|  |  |
| --- | --- |
| **Joint Video Experts Team (JVET)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29**  21st Meeting, by teleconference, 6–15 Jan. 2021 | Document: JVET-U\_Notes\_d6 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Meeting Report of the 21st Meeting of the Joint Video Experts Team (JVET), by teleconference, 6–15 January 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-first meeting during 6–15 January 2021 as an online-only meeting. It had previously been planned to be in Capetown, ZA, 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 second 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 a new standard. The name Versatile Video Coding (VVC) was chosen in April 2018 as the informal nickname for the new standard.

The JVET meeting began at approximately 0500 hours UTC on Wednesday 6 January 2021. Meeting sessions were held on all days except the weekend days of Saturday and Sunday 9 and 10 January 2021, until the meeting was closed at approximately XXXX hours UTC on Friday 15 January 2021. Approximately XXX people attended the JVET meeting, and approximately XX input documents (not counting crosschecks), 11 AHG reports, 2 CE/EE summary reports, and X BoG reports were discussed. The meeting took place in a collocated fashion with a meeting of various SG16 Working Groups – where WG 5 is representing the Joint Video Coding Team(s) and their activities from the SC 29 parent body. The subject matter of the JVET meeting activities consisted of work on further development and maintenance of the twin-text video coding technology standards *Advanced Video Coding* (AVC), *High Efficiency Video Coding* (HEVC), *Versatile Video Coding* (VVC)*, Coding-independent Code Points (Video)* (CICP), and *Versatile Supplemental Enhancement Information Messages for Coded Video Bitstreams* (VSEI), as well as related technical reports, software and conformance packages. As a primary goal, the JVET meeting reviewed the work that was performed in the interim period since the twentieth JVET meeting in producing the following documents:

* JVET-T1000 Meeting report
* JVET-T1003 Revised coding-independent code points for video signal type identification (Draft 2)
* JVET-T1004 Errata report items for HEVC, AVC, Video CICP, and CP usage TR
* JVET-T1005 Shutter interval information SEI message for HEVC (Draft 3)
* JVET-T1006 Annotated regions and shutter interval information SEI messages for AVC (Draft 2)
* JVET-T1008 Usage of video signal type code points (Draft 2 for version 3)
* JVET-T2001 Versatile Video Coding Editorial Refinements on Draft 10
* JVET-T2002 Algorithm description for Versatile Video Coding and Test Model 11 (VTM 11)
* JVET-T2004 Algorithm descriptions of projection format conversion and video quality metrics in 360Lib (Version 12)
* JVET-T2006 Common Test Conditions and evaluation procedures for neural network-based video coding technology
* JVET-T2008 Conformance testing for versatile video coding (Draft 5)
* JVET-T2009 VVC verification test plan (Draft 4)
* JVET-T2010 VTM common test conditions and software reference configurations for SDR video
* JVET-T2011 VTM common test conditions and evaluation procedures for HDR/WCG video
* JVET-T2013 VTM common test conditions and software reference configurations for non-4:2:0 colour formats
* JVET-T2016 Summary information on BD-rate experiment evaluation practices
* JVET-T2017 Additional SEI messages for VSEI (Draft 1)
* JVET-T2018 Common Test Conditions for High Bit Depth and High Bit Rate Coding
* JVET-T2019 New level and additional SEI messages for VVC (Draft 1)
* JVET-T2020 VVC verification test report for UHD SDR video content
* JVET-T2021 Reference software for versatile video coding (Draft 1)
* JVET-T2022 CE on Entropy Coding for High Bit Depth and High Bit Rate Coding
* JVET-T2023 EE on Neural Network-based Video Coding

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 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-T1003 Revised coding-independent code points for video signal type identification (Draft 2)
* JVET-T1004 Errata report items for HEVC, AVC, Video CICP, and CP usage TR
* JVET-T1005 Shutter interval information SEI message for HEVC (Draft 3)
* JVET-T1006 Annotated regions and shutter interval information SEI messages for AVC (Draft 2)
* JVET-T1008 Usage of video signal type code points (Draft 2 for version 3)
* JVET-T2001 Versatile Video Coding Editorial Refinements on Draft 10
* JVET-T2002 Algorithm description for Versatile Video Coding and Test Model 11 (VTM 11)
* JVET-T2004 Algorithm descriptions of projection format conversion and video quality metrics in 360Lib (Version 12)
* JVET-T2006 Common Test Conditions and evaluation procedures for neural network-based video coding technology
* JVET-T2008 Conformance testing for versatile video coding (Draft 5)
* JVET-T2009 VVC verification test plan (Draft 4)
* JVET-T2010 VTM common test conditions and software reference configurations for SDR video
* JVET-T2011 VTM common test conditions and evaluation procedures for HDR/WCG video
* JVET-T2013 VTM common test conditions and software reference configurations for non-4:2:0 colour formats
* JVET-T2016 Summary information on BD-rate experiment evaluation practices
* JVET-T2017 Additional SEI messages for VSEI (Draft 1)
* JVET-T2018 Common Test Conditions for High Bit Depth and High Bit Rate Coding
* JVET-T2019 New level and additional SEI messages for VVC (Draft 1)
* JVET-T2020 VVC verification test report for UHD SDR video content
* JVET-T2021 Reference software for versatile video coding (Draft 1)
* JVET-T2022 CE on Entropy Coding for High Bit Depth and High Bit Rate Coding
* JVET-T2023 EE on Neural Network-based Video Coding

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 Experiment (EE) were defined. The next four JVET meetings were planned for Wed. 21 – Wed. 28 April 2021 under ITU-T SG16 auspices in Geneva, CH, during Fri. 9 – Fri. 16 July 2021 under ISO/IEC SC 29 auspices in Prague, CZ, during Fri. 8 – Fri. 15 October 2021 under ISO/IEC SC 29 auspices in Antalya, TR, and during Wed. XX – Wed. XX January 2022 under ITU-T SG16 auspices in Geneva, CH.

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

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

# Administrative topics

## Organization

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

The Joint Video Experts Team (JVET) of ITU-T WP3/16 and ISO/IEC JTC 1/ SC 29 held its twenty-first meeting during 6–15 January 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 second 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.

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/2021_01_U_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-T1003 Revised coding-independent code points for video signal type identification (Draft 2)
* JVET-T1004 Errata report items for HEVC, AVC, Video CICP, and CP usage TR
* JVET-T1005 Shutter interval information SEI message for HEVC (Draft 3)
* JVET-T1006 Annotated regions and shutter interval information SEI messages for AVC (Draft 2)
* JVET-T1008 Usage of video signal type code points (Draft 2 for version 3)
* JVET-T2001 Versatile Video Coding Editorial Refinements on Draft 10
* JVET-T2002 Algorithm description for Versatile Video Coding and Test Model 11 (VTM 11)
* JVET-T2004 Algorithm descriptions of projection format conversion and video quality metrics in 360Lib (Version 12)
* JVET-T2006 Common Test Conditions and evaluation procedures for neural network-based video coding technology
* JVET-T2008 Conformance testing for versatile video coding (Draft 5)
* JVET-T2009 VVC verification test plan (Draft 4)
* JVET-T2010 VTM common test conditions and software reference configurations for SDR video
* JVET-T2011 VTM common test conditions and evaluation procedures for HDR/WCG video
* JVET-T2013 VTM common test conditions and software reference configurations for non-4:2:0 colour formats
* JVET-T2016 Summary information on BD-rate experiment evaluation practices
* JVET-T2017 Additional SEI messages for VSEI (Draft 1)
* JVET-T2018 Common Test Conditions for High Bit Depth and High Bit Rate Coding
* JVET-T2019 New level and additional SEI messages for VVC (Draft 1)
* JVET-T2020 VVC verification test report for UHD SDR video content
* JVET-T2021 Reference software for versatile video coding (Draft 1)
* JVET-T2022 CE on Entropy Coding for High Bit Depth and High Bit Rate Coding
* JVET-T2023 EE on Neural Network-based Video Coding

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 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 Wednesday, 30 December 2020. Any documents uploaded after 1159 hours Paris/Geneva time on Thursday 31 December 2020 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-U0107 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-U00XX (a proposal on …), uploaded 01-XX.
* …

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

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

* JVET-U00XX (a document on …), uploaded 01-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 (3) contribution registrations were later cancelled, withdrawn, never provided, were cross-checks of a withdrawn contribution, or were registered in error: JVET-U00XX, … .

“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-T1000, the Revised coding-independent code points for video signal type identification (Draft 2) JVET-T1003, the Errata report items for HEVC, AVC, Video CICP, and CP usage TR JVET-T1004, the Shutter interval information SEI message for HEVC (Draft 3) JVET-T1005, the Annotated regions and shutter interval information SEI messages for AVC (Draft 2) JVET-T1006, the Usage of video signal type code points (Draft 2 for version 3) JVET-T1008, the Versatile Video Coding Editorial Refinements on Draft 10 JVET-T2001, the Algorithm description for Versatile Video Coding and Test Model 11 (VTM 11) JVET-T2002, the Algorithm descriptions of projection format conversion and video quality metrics in 360Lib (Version 12) JVET-T2004, the Common Test Conditions and evaluation procedures for neural network-based video coding technology JVET-T2006, the Conformance testing for VVC (Draft 5) JVET-T2008, the VVC verification test plan (Draft 4) JVET-T2009, the VTM common test conditions and software reference configurations for SDR video JVET-T2010, the JVET common test conditions and evaluation procedures for HDR/WCG video JVET-T2011, the VTM common test conditions and software reference configurations for non-4:2:0 colour formats JVET-T2013, the Working practices using objective metrics for evaluation of video coding efficiency experiments (Draft 4) JVET-T2016, the Additional SEI messages for VSEI (Draft 1) JVET-T2017, the Additional SEI messages for VSEI (Draft 1) JVET-T2017, the Common Test Conditions for High Bit Depth and High Bit Rate Coding JVET-T2018, the New level and additional SEI messages for VVC (Draft 1) JVET-T2019, the VVC verification test report for UHD SDR video content JVET-T2020, the Reference software for versatile video coding (Draft 1) JVET-T2021, the Description of the CE on Entropy Coding for High Bit Depth and High Bit Rate Coding JVET-T2022, and the Description of the EE on Neural Network-based Video Coding JVET-T2023 had been completed and were approved. The software implementations of VTM (versions 11.0) and 360lib (version 12.0) were also approved.

The group was initially asked to review the meeting reports of the previous JVET meeting for finalization. The meeting reports were later approved with a minor modification of the JVET report to include a missing output document in a list.

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. Due to the difficulty of determining how to send the password only to registered participants, the password was simply embedded in links in the posted meeting logistics information and the calendar of meeting sessions. No particular problems were observed that resulted in interference with the meeting due to the lack of strict access control.

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. We will use the participation list for 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, we do not plan to use 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
* 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.

* 0500–0700 1st “morning” session [break after 2 hours]
* 0720–0920 2nd “morning” session
* [“midday” break – nearly 4 hours]
* 1300–1500 1st “afternoon” session [break after 2 hours]
* 1520–1720 2nd “afternoon” session
  1. ***ISO Code of Conduct reminder***

Participants were reminded of the ISO Code of Conduct, found at

<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.

## 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 1270. Furthermore, the JCT-VC email list currently had 1295 subscribers (as of 5 January 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 Wednesday 6 January at 0500 UTC (chaired by GJS and JRO) were as follows.

* Timing and organization of online meetings, calendar
* Standards publication status
  + VVC and VSEI under ballot in ISO/IEC, published in ITU-T (includes refinements from JVET-T2001), both versions aligned
  + Working practices TR
  + HEVC latest edition in ISO
  + HEVC Amd.1 (additional SEI) and Amd.2 (shutter interval)
  + CICP usage 3rd ed. – was TR last meeting (wrong entry PDTR in MPEG progression database)
* Draft standards progression status
  + AVC additional SEI – CDAM next meeting
  + VSEI extensions – CDAM next meeting, new WD
  + VVC operation range extensions – CDAM next meeting, new WD
  + VVC conformance – DIS?
  + VVC reference SW – DIS?
  + CICP new edition – DIS ongoing, FDIS next meeting
* 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
  + Potential extensions of VVC
    - Neural network tools
* Funding of verification testing activities: Thank resolution, resolution calling for funding wrt upcoming tests.
* Liaison to JPEG on AI activities
* Fewer documents than recently, such that it seemed potentially not necessary to 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 (1 hour behind the time in Geneva, Paris; 8 hours ahead of the time in Los Angeles, etc.). No session should last longer than 2 hrs.

* 0500–0700 1st “morning” session [break after 2 hours]
* 0720–0920 2nd “morning” session
* [“midday” break – nearly 4 hours]
* 1300–1500 1st “afternoon” session [break after 2 hours]
* 1520–1720 2nd “afternoon” session

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

* Wed. 6 January, 1st day
  + Session 1:
    - 0500–0550 Opening remarks, review of practices, agenda, IPR reminder
    - 0550–0710 Reports of AHGs 1–4
  + Session 2:
    - 0720–0920 Reports of AHGs 5–11
  + Session 3:
    - 1300-1500 Review of CE/EE and related (5.1.2, 5.1.3, 5.2.2, 5.2.3)
  + Session 4:
    - 1520-1720 Review of CE/EE and related (5.1.2, 5.1.3, 5.2.2, 5.2.3)
* Thu. 7 January, 2nd day
  + Session 5:
    - 0500–0655 Errata items (4.2)
  + Session 6:
    - 0720-0930 HLS SEI (6.1)
  + Session 7:
    - 1300-1500 Verification test planning (4.4) + other issues related to subjective testing (if time allows)
  + Session 8:
    - 1520-1720 Further review neural network technology 5.2
* Fri. 8 January, 3rd day
  + Session 9:
    - 0500–0710 HLS remaining 6.1
  + Session 10:
    - 0730-0920 HLS 6.2
  + Session 11:
    - 1300-1405 Further review neural network EE contributions 5.2.2
    - 1405-1530 Review neural network EE related contributions 5.2.3
    - 1300-1500 BoG on high bit depth and high bit rate (A. Browne)
  + Session 12:
    - 1550-1720 Review neural network EE related contributions 5.2.3
    - 1520-1640 BoG on high bit depth and high bit rate (A. Browne)
* Mon. 11 January, 4th day
  + 0500–0800 MPEG WGs’ information sharing meeting
  + JVET plenary:
    - 0820-0920 BoG review, further planning, review 4.1
  + Session 13:
    - 1300-1400 Further review neural network technology 5.2.4ff (chaired by Y. Ye)
  + 1300-1400 BoG on high bit depth and high bit rate (A. Browne)
  + Joint meetings:
    - 1520-1550 with VCEG/Systems/Video/3DG: SEI for MIV and V3C
    - 1550-1620 with VCEG/Systems: SEI for picture composition and decoder initialization
    - 1630-1740 with VCEG/Requirements: Profiles, extensions of VVC and other SEI
* Tue. 12 January, 5th day
  + Session 14:
    - 0500-0700 Review 4.6-4.8
  + Session 15:
    - 0720-0920 Review remaining 4.8, 4.9, 5.3
  + Session 16:
    - 1300-1500 Review remaining 5.1.x, 5.2.x
  + Session 17:
    - 1520-1720 Review remaining docs, planning CE/EE
* Wed. 13 January, 6th day
  + 0500–0700 MPEG WGs’ information sharing meeting
  + Session 18:
    - 0720-0920 BoG report, further planning, review remaining docs
  + BoG session:
    - 1300-1500 BoG on high bit depth CE (A. Browne)
  + BoG session:
    - 1520-1720 BoG on additional coding tools EE (V. Seregin)
* Thu. 14 January, 7th day
  + BoG session:
    - 0500-0700 BoG on neural network video coding EE (A. Segall)
  + Session 22:
    - 0720-0920 JVET-0093, revisits, BoG reports
  + Session 23:
    - 1300-1500 Remainders, revisits, AHG planning, potential BoG
  + Session:
    - 1520-1720 Potential BoG work
* Fri. 15 January, 8th day
  + Session XX
    - 0500–0700 General wrap-up and remainders
  + Session XX
    - 0720–0920 Review of outputs, actions and planning
  + Session XX
    - 1300–1500 Optional
  + Parent body meetings
    - 2100–2230 MPEG information sharing (parent body matter)
  + Closing session
    - 2330–2350 JVET closing plenary: Final recommendations approval

## 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) – yellow mark means not reviewed yet:

* AHG reports (section 3)
* Project development (section 4)
  + Deployment status (3)
  + Text development and errata reporting (5)
  + Test conditions (0)
  + Verification test (6)
  + Test Material (1)
  + Conformance test development (2)
  + Software development (2)
  + Implementation studies (2)
  + Complexity analysis (0)
  + Encoder optimization (2)
  + Profile/tier/level specification (1) – Joint Meeting
* Low-level tool technology proposals (section 5) with subtopics
  + AHG8: High bit depth and high bit rate coding (19) (section 5.1) 5.1.1 (2), 5.1.4 (5)
  + AHG11: Neural network-based technology (22) (section 5.2) 5.2.1 (1), 5.2.5 (3), 5.2.6 (3)
  + Other compression technology (3) (section 5.3)
* High-level syntax (HLS) proposals (section 6) with subtopics
  + AHG9: SEI message studies and proposals (6) (section 6.1)
  + HLS signalling for specific tools (3) (section 6.2)
* Joint meetings, plenary discussions, BoG reports (1), 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 (11)

These reports were discussed Wednesday 6 January 2021 during 0550–0700 and 0720–0920 UTC (chaired by GJS & JRO), except as otherwise noted.

[JVET-U0001](https://jvet-experts.org/doc_end_user/current_document.php?id=10623) 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 a moderate number of input documents 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 "Phenix" site (http://phenix.int-evry.fr/jvet/) or the ITU-based JVET site (<http://wftp3.itu.int/av-arch/jvet-site/2020_10_T_Virtual/>). It is however noted that the document distribution was migrated to <https://jvet-experts.org/> , and this site was used for distribution of all documents of the current meeting, as well as updated versions of documents from the previous meeting. The previous 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-T2000) [Posted 2020-12-16]
* Revised coding-independent code points for video signal type identification (Draft 2) (JVET-T1003) [Posted 2021-01-04]
* Errata report items for HEVC, AVC, Video CICP, and CP usage TR (JVET-T1004) [Posted 2020-11-12]
* Shutter interval information SEI message for HEVC (Draft 3) (JVET-T1005) [Posted 2020-10-30, revised as noted below]
* Annotated regions and shutter interval information SEI messages for AVC (Draft 2) (JVET-T1006) [Posted 2020-11-24]
* Usage of video signal type code points (Draft 2 for version 3) (JVET-T1008) [Posted 2021-01-05]
* Versatile Video Coding Editorial Refinements on Draft 10 (JVET-T2001) [Posted 2020-10-30, last update 2020-11-24]
* Algorithm description for Versatile Video Coding and Test Model 11 (VTM 11) (JVET-T2002) [Posted 2021-12-11, last update 2020-12-15]
* Algorithm descriptions of projection format conversion and video quality metrics in 360Lib (Version 12) (JVET-T2004) [Posted 2020-12-16]
* Common Test Conditions and evaluation procedures for neural network-based video coding technology (JVET-T2006) [Posted 2020-10-24, last update 2020-12-15]
* Conformance testing for versatile video coding (Draft 5) (JVET-T2008) [Posted 2020-10-31, last update 2020-11-03]
* VVC verification test plan (Draft 4) (JVET-T2009) [Posted 2020-11-17]
* VTM common test conditions and software reference configurations for SDR video (JVET-T2010) [Posted 2020-10-28, last update 2020-11-19]
* VTM common test conditions and evaluation procedures for HDR/WCG video (JVET-T2011) [Posted 2020-10-24]
* VTM common test conditions and software reference configurations for non-4:2:0 colour formats (JVET-T2013) [Posted 2021-01-05]
* Summary information on BD-rate experiment evaluation practices (JVET-T2016) [Posted 2021-01-04]
* Additional SEI messages for VSEI (Draft 1) (JVET-T2017) [Posted 2020-11-24]
* Common Test Conditions for High Bit Depth and High Bit Rate Coding (JVET-T2018) [Posted 2020-10-16, last update 2020-10-30]
* New level and additional SEI messages for VVC (Draft 1) (JVET-T2019) [Posted 2020-10-31, last update 2020-11-24]
* VVC verification test report for UHD SDR video content (JVET-T2020) [Posted 2020-11-19]
* Reference software for versatile video coding (Draft 1) (JVET-T2021) [Posted 2020-10-30]
* CE on Entropy Coding for High Bit Depth and High Bit Rate Coding (JVET-T2022) [Posted 2020-10-16, last update 2020-12-17]
* EE on Neural Network-based Video Coding (JVET-T2023) [Posted 2020-10-16, last update 2020-10-31]

The eleven *ad hoc* groups had made progress, and reports from those activities had been submitted. Furthermore, on core experiment (CE) on entropy coding for high bit depth and high bit rate coding, and one exploration experiment (EE) on neural network-based video coding 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) 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 60 input contributions to the current meeting (not counting the AHG, CE and EE summary reports and crosschecks) had been registered for consideration at the 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 21st meeting had been made publicly available on the ITU-hosted ftp site <http://wftp3.itu.int/av-arch/jvet-site/2021_01_U_Virtual/>.

[JVET-U0002](https://jvet-experts.org/doc_end_user/current_document.php?id=10624) JVET AHG report: Draft text and test model algorithm description editing (AHG2) [B. Bross, J. Chen, C. Rosewarne, F. Bossen, J. Boyce, V. Drugeon, S. Kim, S. Liu, J.-R. Ohm, K. Sharman, G. J. Sullivan, A. Tourapis, Y.-K. Wang, Y. Ye]

At the 20th JVET meeting, AHG2 on draft text and test model algorithm description editing was established with the following mandates:

• Produce and finalize draft text outputs of the meeting (JVET-T1003, JVET-T1005, JVET-T1006, JVET-T1008, JVET-T2001, JVET-T2016 JVET-T2017 and JVET-T2019).

