no action is needed for HEVC.

It was remarked that nesting is allowed in HEVC. The idea was said to be to make the message parsable without the VPS. Another suggested reason was to support pruning of views. Thus no change should be made that affects HEVC in this regard. However, the changes remained proposed for the VVC context.

Discussion stopped here in session 9 at 1400 and resumed at 1525 (chaired by GJS).

1. Add following constraints on the presence of the MAI, DRI, and ACI SEI messages depending on the SDI SEI message, which seemed straightforward:
   1. When a CVS does not contain an SDI SEI message, the CVS shall not contain an MAI SEI message.
   2. When a CVS does not contain an SDI SEI message with sdi\_aux\_id[ i ] equal to 2 for at least one value of i, no picture in the CVS shall be associated with a DRI SEI message.
   3. When a CVS does not contain an SDI SEI message with sdi\_aux\_id[ i ] equal to 1 for at least one value of i, no picture in the CVS shall be associated with an ACI SEI message.
2. Add following constraints on the decoding order of the MAI, DRI, and ACI SEI messages relative to the SDI SEI message, which seemed straightforward:
   1. When an AU contains both an SDI SEI message and an MAI SEI message, the SDI SEI message shall precede the MAI SEI message in decoding order.
   2. When an AU contains both an SDI SEI message with sdi\_aux\_id[ i ] equal to 2 for at least one value of i and a DRI SEI message, the SDI SEI message shall precede the DRI SEI message in decoding order.
   3. When an AU contains both an SDI SEI message with sdi\_aux\_id[ i ] equal to 1 for at least one value of i and an ACI SEI message, the SDI SEI message shall precede the ACI SEI message in decoding order.
3. The variable NumViews is derived in the semantics of the SDI SEI message and used in the MAI SEI message syntax and semantics, and the use of the array nestingLayerIdList is removed from the semantics of the MAI SEI message.

It was commented that this could create a parsing dependency between SEI messages. There was discussion of whether that is a problem or not. Picture timing depends on buffering period, for example. However, it was agreed to instead simply signal the number of views and constrain it to match the other message content. This is whole-CVS level info, so not so critical for data overhead.

1. The associated primary layers for an auxiliary layer are signalled in the SDI SEI message syntax and specified in the SDI SEI semantics, and the related texts in the semantics of the DRI and ACI SEI messages are changed accordingly based on the established association. This includes enabling an association of an aux layer to multiple non-aux (primary) layers.

The text changes for the proposals in JVET-V0064 (this contribution) are included in the attachment to JVET-V0063.

Decision (cleanup): Adopt with MAI persistence scope modified to only the containing CVS and to signal the number of views in the MAI SEI message.

[JVET-V0065](https://jvet-experts.org/doc_end_user/current_document.php?id=10713) AHG9: On the DRAP and EDRAP indication SEI messages [Y.-K. Wang, L. Zhang, Y. Wang (Bytedance)]

This contribution was discussed in sesssion 9 at 1545 on Thursday 22 April (chaired by GJS).

This contribution proposes the following changes to the dependent random access point (DRAP) indication and the extended DRAP indication SEI messages:

1. The semantics of the DRAP indication SEI message is changed such that the semantics is clear when the SEI message is applied in the context of a multi-layer bitstream.
2. Change edrap\_rap\_id\_in\_clvs to edrap\_rap\_id\_minus1 and the value plus 1 specified the random access point (RAP) picture ID.
3. The semantics of the EDRAP SEI message is changed to allow two IRAP pictures to have the same RapPicId value, and also to allow two EDRAP pictures associated with different IRAP pictures to have the same RapPicId value, while disallowing RAP pictures (IRAP or EDRAP pictures) within an IRAP period to have the same RapPicId value.
4. The semantics of edrap\_ref\_rap\_id[ i ] is changed to clarify that a RAP picture that may be referenced by a current EDRAP picture shall be either the IRAP picture associated with the current EDRAP picture or an EDRAP picture associated with the same IRAP picture as the current EDRAP picture.

The text changes for the proposals in JVET-V0065 (this contribution) are included in the attachment to JVET-V0063.

Decision (cleanup & expression of existing intent): Adopt.

[JVET-V0069](https://jvet-experts.org/doc_end_user/current_document.php?id=10717) AHG9: Decoded GDR clean area hash SEI message [L. Wang, S. Hong, K. Panusopone, M. M. Hannuksela (Nokia)]

This contribution was discussed in sesssion 9 at 1710 on Thursday 22 April 2021 (chaired by GJS).

This contribution asserts that only the clean (or refreshed) areas of a GDR picture and the associated recovering pictures are important to meet the exact match requirement when starting the decoding from the GDR picture. Hence, it is proposed that a decoded GDR clean area hash SEI message is specified so that an encoder can generate and indicate hashes for the clean areas for GDR pictures and recovering pictures. Consequently, when starting the decoding from a GDR picture, a decoder would only need to check hashes for the clean areas for GDR pictures and recovering pictures.

Two options for the decoded GDR clean (or refreshed) area hash are proposed, summarized as follows:

1. Decoded GDR clean area hash SEI message that contains a decoded\_picture\_hash( ) syntax structure derived from the clean area only.

The clean area is defined through the virtual boundary indicated in the picture header: it is either the left side of the picture (when a single vertical virtual boundary is in use in the picture) or the top of the picture (when a single horizontal virtual boundary is in use in the picture).

1. Decoded GDR clean area hash SEI message that contains

* a vertical or horizontal boundary position defining the clean area, and
* a decoded\_picture\_hash( ) syntax structure derived from the clean area only.