• 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-T1004 errata output collection.

• Produce and finalize JVET-T2002 VVC Test Model 11 (VTM 11) Algorithm and Encoder Description.

• Propose improvements to the JCTVC-AN1002 HEVC Test Model 16 (HM 16) Update 14 of Encoder Description

• 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

JVET-T1003 Revised coding-independent code points for video signal type identification (Draft 2)

This document contains draft text changes to Rec. ITU-T H.273 | ISO/IEC 23091-2 as prepared by JVET toward a future revision of these technically aligned twin text standards.

JVET-T1004 Errata report items for HEVC, AVC, Video CICP, and CP usage TR

This document contains a list of reported errata items for 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 JCT-VC and have been agreed to require fixing, while some other items have not yet been confirmed.

JVET-T1005 Shutter interval information SEI message for HEVC (Draft 3)

This document contains the draft text for changes to the High Efficiency Video Coding (HEVC) standard (Rec. ITU-T H.265 | ISO/IEC 23008-2) to specify the shutter interval information SEI message.

JVET-T1006 Annotated regions and shutter interval information SEI messages for AVC (Draft 2)

This document contains the draft text for changes to the Advanced Video Coding (AVC) standard (Rec. ITU-T H.264 | ISO/IEC 14496-10), for specifying the annotated regions and shutter interval information SEI messages and for some technical corrections and editorial improvements.

JVET-T1008 Usage of video signal type code points (Draft 2 for version 3)

This document contains draft revisions for a future third edition of ITU-T H.Sup19 | ISO/IEC 23091-4, in the form of a list of changes to the text of ISO/IEC 23091-4.

JVET-T2001 Versatile Video Coding Editorial Refinements on Draft 10

This document provides editorial refinements on Draft 10 of Versatile Video Coding, i.e., Recommendation ITU-T H.266 | International Standard ISO/IEC 23090-3, including the integration of a number of bug fixes reported in the bug tracking system as well as some other editorial changes, based on the latest JVET output draft VVC text in JVET-S2001-vH.

List of logged changes:

• Replaced Fig. 20 with the correct one (now the same as in the final FDIS text).

• Fixed tickets #1364, #1365, #1366, #1367, #1368, #1369, #1371, #1372, #1379, #1380, #1381, #1386, #1387, #1388, #1410, and #1412.

• Fixed block vector used for IBC-coded chroma blocks in equations (1109, 1110) for 4:2:2 and 4:4:4 9 (Thanks to Li Zhang).

• Fixed typos ‘mesage’ and a couple of wrong minus sign characters for the interpolation filter coefficients.

• Added missing “NAL unit” (3 instances) after “AUD ”.

• Added a dependency on y component to completely fix the issue reported in #1372.

• Added a missing condition for the inference of tu\_y\_coded\_flag to completely fix the issue reported in #1366.

• Rephrased NOTEs 1 and 2 in clause 8.3.1 to take into account the bitstream conformance constraint specified by item 8 in clause C.4. (JVET-T0055 item 1)

• Removed the step 9 of the general sub-bitstream extraction process. (JVET-T0055 item 2)

• Moved step 5 to be immedaitely after step 6 in the subpicture sub-bitstream extraction process. (JVET-T0055 item 3)

• Added the following constraint: When there are multiple SEI messages with a particular value of payloadType not equal to 133 that are associated with a particular AU or DU and apply to a particular OLS or layer, regardless of whether some or all of these SEI messages are scalable-nested, the SEI messages shall have the same SEI payload content. (JVET-T0055 item 4)

• Fixed a CABAC flowchart typo.

• Fixed an apparent small indentation problem (Thanks to Gary J. Sullivan).

• Removed redundant (or misplaced) spaces immediately before a comma or period.

• Fixed a typo on the general SEI constraints on the applicable OLSs or layers of non-scalable-nested SEI messages (Thanks to Yang Wang).

• In clause 7.4.6.3, removed "Cpb" from the variable names "CpbBrVclFactor" and "CpbBrNalFactor" (Thanks to Yago Sánchez).

JVET-T2002 Algorithm description for Versatile Video Coding and Test Model 11 (VTM 11)

The JVET established the Versatile Video Coding (VVC) working draft 11 and the VVC Test Model 11 (VTM11) algorithm description and encoding method at its 20th meeting (7 – 16 Oct. 2020, teleconference). This document serves as a source of general tutorial information on the VVC design and also provides an encoder-side description of VTM11. The VVC has been developed by a joint collaborative team of ITU-T and ISO/IEC experts known as the Joint Video Experts Team (JVET), which is a partnership of ITU-T Study Group 16 Question 6 (known as VCEG) and ISO/IEC JTC 1/SC 29/WG 11 (known as MPEG). This draft new standard has been designed with two primary goals. The first of these is to specify a video coding technology with a compression capability that is substantially beyond that of the prior generations of such standards, and the second is for this technology to be highly versatile for effective use in a broadened range of applications. Some key application areas for the use of this standard particularly include ultra-high-definition video (e.g., with 3840×2160 or 7620×4320 picture resolution and bit depth of 10 or 12 bits as specified in Rec. ITU-R BT.2100), video with a high dynamic range and wide colour gamut (e.g., with the perceptual quantization or hybrid log-gamma transfer characteristics specified in Rec. ITU-R BT.2100), and video for immersive media applications such as 360° omnidirectional video projected using a common projection format such as the equirectangular or cubemap projection format, in addition to the applications that have commonly been addressed by prior video coding standards.

Ed. Notes: VVC Test Model 11 (VTM11) algorithm description and encoding method

• General editorial improvements

• Added description of encoder configuration with GOP size 32 in random access and GOP size 8 in low delay

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

• Editorial fixes in various sections (introduction, wrap around motion compensation, VTM encoder, etc).

• Fixes in block partitioning signalling and picture boundary forced partitioning

JVET-T2016 Working practices using objective metrics for evaluation of video coding efficiency experiments (Draft 4)

This document provides a description of Bjøntegaard Delta rate (BD-rate) measurement practices for video coding experiments. It provides a concept-level overview of recent practices and provides references to technical papers that describe further details. It provides comments on why some of the choices were made and identifies situations where caution must be taken when interpreting the results. Revision marks are shown relative to the July 2020 output document, JVET-S2016-v1.

JVET-T2017 Additional SEI messages for VSEI (Draft 1)

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.264 | ISO/IEC 14496-10) to specify the annotated regions SEI message and for some technical corrections and editorial improvements.

JVET-T2019 New level and additional SEI messages for VVC (Draft 1)

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.

Draft 1 incorporated items:

• Addition of SEI manifest and SEI prefix indication SEI messages (JVET-T0056)

• Addition of ExtensionBitsPresentFlag to the sei\_payload( ) syntax and the vui\_payload( ) syntax (JVET-T0048).

• Addition of Level 6.3 (JVET-T0065)

• Addition of payloadType value etc. for the annotated regions SEI message (JVET-T0053)

Related contributions

JVET-U0049 AHG2: Some errata items for AVC (Y.-K. Wang (Bytedance), G. J. Sullivan (Microsoft))

JVET-U0073 AHG2: Errata on referencing of parameter sets (K. Suehring, R. Skupin, Y. Sanchez, T. Schierl (Fraunhofer HHI))

JVET-U0076 AHG2/AHG3: Proposal to remove some RPL constraints (A. Hallapuro, M. Hannuksela (Nokia))

JVET-U0085 AHG2: Some VVC errata items (Y.-K. Wang, Z. Deng, Y. Wang (Bytedance)

JVET-U0086 AHG2: Some VSEI errata items (Y.-K. Wang (Bytedance))

Recommendations

The AHG recommends to:

• Approve JVET-T1003, JVET-T1004, JVET-T1005, JVET-T1006, JVET-T1008, JVET-T2001, JVET-T2002, JVET-T2016, JVET-T2017, and JVET-T2019 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.

It is mentioned that there are some open tickets on VVC and VSEI. Many of those had been resolved in the last meeting’s outputs which should be closed. One ticket is resolved in the errata report JVET-U0073, and some additional new ones may need further discussion. Revisit in context of presenting errata reports.

[Revisit to discuss JVET-T1005 payloadType 203 🡪 205 issue: revised 2021-01-12]

[JVET-U0003](https://jvet-experts.org/doc_end_user/current_document.php?id=10625) JVET AHG report: Test model software development (AHG3) [F. Bossen, X. Li, K. Sühring, K. Sharman, V. Seregin, A. Tourapis]

The mandates given to the AHG are:

• Coordinate development of test model (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 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.

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

• VTM 11.0 (Nov. 2020)

• HM-16.22 (Jul. 2020)

• HM-16.21+SCM-8.8 (Mar. 2020)

• SHM 12.4 (Jan. 2018) [svn]

• HTM 16.3 (Jul. 2018) [svn]

• JM 19.0

• JSVM 9.19.15

• JMVC 8.5

• 3DV ATM 15.0 (no version history)

• HDRTools 0.19.1 (Sep. 2019)

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 10.1 and 10.2 were tagged on Oct. 20 and Oct. 21, and VTM version 11.0 was tagged on Nov. 30. VTM version 11.1 is expected during the 21st JVET meeting.

VTM 10.1 was tagged on Oct. 20, 2020. Changes include:

• JVET-S0102 Aspect 3: Place PT SEI messages in individual SEI NAL units when general\_same\_pic\_timing\_in\_all\_ols\_flag is equal to 1

• JVET-S0102 Aspect 4: Mandate same value for buffering period syntax elements

• Fix #1103: PPS extension data flag cleanup

• JVET-S0158: On the general sub-bitstream extraction process

• Bugfix and code cleanup for Ticket #1095.

• Refactor/Bugfix: VPS handling

• Fix for Ticket #1319 - encoder is crashed when setting DecodingRefreshType equal to 2

• Fix segmentation fault when decoding a VPS after 07afae5e

• Fix #1320: Add process for decoding NAL\_UNIT\_FD when filler data NAL units are in the bitstream

• JVET-S0097: set default values in VPS

• Fix cfg files for field coding

• Cleanups and fixes for software manual / remove unused encoder option

• Fix intra-layer references in enhancement layer IRAP

• Allow build with OpenMP on macOS/Xcode

• JVET-R0227 aspect 3: Rename gci\_no\_qp\_delta\_constraint\_flag to gci\_no\_cu\_qp\_delta\_constraint\_flag

• Fix decoding if bitstream starts with a CRA picture with RASL

• Fix for ticket #1373: Repeated conversion of the Intra Chroma minQT setting from chroma unit to luma unit for multiple SPSs

• JVET-R0249: Prefix syntax elements in the VPS, SPS, PPS, PH, and SH with vps\_,...

• JVET-R0251 aspect 1: Renaming sps\_XXX\_pic\_present\_flag to...

• JVET-S0172 item 2: Constraints on GCMP

• Fix syntax order of sublayer\_hrd\_parameters() to align with specification

• Bugfix for Ticket #1291: Mismatch between spec and software in BPSEI and PTSEI

• JVET-S0047 and JVET-S0211: Add "\_minus1" to VB pos SEs and signal num VBs with ue(v)

• Fix RPL construction: allow short-term to be marked as long-term (regression after !1868)

• Fix #1382: modify list construction and verification

• Fix decoding when using long term in the SPS RPL

• Fix #1383: ALF APS management in multilayer streams

• Add .layer to filenames when outputting multiple layers

• Add layer ID to output picture log file

• Fix for IBC buffer initialization when CTU size changes

• JVET-S0212 aspect 2: Check that bitdepth\_minus8 is in the range of 0..2, inclusive

• JVET-S0154 and JVET-R0068 Apsect 5: On the subpicture sub-bitstream extraction process

• JVET-S0208 aspect7: Prefix all syntax elements in dpb\_parameters() with 'dpb\_'

• Fix #1148: POC management

• Fix #1385: reset adaptive maximum BT stats on IRAPs

• JVET-S0162: Subpicture merge app

• JVET-S0208: aspects 1 (extra header bits) and 6 (OLS count)

• JVET-S0208 aspect8: Rename SEI syntax elements with descriptive prefixes

• Fix #1395: don't adapt CABAC init flag for IRAPs

• Fix #1396 : Wrong output order of multiple layers

• Fix #1353: Deblocking, ISP, and 4:2:2/4:4:4 chroma formats

• JVET-S0202: At least one picture with PictureOutputFlag equal to 1 in the bitstream

• Fix check for SEI presence in single layer case after !1473

• Fix output picture present in bitstream checking (!1794)

• JVET-S0121 aspect 2: coding of ph\_deblocking\_filter\_override\_flag

• JVET-S0115 aspect 2: rename syntax element in VPS

• JVET-T0061: remove unnecessary call and fix memory release

• various bug fixes and cleanups

VTM 10.2 was tagged on Oct.21, 2020. Relative to VTM 10.1 only macros of the previous cycle are removed.

VTM 11.0 was tagged Nov. 30, 2020. Changes include:

• JVET-S0219 aspect1, JVET-R0193 and JVET-S0141 item47

• Enable temporal prefilter

• Rename RA config file with GOP size 32 to encoder\_randomaccess\_vtm.cfg

• Fix #1408: Replace int with TCoeff to fix zeroing out of coefficients.

• Fix #1409: Don't scale 15/16 bit sequences to lower bit depths if...

• JVET-T0091: LMCS encoder overflow fix at high bit-depth for SDR

• JVET-S0090: All slices shall have the same value of nuh\_layer\_id when there is only one layer

• JVET-T0062: Extension of rate control to support GOP size of 32 (Ticket #1402).

• Fix: set the default for MaxTidILRefPicsPlus1 to 7

• JVET-S0163: On target OLS and sublayers for decoding (Operating Point Information OPI NAL Unit)

• JVET-S0078-Rebased: Handling NoOutputOfPriorPicFlag

• Port MS-SSIM from HM-16.19

• Change default of gci\_present\_flag

• JVET-R0266 proposal 5: no\_gdr\_constraint\_flag semantics

• JVET-S0084 and JVET-S0110: Specify RPL constraint on RADL picture

• JVET-T2018: configuration files for high bit depth CTC

• Disable temporal filtering for high bit depth CTC

• JVET-T0065: Add level 6.3

• JVET-S0096 aspect 1: a RPL constraint

• Fix for ticket #1418: Mismatch between spec and VTM on signalling of sn\_layer\_id

• Temporarily disable JVET\_S0078\_NOOUTPUTPRIORPICFLAG

• Check for value range of dpb\_max\_num\_reorder\_pics

• Encoder improvement: use IDR\_N\_LP NAL unit type if GOP does not specify reordering of pictures

• Fix #1299: handling of independent subpics flag

• Fix #1404: condition for coding weighted prediction num\_l1\_weights.

• Fix handling of RPR with subpictures

• Fix #1405: decoded picture hash SEI message syntax (JVET-R0481)

• Fix #1426: Move initialization of bestBcwCost

• JVET-S0175 aspect 5: use u(8) instead of u(4) for ffi\_display\_elemental\_periods\_minus1 and pt\_display\_elemental\_periods\_minus1

• Fix #1407: Add missing HRD syntax conditions (JVET-S0175 aspect 6 option 2)

• JVET-R0046 aspect2

• Fix potential race condition when frame field information SEI arrives before picture timing SEI

• Update output YUV shift if the bitdepth changes between two sequences.

• Fix #1431: Avoid clipping after inverse transform when RExt\_\_HIGH\_BIT\_DEPTH\_SUPPORT is enabled

• JVET-T0064: Addition of ALF filter strength control to VTM

• Ticket #1433: Mark GDR with RecoveryPocCnt==0 as immediately decodable picture

• Add support for pps\_single\_slice\_per\_subpic\_flag in subpicMergeApp

• Fix for ticket #1429: Temporal Filter does not work with field coding

• various other cleanups and bug fixes

It is expected that VTM 11.1 will be tagged during or short after the 21st JVET meeting. Changes include:

• JVET-R0264: IRAP constraint

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

• Fix access to data that might be deleted

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

• 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

CTC Performance

The following tables show VTM 11.0 performance over HM 16.22:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | **All Intra Main10** |  |  |
|  |  |  | **Over HM 16.22** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 | -29,04% | -32,17% | -34,07% | 1545% | 169% |
| Class A2 | -29,29% | -23,92% | -21,06% | 2505% | 177% |
| Class B | -21,73% | -26,96% | -30,76% | 2780% | 177% |
| Class C | -22,54% | -18,95% | -22,70% | 3886% | 192% |
| Class E | -25,76% | -25,91% | -24,46% | 2249% | 170% |
| **Overall** | -25,06% | -25,37% | -26,85% | 2576% | 178% |
| Class D | -18,47% | -13,31% | -13,42% | 4414% | 182% |
| Class F | -39,33% | -39,73% | -42,22% | 5107% | 176% |
|  |  |  |  |  |  |
|  |  |  | **Random access Main10** |  |  |
|  |  |  | **Over HM 16.22** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 | -41,67% | -43,42% | -49,16% | 675% | 157% |
| Class A2 | -47,76% | -46,20% | -44,93% | 752% | 170% |
| Class B | -41,72% | -53,65% | -51,59% | 754% | 155% |
| Class C | -34,68% | -37,88% | -39,61% | 1033% | 163% |
| Class E |  |  |  |  |  |
| **Overall** | -41,04% | -45,91% | -46,58% | 802% | 161% |
| Class D | -30,84% | -33,63% | -33,40% | 1161% | 164% |
| Class F | -48,00% | -50,91% | -51,69% | 572% | 137% |
|  |  |  |  |  |  |
|  |  |  | **Low delay B Main10** |  |  |
|  |  |  | **Over HM 16.22** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  |  |  |
| Class A2 |  |  |  |  |  |
| Class B | -30,81% | -37,42% | -35,46% | 744% | 152% |
| Class C | -29,13% | -22,62% | -22,41% | 897% | 157% |
| Class E | -33,35% | -40,13% | -34,22% | 357% | 125% |
| **Overall** | -30,88% | -33,16% | -30,80% | 659% | 147% |
| Class D | -26,02% | -16,65% | -15,91% | 932% | 165% |
| Class F | -42,80% | -44,57% | -44,66% | 489% | 130% |
|  |  |  |  |  |  |
|  |  |  | **Low delay P Main10** |  |  |
|  |  |  | **Over HM 16.22** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  |  |  |
| Class A2 |  |  |  |  |  |
| Class B | -35,15% | -39,91% | -37,83% | 691% | 168% |
| Class C | -30,83% | -22,55% | -22,66% | 824% | 167% |
| Class E | -36,05% | -43,41% | -37,33% | 353% | 139% |
| **Overall** | -33,93% | -35,00% | -32,65% | 619% | 160% |
| Class D | -27,47% | -15,71% | -14,93% | 855% | 174% |
| Class F | -42,31% | -43,56% | -44,09% | 524% | 138% |

In random access configuration HM is using a GOP size of 16 pictures compared to VTM using a GOP of 32 pictures. Since random access points are inserted approximately every second aligned with a GOP boundary, intra period differs in some frame rates between 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 11.0** performance compared to **VTM 10.0**:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | **All Intra Main10** |  |  |
|  |  |  | **Over VTM-10.2** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 | 0,00% | 0,00% | 0,00% | 98% | 100% |
| Class A2 | 0,00% | 0,00% | 0,00% | 99% | 101% |
| Class B | 0,00% | 0,00% | 0,00% | 99% | 104% |
| Class C | 0,00% | 0,00% | 0,00% | 101% | 110% |
| Class E | 0,00% | 0,00% | 0,00% | 98% | 108% |
| **Overall** | 0,00% | 0,00% | 0,00% | 99% | 105% |
| Class D | 0,00% | 0,00% | 0,00% | 100% | 106% |
| Class F | 0,00% | 0,00% | 0,00% | 100% | 104% |
|  |  |  |  |  |  |
|  |  |  | **Random access Main10** |  |  |
|  |  |  | **Over VTM-10.2** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 | -4,83% | -11,43% | -10,74% | 94% | 102% |
| Class A2 | -8,31% | -14,93% | -14,18% | 95% | 100% |
| Class B | -10,53% | -15,03% | -15,34% | 94% | 101% |
| Class C | -6,72% | -10,55% | -10,40% | 98% | 108% |
| Class E |  |  |  |  |  |
| **Overall** | -7,93% | -13,09% | -12,87% | 95% | 103% |
| Class D | -3,88% | -9,44% | -9,79% | 99% | 102% |
| Class F | -11,36% | -11,57% | -11,35% | 95% | 103% |
|  |  |  |  |  |  |
|  |  |  | **Low delay B Main10** |  |  |
|  |  |  | **Over VTM-10.2** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  |  |  |
| Class A2 |  |  |  |  |  |
| Class B | 0,00% | -0,10% | -0,04% | 98% | 102% |
| Class C | 0,00% | -0,18% | 0,04% | 101% | 105% |
| Class E | 0,01% | 0,11% | -0,06% | 97% | 102% |
| **Overall** | 0,00% | -0,07% | -0,02% | 99% | 103% |
| Class D | 0,00% | 0,03% | -0,03% | 102% | 104% |
| Class F | 0,01% | -0,23% | 0,07% | 97% | 101% |
|  |  |  |  |  |  |
|  |  |  | **Low delay P Main10** |  |  |
|  |  |  | **Over VTM-10.2** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 |  |  |  |  |  |
| Class A2 |  |  |  |  |  |
| Class B | 0,00% | 0,00% | 0,00% | 96% | 101% |
| Class C | 0,00% | 0,00% | 0,00% | 100% | 105% |
| Class E | 0,00% | 0,00% | 0,00% | 96% | 108% |
| **Overall** | 0,00% | 0,00% | 0,00% | 98% | 104% |
| Class D | 0,00% | 0,00% | 0,00% | 100% | 109% |
| Class F | 0,00% | 0,00% | 0,00% | 97% | 102% |

It is noted that VTM10 used the same intra period as HM16.22, therefore the intra period in RA is not the same as for VTM11.