The clean area is either the left side of the picture (when a vertical boundary is indicated in the SEI message) or the top of the picture (when a horizontal boundary is indicated in the SEI message).

The area is proposed to be defined in one of two possible ways - one option being to use the virtual boundary line and the other is to signal the position of a boundary line in units of 8 luma samples. The clean area is assumed to always be on the top (or left) side of a horizontal (or vertical) boundary.

It was asked whether the purpose of this is for debugging or actual application use. The proponent said it could be for either.

In addition to providing the ability to check the hash, the SEI message would define where the clean area is, such that a decoder might choose to display the clean area and not display the other part of the decoded picture. It was commented that, even without the SEI message, if a virtual boundary is indicated, one might infer the same area to be clean (especially if the virtual boundary is moving in a progressive fashion).

Further study was encouraged.

[JVET-V0113](https://jvet-experts.org/doc_end_user/current_document.php?id=10762) Thoughts on SEI messages [W. Husak (Dolby)]

(Initially rejected due to header problem)

This contribution was discussed in sesssion 15b at 2100 on Friday 23 April 2021 (chaired by GJS).

SEI messages associated with video streams can be used to assist in processes related to decoding, display or other purposes. SEI messages are considered optional for decoder usage, such that a conforming decoder is not required to process the information since the decoder conformance point is at the decoded picture buffer. As has been specified, the expected operation, if one exists, would occur after the output and the storage in the picture buffer and thus is outside the traditional domain of ISO/IEC/ITU coding standards.

It is not unusual for organizations and applications standards development organizations (apps SDOs) to select some SEI messages to transmit necessary information. These SEI messages may be optional or even mandatory depending on the problem to be solved and the reasons for inclusion. The requirement for inclusion of these SEI messages implies the conformance points for the applications-specific decoder is beyond the decoded picture buffer.

Periodically MPEG and VCEG reconsider their positions on SEI messages – specifically whether some SEI messages should be required to be supported or remain optional. This document will raise some issues to help guide the discussion amongst the broader group.

Aspects of discussion:

* Rec. ITU-T T.35 usage for closed captioning
* What is the need for an SEI message(s) to have associated decoder processing requirements? Broad industry support?
* Is the information described in the SEI message a post-process that is required for proper display? (examples – 3D, AR, VR).
* Does the information improve the decoded image? (examples – film grain)
* Would the SEI message(s) be required to be supported by a decoder in a specific profile and/or the broader profile(s)?
* Could the SEI be profiled where some parts are required to be supported and some optional??
* SEI utility discussion appears every standard cycle
* There has already been discussion of the basic issue at recent meetings. In principle, we have said we are willing ton consider such a concept – e.g., when there is a significant differentiation in the user experience based on whether an SEI message is used or not, and an ability to indicate conformance is considered desirable. However, our reflex tends to be not doing that.
* It was suggested that we should look on a case-by-case basis at particular circumstances for particular use cases.
* It was commented that SEI messages provide a toolbox of tools that can be selected by app SDOs, and that can be an appropriate approach – for the App SDOs to choose the features that are important to their applications and use cases.
* VUI is another example – considered more fundamental in some scenarios than in others (e.g., in MIV, the interpretation is not even always to just use the samples as video output).
* Scope is an issue – potentially having system requirements creeping into the video layer.
* Required to be signalled versus required to be used in a decoder.
* There could even be more than one conformance point for a decoder – e.g., one with and one without some extra functionality (not a universally supported idea, but mentioned).
* Conformance would need to be specified for the additional functionality.
* We tend to like bit-accurate specification of results.
* We can provide information to clarify what we intended for how the SEI messages are expected to be used – example use cases, example content, technical reports (e.g., cube map), informative remarks in the standard, informative annexes. In some cases our text could go further in terms of providing examples and information in addition to bare specification requirements.
* Note prior technical reports:
  + [H.Sup15](https://www.itu.int/rec/T-REC-H/recommendation.asp?lang=en&parent=T-REC-H.Sup15) *Conversion and coding practices for HDR/WCG Y'CbCr 4:2:0 video with PQ transfer characteristics*
  + [H.Sup18](https://www.itu.int/rec/T-REC-H/recommendation.asp?lang=en&parent=T-REC-H.Sup18) *Signalling, backward compatibility and display adaptation for HDR/WCG video coding*
* It’s very hard to specify conformance when display is involved - everything gets processed differently and displayed differently using differing and continually evolving technology (frame rate conversion, post-proc, …). On the other hand, it would be possible to make a distinction between the display process and some intermediate point in the chain.

No action taken at this time; further study in particular cases could identify potential needs.

[JVET-V0071](https://jvet-experts.org/doc_end_user/current_document.php?id=10719) AHG9: Temporal sublayer information SEI message [R. Sjöberg, M. Pettersson, M. Damghanian, J. Ström (Ericsson)]

This contribution was discussed in sesssion 15b at 2210 on Friday 23 April 2021 (chaired by GJS).

This contribution is a follow-up on JVET-U0045 and proposes that a new temporal sublayer information SEI message is included in the VSEI version 2 specification and proposed to be supported by VVC version 2.

It is claimed that it is useful to have an SEI message for providing detailed temporal sublayer information. It is also claimed that the proposed SEI message is useful for providing spatial, SNR and multiview scalability support for VVC decoders that support Main 10 (4:2:0 or 4:4:4) but do not support any of the Multilayer Main 10 (4:2:0 or 4:4:4) profiles. Scalability in this proposal is supported by treating temporal sublayers as scalability layers.

The proposed SEI message contains information for temporal sublayers and is described by the proponents to contain the following main elements:

1. *Direct reference sublayer flags.* These flags describe which sublayers are used as direct references for a sublayer. The flags are similar in spirit to the vps\_direct\_ref\_layer\_flag syntax element in VVC but describes sublayers rather than layers.
2. *Output sublayer sets (OSSs).* Similar to output layer sets (OLSs), these describes output sets, but instead of basing them on layers as in VVC, this proposal bases them on sublayers. A syntax element for indicating a preferred OSS is also included in the proposal.
3. *Picture output width and height.* To assist spatial scalability, signaling of an intended picture output width and height for each sublayer is included in the SEI message. The signaling is optional, and when used it also indicates that the bitstream is a spatial scalability bitstream and that the output picture width and height is kept constant within sublayers.
4. *View ID and aux ID.* To support multiview scalability and auxiliary pictures, view ID and aux ID values may be signalled for each sublayer. This is similar to the Scalability dimension information SEI message, but in this proposal the information is signalled per sublayer.

The proponents point out that the SEI message can be viewed as an SEI message that contains VPS layer information, but instead of layer information the SEI message carries sublayer information.

The persistence of the SEI message is claimed to be similar to the design in the Scalability dimension information SEI message using a bitstreamInScope.

The proponents claim that the required level for scalable use-cases does not need to be higher compared to the level needed for the Multilayer Main 10 profile.

This is sort of using temporal sublayers as spatial layers, as a way to “tunnel” scalability through a decoder the does not support a scalable profile.

Prior meeting report content:

This contribution proposes a new picture output suppression SEI message to be included in the VSEI specification and to be supported by VVC version 2.

It is claimed that the proposed SEI message is useful for providing spatial and SNR scalability support for VVC decoders that support the Main 10 (4:2:0 or 4:4:4) profile but do not support any of the Multilayer Main 10 (4:2:0 or 4:4:4) profiles. Spatial and SNR scalability are proposed to be supported by treating temporal sublayers as scalability layers and using the picture output suppression SEI message to indicate which temporal sublayers, or which specific pictures, to not output.

The idea is using temporal sublayers with different resolution for invoking spatial scalability (when using different spatial resolution per sublayer and RPR) or SNR scalability, without using a multilayer profile.

It was commented that the method might have difficulty with more than two layers. It is typical for a receiving system or “middle box” to decide what to trim. Which layer/OLS to output might not be the business of the video bitstream (or SEI message) to determine.

It was commented that suppression of normatively specified output could be interpreted as altering normative behaviour.

It was also commented that this seems to be trying to create an alternative to a functionality already specified in the standard, and that perhaps decoders should just be supporting scalable profiles rather than trying to find a substitute scheme for doing the same thing (with less complete functionality). The proponent said this scheme could be easier to support in a decoder.

It was commented that if we want to do something to provide some such functionality, this might not be the best way to approach it, saying that perhaps something indicating the scalability use case more clearly would be better, rather than just decoding everything and suppressing output of some pictures.

It was pointed out that it is common in scalability applications for it to be decided at the decoder which layer to output.

It was pointed out that VVC already has a picture not output flag.

Furthermore, the decoding complexity would be practically identical to the multi-layer case, but with less flexibility.

An SEI message targeting this should better describe that this is a stream which would have different spatial resolution at different sublayers where the decoder output is not useful for display as a single sequence.

It was also pointed out that this might also be applicable for other use cases such as stereo/multiview.

The output suppression aspect proposed at the last meeting is not in the current contribution. Here there is a description of the bitstream rather than active control.

A key question is whether we would have an interest in a “substitute scalability approach” as an alternative to implementation of the scalable profiles which are already specified.

It was commented that if a system has scalability functionality support, the application should just use the scalable profiles.

The approach is interesting and plausible, but no non-proponent support was expressed for pursuing such a scheme.

[JVET-V0108](https://jvet-experts.org/doc_end_user/current_document.php?id=10756) AHG9: Colour Transform Information SEI message [E. François, M. Radosavljevic, P. de Lagrange, F. Le Léannec (InterDigital)]

This contribution was discussed in sesssion 16b at 2330 on Friday 23 April 2021 (chaired by GJS).

This contribution describes additional study of a new SEI message named colour transform information (CTI) SEI message, proposed in JVET-U0078. The SEI message signals mapping parameters that can apply to the decoded pictures. The mapping can operate intra-component (for luma and chroma) or inter-component (for chroma). In this contribution, the CTI has been tested as a mandatory process for out-of-loop mapping for HDR-PQ content. Tests without and with LMCS were reported. The best results were reportedly obtained with CTI and LMCS both applied. BD-rate variations compared to VTM12.0, for the CTCs HDR PQ content, are reported in the table below, for AI, RA, LB and LP configurations.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | DE100 | PSNR-L | wPsnrY | wPsnrU | wPsnrV | psnrY | psnrU | psnrV | EncT | DecT |
| AI | -0.46% | -0.91% | -0.30% | -5.27% | -6.21% | 0.40% | -1.79% | -2.15% | 106% | 99% |
| RA | -2.30% | -3.31% | -1.74% | -5.65% | -7.35% | -1.06% | -1.32% | -1.99% | 102% | 96% |
| LB | -5.85% | -3.96% | -2.57% | -9.52% | -14.06% | -1.89% | 0.69% | -3.25% | 100% | 92% |
| LP | -6.68% | -3.83% | -2.49% | -9.90% | -14.50% | -1.86% | 1.18% | -3.40% | 103% | 100% |

\* LB and LP configuration is not part of the HDR CTCs. (This was suggested to be just to make the simulations faster.) It was pointed out that the two LD configurations are not included in some of the other “specialized” CTC, such as 360, 444.