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

Issues in VTM 11.0 affecting conformance

The following issues in VTM 11.0 affect conformance:

• Decoding of multi-layer bitstreams broken (issues #1438, #1444)

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

• Crashes related to CRA and RASL pictures (issues #1413 and #1414)

• Missing HLS features

Some HLS features are still missing (see sections below) and there are several tickets related to multilayer coding that are still open.

It is mentioned that there may also be a problem with GDR implementation which needs further investigation (see AHG5 report).

It should be noted that the syntax of the decoded picture hash SEI message was corrected between VTM versions 10.1 and 11.0 (issue #1405 / JVET-R0481). Thus, VTM 11.0 is incompatible with “draft5” conformance bitstreams provided at

https://www.itu.int/wftp3/av-arch/jvet-site/bitstream\_exchange/VVC/draft\_conformance/draft5/.

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

• JVET-Q0164

• JVET-Q0402

• JVET-Q0443

• JVET-R0073

• JVET-R0070

• JVET-R0178

• JVET-R0221

• JVET-R0046

• JVET-R0065

• JVET-R0191

• JVET-R0222 aspect 1

• JVET-R0255 item3

• JVET-S0196 (JVET-S0144 item 17)

• JVET-S0227 (JVET-S0144 item 22)

• JVET-S0077 (JVET-S0139 item 5)

• JVET-S0139 item 18.b

• JVET-S0139 item 21

• JVET-S0139 item 26

• JVET-S0139 item 28

• JVET-S0139 item 40

• JVET-S0210

• JVET-S0050 aspect 4 / S0054 aspect 4 (JVET-S0138 item 31)

• 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-S0141 item 7

• JVET-S0141 item 9.a

• JVET-S0141 item 11

• JVET-S0141 item 13

• JVET-S0141 item 14

• JVET-S0141 item 15

• JVET-S0141 item 16

• JVET-S0141 item 17

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

• JVET-S0141 item 19

• JVET-S0141 item 24

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Status of proposals of the 20th JVET meeting (Online)

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

• JVET-T0055 aspect 4

• JVET-T0055 item 2

It should be noted that pending merge requests are available for both items.

HM related activities

There had not been any further developments to the HM software during this meeting cycle.

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

Merge requests 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,

• 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

There had not been any updates of the HDRTools.

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

On the GitLab server the different reference software was assigned to either the JVET group or the JCT-VC group resembling the organization structure. With the merge of JCT-VC into JVET, the repositories should also be moved into the JVET server location. This changes the URLs of software locations and there is no forwarding available from the old to the new location. The change will be made during or shortly after the 21st JVET meeting and announced to the JVET reflector.

The remaining subversion repositories should be converted to git and 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:

- In random access configuration VTM uses a different intra period due to the change to GOP 32

It is suggested that the HM intra period should be aligned with the one that VVC uses in GOP 32 (this particularly applies to low frame rate sequences). Also check for verification test setup.

- 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.

To be further considered when the new proposal on an 8-bit profile for VVC is discussed.

Recommendations

The AHG recommends to:

- Continue to develop reference software

- Improve documentation, especially the software manual

- Resolve any normative issues resulting from the large number of integrations in the most recent development cycle

- 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 reduce encoder run time without significantly sacrificing compression performance

- 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.

Many of the not-yet-implemented issues relate to decoder-side checks for illegal bitstreams.

It is noted that scalable bitstreams can be generated (including spatial scalablity and multiview), but it is not known how efficient those are.

[JVET-U0004](https://jvet-experts.org/doc_end_user/current_document.php?id=10626) JVET AHG report: Test material and visual assessment (AHG4) [V. Baroncini, T. Suzuki, M. Wien, E. François, A. Norkin, A. Segall, P. Topiwala, S. Wenger, Y. Ye]

The mandates of this AHG were:

• Produce the draft verification test plan JVET-T2009 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.

• 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.

• 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, particularly including Rep. ITU-R BT.2245.

• Prepare availability of viewing equipment and facilities arrangements for future meetings.

Verification test

Two online AHG meetings were held related to the preparation of the VVC verification tests in the SDR HD, 360° video and HDR categories on 2020-12-09 and 2020-12-10, respectively. The reports of these meetings are provided in the input documents JVET-U0044 and JVET-U0046.

For SDR HD low delay verification tests, a new set of 4 conversational content has been identified and visually inspected by JVET experts. The status before the AHG meeting including activities on SDR HD and 360° video verification test preparation are reported in JVET-U0041. Candidate gaming content has been proposed in input document JVET-U0043 which has been reviewed during the AHG meeting on SDR and 360° video. At the 20th JVET meeting, the SDR HD random access configuration has also been added to the planned verification tests. A set of 7 candidate sequences has been identified. Regarding 360° video, two projection formats are considered, PERP and cube map. For cube map projection, the improved blending method proposed in JVET-T0118 has been validated by expert viewing session. For the verification tests, 4 sequences are considered for both projection formats.

Input document JVET-U0119 summarizes the results of dry-run activities in the SDR HD and 360 video categories which have been conducted after the AHG meetings. The document further includes results for additional gaming-type test sequences which have been proposed after the AHG meetings.

For the HDR category, verification tests are focused on HDR-HLG and HDR-PQ categories. Input document JVET-U0042 summarizes the status in the HDR category and has been reviewed in the AHG meeting on HDR. For HDR-HLG, a set of 3 test sequences has been identified and visually inspected by JVET experts. Additional test sequences were considered but color artefact issues have been observed and are being investigated. For HDR-PQ among the 5 sequences initially considered, only 2 are considered as suitable for verification tests. Several new candidate sequences are being considered and investigated.

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-U0041 “Status Report on SDR HD and 360 Video Verification Test Preparation” [M. Wien, V. Baroncini, Y. Ye (AHG coordinators)]

JVET-U0042 “Status Report on HDR Verification Test Preparation” [M. Wien, V. Baroncini, A. Segall (AHG coordinators)]

JVET-U0043 “Game video sequences proposal for the SDR HD low delay VVC verification test” [F. Le Léannec, G. Martin-Cocher (InterDigital)]

JVET-U0044 “AHG4: Agenda and report of the AHG meeting on the SDR HD and 360 video verification test preparation on 2020-12-09” [M. Wien, V. Baroncini, Y. Ye (AHG Coordinators)]

JVET-U0046 “AHG4: Agenda and report of the AHG meeting on the HDR verification test preparation on 2020-12-10” [A. Segall, M. Wien, V. Baroncini (AHG Coordinators)]

JVET-U0119 “Dry run subjective assessment of SDR HD and 360 video verification tests” [V. Baroncini, M. Wien] (late)

The AHG reports JVET-U0044 and JVET-U0046 include a summary of the discussion of JVET-U0041, JVET-U0042, and JVET-U0043. 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.

[JVET-U0005](https://jvet-experts.org/doc_end_user/current_document.php?id=10627) JVET AHG report: Conformance testing (AHG5) [J. Boyce, W. Wan, E. Alshina, F. Bossen, I. Moccagatta, K. Kawamura, K. Sühring, X. Xu]

At the 20th JVET meeting, the AHG on Conformance testing was established with the following mandates:

• Produce the JVET-T2008 draft conformance testing specification and develop proposed improvements.

• Study the requirements of VVC, HEVC, and AVC conformance testing to ensure interoperability.

• Maintain and update the conformance bitstream database.

• Study additional testing methodologies to fulfil the needs for VVC conformance testing.

Timeline

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:

• 101 bitstream categories have been identified

• Volunteers have been identified for all categories

• 256 bitstreams in 94 bitstream categories have been provided for VTM 11.0

• 5 bitstream categories have no VTM 11.0 bitstreams

• 2 additional bitstream categories have no provided bitstreams (for any VTM version)

Activities and Discussion

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

JVET-T2008 “Conformance testing for Versatile Video Coding (draft 5)” was published on 3 Nov 2020. Committee Draft in document WG05 N00009 was issued, based on JVET-T2008.

Most VTM 10.0 based bitstream packages were modified via a script to be compliant with VTM 11.0, which corrected the syntax of the decoded picture hash SEI message.

Several new bitstream packages were provided to improve coverage of the VVC standard.

Several conformance tickets were opened regarding provided conformance bitstreams (see “Conformance tickets” tab on https://jvet.hhi.fraunhofer.de/trac/vvc/ ). Updated bitstream packages were submitted to correct some of the reported issues. Bitstream packages that have not been updated as of Jan 4, 2021 include:

– 10b444\_A\_Kwai

– ILRPL\_A\_Huawei

– OLS\_B\_Tencent

– OLS\_C\_Tencent

– VPS\_A\_Intel

– DCI\_A\_Tencent

– DCI\_B\_Tencent

– OLS\_A\_Tencent

– TEMPSCAL\_B\_Panasonic

In addition, there are currently 7 bitstream categories missing bitstreams for VTM 11.0, most of which are related to high-level syntax.

Two bitstream categories have no submitted bitstreams for any VTM version:

– Suffix APS NAL units (SUFAPS)

– Operating Point Information (OPI)

Five additional bitstream categories have not been updated to VTM11.0:

– 10-bit 4:4:4 w/o specific tools (10b444)

– Decoding Capability Indication (DCI)

– Inter-layer ref pic list (ILRPL)

– Layered coding with Output Layer Set (OLS)

– 8-bit 4:4:4 (8b444)

In addition, there has been discussion for the GDR category about how to handle output of pictures prior to the recovery point.

It is agreed that only the behaviour starting from the recovery point is normative. VTM software needs modification such that nothing is output prior to the recovery point (as an option in decoder setting). A ticket should be filed on this. Ericsson will provide the software update.

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.

Contributions

There are two related contributions. The first contribution discusses gaps in the existing conformance bitstream suite.

JVET-U0108, AHG5: On gaps in conformance bitstreams, [F. Bossen (Sharp)]

The second is an editor’s update to the conformance testing specification.

JVET-U0120, AHG5: Editors update on VVC conformance testing, [J. Boyce, E. Alshina, F. Bossen, K. Kawamura, I. Moccagatta, K. Suehring, W. Wan, X. Xu]

Ftp site information

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

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

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

The ftp site for uploading bitstream file is as follows.

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

(user id: avguest, passwd: Avguest201007)

If using FileZilla, the following configuration is suggested:



Recommendations

The AHG recommends the following:

• Encourage volunteers of missing or deficient conformance bitstreams to provide them quickly

• Encourage conformance bitstream providers to provide text descriptions of provided bitstreams for the Conformance specification online at http://mpeg.expert/live/nextcloud/index.php/f/37368

• Discuss and refine the list of conformance bitstreams and the conformance specification solicit volunteers

• Review submitted bitstreams and consider if the flexibility of the tested tool is sufficiently exercised, including consideration of input contribution JVET-U0108

• Proceed with DIS issuance as output of this meeting

The July meeting seems to be unrealistic for reaching FDIS in ISO. Better target October. Better have a longer editing period for DIS and make sure that all streams are provided for that.

It would be desirable to have mechanisms in the reference decoder to check for bitstream compliance. Some of these are in the list of software implementation tasks. This situation is already much better than in previous standards.

Otherwise, if it is later detected that certain bitstreams of the conformance test set would not comply to the spec, they should be replaced.

[JVET-U0006](https://jvet-experts.org/doc_end_user/current_document.php?id=10628) JVET AHG report: 360° video coding, software and test conditions (AHG6) [J. Boyce, Y. He, K. Choi, J.-L. Lin, Y. Ye]

The mandates of this AHG are as follows:

• Study the effect on compression and subjective quality of different projections formats, resolutions, and packing layouts.

• Discuss refinements of common test conditions, test sequences, and evaluation criteria.

• Solicit additional test sequences, and evaluate suitability of test sequences on head-mounted displays and normal 2D displays.

• Study coding tools dedicated to 360° video, their impact on compression, and implications to the core codec design, including consideration of subpicture segmentations and adaptive viewport usage.

• Study the effect of viewport resolution, field of view, and viewport speed/direction on visual comfort.

• Study complexity of GPU rendering of projection formats.

• Study syntax for signalling of projection formats, cubeface layouts, spherical rotations.

• Generate CTC anchors and PERP results for the VTM according to JVET-L1012 within two weeks of availability of SDR CTC anchors.

• Coordinate with AHG4 in preparation for verification testing for 360° video content.

• Produce documentation of software usage for distribution with the software.

Activities

The 360Lib-12.0 software package released on Dec. 29, 2020 included following changes:

(1) Support blending for GCMP with padding when GCMP is converted to other formats or for rendering (from JVET-T0118)

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 can be found at:

ftp.ient.rwth-aachen.de/ahg/testresults/360Lib-12.0

360Lib bug tracker

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

360Lib-12.0 results

Table 1 is for the projection formats comparison using VTM-11.0 according to 360-degree video CTC (JVET-L1012) compared to that using VTM-10.0. Table 2 compares padded hybrid equi-angular cubemap (PHEC) coding and padded equi-rectangular projection (PERP) coding using VTM-11.0.

Table 3 is for PERP coding comparison between VTM-11.0 and HM-16.16. Table 4 is to compare PHEC coding with VTM-11.0 with and CMP coding with HM-16.16.

**Table 1. VTM-11.0 vs VTM-10.0 (VTM-10.0 as anchor)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **PERP: VTM-11.0 over VTM-10.0** | | | | | | **PHEC: VTM-11.0 over VTM-10.0** | | | | | |
|  | **End-to-end**  **WS-PSNR** | | | **End-to-end**  **S-PSNR-NN** | | | **End-to-end**  **WS-PSNR** | | | **End-to-end**  **S-PSNR-NN** | | |
|  | Y | U | V | Y | U | V | Y | U | V | Y | U | V |
| Class S1 | -6.09% | -7.25% | -7.06% | -6.09% | -7.27% | -7.07% | -6.26% | -7.29% | -7.34% | -6.24% | -7.31% | -7.35% |
| Class S2 | -1.88% | -6.57% | -6.66% | -1.87% | -6.58% | -6.67% | -1.62% | -6.20% | -6.42% | -1.61% | -6.21% | -6.43% |
| **Overall** | -4.41% | -6.98% | -6.90% | -4.40% | -6.99% | -6.91% | -4.41% | -6.86% | -6.97% | -4.39% | -6.87% | -6.98% |

**Table 2. VTM-11.0 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 | -12.26% | -6.82% | -7.52% | -12.17% | -6.75% | -7.47% |
| Class S2 | -5.29% | -1.34% | -1.10% | -5.28% | -1.24% | -1.03% |
| **Overall** | -9.47% | -4.63% | -4.95% | -9.41% | -4.54% | -4.89% |

**Table 3. VTM-11.0 PERP vs HM-16.16 PERP (HM-16.16 PERP as anchor)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **VTM-11.0 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-11.0 PHEC vs HM-16.16 CMP (HM-16.16 CMP as anchor)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **VTM-11.0 PHEC - Over HM-16.16 CMP** | | | | | |
|  | **End-to-end WS-PSNR** | | | **End-to-end S-PSNR-NN** | | |
|  | Y | U | V | Y | U | V |
| Class S1 | -34.56% | -40.04% | -42.29% | -34.45% | -40.02% | -42.26% |
| Class S2 | -38.74% | -39.28% | -41.29% | -38.73% | -39.27% | -41.32% |
| **Overall** | -36.23% | -39.73% | -41.89% | -36.16% | -39.72% | -41.89% |

Recommendations

The AHG recommends:

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

• To discuss whether to align 360 CTC projection formats with verification test projection formats.

It is noted that the implementation of PCMP (for usage with HM) is not fully integrated yet.

It is agreed that the CTC should be updated.

[JVET-U0007](https://jvet-experts.org/doc_end_user/current_document.php?id=10629) JVET AHG report: Coding of HDR/WCG material (AHG7) [A. Segall, E. François, W. Husak, S. Iwamura, D. Rusanovskyy]

The AHG was established with the following mandates:

• Study and evaluate available HDR/WCG test content.

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

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

• Generate CTC anchors for the VTM according to JVET-T2011 within two weeks of availability of SDR CTC anchors.

• 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.

Activities

The primary activity of the AhG was related to the mandates of (i) study and evaluate available HDR/WCG test content, (ii) comparing the performance of the VTM for HDR/WCG content, and (iii) 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 11.0 versus VTM 10.0**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **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% | 102% | 101% |
| Class H2 |  |  |  |  |  | 0.00% | 0.00% | 0.00% | 103% | 104% |
| **Overall** | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% | 102% | 102% |
|  |  |  |  |  |  |  |  |  |  |  |
|  | **Random Access** | | | | | | | | | |
|  | **Over VTM-10.0** | | | | | | | | | |
|  |  |  | **wPSNR** | | | **PSNR** | | |  |  |
|  | DE100 | PSNR-L100 | Y | U | V | Y | U | V | EncT | DecT |
| Class H1 | -19.12% | -15.43% | -9.47% | -11.80% | -11.26% | -9.49% | -11.54% | -11.01% | 100% | 101% |
| Class H2 |  |  |  |  |  | -7.58% | -9.84% | -11.45% | 99% | 103% |
| **Overall** | -19.12% | -15.43% | -9.47% | -11.80% | -11.26% | -8.79% | -10.92% | -11.17% | 99% | 101% |

**VTM 11.0 versus HM 16.18**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **All Intra** | | | | | | | | | |
|  | **Over HM-16.18** | | | | | | | | | |
|  |  |  | **wPSNR** | | | **PSNR** | | |  |  |
|  | DE100 | PSNR-L100 | Y | U | V | Y | U | V | EncT | DecT |
| Class H1 | -41.23% | -26.74% | -26.21% | -56.67% | -52.06% | -23.54% | -52.57% | -45.24% |  |  |
| Class H2 |  |  |  |  |  | -19.90% | -53.76% | -48.39% |  |  |
| **Overall** | -41.23% | -26.74% | -26.21% | -56.67% | -52.06% | -22.22% | -53.00% | -46.39% |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  | **Random Access** | | | | | | | | | |
|  | **Over HM-16.18** | | | | | | | | | |
|  |  |  | **wPSNR** | | | **PSNR** | | |  |  |
|  | DE100 | PSNR-L100 | Y | U | V | Y | U | V | EncT | DecT |
| Class H1 | -44.26% | -41.84% | -37.61% | -53.40% | -46.46% | -34.85% | -48.93% | -39.86% |  |  |
| Class H2 |  |  |  |  |  | -31.84% | -66.00% | -60.17% |  |  |
| **Overall** | -44.26% | -41.84% | -37.61% | -53.40% | -46.46% | -33.76% | -55.14% | -47.24% |  |  |

2.2 Study and Evaluate HDR/WCG Content

The HLG content in the CTC was updated during the October meeting to include type-2 versions of the HLG content that was provided in JVET-S0218.

During the interim period, a chroma artifact was identified in two sequences provided in JVET-S0218 that are not in the HDR CTC but are being considered for the verification test. These chroma artifacts had not appeared in the original, type-0 versions. The proponents of JVET-S0218 graciously regenerated the type-2 versions of the two sequences from the original content. And, the chroma artifacts did not appear in the regenerated content. Thus, the working conclusion is that there had been some issue during the file conversion.

Several sequences from JVET-S0218 are also used in the HDR CTC. Due to the previously described chroma artifact, these CTC sequences were studied and visually evaluated for chroma artifacts. Currently, no visual artifact has been identified.

Coordinating with AHG4

The AHG had significant coordination with AHG4, including sharing results on the new HDR content and expert viewing procedures.

Contributions

There are two contributions related to HDR video coding:

JVET-U0042 Status Report on HDR Verification Test Preparation M. Wien, V. Baroncini and A. Segall

JVET-U0046 AHG4: Agenda and report on the AHG meeting on the HDR verification test preparation on 2020-12-10 A. Segall, M. Wien and V. Baroncini

Recommendations

The AHG recommends the following:

• Review all input contributions

[JVET-U0008](https://jvet-experts.org/doc_end_user/current_document.php?id=10630) JVET AHG report: High bit depth, high bit rate, and high frame rate coding (AHG8) [A. Browne, T. Ikai, M. Sarwer, X. Xiu]

The AHG was established with the following mandates:

• 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-T2018 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-T2022.

• Identify suitable test material for testing of high bit depth, high bit rate, and high frame rate coding in coordination with AHG 4.

Activities

The AHG used the main JVET reflector, jvet@lists.rwth-aachen.de, with [AHG8] in message headers. The only messages using AHG8 headers consisted of a discussion about the need for an interim on-line AHG meeting, after which an announcement was sent on 18th November stating that no meeting would be held.

The primary activities of the AhG were the finalisation of the high bit depth CTC (JVET-T2018) and the core experiment on entropy coding for high bit depth and high bit rate coding (JVET-U0022).

In total, there are 33 high bit depth related contributions, but none related to high frame rate. The following section lists these contributions.