The proponent emphasized that low delay is an important use case.

No benefit (or not much benefit) was reported for HLG content and SDR content.

This uses a static mapping function (non-varying in time). LMCS, as tested, has a mapping function that varies over time. This is operated outside the loop.

It was noted that we do not have the CRI SEI message in VVC/VSEI. This proposal is inspired from CRI but includes cross-component operations that are not available in CRI. It is somewhat of a superset of CRI.

The CTI mapping function used that was designed for HDR PQ in LMCS. It uses a static mapping function based on a QP mapping function in the HM (since 2015). The LMCS mapping function in HDR mode was reported to be transfer function based (BT.2020 to SMPTE 2084).

When tested in combination with LMCS, the LMCS is configured in SDR mode, reportedly since the preprocessing reportedly converts the signal statistics to something resembling SDR.

The joint session at the previous meeting concluded that, regarding extensions, e.g., normative post-processing (e.g., JVET-U0078, JVET-U0100) – investigation can proceed in JVET.

It was commented that this seems really more like a coding tool proposal than an SEI message proposal.

The proposal could also be used without the preprocessing. It is somewhat analogous to CRI.

There are two possible applications:

* Display adaptation, e.g., SDR to HDR mapping or vice versa (e.g. as investigated with somewhat different technology in 2016 – e.g., as with CRI already standardized)
* Coding gain with preprocessing (in which scenario, it becomes basically mandatory for the decoder to support – this was also somewhat discussed circa 2016).

The non-technical aspects were further discussed in joint meeting on Tuesday.

It was agreed that we could specify such an SEI message with a similar status and usage description as the CRI SEI message in HEVC - described as a post-decoding remapping for such purposes as display adaptation without describing a coding efficiency transformation usage.

Future study would be needed to consider whether it could be part of a profile requirement.

Follow-up discussion in JVET session 23:

The SEI proposal is considered technically mature and provides benefit beyond the CRI SEI. Additional question that had been put by the last meeting (combination with in-loop LMCS, complexity analysis) have been answered satisfactory.

Support was also expressed by non-proponents

Decision: Adopt JVET-V0108 for VSEI v2. Payload type might be considered ro signal this as a superset of CRI. Software is attached with the proposal, this should include all components required to make it “ready to use”.

[JVET-V0152](https://jvet-experts.org/doc_end_user/current_document.php?id=10813) Crosscheck of JVET-V0108 (AHG9: Colour Transform Information SEI message) [F. Pu (Dolby)] [late]

## Non-SEI HLS aspects (4)

Contributions in this area were initially discussed in session 16b at 0030–0120 UTC on Saturday 24 April 2021 (chaired by GJS).

[JVET-V0060](https://jvet-experts.org/doc_end_user/current_document.php?id=10708) Constrained RASL encoding for bitstream switching [R. Skupin, C. Bartnik, A. Wieckowski, K. Suehring, Y. Sanchez, B. Bross (HHI)]

This contribution was discussed in sesssion 16b at 0030 on Saturday 24 April 2021 (chaired by GJS).

VVC features reference picture resampling (RPR) which was investigated in a CE for conversational use cases with LD configuration. In such a setup, RPR allows avoiding sending bitrate-costly Intra coded pictures for resolution changes.

This contribution investigates the benefit of applying RPR for resolution switching in HTTP streaming use cases, also known as open GOP drift switching, by encoding RASL pictures in a manner allowing client-side RPR with other reference pictures than used on encoder side. The coding efficiency benefit of unconstrained open GOP coding over closed GOP coding is reported to be between -9,22% BD-rate for short IRAP periods and -2.35% for longer IRAP periods.

It is reported that CCLM introduces noticeable artefacts in the chroma components when reference pictures are exchanged due to bitrate or resolution switching which is assumed to be attributable to the underlying linear model relying on extrema of neighbouring samples for its parameter derivation. CCLM can only be indicated to be inactive through sequence-scope signalling and deactivating CCLM for all pictures leads to substantial coding efficiency losses (3.43% BR-rate in RA according to AHG13).

Therefore, it is proposed to add a GCI flag that indicates that RASL pictures adhere to the RPR related constraints as well as that RASL pictures do no use CCLM. The coding efficiency benefit of the proposed constrained open GOP RASL picture coding over closed GOP coding is reported to be between -8.69% BD-rate for short IRAP periods and -2.20% for longer IRAP periods while avoiding severe artefacts in open GOP drift switching.

The primary issue is CCLM in regard to its signalling only at the SPS and CU level, not at the picture level. Disabling CCLM results in a 3.4% loss.

It was commented that interop would still function without this, although it might generate some artefacts, and that some encoders might have another way to avoid such artefacts, and that encoders could self-impose such constraints even without a flag to indicate that they are doing so.

It was commented that we could add information somewhere (e.g., as informative content in the standard) to explain that this set of constraints can be used mitigate such artefacts for open GOP resolution switching. An alternative phrasing in a similar spirit could be to warn that if these constraints are not applied, there could be problems.

Another participant said that such constraints could be documented in implementation guidelines, e.g., in DASH-IF, and that this would be a good place for such issues to be investigated and documented. The proponent said this could be done but they believe it is desirable to have a way to indicate this in the bitstream. If we put the flag in the spec, we would definitely also add an explanation of why we defined this particular combination of constraints.

It was commented that this is a very good study of how resolution switching can be achieved.

An SEI message would also be a possible way to provide such an indication.