Contributions

The contributions can be sub-divided as follows:

a) CE and CE related (Golomb-Rice coding)

JVET-U0022, “CE Summary Report: Entropy Coding for High Bit Depth and High Bit Rate Coding”, A. Browne, T. Hashimoto, H.-J. Jhu, D. Rusanovskyy (CE coordinators)

JVET-U0050, “CE-1.3 and CE-3.1: Rice parameter derivation for high bit-depth coding”, T. Hashimoto, T. Ikai (Sharp)

JVET-U0051, “Non-CE: Rice parameter derivation for high bit-depth coding with state value”, T. Hashimoto, T. Ikai (Sharp)

JVET-U0057, “CE-1.4, CE-1.5, CE-2.2 and CE-2.3: Rice parameter selection for high bit depths”, A. Browne, S. Keating, K. Sharman (Sony)

JVET-U0058, “CE-3.2: Combination of CE-1.5 and CE-2.1”, A. Browne, S. Keating, K. Sharman (Sony)

JVET-U0059, “CE-3.3: Combination of CE-1.5 and CE-2.3”, A. Browne, S. Keating, K. Sharman (Sony)

JVET-U0062, “CE related: On Rice Parameter Derivation with Content Adaptation”, K. Kawamura, K. Unno (KDDI)

JVET-U0064, “CE-1.1 and CE-1.2: On the Rice parameter derivation for high bit-depth coding”, D. Rusanovskyy, L. P. Van, M. Coban, M. Karczewicz (Qualcomm)

JVET-U0065, “CE-3.4: Combination of CE-1.2 and CE-2.1”, D. Rusanovskyy, L. P. Van, M. Coban, M. Karczewicz (Qualcomm)

JVET-U0066, “CE-3.5 and CE-3.6: Combination of CE-1.2, CE-1.4/1.5 and CE-1.3”, D. Rusanovskyy, L. P. Van, M. Coban, M. Karczewicz (Qualcomm)

JVET-U0070, “CE Related: On signalling and encoder optimization for Rice parameter derivation”, D. Rusanovskyy, M. Coban, M. Karczewicz (Qualcomm)

JVET-U0075, “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-U0090, “Crosscheck of JVET-U0064 (CE-1.1 and CE-1.2: On the Rice parameter derivation for high bit-depth coding)”, A. Browne (Sony)

JVET-U0095, “Crosscheck of JVET-U0065 (CE-3.4: Combination of CE-1.2 and CE-2.1)”, A. Browne (Sony)

JVET-U0106, “Crosscheck of JVET-U0057: CE-1.4, CE-1.5: Rice parameter selection for high bit depths”, D. Rusanovskyy (Qualcomm)

JVET-U0107, “Cross-check of JVET-U0059: Combination of CE-1.5 and CE-2.3”, D. Rusanovskyy

JVET-U0109, “Crosscheck of JVET-U0050 (CE-1.3: Rice parameter derivation for high bit-depth coding)”, H.-J. Jhu (Kwai)

JVET-U0110, “Crosscheck of JVET-U0057 (CE-2.2 and CE-2.3: Rice parameter selection for high bit depths)”, H.-J. Jhu (Kwai)

JVET-U0111, “Crosscheck of JVET-U0066 (CE-3.5 and CE-3.6: Combination of CE-1.2, CE-1.4/1.5 and CE-1.3) “, H.-J. Jhu (Kwai)

JVET-U0112, “Crosscheck of JVET-U0051 (Non-CE: Rice parameter derivation for high bit-depth coding with state value)”, H.-J. Jhu (Kwai)

JVET-U0113, “Crosscheck of JVET-U0070 (CE Related: On signalling and encoder optimization for Rice parameter derivation)”, H.-J. Jhu (Kwai)

JVET-U0117, “Crosscheck of JVET-U0066 (CE-3.6)”

b) Test conditions and sequences

JVET-U0072, “AHG8: Tool Off Tests for High Bit-depth”, S. Keating, K. Kondo (Sony)

c) Tools and transforms

JVET-U0052, “AHG8: Transform coefficients range extension for high bit-depth coding”, T. Zhou, T. Chujoh, E. Sasaki, T. Ikai (Sharp)

JVET-U0063, “A constraint of max transform size for high bit depth”, K. Kondo, M. Ikeda (Sony)

JVET-U0067, “AHG8: On ALF clipping of high bit-depth coding”, M. G. Sarwer, Y. Ye (Alibaba)

JVET-U0069, “AHG8: CABAC-bypass alignment for high bit-depth coding”, M. G. Sarwer, J. Chen, Y. Ye, R. -L. Liao (Alibaba)

JVET-U0103, “AHG8: SIMD support for VTM software at high bit-depth coding”, X. Xiu, H.-J. Jhu, Y.-W. Chen, W. Chen, C.-W. Kuo, X. Wang (Kwai)

JVET-U0114, “Crosscheck of JVET-U0069 (AHG8: CABAC-bypass alignment for high bit-depth coding)”, H.-J. Jhu (Kwai)

JVET-U0121, “AHG8: Combination of JVET-U0069 and CE-2.1”, M. G. Sarwer, J. Chen, Y. Ye, R.-L. Liao (Alibaba), H.-J. Jhu, X. Xiu, Y.-W. Chen, W. Chen, C.-W. Kuo, X. Wang (Kwai Inc.) (late)

JVET-U0122, “Crosscheck of JVET-U0103 (AHG8: SIMD support for VTM software at high bit-depth coding)”, M.G. Sarwer (Alibaba)

JVET-U0123, “Cross-check of JVET-U0052 “, D. Rusanovskyy (Qualcomm)

JVET-U0124, “Cross-check on JVET-U0067”, D. Rusanovskyy (Qualcomm)

Recommendations

The AHG recommends the following:

• To review all related contributions

• To continue studying the benefits and characteristics of VVC coding tools for high bit depth and high bit rate coding.

• To continue investigating the requirements for future extensions of VVC to support high bit depth and high bit rate coding

• To review and refine the test conditions for high bit depth and high bit rate coding including the evaluation of new and modified test sequences

• To identify new test sequences for high frame rate coding which might be included in future test conditions

[JVET-U0009](https://jvet-experts.org/doc_end_user/current_document.php?id=10631) 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 the 20th JVET meeting, the AHG on SEI message studies was established with the following mandates:

• Study the SEI messages in VSEI, VVC, HEVC and AVC.

• Collect software and SEI showcase and usage information for SEI messages, including encoder and decoder implementations and bitstreams for demonstration and testing.

• Identify potential needs for additional 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.

Related contributions

The following contributions were identified as related to AHG9. It is noted that some of the contributions were not identified as AHG9.

JVET-U0045 AHG9: Picture output suppression SEI message [M. Pettersson, R. Sjöberg, M. Damghanian, J. Ström (Ericsson)]

JVET-U0049 AHG2: Some errata items for AVC [Y.-K. Wang (Bytedance), G. J. Sullivan (Microsoft)] (Errata item 1 relates to AHG9)

JVET-U0053 AHG9/AHG11: Level information for super-resolution neural network [T. Chujoh, E. Sasaki, T. Suzuki, T. Ikai (Sharp)]

JVET-U0078 AHG9: Out-of-loop luma mapping with chroma scaling using APS or SEI message parameters signalling [E. François, P. de Lagrange, F. Le Léannec (InterDigital)]

JVET-U0082 AHG9: Scalability dimension SEI message and three HEVC SEI messages [Y.-K. Wang, L. Zhang, K. Zhang, Z. Deng, Y. Wang (Bytedance), A. Vetro (MERL), M. Mrak, S. Blasi (BBC)]

JVET-U0083 Signalling of decoder initialization information [Y.-K. Wang, K. Zhang, L. Zhang, Y. Wang, J. Xu, Z. Deng (Bytedance)] (Aspect 2.2 DII signalling in SEI relates to AHG9)

JVET-U0084 AHG9: Cross RAP referencing (CRR) SEI message [Y.-K. Wang, Y. Wang, L. Zhang, K. Zhang, Z. Deng (Bytedance)]

JVET-U0085 AHG2: Some VVC errata items [Y.-K. Wang, Z. Deng, Y. Wang (Bytedance)] (Errata items 1, 2, 5, and 6 relate to AHG9)

JVET-U0086 AHG2: Some VSEI errata items [Y.-K. Wang (Bytedance)]

JVET-U0091 AHG9/AHG11: SEI message for carriage of neural network information for post filtering [B. Choi, Z. Li, W. Wang, W. Jiang, X. Xu, S. Wenger, S. Liu (Tencent)]

JVET-U0092 AHG9: Allocation of SEI message payload type for MPEG-I MIV/V3C carriage [J. Boyce (Intel)]

JVET-U0097 GDR Software [S. Hong, L. Wang, K. Panusopone (Nokia)] (Software and configuraton relate to AHG9)

JVET-U0098 AHG9: Composite Picture Information (CPI) SEI Message [Hendry, H. Jang, S. Kim, J. Lim (LGE)]

JVET-U0118 Crosscheck of JVET-U0078 (AHG9: Out-of-loop luma mapping with chroma scaling using APS or SEI message parameters signalling) [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-U0010](https://jvet-experts.org/doc_end_user/current_document.php?id=10632) JVET AHG report: Encoding algorithm optimization (AHG10) [A. Duenas, A. Tourapis, A. Norkin, R. Sjöberg]

At the 20th JVET meeting, the AHG on Encoding algorithm optimizations was established with the following mandates:

• Study the impact of using techniques such as GOP structures, GDR, LMCS and perceptually optimized adaptive quantization for encoder optimization.

• Study encoding techniques of optimization for objective quality metrics and their relationship to subjective quality.

• Particularly consider neural network-based encoding optimization technologies.

• Study the impact of adaptive quantization.

• 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.

The regular JVET e-mail reflector was used for discussions (jvet@lists.rwth-aachen.de). No e-mail related to AHG10 activity was sent to the JVET reflector during the AHG period.

Overview of input documents related to the AHG

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

JVET-U0056: [AHG10] GOP-based temporal filter improvements

This contribution proposes a set of changes to the GOP-based temporal filter that is included in the VTM that aim at improving PSNR BD-rate performance.

The proposed changes were reportedly tested under the VTM-11.0 RA common test conditions (CTC) The filter is not used for AI or LD test conditions. On average a Y/U/V BD-rate performance improvement of −1.3%/−1.0%/−1.1% for the RA CTC compared to the current filtering method and an average BD-rate improvement of -5.4%, -8.2%, and 8.0% compared to no use of filtering are reported. More detailed results are shown in the next section.

Results

The proposal was implemented on top of VTM-11.0. The method is currently applied for the RA coding conditions only. It would be of interest if the scheme could be extended to other configurations or if finer control of the filtering parameters was included.

BD-rate performance relative to the current scheme in VTM-11.0 (MCTF = 1) is as follows

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Over VTM-11.0** | | | | |
|  | Y | U | V | EncT | DecT |
| Class A1 | -1.79% | -1.77% | -2.56% | 89% | 104% |
| Class A2 | -1.03% | -0.70% | -0.76% | 94% | 120% |
| Class B | -1.32% | -1.08% | -1.07% | 87% | 100% |
| Class C | -1.04% | -0.53% | -0.34% | 98% | 108% |
| Class E |  |  |  |  |  |
| **Overall** | -1.28% | -0.99% | -1.11% | 92% | 107% |
| Class D | -1.31% | 0.01% | 0.25% | 100% | 111% |
| Class F | 0.00% | 0.00% | 0.00% | 96% | 119% |

In addition, the contribution includes the following results when compared to VTM-11.0 without temporal filtering:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Over VTM-11.0 with MCTF = 0** | | | | |
|  | Y | U | V | EncT | DecT |
| Class A1 | -4.50% | -8.81% | -8.66% | 86% | 89% |
| Class A2 | -6.17% | -10.76% | -10.56% | 89% | 101% |
| Class B | -7.60% | -9.30% | -9.47% | 87% | 92% |
| Class C | -2.68% | -4.33% | -3.89% | 98% | 94% |
| Class E |  |  |  |  |  |
| **Overall** | -5.38% | -8.17% | -8.04% | 90% | 94% |
| Class D | -0.64% | -3.44% | -3.09% | 99% | 96% |
| Class F | 0.00% | 0.00% | 0.00% | 99% | 100% |

It is interesting to note that the proposed method shows a complexity reduction, in addition to the coding benefits such a scheme is introducing, compared to the use of no filtering. Furthermore, it is indicated also in the text that there seems to be no discernible impact on encoding complexity. It would be of interest to analyze such behavior and see if such relates to the improved temporal characteristics, and thus implications in motion estimation and mode decision, of the coded sequences. Additional analysis at different operation/bitrate points may also be advisable.

JVET-U0081: [AHG10] ALF filter optimization with filter strength target

The VTM encoder currently has the possibility to control the amount of ALF filtering by changing the filter strength parameters. The current VTM behavior is that the optimal filters are calculated and are then scaled by the filter strength parameters (decreasing the magnitude) as part of the quantization. That approach can, together with disabling pre-defined filters and omitting the refinement step of the quantized coefficients, reduce the filter strength for ALF according to the filter strength parameters. This proposal suggests including the filter strength target parameter in the ALF filter optimization. This approach favors filters that have a reduced strength according to the filter strength target in all stages of the VTM’s filter optimization, i.e., during merging, refinement of quantized coefficients and when using pre-defined filter coefficients, but does not guarantee a reduction of filter strength according to the target parameter. Separate parameters for luma and chroma ALF are also suggested.

A BD-PSNR rate impact with filter strength target parameters of 0.875 for both luma and chroma is reported to be equal to 0.04%/-0.03%/-0.06% for the AI/RA/LDB test conditions respectively. A BD-PSNR rate impact with filter strength target parameter of 0.875 only for luma of 0.04%/-0.02%/-0.09% for the AI/RA/LDB test conditions respectively is also reported. The authors also provide results for the current approach in the VTM for the case of ALFStrength equal to 0.875. In this case the BD-PSNR rate impact is equal to 0.10% for AI and 0.07% for RA test conditions. LDB results were not included in the latest version.

It was claimed that the proposed approach can favor filters according to the filter strength target and that in some cases it can improve the subjective quality of VTM for inter predictive coding.

JVET-U0097: GDR Software

This contribution includes a software update for includes gradual decoder refresh (GDR) support in VTM11.0. The authors claim that since the VVC specification has now been finalized, it was also possible to make the corresponding GDR software, based upon VTM11.0, also more stabilized. Therefore, they propose integrating this implementation into the VTM source code package.

The authors mention that the following features, relating to GDR, are supported by this implementation:

1. Encoding input video sequences into GDR bitstreams with the following features:

• Flexible GDR period configured through parameter set by user,

• The first picture can be either IDR or GDR picture (configurable),

• SPS/PPS/APS are signaled at each GDR picture,

• Similar number of bits per picture, implying the delay is as low as one frame interval,

• Progressive intra refresh over a GDR period with even distribution of the forced intra areas over pictures within the GDR period using virtual boundary syntax in picture header,

• Necessary (encoding) constraints on coding tools to prevent the leaks,

• Exact match at the recovery point (or leak-free).

• MD5 can be turned on or off for GDR pictures and recovering pictures (configurable),

• ph\_pic\_output\_flag may be signaled for GDR pictures and recovering pictures.

2. Decoding the leak-free GDR bitstream using the VTM decoder.

• If MD5 is off for GDR pictures and recovering pictures, decoder will ignore hash check for those pictures and only check the hashes for non GDR/recovering pictures.

There is no mention if a similar implementation could also be provided for previous standards and their reference implementations (e.g. the HM reference software of HEVC). Such implementations may, however, be desirable.

It is noted that this contribution is only for encoder improvement, not related to the GDR conformance issue discussed under AHG5.

JVET-U0103: AHG8: SIMD support for VTM software at high bit-depth coding

This contribution provides software a patch that enables single instruction multiple data (SIMD) support for the VTM-11.0 software when RExt\_\_HIGH\_BIT\_DEPTH\_SUPPORT is enabled for high bit-depth. Compared to the VTM-11.0 12/16-bit anchor, simulation results reportedly show that both the encoding and decoding times are reduced by about 40% for RA and LD configurations, with bit-exact BD-rate results.

Recommendation from AHG

The AHG recommends that the related input contributions are reviewed and to further continue the study of encoding algorithm optimizations in JVET. In addition, it is recommended to coordinate with the appropriate CE group for high bit-depth coding, the review of contribution JVET-U0070 on encoder optimization for Rice parameter derivation.

[JVET-U0011](https://jvet-experts.org/doc_end_user/current_document.php?id=10633) 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 was established with the following mandates:

• 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-T2023.

• Solicit input contributions on NN based video coding technologies.

• Continue to refine the test conditions for neural network-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 on the performance of candidate technology.

• Analyse complexity characteristics, perform complexity analysis, and develop complexity reductions of candidate technology.

• Identify video test materials, training set materials, and testing methods for assessment of the effectiveness and complexity of considered technology.

• Generate and distribute anchor encodings and develop improvements of the JVET-T2006 common test conditions for NNVC technology.

• Particularly consider the suitability of sequences from the YouTube UGC data set for future inclusion in the test set.

• Coordinate with other relevant groups, including SC29/AG5 on visual quality assessment.

2 Activities

The AHG used the main JVET reflector, jvet@lists.rwth-aachen.de, for email exchange with AHG11 included in the subject lines. A kick-off message was sent on December 7, 2020, and four additional emails were exchanged on the reflector. While not official AHG activity, the AHG11 coordinators also observed significant side activity to progress the work of the AHG and the corresponding EE on Neural Network-based video coding.

2.1 Test and Training Material

The AHG made meaningful progress on the mandate to identify training content for the NNVC activity. 102 sequences were identified from the JVET ftp site that are asserted to be available for use by the NNVC activity. Additionally, another 764 sequences were identified from the BVI-DVC dataset that are asserted to be available for use by the NNVC activity.

To facilitate the study and use of the training data, a git repository was created. The repository can be accessed at https://vcgit.hhi.fraunhofer.de/jvet-ahg-nnvc/nnvc-ctc, and it includes the URL for each file in the dataset, as well as information about how to extract the training data. Example screen shots are shown below. It is anticipated that the git repository will be used for ongoing management of the test and training material.

Graphical user interface, text, application, email

Description automatically generated

*Figure 1 – Screen shot of the NNVC git repository providing information about the training data and extraction procedures.*

Graphical user interface, text

Description automatically generated

*Figure 2 – Screen shot of the NNVC git repository providing information about the training data URL and filename.*

2.2 Finalize and Conduct EE on NNVC

The EE on NNVC was finalized and conducted during the AHG period. And, a summary report is provided separately at this meeting as:

JVET-U0023 EE Summary Report: Neural Network-based Video Coding [E. Alshina, S. Liu, W. Chen, Y. Li, R.-L. Liao, Z. Ma, H. Wang]

It is noted that a primary goal of the EE was to test the NNVC Common Test Condition procedure provided in JVET-T2006. During the evaluation process, at least three issues were identified:

1. The NNVC CTC defines that VTM10.0 shall be used for the anchor, while the use of more recent versions of the VTM are encouraged. However, at the last meeting, it was decided to make non-normative changes to the VTM that were included in the VTM11.0 release. These non-normative changes include using a larger GOP size and enabling the temporal filter. The NNVC CTC is unclear if these non-normative changes should be included in the anchor.

2. The NNVC CTC requests reporting of MS-SSIM using HDRTools. This currently requires storing the decoded YUV file and then generating the MS-SSIM data in a separate processing pass. This did not directly fit into many of the proponents workflows.

3. Additionally, while the NNVC requests reporting MS-SSIM using HDRTools, VTM11.0 now supports the calculation of MS-SSIM directly. However, the MS-SSIM calculation used in VTM11.0 is not the same as what is used in HDRTools. This created additional confusion in the EE process.

As a results of these issues, it is noted that the simulations reported in the EE contain a mixture of GOP size configurations (16 vs 32). And, they do not report MS-SSIM data. To be clear, this is explicitly captured in the summary report. And, it is asserted to still allow for comparison and evaluation of the studied methods.

Using the obsersations of the EE process, it is recommended to update the NNVC Common Test Conditions at the January meeting to:

1. Clarify that VTM11.0 (or later) should be used as the anchor

2. Clarify that the GOP size should be 32

3. Clarify that temporal filtering should be enabled

4. Change the MS-SSIM calculation to use the algorithm provided in VTM11.0

During the presentation, it is agreed that the test conditions shall be aligned with VTM11. Regarding the MS-SSIM calculation, different methods exist (the one used in the EE was coming from HDRtools) which produce different results. Not clear which one is the best (also in terms of matching visual quality in case of video), but it would be desirable that only one method would be used in JVET.

Additionally, volunteers are requested to create a stand alone implementation of the MS-SSIM algorithm that appears in VTM11.0 to better support proposals considering end-to-end solutions (and, thus, not using the VTM11.0 code base.)

2.3 Anchor Encoding

Due to the three issues identified in the previous sub-section, the generation and cross-check of anchor encodings was delayed. However, the AHG was ultimately able to complete the anchor generation and cross-check. The data for VTM10.0 (GOP Size = 16) and VTM10.0 (GOP Size = 32) are attached to this report.

As part of the anchor encoding process, the AHG was also able to study the impact of the GOP size change and temporal filter changes incorporated into VTM11.0. This is summarized in the two tables below. In the first table, the VTM10.0 (GOP Size=16) anchor is compared to the VTM10.0 (GOP Size = 32) anchor. As can be seen in Table 1, the increase in GOP size results in an average gain of approximately 3.8% for the Random Access configuration.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Random access Main10 (QP22,27,32,37)** | | | | | | | |  |
|  | **BD-rate Over** | | | | | | | |  |
|  | Y-PSNR | U-PSNR | V-PSNR | Y-MSIM | U-MSIM | V-MSIM | EncT | DecT | bit DIFF |
| Class A1 | -1.94% | -4.82% | -5.04% | -2.60% | -7.34% | -9.57% | #DIV/0! | #DIV/0! | #DIV/0! |
| Class A2 | -3.32% | -6.67% | -5.88% | -2.35% | -6.41% | -5.93% | #DIV/0! | #DIV/0! | #DIV/0! |
| Class B | -4.35% | -7.77% | -7.97% | -5.34% | -8.30% | -8.48% | #DIV/0! | #DIV/0! | #DIV/0! |
| Class C | -5.08% | -7.02% | -7.12% | -5.77% | -7.13% | -7.25% | #DIV/0! | #DIV/0! | #DIV/0! |
| Class E |  |  |  |  |  |  |  |  |  |
| **Overall** | -3.85% | -6.76% | -6.74% | -4.31% | -7.41% | -7.86% | #DIV/0! | #DIV/0! | #DIV/0! |
| Class D | -4.50% | -6.22% | -6.65% | -5.28% | -6.37% | -7.09% | #DIV/0! | #DIV/0! | #DIV/0! |
| Class F | -11.36% | -11.56% | -11.36% | -11.30% | -15.58% | -13.45% | #DIV/0! | #DIV/0! | #DIV/0! |
| Class H | #VALUE! | #VALUE! | #VALUE! | #VALUE! | #VALUE! | #VALUE! | #DIV/0! | #DIV/0! | #DIV/0! |
|  |  |  |  |  |  |  |  |  |  |

*Table 1 – Comparison of VTM10.0 (GOP Size =32) to VTM10.0 (GOP Size = 16). As can be seen from the Table, the change in GOP Size results in approximately a 3.8% improvement in BD-rate for the random access configuration.*

Additionally, it was noted that the comparison of VTM-10.0 vs VTM-11.0 was shared on the JVET reflector on December 1, 2020, and re-produced in Table 2. As can be seen in the Table, VTM-11.0 provides approximately 7.9% improvement in BD-rate for the random access configuration. Since the major changes between VTM-10.0 and VTM-11.0 are (i) an increase in GOP Size to 32 and (ii) enabling the temporal filter, the following was concluded:

1. Approximately 3.8% of the gain can be attributed to the change in GOP Size.
2. The remainder (approximately 4%) of the gain can likely be attributed to enabling the temporal filter.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  | **Random access Main10** |  |  |
|  |  |  | **Over VTM-10.0** |  |  |
|  | Y | U | V | EncT | DecT |
| Class A1 | -4.83% | -11.43% | -10.74% | 93% | 101% |
| Class A2 | -8.31% | -14.93% | -14.18% | 92% | 99% |
| Class B | -10.53% | -15.03% | -15.35% | 96% | 99% |
| Class C | -6.72% | -10.55% | -10.40% | 100% | 98% |
| Class E |  |  |  |  |  |
| **Overall** | -7.93% | -13.09% | -12.87% | 96% | 99% |
| Class D | -3.88% | -9.44% | -9.79% | 100% | 97% |
| Class F | -11.36% | -11.57% | -11.35% | 99% | 99% |
|  |  |  |  |  |  |

*Table 2 – Comparison of VTM11.0 (Temporal Filter Enabled and GOP Size = 32) to VTM10.0 (Temporal Filter Disabled and GOP Size = 16). As can be seen from the Table, VTM11.0 results in approximately a 7.9% improvement in BD-rate for the random access configuration.*

2.4 Software Support

An implementation of MS-SSIM was ported to the VTM to better support the NNVC activity. The commit id is 18f28ecf5c4de1548612f858284606e57905b797, and the implementation is reported to correspond to the MS-SSIM implementation available in HM-16.19.

2.5 Evaluation of YouTube UGC Content

The AHG performed further evaluation of the YouTube User Generated Content (UGC) dataset to determine its suitability for future inclusion in the test (or training) set. The dataset is available at https://media.withyoutube.com and contains 1380 sequences, each with a duration of 20 seconds. The sequences in the dataset are asserted “to be distributed under the Creative Commons license”, according to the website description. The dataset is divided into 15 categories on content type, as shown below.

A picture containing text, screenshot, different

Description automatically generated

*Figure 3 – Example clips corresponding to each category in the YouTube UGC dataset.*

The dataset also includes Mean Opinion Scores for all of the video clips. Each sequence is reported to be rated by 100+ subjects using crowsourcing methods. These scores are reported for each complete sequence, as well as 10 second sub-sections of the sequence.

The AHG studied the MOS scores for the full sequences, and the distribution for the dataset is as follows:

Chart, histogram

Description automatically generated

*Table 3 – Histogram of MOS scores in the YouTube UGC data set. Each MOS scores correspond to a full, 20s sequence in the dataset.*

It is observed that approximately 185 of the sequences in the dataset have a MOS score of 4.3 or larger, which is anticipated to correspond to a high quality video suitable for the NNVC activity.

A study of the distribution of these potential high quality videos is shown in Table 4. As can be seen from the Table, a larger number of the high quality videos correspond to the Gaming and Sports categories.

Chart, histogram

Description automatically generated

*Table 4 – Histogram of the categories of sequences in the YouTube UGC data set with a MOS score greater than or equal to 4.3, which is asserted to be high quality.*

It is anticipated that the above information will be useful for future discussion of the NNVC CTC during the January meeting.

There are 20 input contriubtions related to the AHG mandates. Eight of the contributions are directly relatd to the EE activity, while the remaining 12 contributions are related to AHG11 but not directly part of the EE. (As a reminder, the purpose of the EE was mainly to evaluate the NNVC CTC, and so there is not an asserted difference in status of the doucments.) The list of input contributions is provided below.

***EE Related Input Contributions***

|  |  |  |
| --- | --- | --- |
| [JVET-U0023](https://jvet-experts.org/doc_end_user/current_document.php?id=10634) | EE Summary Report: Neural Network-based Video Coding | E. Alshina, S. Liu, W. Chen, Y. Li, R.-L. Liao, Z. Ma, H. Wang |
| [JVET-U0054](https://jvet-experts.org/doc_end_user/current_document.php?id=10568) | EE: Neural network based in-loop filtering | [Z. Wang](mailto:baixiu.wz@alibaba-inc.com), [R.-L. Liao](mailto:ruling.lrl@alibaba-inc.com), [C.Y. Ma](mailto:changyue.mcy@alibaba-inc.com), [Y. Ye (Alibaba)](mailto:yan.ye@alibaba-inc.com) |
| [JVET-U0060](https://jvet-experts.org/doc_end_user/current_document.php?id=10574) | EE-1.1: A comparison of depthwise separable convolution and regular convolution with the JVET-T0057 neural network based in-loop filter | [C. Auyeung](mailto:cauyeung@tencent.com), [W. Wang](mailto:rickweiwang@tencent.com), [W. Jiang](mailto:vwjiang@tencent.com), [X. Li](mailto:xlxiangli@tencent.com), [S. Liu (Tencent)](mailto:shanl@tencent.com) |
| [JVET-U0061](https://jvet-experts.org/doc_end_user/current_document.php?id=10575) | EE-1.1-related: BD-Rate improvements to JVET-T0057 neural network based in-loop filter using depthwise separable convolution and regular convolution | [C. Auyeung](mailto:cauyeung@tencent.com), [X. Li](mailto:xlxaingli@tencent.com), [S. Liu (Tencent)](mailto:shanl@tencent.com) |
| [JVET-U0074](https://jvet-experts.org/doc_end_user/current_document.php?id=10588) | EE: SSIM based CNN model for in-loop filtering | [T. Ouyang](mailto:oyjiyu@whu.edu.cn), [H. Zhu](mailto:zhuhanlyx@whu.edu.cn), [Z. Chen (Wuhan Unvi.)](mailto:zzchen@whu.edu.cn), [X. Xu](mailto:xiaozhongxu@tencent.com), [S. Liu (Tencent)](mailto:shanl@tencent.com) |
| [JVET-U0094](https://jvet-experts.org/doc_end_user/current_document.php?id=10609) | EE: Tests on Neural Network-based In-Loop Filter | [H. Wang](mailto:hongtaow@qti.qualcomm.com), M. Karczewicz, J. Chen, A. M. Kotra (Qualcomm) |
| [JVET-U0096](https://jvet-experts.org/doc_end_user/current_document.php?id=10611) | EE: Tests on Decomposition, Compression and Synthesis (DCS)-based Technology | [M. Lu](mailto:luming@smail.nju.edu.cn), [Z. Ma (Nanjing Univ.)](mailto:mazhan@nju.edu.cn), L. Xu, D. Wang (OPPO) |
| [JVET-U0101](https://jvet-experts.org/doc_end_user/current_document.php?id=10616) | EE-2.1.5: In-loop filtering based on neural network | [W. Chen](mailto:chenwei06@kwai.com), X. Xiu, Y.-W. Chen, H.-J. Jhu, C.-W. Kuo, X. Wang (Kwai) |

***Non-EE Input Contributions***

|  |  |  |
| --- | --- | --- |
| Loop Filtering | | |
| [JVET-U0055](https://jvet-experts.org/doc_end_user/current_document.php?id=10569) | AHG11: Multi-density network for in-loop filtering | [Z. Wang](mailto:baixiu.wz@alibaba-inc.com), [R.-L. Liao](mailto:ruling.lrl@alibaba-inc.com), [C.Y. Ma](mailto:changyue.mcy@alibaba-inc.com), [Y. Ye (Alibaba)](mailto:yan.ye@alibaba-inc.com) |
| [JVET-U0068](https://jvet-experts.org/doc_end_user/current_document.php?id=10582) | AHG11: Convolutional Neural Network-based 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-U0077](https://jvet-experts.org/doc_end_user/current_document.php?id=10591) | AHG11: Revisiting 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-U0104](https://jvet-experts.org/doc_end_user/current_document.php?id=10619) | AHG11: In-loop filtering with convolutional neural network and large activation | [J. Chen](mailto:cjianle@qti.qualcomm.com), [H. Wang](mailto:hongtaow@qti.qualcomm.com), [A. M. Kotra](mailto:akotra@qti.qualcomm.com), [M. Karczewicz (Qualcomm)](mailto:martak@qti.qualcomm.com) |
| [JVET-U0115](https://jvet-experts.org/doc_end_user/current_document.php?id=10642) | AHG11: Neural Network-based In-Loop Filter Performance 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), [M. Karczewicz (Qualcomm)](mailto:martak@qti.qualcomm.com) |
| Super Resolution | | |
| [JVET-U0053](https://jvet-experts.org/doc_end_user/current_document.php?id=10565) | AHG9/AHG11: Level information for super-resolution neural network | [T. Chujoh](mailto:chujoh.takeshi@sharp.co.jp), [E. Sasaki](mailto:eiichi.sasaki@sharp.co.jp), [T. Suzuki](mailto:takuya.suzuki@sharp.co.jp), [T. Ikai (Sharp)](mailto:ikai.tomohiro@sharp.co.jp) |
| [JVET-U0099](https://jvet-experts.org/doc_end_user/current_document.php?id=10614) | AHG11: Neural Network-based Super Resolution | [A. M. Kotra](mailto:akotra@qti.qualcomm.com), K. Reuzé, J. Chen, H. Wang, M. Karczewicz, J. Li (Qualcomm) |
| Inter Prediction | | |
| [JVET-U0087](https://jvet-experts.org/doc_end_user/current_document.php?id=10601) | AHG11: Updated information on inter-prediction coding tool with deep neural network | [Z. Li](mailto:zeqiangli@tencent.com), [B. Choi](mailto:bdchoi@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) |
| Intra Prediction | | |
| [JVET-U0105](https://jvet-experts.org/doc_end_user/current_document.php?id=10620) | AHG11-related: Investigation on CNN-based Intra Prediction | [M. Meyer](mailto:meyer@ient.rwth-aachen.de), C. Rohlfing (RWTH Aachen Univ.) |
| End-to-End | | |
| [JVET-U0102](https://jvet-experts.org/doc_end_user/current_document.php?id=10617) | AHG11: Variable rate end-to-end image compression | [C. Lin](mailto:linchaoyi@hikvision.com), [F. Chen](mailto:chenfangdong@hikvision), [L. Wang (Hikvision)](mailto:wangli7@hikvision.com) |
| High Level Syntax | | |
| [JVET-U0091](https://jvet-experts.org/doc_end_user/current_document.php?id=10605) | AHG9/AHG11: SEI message 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) |
| Test Procedures and Content | | |
| [JVET-U0116](https://jvet-experts.org/doc_end_user/current_document.php?id=10643) | A video dataset for training in neural network based video coding | [X. Xu](mailto:xiaozhongxu@tencent.com), [S. Liu](mailto:shanl@tencent.com), R. Yao, L. Wang, S. Tian (Tencent), D. Wu (Shenzhen Boyan Technology Ltd.), Y. Hu, J. Li, J. Xia, W. Qi, J. Zhang, J. Wen (Tsinghua University), |

The AHG recommends:

* Review all input contributions
* Release an updated version of the NNVC CTC clarifying the GOP size and temporal filter configuration for the anchor.
* Release an update version of the NNVC CTC modifying the MS-SSIM calculation procedure to use the method available in VTM11.0.
* Further discussion and definition of common test conditions and exploration experiments as appropriate
* Continue investigating neural-network-based video coding tools, including coding performance and complexity;

# Project development (XX)

## Deployment status (3)

Contributions in this area were discussed in plenary session at 0900-0930 UTC on Monday 11 January 2021 (chaired by JRO).

[JVET-U0020](https://jvet-experts.org/doc_end_user/current_document.php?id=10567) Deployment status of the HEVC standard [G. J. Sullivan (Co-chair)]

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 JCTVC-AN0020-v1 of June 2020.

New developments:

As of September 2020, a developer survey (conducted from 29 June to 9 August 2020) by Bitmovin with 792 respondents from 87 countries (primarily with technical roles) reported [23]:

* 1. 42% of video developers “currently using” HEVC
  2. 47% of video developers “planning to implement in the next 12 months”

As of December 2020, nearly all of the Ultra HD TV services identified in the UHD service tracker published by the Ultra HD Forum (at <https://ultrahdforum.org/uhd-service-tracker/>) are reportedly using HEVC. Of the 60 such services that have a coding format explicitly identified on the site, 58 of them are using HEVC.

Sony Alpha 7S III: Mirrorless camera with 4K video at 120 fps and Full HD at 240 fps (June 2020) [148][149].

[JVET-U0021](https://jvet-experts.org/doc_end_user/current_document.php?id=10566) Deployment status of the VVC standard [G. J. Sullivan (Co-chair)]

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

**Publicly available software source code**

1. **JVET** has developed the **VVC Test Model (VTM)** as its reference software encoder and decoder codebase [5]. It is intended primarily to demonstrate coding efficiency capability and proper interpretation of the syntax and decoding process specified in the standard (but not as a speed-optimized implementation), and is intended to be usable as a starting basis for product implementations. The software is available under a BSD copyright licence.
2. **InterDigital** developed a multi-threaded VTM decoder, and reported 6–10× speed-up relative to the single-threaded reference software [6]. It is intended to support all features of the VTM. The software was later placed in an accessible repository, and it is available under the same BSD copyright licence as the VTM software [7].
3. **Fraunhofer HHI** announced the **VVenC** encoder and **VVdeC** decoder open-source software (release 0.1) in September 2020 [8][9][10][11]. It includes support for multithreading operation, single-pass rate control, perceptual QP adaptation, and motion-compensated temporal filtering (MCTF). The software has four defined presets for quality/speed traedoff (called slow, medium, fast, and faster). Subjective testing reported in October 2020 indicated that the VVenC encoder had about the same or better subjective compression performance as the VTM encoder when operating in its “medium” speed configuration (operating with MCTF and QP adaptation disabled in the VTM and enabled in VVenC and with rate control disabled in both) with encoding speed more than 100× that of the VTM, for 4K UHD SDR video content [11][12]. As of December 2020, with release 0.2, the software is available under a BSD copyright licence.
4. **Friedrich–Alexander University** Erlangen–Nürnberg released an open-source bitstream analyser as an add-on for the VTM decoder [13][14]. The analyzer counts the occurrence of coding tools and coding modes used in a decoded bitstream and can be used for evaluating the decoding energy and time demands of VVC features. The software is available under a BSD copyright licence.

**Software decoders**

1. **Sharp** announced a real-time software decoder in June 2020, and issued a corresponding press release in December 2020 [15][16]. As of June 2020, it was reportedly capable of decoding 4K CTC UHD bitstreams at up to 40 Mbps at more than 60 fps.
2. **Tencent** announced its **O266dec** software decoder with SIMD and multithreading support and an associated FFmpeg/VLC-based video player in October 2020 [11][17][18]. As of December 2020, it is reportedly more than 3× the speed of the VTM reference software decoder when tested under VVC common test conditions (CTC) in single-threaded operation and about 20× the VTM decoder speed in 8-thread operation. It could reportedly decode UHD video at more than 60 fps at up to 40 Mbps and decode full HD video at more than 200 fps. In December 2020, a version with mobile platform support based on Arm Neon technology was reported. On an Apple A14 processor (iPhone 12pro) in single-threaded operation, it could reportedly decode 8-bit 1080p CTC bitstreams at more than 50 fps, and in multi-threaded operation it could decode such bitstreams at more than 100 fps and could decode 8-bit 4K UHD bitstreams at more than 30 fps in the RA configuration [19].

**Bitstream analyser products**

1. **Elecard** announced support for VVC in its StreamEye and StreamAnalyzer products in April 2020 [20].
2. **ViCueSoft** supports VVC in its VQ Analyzer bitstream analysis product, as of late 2020 [21].

**Encoding products and services**

1. **Bitmovin**, in partnership with Fraunhofer HHI based on VVencC as described in item 3), announced support of VVC in its video encoding platform in November 2020 [22].

Experts are encouraged sending information about other public announcements of products and services based on VVC to Gary Sullivan.

[JVET-U0128](https://jvet-experts.org/doc_end_user/current_document.php?id=10655) 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 aims at informing the JVET group about the Call for Proposals (CfP) issued by the Brazilian Digital Terrestrial TV Forum (SBTVD Forum) in July 2020. Reference documents of the CfP are indicated. A short description of the requirements related to video coding and of the timeline of the CfP are provided. It is reported that VVC has been proposed as video coding format to this CfP by several companies, including Ateme, Fraunhofer HHI, InterDigital.

Section 4.4 of the requirements document provides a detailed table of the requirements related to video coding, from which the key points can be summarized as follows:

* Spatial resolution up to 8K (7680x4320)
* Temporal resolution up to 120 fps
* Support of HDR ITU-R BT.2100 format (HLG and PQ formats)
* Support of spatial scalability
  + “Enable scalability (e.g. to improve over-the-air video quality with an Internet-delivered enhancement layer) and extensibility (support new settings and/or features in the future, in a backward-compatible way).”

The proponents would ask for

* The HDR verification tests for HDR PQ and HLG can be available before May 14, 2021, which will enable submitting them to SBTVD,
* The amendment for Level 6.3 (8K/4K scalable profile) is published not later than end of 2021. A draft would need to be shared with SBTVD by April 9th.

Bullet 1 appears realistic for HLG, PQ still may still require more effort.

Bullet 2 could have reached DIS level.

## Text development and errata reporting (5)

Contributions in this area were discussed in Session 5 at 0500-0655 UTC on Thursday 7 January 2021 (chaired by JRO and GJS).

[JVET-U0049](https://jvet-experts.org/doc_end_user/current_document.php?id=10561) AHG2: Some errata items for AVC [Y.-K. Wang (Bytedance), G. J. Sullivan (Microsoft)]

All AVC errata items in JVET-T1004 (Errata report items for HEVC, AVC, Video CICP, and CP usage TR) except the following two have been incorporated into JVET T1006 (Annotated regions and shutter interval information SEI messages for AVC (Draft 2)):

1. Fix the description of the recovery point picture to consider the case with a recovery POC distance of 0.

2. Change the range notation from the following (from JVET-T0048, agreed at the 20th JVET meeting (i.e., the JVET-T meeting)):

x = y..z x takes on integer values starting from y to z, inclusive, with x, y, and z being integer numbers and z being greater than y.

to be as follows (same as in VVCv1):

x = y..z x takes on integer values starting from y to z, inclusive, with x, y, and z being integer numbers and z being greater than or equal to y.

Reasons:

1. For the first item above, in the semantics of the recovery point SEI message in AVC, there is no definition of the term "recovery point picture", and the semantics reads OK as the case of recovery\_frame\_cnt equal to 0 is covered.

2. For the second item above, the current text is actually as follows: " x = y..z x takes on integer values starting from y to z, inclusive, with x, y, and z being integer numbers." In other words, even the condition "the requirement of z being greater than y" is not included. Without carefully checking to make sure that there is no use of the range notation with z less than y throughout the spec, I took the cautious approach to not integrate this change.

Suggestions:

1. Just drop the AVC errata item on the recovery point SEI message.

2. For the AVC errata item on the range notation definition, it seems that we should either just drop it, or do a systematic change of all the uses of the range notation, and change those where z is less than y, and then add z being greater than or equal to y. However, the latter would need quite some effort, and it is not sure that is worth to do for AVC at this stage (17+ years after the finalization of the first version).

3. Discuss the following: In the next version of the errata report, probably JVET-U1004, we probably should remove all the AVC errata items, as they are to be included into JVET U1006 (Annotated regions and shutter interval information SEI messages for AVC (Draft 3)).

Furthermore, it is suggested to correct wrong prime symbols.

Decision: Adopt - perform the suggested actions.

[JVET-U0073](https://jvet-experts.org/doc_end_user/current_document.php?id=10587) AHG2: Errata on referencing of parameter sets [K. Sühring, R. Skupin, Y. Sanchez, T. Schierl (Fraunhofer HHI)]

In the discussion of merge request !1971 it was noted that the specification text was unclear regarding the availability of a suffix APS to the decoding process. It seems the specification text does not clearly express that a picture cannot refer to a suffix APS in the same PU. Ticket #1443 was filed for this issue. During further investigation it was found that there may be some unclarity about references to parameter sets in general. This contribution proposes text improvements to express previous meeting decisions more clearly.

It is proposed to add the following sentence to the semantics of each syntax element that is a reference to a parameter set ID (adapted to the appropriate syntax elements). For example, for sh\_alf\_cc\_cr\_aps\_id:

If multiple APS with the same value of aps\_adaptation\_parameter\_set\_id exist in the bitstream, the PH refers to the APS with aps\_adaptation\_parameter\_set\_id equal to sh\_alf\_cc\_cr\_aps\_id, that most closely precedes the VCL NAL unit in decoding order, which refers to the PH.

It is generally agreed that the specification needs to clarify this item. A problem may be that the spec could be interpreted such that first all related NAL units are parsed before reference is made. It might not be clear that the second parameter set with same ID which is not used yet should not be parsed before the decding of the current picture is done.

It is pointed out that currently a note exists which clarifies that the suffix APS NAL unit is only for usage with later pictures. This however is not normative.