It was agreed to further study whether we can identify any other reasonable approach for achieving efficient open-GOP operation with resolution switching, and to consider adopting the flag if no such approach.

[JVET-V0081](https://jvet-experts.org/doc_end_user/current_document.php?id=10729) AHG9: On the decoder initialization information [Hendry, S. Lee, S. Kim (LGE), Y.-K. Wang, K. Zhang, L. Zhang, Y. Wang, J. Xu, Z. Deng (Bytedance)]

This contribution was discussed in sesssion 16b at 0115 on Saturday 24 April 2021 (chaired by GJS).

This contribution proposed modifications to the decoder initialization information that were proposed by JVET-U0083 at the 21st JVET meeting. It was asserted that the proposed signalling in JVET-U0083 helps for decoder to avoid re-initialization; however, it might not be complete.

The proposed modification for the previously proposed decoder initialization information is as follows:

* Instead of just signalling one set of decoder initialization information, which comprises of maximum values of parameters (i.e., max DPB size, max picture size, max bit depth, and max chroma format), allow signalling multiple sets of decoder initialization information.
* For each set of decoder initialization information, include the level information.

The proposed decoder initialization information may be signalled as an extension to the DCI NAL unit or as a new SEI message as suggested by JVET-U0083, although the authors of this contribution prefer that the signalling is conveyed in a new SEI message.

It was commented that the main interop point for a decoder is the level, and was suggested that not too many details should be involved in establishing decoder capabilities.

For JVET-U0083 at the previous meeting, no action by JVET was recommended on the proposed approach at that time, pending study in MPEG Systems.

JVET-V0112 is closely related; see the notes for that contribution.

Sesssion 16b stopped here.

[JVET-V0112](https://jvet-experts.org/doc_end_user/current_document.php?id=10761) AHG9: On Bitstream Properties Signalling for Decoder Initialization [S. Deshpande (Sharp)]

This contribution was discussed in sesssion 18b at 1530 on Monday 26 April 2021 (chaired by GJS).

It is proposed to signal picture storage and picture format related information about CVSs in the bitstream to help the decoder initialization.

It is proposed to signal picture storage and picture format related information about CVSs in the bitstream to help the decoder initialization. In this document this information is called bitstream properties information.

It is asserted that signaling information which provides maximum required decoder resources for decoding a bitstream allows a decoder to allocate enough resources (e.g. memory/ DPB space) which can be used to decode the entire bitstream. Thus, need to deallocate and reallocate memory when decoding each CVS of a bitstream may be avoided. This can also result in lowering the decoder delay when switching decoding from one CVS to another CVS, including in streaming environments. Additionally, this may result in requiring allocation of less memory compared to the maximum required from the signalled profile-tier-level. JVET-U0083 also provides a summary of issues related to decoder initialization and additionally refers to the related discussions in systems, including in Systems for Video AHG.

Two alternative options are proposed in this document for helping decoder initialization:

* Option 1: It is proposed to signal following information in decoding capability information (DCI) (or in a new SEI):
  + Maximum number of luma samples for a picture
  + Maximum number of (luma and chroma) samples for a picture
  + Maximum number of luma samples multiplied by bit depth for a picture
  + Maximum number of (luma and chroma) samples multiplied by bit depth for a picture
  + Maximum value for - number of luma samples for a picture multiplied by maximum required DPB size
  + Maximum value for - number of (luma and chroma) samples for a picture multiplied by maximum required DPB size
  + Maximum value for - number of luma samples for a picture multiplied by maximum required DPB size multiplied by bit depth
  + Maximum value for - number of (luma and chroma) samples for a picture multiplied by maximum required DPB size multiplied by bit depth
  + Maximum picture width
  + Maximum picture height
  + Maximum value of required DPB size
  + Maximum bit depth (minus 8)
  + Maximum chroma format
* Option 2: A list of unique sets of storage and format information is specified from all the OLSs in the bitstream.

Relative to option 2, JVET-V0081 also includes level information.

There was discussion of JVET-V0112 with JVET-V0081 in joint discussion Tuesday at 1315:

* Does additional decoder initialization information of some sort (beyond PTL, VPS, DCI, subprofiles, general constraint flags, as they exist) need to be specified?
* If so, should that be carried in the bitstream (e.g., in DCI or SEI)?

The main issue is said to be decoder configuration initialization, to ensure that decoders are able to allocate sufficient resources without a reconfiguration interruption. One participant said the existing syntax should be enough and adding more syntax features may be complicated and difficult to understand and might not be used by decoding systems, also saying that if such a thing is defined, it should be minimal.

A participant said decoders should not need to allocate maximum resources (e.g., due to picture size); that if resources less than necessary for the maximum capability of the level are sufficient, then having such information provided could save resources.

It was questioned whether the amount of resource savings would really be sufficient to justify trying to add more detailed resource descriptions into the bitstream.

Further study is needed to reach agreement on the need; such study is needed among both systems and JVET experts to determine the potential requirement.

[JVET-V0109](https://jvet-experts.org/doc_end_user/current_document.php?id=10757) Early access to decoded samples for cloud rendering/gaming applications [E. Thomas, A. Gabriel, Y. Shiferaw (TNO)]

This contribution was discussed in sesssion 18b at 1605 on Monday 26 April 2021 (chaired by GJS).

Emerging video applications such as cloud gaming and cloud rendering rely on time critical decoding operations so that user interactions are reflected on the user’s screen as soon as possible. Under such low delay requirement, it asserted that applications could achieve higher performance if decoded samples would be released at an earlier point in time, i.e., prior to the complete decoding of a coded picture. To this end, the contribution proposes to allow the output of partially decoded pictures. A modified VTM decoder is also provided implementing the output of partial decoded pictures.