Adopt in spirit. Revisit: The proposed text should be further improved (action item for proponents and editors)

[JVET-U0076](https://jvet-experts.org/doc_end_user/current_document.php?id=10590) AHG2/AHG3: Proposal to remove some RPL constraints [A. Hallapuro, M. M. Hannuksela (Nokia)]

This contribution asserts that there is a problem in RPL constraints in VVC specification. The constraints 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 many 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. The contribution proposes to remove the constraints.

Not obvious that there would be a bug. This would be a normative change.

As an alternative, it was suggested to replace “entry” by “active entry” in bullet items 3 and 4. It is however not clear if that would resolve the intent of the proposal (which is about mixing pictures of different types as subpictures in a bitstream), or might cause more problems.

Some reservation is expressed doing such a change. Further study recommended – revisit after offline clarification with other experts.

Bullets 3 and 4 did not actually seem redundant.

The possibility of changing bullets 3 and 4 to prohibit only *active* entries rather than all entries was discussed. It was commented that if this had been proposed before approval of VVC v1, it would be agreed. Some participants said the current text seemed to be an oversight, and agreed that changing it should be considered. Another participant said they were concerned about potentially introducing a problem.

Revisit after offline study.

[JVET-U0085](https://jvet-experts.org/doc_end_user/current_document.php?id=10599) AHG2: Some VVC errata items [Y.-K. Wang, Z. Deng, Y. Wang (Bytedance)]

This contribution reports some errata items in VVC.

1. In clause D.2.2, the list VclAssociatedSeiList should include the value 132 (decoded picture hash). In HEVC, VclAssociatedSeiList includes the value 132.
2. In clause D.2.2, the user\_data\_registered\_itu\_t\_t35 and user\_data\_unregistered SEI messages should be handled the same way as 133 (scalable nesting) in the following constraint:

When there are multiple SEI messages with a particular value of payloadType not equal to 133 that are associated with a particular AU or DU and apply to a particular OLS or layer, regardless of whether some or all of these SEI messages are scalable-nested, the SEI messages shall have the same SEI payload content.

I.e., change it to be as follows (i.e., add the yellow-highlighted texts), such that specific rules are not specified for user-defined SEI messages:

When there are multiple SEI messages with a particular value of payloadType not equal to 4, 5, or 133 that are associated with a particular AU or DU and apply to a particular OLS or layer, regardless of whether some or all of these SEI messages are scalable-nested, the SEI messages shall have the same SEI payload content.

1. Search "with sn\_ols\_flag is equal to 0" and remove "is".
2. Search "recovery point point" and replace with "recovery point picture".
3. Add the following clarification regarding SEI messages allowed to be nested in a scalable nesting SEI message in a suffix SEI NAL unit, e.g., as a NOTE in clause D.2.2:

NOTE: Only filler\_payload, decoded\_picture\_hash, and scalable\_nesting may be included in a suffix SEI NAL unit, all other SEI messages are not allowed to be included in a suffix SEI NAL unit. When there is a scalable nesting SEI included in a suffix SEI NAL unit, it might only contain those SEI messages that are allowed to be included in a suffix NAL unit.

1. Make the following changes (i.e., add the yellow-highlighted texts):

It is a requirement of bitstream conformance that, when an SLI SEI message applicable to an OLS is present for a CVS, for all the SPSs referenced by the pictures in the multiSubpicLayers in the OLS, the value of sps\_num\_subpics\_minus1 shall be the same and the value of sps\_subpic\_treated\_as\_pic\_flag[ i ] shall be equal to 1 for each value of i in the range of 0 to sps\_num\_subpics\_minus1, inclusive.

Items 1-4 are obvious errors.

It is pointed out that the note in item 5 would be somewhat redundant, and also the wording is inappropriate for a note (according to ISO rules).

It is agreed to better add this as body text (normative), and also refer to prefix SEI messages of the given types. Further editing was performed during the session – to be uploaded as a new version 2.

Item 6 is a desirable clarification.

Decision: Adopt with changes in item 5 as noted above.

[JVET-U0086](https://jvet-experts.org/doc_end_user/current_document.php?id=10600) AHG2: Some VSEI errata items [Y.-K. Wang (Bytedance)]

This contribution reports some errata items in the VSEI specification.

1. Add "e.g.," as shown below:

Technical specifications that reference this Specification for carrying SEI messages shall specify a way to carry the payload syntax of each specified SEI message, to identify which SEI message is conveyed, and to identify the length in bits of the SEI message syntax structure, e.g., the sei\_payload( ) syntax structure specified in Rec. ITU-T H.266 | ISO/IEC 23090-3 and Rec. ITU-T H.265 | ISO/IEC 23008-2.

1. Make the following change:

The design of the container ~~needs~~ should be able to provide the ability to detect the number of bits in an SEI message and to allow the number of bits to be increased in future versions of this Specification, thus enabling this Specification to provide extensibility by directly appending additional syntax elements to the end of the SEI message syntax structure in future versions of this Specification.

It is suggested to add a note for the second item saying that SEI messages in VSEI which are intended to be used by AVC should be restricted by not extending the syntax.

Decision: Adopt with the suggested change on the note.

## Test conditions (0)

Contributions in this area were discussed in Session X at XXXX UTC on XXday X January 2021 (chaired by XXX).

## Verification test (6)

Contributions in this area were discussed in Session 7 at 1300 UTC on Thursday 7 January 2021 (chaired by JRO and GJS).

See also JVET-U0043 in section 4.5 on test material.

Also related to subjective testing, potential subjective testing of neural network techniques was also briefly discussed (see further notes under 5.2.2 – potentially move this paragraph).

There was discussion of potentially using remote availability of the coded material for this purpose (perhaps with “light” compression to ease distribution and viewing). It was noted that such methods have been explored recently in several efforts.was suggested to consider asking JPEG for information about their subjective viewing methodology and types of artefacts observed, although they may not have considered temporal effects and may have less of a need for compression for distribution.

It was commented that the artefacts for end-to-end NN methods can be very different from each other and from conventional methods, and can sometimes be surprising, and that the characteristics of the artefacts may be more manageable for the NN filtering tool methods and are also different for the super-resolution methods than from other techniques.

[JVET-U0041](https://jvet-experts.org/doc_end_user/current_document.php?id=10553) Status Report on SDR HD and 360° Video Verification Test Preparation [M. Wien, V. Baroncini, Y. Ye]

Was presented in interim AHG4 meetingon Dec. XX, 2020.

This document summarizes the status of test sequence selection and rate point determination for the VVC verification test categories SDR HD and 360° video. For SDR HD low delay, a set of test sequences with conversational content has been identified and visually inspected by JVET experts. The acquisition of gaming-type test sequences this category has not been completed yet. For SDR HD random access, an initial set of test sequences is proposed. For 360° video, test sequences and rate points for the test sequence encoded using equirectangular and cubemap projection formats are proposed.

[JVET-U0042](https://jvet-experts.org/doc_end_user/current_document.php?id=10554) Status Report on HDR Verification Test Preparation [M. Wien, V. Baroncini, A. Segall]

This document summarizes the status of test sequence selection and rate point determination for the VVC verification test of the HDR-HLG and HDR-PQ categories. For HDR-HLG, a set of test sequences has been identified and visually inspected by JVET experts. Rate points have been identified for three of the sequences. During evaluation, a color artifact was identified in two additional test sequences and is currently being studied and corrected. For HDR-PQ, a set of test candidates has been visually inspected by JVET experts. Unfortunately, only one or two of the available sequences were determined to be suitable for further consideration. Additional sources for sequences have been identified, and these sequences are currently under study.

JVET-U0042-v1 was presented during the AHG4 meeting on December 10, 2020. Later revisions of the report provide an update on the HDR related activities. They are kept in the same document to maintain context of the work. The update (v2) was provided for the JVET meeting and presented on Thursday 7 Jan.

A previous problem with chroma artifacts was resolved by using the new versions of HLG sequences with modified chroma position.

For HLG, HM shows some more severe chroma artifacts at low rate points (when using default configuration). This may be due to the fact that HM was found to be suitable for HLG in default comfiguration (see JCTVC-AB0041 and JCTVC-AB0042), whereas additional optimization was found to be necessary for PQ. It could however be inconsistent to further optimize HM wrt to fine tuning of QP offset, when still more optimization could be done with VTM. It is concluded to leave the HM setting as is.

It is noted that VVC has more options for chroma tuning than HEVC, even though they are not even fully exercised.

It was reportedly planned to conduct a dry run 3-4 weeks after the meeting. It was agreed that the precise schedule of testing shortly after this meeting is not so important and can be worked out by the test coordinators.

[JVET-U0044](https://jvet-experts.org/doc_end_user/current_document.php?id=10556) AHG4: Agenda and report of the AHG meeting on the SDR HD and 360° video verification test preparation on 2020-12-09 [M. Wien, V. Baroncini, Y. Ye]

A meeting was held online on Wed, 2020-12-09, 17:00h UTC with the connection information provided via the JVET mailing list.

* Review of contributions JVET-U0041 and JVET-U0043
* Discussion on next steps

Was shortly presented. Update since the interim Adhoc meeting: Progress in investigating the new gaming sequences, preselection of suitable parts (see JVET-U0127). Dry run was done for SDR HD (RA) and 360° video viewports (see JVET-U0119). No progress on SDR HD LD (investigate GOP4 vs. GOP8) yet.

It was also confirmed that for SDR HD RA and 360 HM16.22 and VTM11 use same number of IDR pictures (omly 30 and 60 fps).

[JVET-U0046](https://jvet-experts.org/doc_end_user/current_document.php?id=10558) AHG4: Agenda and report of the AHG meeting on the HDR verification test preparation on 2020-12-10 [A. Segall, M. Wien, V. Baroncini]

A meeting was held online on Thurs, 2020-12-09, 17:00h UTC with the connection information provided via the JVET mailing list.

* Review of contributions JVET-U0042
* Discussion on following steps

[JVET-U0119](https://jvet-experts.org/doc_end_user/current_document.php?id=10646) Dry run subjective assessment of SDR HD and 360° video verification tests [V. Baroncini, M. Wien] [late]

This document provides a report on results of subjective quality evaluation for test sequences under consideration for the VVC verification test. The QP selection for the test sequences under consideration in the SDR HD and 360° video categories as proposed in JVET-U0041 is tested for suitability for use in the VVC verification test according to the verification test plan JVET-T2009. Furthermore, candidate QPs for gaming-type test sequences are proposed and evaluated. In the SDR HD category, subset of the candidate test sequences and candidate rate points for the low delay and random access configurations are evaluated. In the 360° video category, mp4 files of viewports of the test sequences encoded using the PERP projection format are assessed. Here, the full set of rate points was included in the tests. The results indicate that the candidate QPs provide a coverage of the quality range from quite bad to very good quality and some adjustments of the QPs are suggested. In the SDR HD category, overall rate saving estimates cannot be provided from the data. Only potential isolated rate savings are estimated for rate points with overlapping confidence intervals. In the 360° video category, estimated rate savings between approximately 30% and 60% are reported.

GOP size 32 is being used for VVC RA.

Open issue: GOP size for LD could be 4 or 8.

MCTF was not planned to be used for LD and 360° since not confident about the effects in these cases.

It was reported that it may be possible perform remaining planned tests by the period until next meeting.

[JVET-U0127](https://jvet-experts.org/doc_end_user/current_document.php?id=10654) Coding results of DERF-TWITCH game sequences for the SDR HD low delay VVC verification test [F. Le Léannec, G. Martin-Cocher, E. Francois (InterDigital), M. Wien (RWTH)] [late]

Contribution JVET-U0043 proposed three game video sequences for the SDR HD low delay VVC verification tests and was discussed during the AHG4 meeting of December 9, 2020.

Following this meeting, a number of chunks in game sequences, freely available on the DERF-TWITCH site, have been pre-selected and coded at various QPs. The goal is to provide candidate sequences to be used for the SDR HD low delay VVC verification tests.

This informational contribution describes the selected pieces of DERF-TWITCH sequences. It reports the HM-16.22 and VTM-11.0 rate distortion points that have been generated for each of the selected sequences in low delay configuration. BD-rate performances of VTM-11.0 over HM-16.22 with default CTC QPs are also provided.

The coded bitstreams were available on the JVET ftp site for further use in the VVC verification test, SDR HD low-delay.

## Test material (1)

See also JVET-U0127 in section 4.4 on DERF gaming sequences.

Contributions in this area were discussed in Session X at XXXX UTC on XXday X January 2021 (chaired by XXX).

[JVET-U0043](https://jvet-experts.org/doc_end_user/current_document.php?id=10555) Game video sequences proposal for the SDR HD low delay VVC verification test [F. Le Léannec, G. Martin-Cocher (InterDigital)]

Was presented in interim AHG meeting, and related contribution JVET-U0127 was presented in JVET meeting.

Resolution of thanks for sequences

## Conformance test development (2)

Contributions in this area were discussed in Session 14 at 0500 UTC on Tuesday 12 January 2021 (chaired by JRO and GJS).

[JVET-U0108](https://jvet-experts.org/doc_end_user/current_document.php?id=10635) AHG5: On gaps in conformance bitstreams [F. Bossen (Sharp)] [late]

An independent implementation of a VVC decoder is used to analyse bitstreams submitted for conformance. Several gaps in the conformance set are identified, including syntax elements that never appear, syntax elements that exhibit little variety, and combinations of syntax element values that are never exercised. It is noteworthy that there are fewer gaps than in a previous report (JVET-T0100).

Upon request, Mediatek takes responsibility on providing bitstreams for exercising parallel merge and enabling loop filter across subpictures.

Bitstreams are missing for the still picture profiles, and multi-layer 4:4:4 (in particular for spatial scalability). Still picture bitstreams should exercise the fact that level constraints are different.

No volunteer was identified to provide these bitstreams. For multi-layer 4:4:4 a placeholder exists in the conformance spec, for the still picture profiles placeholders should be added.

Revisit.

[JVET-U0120](https://jvet-experts.org/doc_end_user/current_document.php?id=10647) AHG5: Editors update on VVC conformance testing [J. Boyce, E. Alshina, F. Bossen, K. Kawamura, I. Moccagatta, K. Sühring, W. Wan, X. Xu] [late]

No need for detailed presentation.

Contributors of bitstreams are asked to check the text descriptions of their streams and confirm that it is precise.

Some bitstreams have been recently provided (see AHG report). There seems to be no major problem for all cases where contributors are identified and willing to do the work.

## Software development (2)

Contributions in this area were discussed in Session 14 at 0540 UTC on Tuesday 12 January 2021 (chaired by JRO and GJS).

[JVET-U0047](https://jvet-experts.org/doc_end_user/current_document.php?id=10559) Updating StreamMergeApp to VTM 11.0 [E. Thomas, K. El Assal (TNO)]

The StreamMergeAp allows to merge multiple single layer bitstreams into a single output multi-layer bitstream.

This contribution provides software updates to make the StreamMergeApp functional again with VTM 11.0.

The contribution also describes the decisions that were made in the software implementation to ensure the validity of the generated output bitstream.

There are some limitations, e.g. each layer has independent SPS such that the number of layers is restricted to 16.

It is suggested to provide a documentation (including the limitations) to the software manual.

Decision (SW/BF): Adopt.

[JVET-U0097](https://jvet-experts.org/doc_end_user/current_document.php?id=10612) GDR Software [S. Hong, L. Wang, K. Panusopone (Nokia)]

This contribution includes GDR software updated on VTM11.0. VVC has been finalized, and the GDR software based upon VTM11.0 is also stabilized. It is therefore proposed to integrate the GDR software into the VTM source code package. The GDR software offers the following setups:

1. Encoding input video sequences into GDR bitstreams with the following features:

o Flexible GDR period configured through parameter set by user,

o The first picture can be either IDR or GDR picture (configurable),

o SPS/PPS/APS are signaled at each GDR picture,

o Similar number of bits per picture, implying the delay is as low as one frame interval,

o Progressive intra refresh over a GDR period with even distribution of the forced intra areas over pictures within the GDR period using virtual boundary syntax in picture header,

o Necessary (encoding) constraints on coding tools to prevent the leaks,

o Exact match at the recovery point (or leak-free).

o MD5 can be turned on or off for GDR pictures and recovering pictures (configurable),

o ph\_pic\_output\_flag may be signaled for GDR pictures and recovering pictures.

2. Decoding the leak-free GDR bitstream using the VTM decoder.

o If MD5 is off for GDR pictures and recovering pictures, decoder will ignore hash check for those pictures and only check the hashes for non GDR/recovering pictures.

It is reported that when enabling GDR functionality, a bit rate increase of 2.3% is observed on average (LD B conf.). This is explained to be due to additional insertion of refresh areas, and some limitation of tool usage (e.g. boundary between refresh and non-refresh areas). LMCS also was disabled.

It is generally agreed that this provides useful GDR functionality.

Decision (SW): Adopt. Further cleanup to be performed together with SW coordinators before putting it to main branch of VTM.

[JVET-U0132](https://jvet-experts.org/doc_end_user/current_document.php?id=10659) Crosscheck of JVET-U0097 (GDR Software) [Jack Enhorn (Ericsson)] [late]

## Implementation studies (3)

Contributions in this area were discussed in Session 14 at 0620 UTC and session 15 at 0730 UTC on Tuesday 12 January 2021 (chaired by JRO and GJS).

[JVET-U0071](https://jvet-experts.org/doc_end_user/current_document.php?id=10585) Performance of a VVC Software Decoder on Mobile Platform [Y. Li, S. Liu, Y. Chen, Y. Zheng, S. Chen, B. Zhu, J. Lou (Tencent)]

An independent VVC software decoder implemented by Tencent was demonstrated in [1]. Based on this version, an optimized implementation based on Arm Neon technology is presented in this contribution. Currently, this VVC software decoder can support the real-time decoding on the mobile platform. The performance is provided on an Apple A14 processor (iPhone 12pro) by decoding the 8bit CTC bitstreams generated by VTM-11.0 encoder. It is reported that on the single thread, the average decoding speed is 53.0 fps (RA)/ 50.3 fps (LB) for full HD (1080p) bitstreams; 81.5 fps (RA)/ 88.4 fps (LB) for full HD (1080p) SCC bitstreams. When multithreading is enabled, an average of 31.6 fps (RA) can be achieved when decoding 4K bitstreams. When decoding 2K bitstreams, an average of 129.0 fps (RA)/ 102.1 fps (LB) can be achieved for HD (1080p) bitstreams; an average of 180.0 fps (RA)/ 168.0 fps (LB) can be achieved for full HD (1080p) SCC bitstreams. Test results based on different mobile platforms are also provided.

Speedup approx. 3.5x by using Neon SIMD acceleration.

Numbers above are average. Decoding speed depending on sequence, and by tendency lower for higher bit rates. It is also pointed out that the speed might change for longer streams when it is tried to play out 60 fps continuously.

Results are for 8 bit. 10 bit is working, but not optimized yet.

No detailed analysis about memory bandwidth / consumption.

[JVET-U0088](https://jvet-experts.org/doc_end_user/current_document.php?id=10602) VVC software decoder implementation for mobile devices [J. Chen, L. Wang, R.-L. Liao, Y. Ye (Alibaba)]

This contribution provides the performance of Ali266, a software VVC decoder optimized specifically for the mobile platform by Alibaba. Ali266 was written from scratch following the VVC specification. Our optimization focused on four aspects: multi-threading, ARM assembly, cache efficiency, and memory usage. The result is a thin decoder that occupies only 30 MB of memory when decoding 720p video in real time. Two sets of content are used to test the Ali266 decoding speed: the JVET CTC content coded with fixed QPs, and e-commerce content at 720p resolution coded with three bit rates matching real-world application needs. For the JVET CTC content, the Ali266 decoder can 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. For the e-commerce content, Ali266 can achieve real-time decoding using one or two threads on a wide variety of phones, ranging from the latest models to the most affordable models. Using fewer threads reduces CPU usage and consequently lowers power consumption. The Ali266 decoder was proven to be robust against erroneous and corrupted bitstreams through comprehensive robustness tests. Based on such data, it is our conclusion that software VVC decoder is suitable for deployment on most types of mobile devices in the near future.

Speedup approx. 2.5x by using Neon SIMD acceleration.

Memory footprint approx. 30 MB for 720p decoding. Internal bitdepth 8.

Average decoding speed >30 fps for 4K with max. number of threads (but depending on sequence and bit rate).

Functionality on handling corrupted bitstream, further work on this ongoing.

ALF is disabled. Implementation ALF / CCALF ongoing.

Some slowdown is observed for long streams (tested 1 hr). Slowdown is more noticeable for multi thread.

Portrait format faster than landscape. Reason is that better parallelisation can be achieved with portrait.

Power consumption larger for more cores.

[JVET-U0135](https://jvet-experts.org/doc_end_user/current_document.php?id=10662) Update on open, optimized VVC implementations VVenC and VVdeC [A. Wieckowski, J. Brandenburg, C. Bartnik, V. George, J. Güther, G. Hege, C. Helmrich, A. Henkel, T. Hinz, C. Lehmann, C. Stoffers, I. Zupancic, B. Bross, H. Schwarz, D. Marpe (HHI)]

This document provides updated information on features, coding efficiency and runtime for version 0.2.1 of VVenC released in December 2020. Main changes for VVenC since version 0.1 include:

* **License**: SW license changed to a copyright 3-clause BSD license
* **Presets**: Increased coding efficiency and speed for existing presets (faster, fast, medium, slow) and added a new preset “slower” to achieve VTM-11 CTC coding efficiency.
* **New coding tools**: Intra Sub-Partitions (ISP), Transform Skip Residual Coding (TSRC), Block-level Differential Pulse Code Modulation (BDPCM)
* **Rate Control (RC)**: improved 1-pass RC and added new 2-pass RC.

Without QP adaptation for subjective optimization and 6 threads the following PSNR-based YUV BD-rates and speedup factors compared to HM and VTM-11.0 (GOP32+MCTF) are reported for different presets:

Faster HD: −6.0%, 33x (HM), 250x (VTM) UHD: −15.7%, 54x (HM), 390x (VTM)

Fast HD: −26.2%, 21x (HM), 160x (VTM) UHD: −29.4%, 32x (HM), 230x (VTM)

Medium HD: −37.2%, 9.3x (HM), 71x (VTM) UHD: −39.3%, 15x (HM), 110x (VTM)

Slow HD: −40.3%, 3.6x (HM), 27x (VTM) UHD: −42.1%, 5.9x (HM), 43x (VTM)

Slower HD: −43.6%, 0.7x (HM), 5.2x (VTM) UHD: −45.4%, 1.2x (HM), 8.8x (VTM)

With QP adaptation for subjective optimization and 6 threads, the following MS-SSIM-based YUV BD-rates and speedup factors compared to HM and VTM-11.0 (GOP32+MCTF) are reported:

Faster HD: −19.0%, 32x (HM), 250x (VTM) UHD: −18.1% 51x (HM), 410x (VTM)

Fast HD: −30.0%, 21x (HM), 160x (VTM) UHD: −30.6% 31x (HM), 250x (VTM)

Medium HD: −40.5%, 9.2x (HM), 73x (VTM) UHD: −41.4%, 15x (HM), 120x (VTM)

Slow HD: −43.4%, 3.6x (HM), 28x (VTM) UHD: −44.4%, 5.8x (HM), 47x (VTM)

Slower HD: −46.0%, 0.7x (HM), 5.5x (VTM) UHD: −47.5%, 1.2x (HM), 9.7x (VTM)

Main changes for the VVdeC software decoder since version 0.1 include a number of bug fixes as well as its SW license changed to a copyright 3-clause BSD license.

## Complexity analysis (0)

Contributions in this area were discussed in Session X at XXXX UTC on XXday X January 2021 (chaired by XXX).

## Encoder optimization (2)

Contributions in this area were discussed in Session 15 at 0745 UTC on Tuesday 12 January 2021 (chaired by JRO and GJS).

[JVET-U0056](https://jvet-experts.org/doc_end_user/current_document.php?id=10570) 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.

The method was reportedly tested under VTM-11.0 RA test configurations. The filter is not used for AI or LD.

The average Y/U/V BD-rates for the common test conditions (CTC) are reported to be −1.3%/−1.0%/−1.1% for RA. All BDR numbers were computed using unfiltered source sequences as references.

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

One suggested change is to make the strength of temporal filtering depending on local spatial characteristics. Furthermore, reference frame weighting is modified. Various other optimizations, e.g. making motion estimation faster.

Consideration of spatial characteristics might cause by tendency slightly more smoothing. By tendency, bitrate is reduced (which also leads to reduced encoder run time).

Overall gain (compared to not using MCTF) is approx. 5.3%.

Question: Was it checked perceptually? No. Proponents assume that the suggested changes would not have impact on subjective quality, and generally MCTF is known to be beneficial.

Measurement of decoding time seems inaccurate.

Can this also be used for HM? Yes, but has not been fully optimized yet.

Further study for another meeting cycle, investigate possible impact on subjective quality.

Agreed to put into VTM branch for giving opportunity studying it.

It is noted that generally, CTC for HM and VTM should be aligned. When the method is implemented equivalently for both, it should be adopted for CTC.

[JVET-U0129](https://jvet-experts.org/doc_end_user/current_document.php?id=10656) Crosscheck of JVET-U0056 ([AHG10] GOP-based temporal filter improvements) [A. Wieckowski (HHI)] [late]

A discrepancy was supported for one test sequence. This was said to probably be due to the handling of the last few frames of the test sequence.

[JVET-U0081](https://jvet-experts.org/doc_end_user/current_document.php?id=10595) AHG10: ALF filter optimization with filter strength target [K. Andersson, J. Ström (Ericsson)]

The VTM encoder currently has the possibility to control the amount of ALF filtering by changing the filter strength parameters. The current VTM behavior is that the optimal filters are calculated and are then scaled by the filter strength parameters (decreasing the magnitude) as part of the quantization. That approach can, together with disabling pre-defined filters and omitting the refinement step of the quantized coefficients, reduce the filter strength for ALF according to the filter strength parameters. This proposal suggests including the filter strength target parameter in the ALF filter optimization. This approach favors filters that have a reduced strength according to the filter strength target in all stages of VTM’s filter optimization, i.e., during merging, refinement of quantized coefficients and when using pre-defined filter coefficients, but does not guarantee a reduction of filter strength according to the target parameter. Separate parameters for luma and chroma ALF are also suggested.

BDR impact with filter strength target parameters of 0.875 for both luma and chroma:

SDR CTC: AI/RA/LDB: 0.04%/-0.03%/-0.06%

HDR CTC RA: -0.20% -0.06%, -0.05%, -0.03% (deltaE, psnrL, wpsnrY, psnrY)

BDR impact with filter strength target parameter of 0.875 only for luma:

SDR CTC: AI/RA/LDB: 0.04%/-0.02%/-0.09%

HDR CTC RA: -0.03% -0.08%, -0.08%, -0.04% (deltaE, psnrL, wpsnrY, psnrY)

We also provide results for the current approach in VTM for the case of ALFStrength equal to 0.875:

SDR CTC: AI/RA/LDB: 0.10, 0.07%, 0.08%

It is claimed that the proposed approach favor filters according to the filter strength target and that in some cases it can improve the subjective quality of VTM for inter predictive coding.

Follow-up and further refinement of method from JVET-T0064. F. Bossen reports having cross-checked the results (JVET-U0136).

Minor change in encoder software

Decision(SW): Adopt (not in CTC)

[JVET-U0136](https://jvet-experts.org/doc_end_user/current_document.php?id=10663) Crosscheck of JVET-U0081 (AHG10: ALF filter optimization with filter strength target) [F. Bossen (Sharp)]

## Profile/tier/level specification (1)

Contributions in this area were discussed in Session X at XXXX UTC on XXday X January 2021 (chaired by XXX).

[JVET-U0089](https://jvet-experts.org/doc_end_user/current_document.php?id=10603) 8-bit profiles for VVC [Y. Ye, G. Wu, L. Wang, J. Chen (Alibaba), L. Zhang, Y.-K. Wang, K. Zhang (Bytedance), M. Karczewicz (Qualcomm), Y.-W. Huang, S.-M. Lei (MediaTek), X. Wang (Kwai), D. Wang (OPPO), W. Ding (Baidu), Y.-P. Hsiao (Vivo), P. Wu (ZTE), M.-L. Champel (Xiaomi)]

# Low-level tool technology proposals (XX)

## AHG8: High bit rate and high bit depth coding for VVC (19)

### General (3)

Contributions in this area were discussed in Session 18 at 0845 UTC on Wednesday 13 January 2021 (chaired by JRO).

[JVET-U0072](https://jvet-experts.org/doc_end_user/current_document.php?id=10586) AHG8: Tool Off Tests for High Bit-depth [S. Keating, K. Kondo (Sony)]

This contribution reports the test results of various tools at high bit depths / rates. The reported gains and run times may help in considering disabling tools for future high bit depth / high bit rate simulations. It is also proposed that LMCS be disabled for PQ content in the CTC due to its BD-rate loss.

Table 3. Simulation results in all intra configuration (AI). (VTM anchor)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **AI HDR PQ** | | | | |
|  |  | psnrY | psnrU | psnrV | EncT | DecT |
| **LMChroma** | PQ444 | 1.74% | 2.90% | 2.39% | 90% | 100% |
| PQ422 | 1.25% | 1.81% | 1.73% | 96% | 99% |
| Overall | 1.49% | 2.36% | 2.06% | 93% | 100% |
| **MTS** | PQ444 | 1.28% | 0.43% | 0.43% | 77% | 99% |
| PQ422 | 1.31% | 0.52% | 0.50% | 64% | 99% |
| Overall | 1.30% | 0.47% | 0.46% | 70% | 99% |
| **JCbCr** | PQ444 | 0.82% | 0.14% | 0.06% | 89% | 100% |
| PQ422 | 0.67% | -0.09% | -0.17% | 94% | 101% |
| Overall | 0.75% | 0.03% | -0.05% | 92% | 101% |
| **MIP** | PQ444 | 0.17% | 0.08% | 0.10% | 90% | 101% |
| PQ422 | 0.21% | 0.12% | 0.11% | 85% | 101% |
| Overall | 0.19% | 0.10% | 0.10% | 87% | 101% |
| **ISP** | PQ444 | 0.02% | 0.04% | 0.05% | 96% | 98% |
| PQ422 | 0.03% | 0.04% | 0.04% | 95% | 99% |
| Overall | 0.03% | 0.04% | 0.04% | 95% | 99% |
| **LFNST \*** | PQ444 | -0.33% | -0.02% | 0.03% | 93% | 100% |
| PQ422 | -0.15% | -0.14% | 0.12% | 96% | 100% |
| Overall | -0.24% | -0.08% | 0.07% | 94% | 100% |
| **ALF\_CCALF** | PQ444 | 0.10% | 0.39% | 0.80% | 97% | 91% |
| PQ422 | 0.10% | 0.37% | 0.70% | 96% | 91% |
| Overall | 0.10% | 0.38% | 0.75% | 97% | 91% |
| **LMCS \*\*** | PQ444 | -5.08% | -4.32% | -4.74% | 92% | 94% |
| PQ422 | -4.69% | -3.86% | -4.19% | 95% | 94% |
| Overall | -4.88% | -4.09% | -4.46% | 93% | 94% |
| **MRL** | PQ444 | 0.20% | 0.08% | 0.09% | 99% | 100% |
| PQ422 | 0.21% | 0.11% | 0.11% | 100% | 101% |
| Overall | 0.21% | 0.09% | 0.10% | 100% | 101% |
| **DepQuant** | PQ444 | -0.24% | 2.85% | 2.60% | 76% | 100% |
| PQ422 | 0.31% | 3.38% | 3.42% | 80% | 101% |
| Overall | 0.04% | 3.11% | 3.01% | 78% | 101% |
| **DualITree** | PQ444 | 1.02% | 1.08% | 1.07% | 80% | 89% |
| PQ422 | 0.63% | 0.48% | 0.54% | 102% | 94% |
| Overall | 0.82% | 0.78% | 0.80% | 91% | 91% |

\*Tool on test

\*\*A bug was reported with LMCS in VTM10.2, the test has been repeated with VTM11.0rc1 but the results were unchanged.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **AI HDR HLG** | | | | |
|  |  | psnrY | psnrU | psnrV | EncT | DecT |
| **LMChroma** | HLG444 | 1.52% | 1.65% | 1.92% | 90% | 101% |
| HLG422 | 0.48% | 0.56% | 0.62% | 95% | 99% |
| Overall | 1.00% | 1.10% | 1.27% | 93% | 100% |
| **MTS** | HLG444 | 1.77% | 0.50% | 0.50% | 77% | 99% |
| HLG422 | 1.72% | 0.58% | 0.57% | 66% | 98% |
| Overall | 1.75% | 0.54% | 0.54% | 72% | 99% |
| **JCbCr** | HLG444 | 0.02% | 0.00% | -0.01% | 87% | 100% |
| HLG422 | 0.02% | -0.01% | -0.02% | 92% | 99% |
| Overall | 0.02% | -0.01% | -0.01% | 89% | 100% |
| **MIP** | HLG444 | 0.19% | 0.09% | 0.09% | 90% | 101% |
| HLG422 | 0.19% | 0.11% | 0.11% | 85% | 100% |
| Overall | 0.19% | 0.10% | 0.10% | 88% | 100% |
| **ISP** | HLG444 | 0.02% | 0.01% | 0.01% | 96% | 99% |
| HLG422 | 0.02% | 0.01% | 0.01% | 93% | 98% |
| Overall | 0.02% | 0.01% | 0.01% | 95% | 99% |
| **LFNST \*** | HLG444 | -0.54% | 0.06% | 0.08% | 89% | 100% |
| HLG422 | -0.43% | 0.09% | 0.12% | 91% | 100% |
| Overall | -0.48% | 0.08% | 0.10% | 90% | 100% |
| **ALF\_CCALF** | HLG444 | 0.24% | 1.57% | 2.15% | 98% | 91% |
| HLG422 | 0.24% | 0.85% | 1.12% | 98% | 92% |
| Overall | 0.24% | 1.21% | 1.63% | 98% | 91% |
| **LMCS** | HLG444 | 0.00% | 0.00% | 0.00% | 100% | 100% |
| HLG422 | 0.00% | 0.00% | 0.00% | 100% | 99% |
| Overall | 0.00% | 0.00% | 0.00% | 100% | 100% |
| **MRL** | HLG444 | 0.10% | 0.03% | 0.03% | 99% | 100% |
| HLG422 | 0.10% | 0.04% | 0.04% | 99% | 99% |
| Overall | 0.10% | 0.04% | 0.03% | 99% | 100% |
| **DepQuant** | HLG444 | 0.25% | 4.71% | 4.82% | 76% | 102% |
| HLG422 | 0.98% | 4.93% | 5.21% | 79% | 105% |
| Overall | 0.61% | 4.82% | 5.01% | 77% | 104% |
| **DualITree** | HLG444 | 2.73% | 0.63% | 0.82% | 71% | 91% |
| HLG422 | 1.17% | 0.75% | 0.78% | 93% | 100% |
| Overall | 1.95% | 0.69% | 0.80% | 82% | 95% |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **AI SVT RGB** | | | | |
|  |  | psnrG | psnrB | psnrR | EncT | DecT |
| **LMChroma** | SVT16 | -0.37% | 2.06% | 2.11% | 95% | 101% |
| SVT12 | -0.16% | 1.73% | 1.74% | 96% | 100% |
| Overall | -0.27% | 1.89% | 1.92% | 95% | 101% |
| **ACT \*\*** | SVT16 | 1.26% | -0.42% | -0.46% | 72% | 119% |
| SVT12 | 0.77% | -0.08% | -0.08% | 65% | 115% |
| Overall | 1.02% | -0.25% | -0.27% | 68% | 117% |
| **MTS** | SVT16 | 1.04% | 0.57% | 0.56% | 66% | 98% |
| SVT12 | 0.98% | 0.28% | 0.28% | 66% | 99% |
| Overall | 1.01% | 0.43% | 0.42% | 66% | 98% |
| **JCbCr** | SVT16 | 0.00% | 0.00% | 0.00% | 88% | 100% |
| SVT12 | -0.01% | 0.00% | 0.00% | 88% | 100% |
| Overall | -0.01% | 0.00% | 0.00% | 88% | 100% |
| **MIP** | SVT16 | 0.58% | 0.30% | 0.30% | 79% | 99% |
| SVT12 | 0.46% | 0.17% | 0.16% | 80% | 100% |
| Overall | 0.52% | 0.23% | 0.23% | 79% | 100% |
| **ISP** | SVT16 | 0.00% | 0.04% | 0.05% | 97% | 100% |
| SVT12 | 0.00% | 0.01% | 0.01% | 97% | 99% |
| Overall | 0.00% | 0.03% | 0.03% | 97% | 100% |
| **LFNST \*** | SVT16 | -0.51% | -0.19% | -0.19% | 68% | 100% |
| SVT12 | -0.57% | 0.12% | 0.14% | 69% | 99% |
| Overall | -0.54% | -0.03% | -0.02% | 68% | 99% |
| **ALF\_CCALF** | SVT16 | 0.00% | 0.00% | 0.00% | 98% | 99% |
| SVT12 | 0.30% | 0.00% | 0.00% | 99% | 99% |
| Overall | 0.15% | 0.00% | 0.00% | 98% | 99% |
| **LMCS \*** | SVT16 | 0.00% | 0.00% | 0.00% | 100% | 100% |
| SVT12 | -1.66% | -0.02% | -0.02% | 105% | 100% |
| Overall | -0.83% | -0.01% | -0.01% | 103% | 100% |
| **MRL** | SVT16 | 0.04% | 0.03% | 0.03% | 116% | 100% |
| SVT12 | -0.10% | -0.02% | -0.02% | 108% | 100% |
| Overall | -0.03% | 0.01% | 0.01% | 112% | 100% |
| **DepQuant** | SVT16 | 2.04% | 2.45% | 2.52% | 94% | 98% |
| SVT12 | 1.97% | 2.75% | 2.81% | 85% | 99% |
| Overall | 2.00% | 2.60% | 2.67% | 89% | 98% |
| **DualITree** | SVT16 | 0.00% | 0.00% | 0.00% | 116% | 100% |
| SVT12 | -0.12% | -0.03% | -0.03% | 108% | 99% |
| Overall | -0.06% | -0.02% | -0.02% | 112% | 99% |

\* Tool on test

\*\* Note that enabling ACT requires DualITree to be disabled.

Table 4. Simulation results in low delay B configuration (LDB). (VTM anchor)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **LDB HDR PQ444** | | | | |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| **IMV** | 0.11% | 0.08% | 0.09% | 88% | 99% |
| **CIIP** | 0.33% | 0.54% | 0.59% | 96% | 100% |
| **BCW** | 0.06% | 0.03% | 0.01% | 94% | 101% |
| **MMVD** | 0.21% | 0.01% | 0.00% | 93% | 99% |
| **Geo** | 0.12% | 0.15% | 0.10% | 94% | 100% |
| **SBT** | -0.02% | 0.24% | 0.12% | 96% | 100% |
| **Affine-PROF** | 0.00% | -0.03% | 0.00% | 90% | 98% |
| **ALF\_CCALF** | 1.52% | 3.11% | 3.36% | 97% | 84% |
| **DepQuant** | 2.38% | 3.24% | 2.94% | 89% | 101% |
| **DualITree** | 0.05% | 0.06% | 0.05% | 99% | 100% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **LDB HDR HLG444** | | | | |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| **IMV** | 0.06% | 0.28% | 0.28% | 88% | 101% |
| **CIIP** | 0.24% | 0.19% | 0.16% | 93% | 100% |
| **BCW** | 0.03% | 0.10% | 0.10% | 90% | 99% |
| **MMVD** | -0.05% | -0.01% | -0.01% | 95% | 100% |
| **Geo** | 0.02% | 0.01% | 0.01% | 94% | 100% |
| **SBT** | 0.00% | 0.00% | 0.00% | 97% | 100% |
| **Affine-PROF** | 0.00% | 0.05% | 0.07% | 89% | 97% |
| **ALF\_CCALF** | 2.25% | 6.14% | 7.37% | 99% | 82% |
| **DepQuant** | 3.61% | 4.96% | 5.22% | 84% | 100% |
| **DualITree** | 0.21% | 0.05% | 0.06% | 97% | 99% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **LDB SVT RGB12** | | | | |
|  | psnrG | psnrB | psnrR | EncT | DecT |
| **IMV** | -0.01% | 0.00% | 0.00% | 92% | 98% |
| **CIIP** | 0.07% | 0.03% | 0.03% | 103% | 98% |
| **BCW** | 0.00% | 0.01% | 0.01% | 93% | 101% |
| **MMVD** | 0.07% | 0.09% | 0.09% | 103% | 99% |
| **Geo** | 0.04% | 0.05% | 0.05% | 96% | 100% |
| **SBT** | 0.00% | 0.00% | 0.00% | 99% | 100% |
| **Affine-PROF** | -0.01% | 0.03% | 0.03% | 100% | 98% |
| **ALF\_CCALF** | 0.38% | 0.00% | 0.00% | 98% | 98% |
| **DepQuant** | 2.46% | 3.13% | 3.22% | 87% | 96% |
| **DualITree** | 0.00% | 0.00% | 0.00% | 106% | 99% |

Table 5. Simulation results in random access configuration (RA). (VTM anchor)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **RA HDR PQ444** | | | | |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| **IMV** | 0.10% | 0.03% | 0.06% | 95% | 109% |
| **CIIP** | 0.26% | 0.44% | 0.49% | 96% | 99% |
| **BCW** | 0.03% | 0.04% | 0.06% | 96% | 100% |
| **MMVD** | 0.01% | -0.19% | -0.20% | 92% | 99% |
| **SMVD** | 0.08% | 0.01% | -0.01% | 97% | 100% |
| **Geo** | 0.04% | 0.08% | 0.09% | 97% | 100% |
| **BIO** | 0.02% | 0.03% | 0.03% | 98% | 99% |
| **SBT** | -0.11% | 0.45% | 0.16% | 96% | 100% |
| **DMVR** | 0.02% | 0.03% | 0.02% | 100% | 99% |
| **Affine-PROF** | 0.07% | -0.06% | -0.24% | 93% | 98% |
| **ALF\_CCALF** | 1.26% | 2.74% | 3.10% | 97% | 84% |
| **DepQuant** | 2.55% | 2.60% | 2.50% | 90% | 100% |
| **DualITree** | 0.07% | 0.07% | 0.11% | 99% | 98% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **RA HDR HLG444** | | | | |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| **IMV** | 0.12% | 0.47% | 0.49% | 86% | 101% |
| **CIIP** | 0.13% | 0.10% | 0.09% | 96% | 100% |
| **BCW** | 0.04% | 0.13% | 0.15% | 93% | 102% |
| **MMVD** | 0.02% | 0.03% | 0.04% | 92% | 100% |
| **SMVD** | 0.06% | 0.07% | 0.07% | 95% | 99% |
| **Geo** | 0.01% | 0.01% | 0.01% | 95% | 99% |
| **BIO** | -0.01% | -0.02% | -0.01% | 96% | 98% |
| **SBT** | 0.01% | -0.01% | 0.00% | 97% | 100% |
| **DMVR** | 0.00% | -0.01% | 0.00% | 98% | 98% |
| **Affine-PROF** | 0.01% | 0.03% | 0.04% | 91% | 99% |
| **ALF\_CCALF** | 2.25% | 6.14% | 7.37% | 99% | 82% |
| **DepQuant** | 3.69% | 5.26% | 5.51% | 81% | 101% |
| **DualITree** | 0.13% | 0.02% | 0.03% | 98% | 101% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **RA SVT RGB12** | | | | |
|  | psnrG | psnrB | psnrR | EncT | DecT |
| **IMV** | 0.01% | 0.03% | 0.03% | 96% | 111% |
| **CIIP** | 0.04% | 0.04% | 0.04% | 98% | 100% |
| **BCW** | 0.01% | 0.00% | 0.00% | 95% | 100% |
| **MMVD** | 0.05% | 0.08% | 0.08% | 97% | 101% |
| **SMVD** | 0.02% | 0.01% | 0.01% | 99% | 106% |
| **Geo** | 0.02% | 0.01% | 0.01% | 98% | 100% |
| **BIO** | -0.05% | 0.01% | 0.01% | 99% | 99% |
| **SBT** | 0.00% | 0.00% | 0.00% | 99% | 100% |
| **DMVR** | -0.03% | 0.02% | 0.02% | 100% | 99% |
| **Affine-PROF** | 0.01% | 0.03% | 0.03% | 99% | 99% |
| **ALF\_CCALF** | 0.43% | 0.00% | 0.00% | 99% | 97% |
| **DepQuant** | 2.41% | 3.11% | 3.21% | 84% | 97% |
| **DualITree** | 0.00% | 0.00% | 0.00% | 106% | 99% |

Figure 1. PQ PSNR-Y,U,V vs encoding runtime ratio of VTM (AI) (VTM anchor)

Figure 2. HLG PSNR-Y,U,V vs encoding runtime ratio of VTM (AI) (VTM anchor)

Figure 3. SVT PSNR-G,B,R vs encoding runtime ratio of VTM (AI) (VTM anchor)

Figure 4. PQ PSNR-Y,U,V vs encoding runtime ratio of VTM (LDB) (VTM anchor)

Figure 5. HLG PSNR-Y,U,V vs encoding runtime ratio of VTM (LDB) (VTM anchor)

Figure 6. SVT PSNR-G,B,R vs encoding runtime ratio of VTM (LDB) (VTM anchor)

Figure 7. PQ PSNR-Y,U,V vs encoding runtime ratio of VTM (RA) (VTM anchor)

Figure 8. HLG PSNR-Y,U,V vs encoding runtime ratio of VTM (RA) (VTM anchor)

Figure 9. SVT PSNR-G,B,R vs encoding runtime ratio of VTM (RA) (VTM anchor)

LMCS gives loss for 12 bit PQ content. Disabled for SVT, as it would not work for 16 bit.