It is asserted that more tools beyond the proposed technique would be desirable to improve the support of cloud gaming and cloud rendering applications in meeting the strict low delay constraints.

A change of the system interface is proposed. Syntax and semantics are not proposed. High-level syntax could be a topic later.

The primary consideration to address is a systems interface and requirements matter in regard to application needs.

It was asked whether display with “tearing” would actually be envisioned. An example brought up by proponent was tile-adaptive streaming (e.g. for 360° immersive apps) where not all of the video is immediately rendered. This could be a matter of differences in use of the system interface.

In-loop filtering might/would need to be disabled at boundaries to ensure that the decoded samples are final at the time of output.

Treatment of regions as picture boundaries is proposed.

Would there be an expectation of output order of regions? The expectation in the proposal is that output order is the same as slice order.

It was asked whether there would be actual profiles & levels associated with this requirement, or only some SEI metadata. It was commented that “decoder hint” information might be a better approach at this time, as requesting support of new profiles is a major step.

There was discussion of providing this functionality in reference software. It was commented that this might increase maintenance difficulty of the software. The proponent did not suggest integrating it in the the main branch of the software, but rather to potentially make it a subject of further exploration.

There was some questioning of the general applicability of this functionality, as the main benefit is to reduce decode-to-display latency, which is often an implementation matter and involves a number of issues and complications.

It was commented that this capability could provide an additional benefit of pipelining of post-decoding processing.

It was commented that subpicture-based output rather than slice-based output could be an alternative model.

The proponent suggested that there may be more than can be done along the lines suggested in this contribution.

This was discussed in joint discussion Tuesday. Move to section 4 of notes.

Further study is needed for the potential requirement to support such an alternative output model in systems (either as a normatively specified decoding interface change or alternative model supported by supplemental metadata).

# 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.

Monday 26 April 1300-1500:

* Planning of output documents: Standard parts & DoCs, CE & EE descriptions, verification test plan & report?
  + Summary of voting 23091-2: m56377 – Gary looks into DoC
  + Liaison output to JPEG about progress in NN video coding – Gary/Elena, going out via SG16.
  + White paper on VVC – in closing plenary it was concluded to delay to next meeting, work until then on preparing something. It is mentioned that various IEEE papers have been worked on, including a special issue in TCSVT. These will be referenced in the white paper, anyway. Action item: Send link to email reflector. It is mentioned that the MCIF and HHI web pages also maintain lists of references.
* Hybrid meeting in October? An informative poll gave 22% who would consider travelling to participate physically, 50% who would not, and 28% who are undecided.
* …

## Information sharing meetings

Beyond the joint meetings listed below, information sharing sessions with other WGs of the MPEG community were held on Monday 26 April 0500–XXXX, Wednesday 28 April 0500–XXXX, and Friday 30 April 2100–XXXX. The status of the work in the MPEG WGs was reviewed at these information sharing sessions.

## Joint meeting with Q6/16 (VCEG), WG2 MPEG Requirements and WG3 MPEG Systems 1300-1400 Tuesday 27 April

The following topics were discussed in this joint session.

* Decoder initialization information for systems (JVET-V0081, JVET-V0112)
* Output of partially decoded pictures (JVET-V0109)
* Possibility of extending and nesting SEI message (esp. post-filter hint JVET-V0058)
* Colour transform information SEI message (JVET-V0108) – similar issue to prior msgs (FPA, CRI)
* Compositing related SEI message for video decoding interface (VDI) input formatting function
  + No contribution to this meeting; potential study for a future SEI message that would refer to another standard (similar as with green metadata). Study in the Systems context is needed.
* Other SEI - no particular concerns were expressed.

## Joint meeting with AG5 MPEG Visual Quality Assessment 1400-1500 Tuesday 27 April

Topics of discussion:

* Visual quality metrics
  + m56636 on AI-based visual quality metric; further exploration was encouraged as an AG5 activity
  + [JVET-V0062](https://jvet-experts.org/doc_end_user/current_document.php?id=10710) AHG9: Picture quality metrics SEI
* VVC verification testing
  + JVET-V0174 registered but not yet available at time of joint discussion

Visual tests had been conducted over the past month for LD and RA SDR HD and 360° PERP and CMP (GCMP for VVC, PCMP for HM).

The test results only became available during the meeting but were presented.

The configuration for 3 of 4 test sequences had not matched what was planned, in that MCTF had accidentally not been enabled for the HM for these sequences. This was said to be unlikely to have a large effect. It was said that it should not be difficult to re-run these tests to correct this within a few days after the meeting.

It was agreed to do this and include the results as supplemental to the results previously measured.

The testing process had been difficult due to Covid-related restrictions on lab use.

Two labs, more than 50 naïve viewers over 9 days of testing for the HD RA testing. Three test sessions for LD (30, 50 and 60 fps), three for RA (60 fps).

Three test sessions for 360° video.

MOS graphs were shown.

VVenC was also tested in addition to the VTM, and it appeared to show very strong performance. The VTM used CTC settings.

For summary-level results, an RD Plot package available on GitHub (at <https://github.com/IENT/RDPlot>) was used for some of the summary calculations.

The test coordinators, test labs, and others who helped with bitstreams and tools were thanked for their efforts in preparing and conducting these tests.

Further testing is planned for HDR. Work on testing of scalability, screen content, and 4:4:4 was also desired.

## BoGs (5)

The following break-out groups were established at this meeting and produced the below-listed reports.

## Liaison communications

### Communication with XXXX

An incoming liaison letter was received by JVET as ISO/IEC JTC 1/SC 29/WG 5 and by VCEG, as follows:

# Project planning

## Software timeline

VTM13.0 including the adoptions from JVET-V0047, JVET-V0054, JVET-V0056, JVET-V0106: 2021-05-21. Needed for CE

HM16.24 including the adoption from JVET-V0056: 2021-05-21.

VTM13.1 as appropriate date t.b.d. with remaining adoptions of encoder optimization, SEI messages.

## 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-V2005 and JVET-V2006). * 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-V1004 errata output collection. * Produce and finalize JVET-V1002 HEVC Test Model 16 (HM 16) Update 15 and JVET-V2002 VVC Test Model 13 (VTM 13) 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, 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, 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. * Coordinate with AHG6 for integration with 360lib software. * Investigate on the possibility of merging CTC documents for HM and VTM. | F. Bossen, X. Li, K. Sühring (co-chairs), 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-V2020 and the draft test plan JVET-V2021, and develop proposed improvements for 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) | Tel.  2 weeks notice |
| **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 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, K. Sühring, X. Xu (vice-chairs) | N |
| **360° video coding, software and test conditions (AHG6)**  ([jvet@lists.rwth-aachen.de](mailto:jvet@lists.rwth-aachen.de))   * Study the effect on compression and subjective quality of different projections formats, resolutions, and packing layouts. * Solicit additional test sequences, and evaluate suitability of test sequences on head-mounted displays and normal 2D displays. * Study the effect of viewport resolution, field of view, and viewport speed/direction on visual comfort. * Prepare and deliver a 360Lib software version enabling the usage of PCMP with HM, and provide common test condition configuration files. * Generate CTC anchors and PERP results for the VTM according to JVET-U2012. * Coordinate with AHG4 in preparation for verification testing for 360° video content. * Produce documentation of 360° software usage for distribution with the software. * Prepare a cleaned-up version of JVET-T2004, reducing it to the elements which are relevant to enable 360° video applications based on AVC, HEVC and VVC, and related SEI messages. | J. Boyce and Y. He (co-chairs), K. Choi, Y. Ye (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. * 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. | 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. * Discuss and refine the JVET-U2018 testing conditions for high bit depth, high bit rate, and high frame rate coding. * Finalize, conduct and coordinate the work on the core experiment JVET-V2022. * 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 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) | Tel.  2 weeks notice |
| **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. * 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. | A. Duenas, 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-V2023. * 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. * 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-V2016 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) | Tel.  2 weeks notice |
| **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. * Study the performance and complexity tradeoff of these video coding tools. * Define a common software platform for developing and evaluating video coding tools with promising compression performance. * Refine test conditions in JVET-V2017, generate anchors, identify new test sequences to be added, especially high resolution ones in 8K, in coordination with AHG4. * Investigate methods to reduce simulation time. * Analyse the results of exploration experiments described in JVET-V2024 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) | Tel.  2 weeks notice |

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.

For outgoing liaison letters, see section 7.7. 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-V1000](https://jvet-experts.org/doc_end_user/current_document.php?id=10673) Meeting Report of the 22nd JVET Meeting [G. J. Sullivan, J.-R. Ohm] [WG 5 N 30] (2021-05-26)

Initial versions of the meeting notes (d0 … d6) 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)]

[JVET-V1002](http://phenix.int-evry.fr/jct/doc_end_user/current_document.php?id=11000) 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)] [WG 5 N 60] (2021-06-23)

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] [WG 5 DIS N 12)]

Resolution in WG5 to proceed to publication. Only editorial suggestions by ISO secretariat which are dealt with in publication proceed. No Disposition of comments necessary.

To be Consented in ITU-T SG 16.

[JVET-V1004](https://jvet-experts.org/doc_end_user/current_document.php?id=10674) Errata report items for VVC, HEVC, AVC, Video CICP, and CP usage TR [C. Rosewarne, G. J. Sullivan, Y. Syed, Y.-K. Wang] (2021-06-23, near next meeting)

Remains valid – not updated: [JVET-T1005](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10537) Shutter interval information SEI message for HEVC (Draft 3) [S. McCarthy, G. J. Sullivan, Y.-K. Wang] [WG 5 FDAM N 8]

To be Consented by ITU-T SG 16.

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] [WG 5 CDAM N 50]

To be Consented by ITU-T SG 16.

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: [JVET-T1008](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10539) Usage of video signal type code points (Draft 2 for version 3) [W. Husak, G. J. Sullivan, Y. Syed, A. Tourapis] [WG 5 TR N 14]

To be Approved by ITU-T SG 16 in April 2021.

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)

Reserved for future use (new edition): [JVET-T2001](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10540) Versatile Video Coding Editorial Refinements on Draft 10 [B. Bross, J. Chen, S. Liu, Y.-K. Wang] (2020-10-30)

[JVET-V2002](https://jvet-experts.org/doc_end_user/current_document.php?id=10676) Algorithm description for Versatile Video Coding and Test Model 13 (VTM 13) [J. Chen, Y. Ye, S. Kim] [WG 5 N 56] (2021-06-30, near next meeting)

Further editorial improvements.

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. Add as a mandate to AHG6.

[JVET-V2005](https://jvet-experts.org/doc_end_user/current_document.php?id=10677) VVC operation range extensions (Draft 3) [F. Bossen, B. Bross, T. Ikai, D. Rusanovskyy, Y.-K. Wang] [WG 5 CDAM N 52] (2021-05-07)

See adoption notes elsewhere in the meeting report.

[JVET-V2006](https://jvet-experts.org/doc_end_user/current_document.php?id=10678) Additional SEI messages for VSEI (Draft 3) [J. Boyce, Y.-K. Wang] [WG 5 CDAM N 51] (2021-05-07)

For newly adopted SEI messages at the current meeting, see notes elsewhere in the meeting report.

Reserved for future use (new edition): [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] [WG 11 N 19472]

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] [WG 5 DIS N 37] (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] [WG 5 DIS N 39] (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]

[JVET-V2011](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10533) VTM common test conditions and evaluation procedures for HDR/WCG video [A. Segall, E. François, W. Husak, S. Iwamura, D. Rusanovskyy] (2021-05-14)

Include the change suggested in JVET-V0107

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-V2016](https://jvet-experts.org/doc_end_user/current_document.php?id=10669) Common Test Conditions and evaluation procedures for neural network-based video coding technology [S. Liu, A. Segall, E. Alshina, R.-L. Liao] (2021-05-21)

Updates to template, change in anchor for EE1.

[JVET-V2017](https://jvet-experts.org/doc_end_user/current_document.php?id=10682) Common Test Conditions and evaluation procedures for enhanced compression tool testing [M. Karczewicz and Y. Ye] (2021-05-14)

Changes to LD config and screen content coding are needed for EE2.

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] (2021-01-29)

Updates of anchor via config file in the context of CE document.

[JVET-V2020](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10550) VVC verification test report for HD SDR and 360° video content [V. Baroncini, M. Wien] [WG 5 N 54] (2021-06-11)

[JVET-V2021](https://jvet-experts.org/doc_end_user/current_document.php?id=10684) VVC verification test plan (Draft 6 [M. Wien, V. Baroncini, A. Segall, Y. Ye] [WG 5 N 55] (2021-06-11)

Changes: Remove SDR HD & 360, finalization of sequences and settings for HDR, timeline for dry run and testing for HDR.

[JVET-V2022](https://jvet-experts.org/doc_end_user/current_document.php?id=10670) Core Experiment on Entropy Coding for High Bit Depth and High Bit Rate Coding [K. Naser, D. Rusanovskyy, M. G. Sarwer, F. Wang] [WG 5 N 53] (2021-05-14)

An initial draft was reviewed and approved. One expert expressed the opinion that it might be desirable testing more options in CE3.x, e.g. for bypass alignment. To be further discussed in CE definition period.

[JVET-V2023](https://jvet-experts.org/doc_end_user/current_document.php?id=10672) 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 57] (2021-05-14)

An initial draft was reviewed and approved. The technology elements investigated are the same as in the last round. Extended sequences for subjective testing, two different RPR configurations (every frame downsampled, or every second frame). Update of reporting template. A telco for further discussion of viewing preparation was planned.

It was suggested to add an additional RPR method which selectively decides using downsampled coding.

[JVET-V2024](https://jvet-experts.org/doc_end_user/current_document.php?id=10685) 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 58] (2021-05-21)

An initial draft was reviewed and approved. Agreed to shorten sequences for LB and LP to 5 seconds (all classes). It was suggested to call the software ECM (enhanced compression model). V. Seregin and K. Zhang to coordinate the software.

# 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/WG 5 auspices when it meets (ordinarily starting meetings on the Friday prior to such meetings and closing it at lunchtime on the last day of the WG 5 meeting – a total of 7.5 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. 7 – Fri. 16 July 2021, 23rd meeting, online under ISO/IEC SC 29 auspices.
* Fri. 8 – Fri. 15 October 2021, 24th meeting under ISO/IEC SC 29 auspices as a mixed-mode meeting in Antalya, TR.
* During January 2022, 25th meeting under ITU-T SG 16 auspices in Geneva, CH.
* Fri. 22 – Fri. 29 April 2022, 26th meeting under ISO/IEC SC 29 auspices, location t.b.d.
* Fri. 15 – Fri. 22 July 2022, 27th meeting under ISO/IEC SC 29 auspices in Cologne, DE.
* During October 2022, 28th meeting under ITU-T SG 16 auspices in Geneva, CH.
* During January 2023, 29th meeting under ISO/IEC SC 29 auspices, location t.b.d.
* During April 2023, 30th meeting under ISO/IEC SC 29 auspices, location t.b.d.

The agreed document deadline for the 23rd JVET meeting was planned to be Wednesday 30 June 2021.

Alibaba, Broadcom, ByteDance, Chips&Media, DJI, Dolby, Ericsson, Fraunhofer HHI, Huawei, Intel, InterDigital, KDDI, Kwai, LGE, MediaTek, NHK, Nokia, Orange, Panasonic, Qualcomm, Samsung, Sharp, Sony, and Tencent were thanked for their great efforts in generating and testing the VVC conformance bitstreams.

Giacomo Baroncini and Vittorio Baroncini were thanked for conducting, and Mathias Wien was thanked for coordinating the VVC verification test in the categories of HD SDR and 360° video. It is greatly appreciated that this testing was successfully completed despite requiring extraordinary efforts due the complications caused by the pandemic situation.

Bytedance and Tencent were thanked for providing financial support for the VVC verification tests.

Mathias Wien was thanked for organizing and conducting expert viewing sessions related to the exploration experiment on neural network-based video compression.

The 22nd JVET meeting was closed at approximately 1740 hours UTC on Wednesday 28 April 2021.

# Annex A to JVET report: List of documents

# Annex B to JVET report: List of meeting participants

The participants of the twenty-first meeting of the JVET, according to an attendance sheet circulated during the meeting sessions (approximately 349 people in total), were as follows:

# Annex C to JVET report: Recommendations of the 3rd 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 29**