There are a significant number of tools that does not provide any (or minor, or minor loss) for 12and 16 bit cases.

V2 of the contribution also includes results of disabling a whole set of tools: BCW, MMVD, SMVD, Geo, BIO, DMVR, Affine, ISP and LFNST. The loss is marginal for RA, or minor gain in some cases, but significant encoder run time is saved.

It is suggested to also report the WPSNR values for PQ cases.

Further discuss in BoG whether it is useful to disable certain tools in the 12 and 16 bit anchors.

[JVET-U0103](https://jvet-experts.org/doc_end_user/current_document.php?id=10618) AHG8: SIMD support for VTM software at high bit-depth coding [X. Xiu, H.-J. Jhu, Y.-W. Chen, W. Chen, C.-W. Kuo, X. Wang (Kwai)]

This contribution provides software patch to enable single instruction multiple data (SIMD) support for VTM-11.0 software when RExt\_\_HIGH\_BIT\_DEPTH\_SUPPORT is enabled for high bit-depth. Compared to VTM-11.0 12/16-bit anchor, simulation results reportedly show both the encoding and decoding times are reduced by about 40% for RA and LD configurations, with bit-exact BD-rate results.

Cross-checker confirms runtime reduction, and bit-exact matching of bitstreams.

It is commented that the encoder runtime reduction would be highly beneficial for conducting the CE.

It is noted that the software coordinators have not seen the software patch yet. It would be desirable to provide the ptach with the upload.

Decision(SW): Adopt.

[JVET-U0122](https://jvet-experts.org/doc_end_user/current_document.php?id=10649) Crosscheck of JVET-U0103 (AHG8: SIMD support for VTM software at high bit-depth coding) [M.G. Sarwer (Alibaba)] [late]

This was identified as a proposal in the v1 contribution document, but appeared to be an information contribution.

[JVET-U0138](https://jvet-experts.org/doc_end_user/current_document.php?id=10665) Information on VVC coding efficiency in high bit depth coding [T. Ikai (Sharp)]

TBP

### CE contributions: Entropy Coding for High Bit Depth and High Bit Rate Coding (9)

Contributions in this area were initially discussed in Session 3 at 1300-1505 UTC on Wednesday 6 January 2021 (chaired by JRO and GJS).

[JVET-U0022](https://jvet-experts.org/doc_end_user/current_document.php?id=10607) CE Summary Report: Entropy Coding for High Bit Depth and High Bit Rate Coding [A. Browne, T. Hashimoto, H.-J. Jhu, D. Rusanovskyy (CE coordinators)]

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 14 tests have been completed within the CE between the JVET T and U meetings to study and evaluate technologies related to Rice parameter derivation. Both regular residual coding (RRC) (5 tests) and transform skip residual coding (TSRC) (3 tests) have been tested. In addition 6 tests have been completed in a second stage on combinations of the previous tests.

In this report, coding performance and complexity are reported and analyzed. In particular, test results against VTM-11.0rc1 anchors are provided to show the coding efficiency and complexity trade-off of each tool. Where appropriate results are also provided for lossless conditions. Crosschecking results for the performed tests are submitted as separate documents and their summaries are integrated in this contribution.

***CE-1 (RRC)***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test** | **Tester** | **Document** | **Tool description** | **Cross checker** |
| CE-1.1 | Dmytro Rusanovskyy  (Qualcomm) | JVET-U0064 | On the Rice parameter derivation for high bit-depth coding | Adrian Browne  (Sony) |
| CE-1.2 | Dmytro Rusanovskyy  (Qualcomm) | JVET-U0064 | On the Rice parameter derivation for high bit-depth coding | Adrian Browne  (Sony) |
| CE-1.3 | Tomonori Hashimoto  (Sharp) | JVET-U0050 | Rice parameter derivation by using a formula for RRC | Hong-Jheng Jhu  (Kwai) |
| CE-1.4 | Adrian Browne  (Sony) | JVET-U0057 | Modification of Rice parameter selection for high bit depths using the value of previously coded coefficients within the slice. | Dmytro Rusanovskyy  (Qualcomm) |
| CE-1.5 | Adrian Browne  (Sony) | JVET-U0057 | Simplification of CE-1.4 | Dmytro Rusanovskyy  (Qualcomm) |

***CE-2 (TSRC)***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test** | **Tester** | **Document** | **Tool description** | **Cross checker** |
| CE-2.1 | Hong-Jheng Jhu  (Kwai) | JVET-U0075 | Slice based Rice parameter selection for transform skip residual coding | Tomonori Hashimoto  (Sharp) |
| CE-2.2 | Adrian Browne  (Sony) | JVET-U0057 | Rice parameter selection using the value of previously coded coefficients both locally (within the TB) and globally (within the slice). | Hong-Jheng Jhu  (Kwai) |
| CE-2.3 | Adrian Browne  (Sony) | JVET-U0057 | Rice parameter selection using the value of previously coded coefficients within the slice. Simplification of CE-2.2 | Hong-Jheng Jhu  (Kwai) |

***CE-3 (Combinations)***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test** | **Tester** | **Document** | **Tool description** | **Cross checker** |
| CE-3.1 | Tomonori Hashimoto  (Sharp) | JVET-U0050 | Combination of CE-1.3 and CE-2.1 | Kenji Kondo  (Sony) |
| CE-3.2 | Adrian Browne  (Sony) | JVET-U0058 | Combination of CE-1.5 and CE-2.1 | Tomonori Hashimoto  (Sharp) |
| CE-3.3 | Adrian Browne  (Sony) | JVET-U0059 | Combination of CE-1.5 and CE-2.3 | Dmytro Rusanovskyy  (Qualcomm) |
| CE-3.4 | Dmytro Rusanovskyy  (Qualcomm) | JVET-U0065 | Combination of CE-1.2 and CE-2.1 | Adrian Browne  (Sony) |
| CE-3.5 | Dmytro Rusanovskyy  (Qualcomm) | JVET-U0066 | Combination of CE-1.2 and CE-1.4/1.5 | Hong-Jheng Jhu  (Kwai) |
| CE-3.6 | Dmytro Rusanovskyy  (Qualcomm) | JVET-U0066 | Combination of CE-1.3 and CE-3.5 | Hong-Jheng Jhu  (Kwai) |

CE-1.1: Method of JVET-T0105 without locally adaptivity

In this test, method proposed in JVET-T0105 with globally derived adjusment for rice parameter derivation is being evaluated. To extend the range of the Rice parameters, it is proposed to scale values *locSumAbs* with a right shift, if *locSumAbs* exceed certain threshold, such that scaled value would fit in the allowed range of the Table 128 of the VVC specficiation without clipping. Following this, an output of the Table 128 is being adjusted by incrementing with shift value if scaling process was taken a place for a given locSumAbs value.

Adjustment parameters are determined from signalled syntax elements, without local adaptation within a block (no dependency on locSumAbs). Two subtests were planned:

* **Test CE-1.1.a:** SPS-level control information is used to derive adjustment parameters.
* **Test CE-1.1.b:** Syntax elements controlling adjustment parameters are signalled at the sub-SPS levels, e.g. at slice level.

During CE development, CE-1.1.a test was omitted. Only sub-test with slice level signaling of the derivation parameter was conducted. Detailed description of this method and specification text are provided in JVET-U0064.

CE-1.2: Method of JVET-T0105 with local adaptivity

In this test, method proposed in JVET-T0105 with locally adaptive adjustment for Rice parameter derivation is being evaluated. Similarly to the Test CE-1.1, to extend the range of the Rice parameters, it is proposed to scale values *locSumAbs* with a shift, if *locSumAbs* exceed certain threshold, such that scaled value would fit in the allowed range of the Table 128 of the VVC specficiation without clipping. Following this, an output of the Table 128 is being adjusted by incrementing with shift value if scaling process was taken a place for a given locSumAbs value. A shift parameter for locSumAbs scaling and cRiceParam increment is derived at the encoder side and signalled at the slice header. Motivated by the history usage of CE-1.4 and CE-1.5, additional sub-test was conducted (CE-1.2.b) targeting improving accuracy of the Rice derivation for coefficients with no accurate template information available. Detailed description of this method and specification text are provided in JVET-U0064.

CE-1.3: JVET-U0050

The proposal of JVET-U0050 introduces a modification to the VVC Rice parameters derivation method for regular residual coding (RRC). It is proposed to use a formula instead of conventional look-up table. Specifically, the rice parameter value is predicted based on an adjusted value of the *locSumAbs* using linear prediction with log2 operation. The linear prediction calculate floorLog2(a \* *locSumAbs* + b) + c. The linear prediction parameter (a, b and c) depend on syntax (i.e. abs\_reminder or dec\_abs\_level). An enabled flag for the proposed method is signalled in SPS. This flag doesn’t affect the process if bitdepth is equal or less than 10.

CE-1.4: RRC technique in JVET-U0057

The RRC component of the method previously proposed in JVET-T0072. For this experiment the TSRC component of the modification described in JVET-T0072 is disabled and standard VVC TSRC is used.

The implementation has been improved from that in JVET-T0072 by computing the parameters for updating counters and deriving Rice parameters at the point the counters are updated. In addition, the counters and associated parameters are stored within the structure used for CABAC contexts (as proposed but not implemented in JVET-T0072). Finally a fix has been implemented to automatically adjust the Rice parameter selected based on whether a parity bit has been coded.

CE-1.5: RRC simplification in JVET-U0057

The RRC component of the simplification previously proposed in JVET-T0072. For this experiment the TSRC component of the modification described in JVET-T0072 is disabled and standard VVC TSRC is used.

This experiment simplifies CE-1.4 as the parameters for updating counters and deriving Rice parameters are computed at the end of each transform block. In addition, the counters are only updated when the coefficient being coded is on the leading diagonal of the transform block.

CE-2.1: TSRC technique in JVET-U0075

Method proposed in JVET-U0075, it is proposed to explicitly signal the Rice parameter for each slice to indicate the Rice parameter for the binary codewords of abs\_remainder. In the proposed method, one control flag is signaled in sequence parameter set to indicate the signaling of Rice parameter for the transform skip blocks is enabled or disabled. When the control flag is signaled as enabled, one syntax element is further signaled for each transform skip slice to indicate the Rice parameter of that slice. When the control flag is signaled as disabled (e.g. set equal to “0”), no further syntax element is signaled at lower level to indicate the Rice parameter for the transform skip slice and a default Rice parameter (e.g. 1) is used for all the transform skip slice. Detailed description of this method and specification text are provided in JVET-U0075.

CE-2.2: TSRC technique in JVET-U0057

The TSRC component of the method previously proposed in JVET-T0072. For this experiment the RRC component of the modification described in JVET-T0072 is disabled and standard VVC RRC is used.

The implementation has been improved from that in JVET-T0072 by computing the parameters for updating counters and deriving Rice parameters at the point the counters are updated. In addition, the counters and associated parameters are stored within the structure used for CABAC contexts (as proposed but not implemented in JVET-T0072). Finally a fix has been implemented to automatically adjust the Rice parameter selected based on whether a parity bit has been coded.

CE-2.3: TSRC simplification in JVET-U0057

The TSRC component of the simplification previously proposed in JVET-T0072. In addition, this experiment replaces the usage of locSumAbs and *g\_auiGoRiceParsCoeff* with an offset derived from the value of the selected counter. For this experiment the RRC component of the modification described in JVET-T0072 is disabled and standard VVC RRC is used.

This experiment simplifies CE-1.4 as the parameters for updating counters and deriving Rice parameters are computed at the end of each transform block. In addition, the counters are only updated when the coefficient being coded is on the leading diagonal of the transform block.

CE-3.1: Combination of CE-1.3 and CE-2.1

This experiment combines the RRC coding modifications from CE-1.3 with the TSRC coding modifications in CE-2.1 and is described in JVET-U0050.

CE-3.2: Combination of CE-1.5 and CE-2.1

This experiment combines the RRC coding modifications from CE-1.5 with the TSRC coding modifications in CE-2.1 and is described in JVET-U0058.

CE-3.3: Combination of CE-1.5 and CE-2.3

This experiment combines the RRC coding modifications from CE-1.5 with the TSRC coding modifications in CE-2.3 and is described in JVET-U0059.

CE-3.4: Combination of CE-1.2 and CE-2.1

This experiment will blend the RRC coding modifications from CE-1.2.a with the TSRC coding modifications in CE-2.1. Detailed description of this method and specification text are provided in JVET-U0065.

CE-3.5: Combination of CE-1.2 and CE-1.4/1.5

This experiment will investigate integration of the classification elements of CE-1.4/1.5 with rice parameter inheritance mechanism and the Rice parameter derivation of the CE-1.2.b. Detailed description of this method and specification text are provided in JVET-U0066.

CE-3.6: Combination of CE-1.3 and CE-3.5

This experiment will investigate integration of Rice parameter derivation method of CE1.3 with Rice parameter inheritance mechanism of the CE1.2.b modified with classification elements of CE-1.4/1.5. Detailed description of this method and specification text are provided in JVET-U0066.

Tables 1 -3 show simulation results conducted with CTC for HBD/HBR lossy coding (JVET-T2018)

Table 1. Simulation results for AI configuration

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **AI** | **HDR PQ** | | | | | **HDR HLG** | | | | | **SVT RGB** | | | | |
| wY | wU | wV | Enc | Dec | Y | U | V | Enc | Dec | R | G | B | Enc | Dec |
| **CE-1.1** | -0.60% | -0.67% | -0.67% | 82% | 90% | -0.22% | -0.24% | -0.24% | 83% | 87% | -13.86% | -13.23% | -13.23% | 90% | 83% |
| **CE-1.2a** | -0.76% | -0.79% | -0.79% | 92% | 83% | -0.31% | -0.31% | -0.32% | 90% | 82% | -14.54% | -13.82% | -13.82% | 97% | 91% |
| **CE-1.2b** | -0.76% | -0.79% | -0.79% | 92% | 83% | -0.31% | -0.31% | -0.32% | 90% | 82% | -14.79% | -13.97% | -13.98% | 98% | 91% |
| **CE-1.3** | -0.92% | -0.91% | -0.90% | 100% | 96% | -0.50% | -0.41% | -0.41% | 102% | 99% | -14.33% | -13.76% | -13.78% | 102% | 87% |
| **CE-1.4** | -0.94% | -0.89% | -0.91% | 104% | 100% | -0.60% | -0.45% | -0.45% | 103% | 101% | -16.00% | -14.70% | -14.71% | 107% | 106% |
| **CE-1.5** | -0.92% | -0.85% | -0.85% | 102% | 100% | -0.63% | -0.47% | -0.46% | 102% | 102% | -15.96% | -14.57% | -14.60% | 107% | 110% |
| **CE-2.1** | -0.12% | -0.09% | -0.08% | 99% | 97% | -0.10% | -0.05% | -0.05% | 100% | 100% | -11.84% | -11.48% | -11.43% | 101% | 96% |
| **CE-2.2** | -0.18% | -0.20% | -0.19% | 102% | 102% | -0.07% | -0.05% | -0.05% | 101% | 101% | -11.99% | -11.58% | -11.53% | 106% | 104% |
| **CE-2.3** | -0.15% | -0.14% | -0.13% | 101% | 100% | -0.17% | -0.12% | -0.12% | 100% | 100% | -8.74% | -8.30% | -8.26% | 103% | 99% |
| **CE-3.1** | -0.92% | -0.87% | -0.86% | 108% | 103% | -0.54% | -0.43% | -0.43% | 104% | 100% | -14.62% | -13.95% | -13.94% | 106% | 90% |
| **CE-3.2** | -0.92% | -0.81% | -0.81% | 103% | 100% | -0.63% | -0.47% | -0.46% | 102% | 101% | -15.54% | -14.39% | -14.37% | 105% | 106% |
| **CE-3.3** | -0.92% | -0.84% | -0.84% | 102% | 100% | -0.67% | -0.51% | -0.51% | 101% | 102% | -15.96% | -14.60% | -14.61% | 107% | 110% |
| **CE-3.4** | -0.75% | -0.74% | -0.75% | 88% | 99% | -0.36% | -0.33% | -0.33% | 85% | 96% | -14.66% | -13.95% | -13.93% | 95% | 89% |
| **CE-3.5** | -0.95% | -0.87% | -0.88% | 105% | 104% | -0.46% | -0.32% | -0.32% | 104% | 102% | -15.88% | -14.77% | -14.77% | 98% | 97% |
| **CE-3.5.b** | -0.99% | -0.90% | -0.92% | 104% | 107% | -0.55% | -0.41% | -0.41% | 105% | 105% | -15.86% | -14.74% | -14.74% | 113% | 102% |
| **CE-3.6** | -1.05% | -0.92% | -0.91% | 103% | 101% | -0.56% | -0.37% | -0.37% | 103% | 101% | -15.76% | -14.79% | -14.80% | 115% | 96% |

Table 2. Simulation results for LDB configuration

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **LDB** | **HDR PQ** | | | | | **HDR HLG** | | | | | **SVT RGB** | | | | |
| wY | wU | wV | Enc | Dec | Y | U | V | Enc | Dec | R | G | B | Enc | Dec |
| **CE-1.1** | -0.05% | -0.06% | -0.07% | 99% | 95% | -0.03% | -0.04% | -0.04% | 94% | 101% | -8.43% | -7.79% | -7.92% | 101% | 94% |
| **CE-1.2a** | -0.03% | -0.04% | -0.04% | 89% | 89% | 0.03% | 0.02% | 0.03% | 99% | 99% | -8.85% | -8.37% | -8.42% | 103% | 100% |
| **CE-1.2b** | -0.03% | -0.04% | -0.04% | 89% | 89% | 0.03% | 0.02% | 0.03% | 99% | 99% | -8.94% | -8.57% | -8.58% | 114% | 106% |
| **CE-1.3** | 0.07% | 0.08% | 0.09% | 89% | 86% | -0.03% | 0.01% | 0.01% | 101% | 99% | -8.86% | -8.15% | -8.31% | 104% | 94% |
| **CE-1.4** | -0.09% | -0.10% | -0.13% | 103% | 101% | -0.08% | -0.08% | -0.07% | 103% | 102% | -9.38% | -8.96% | -8.97% | 107% | 109% |
| **CE-1.5** | -0.09% | -0.10% | -0.12% | 101% | 100% | -0.09% | -0.07% | -0.07% | 102% | 101% | -9.40% | -8.95% | -8.96% | 104% | 108% |
| **CE-2.1** | 0.01% | -0.01% | 0.01% | 98% | 100% | 0.00% | -0.01% | -0.01% | 99% | 100% | -7.19% | -6.91% | -6.89% | 99% | 100% |
| **CE-2.2** | -0.02% | -0.03% | -0.06% | 102% | 101% | 0.00% | -0.01% | -0.01% | 102% | 100% | -7.10% | -6.79% | -6.78% | 101% | 104% |
| **CE-2.3** | -0.01% | -0.01% | -0.01% | 100% | 100% | 0.00% | -0.01% | -0.01% | 101% | 100% | -4.65% | -4.37% | -4.37% | 100% | 99% |
| **CE-3.1** | 0.06% | 0.08% | 0.10% | 105% | 104% | -0.03% | 0.01% | 0.01% | 104% | 102% | -8.99% | -8.60% | -8.62% | 109% | 97% |
| **CE-3.2** | -0.07% | -0.09% | -0.12% | 101% | 104% | -0.09% | -0.07% | -0.07% | 102% | 101% | -9.43% | -8.96% | -8.96% | 103% | 107% |
| **CE-3.3** | -0.10% | -0.10% | -0.11% | 101% | 101% | -0.09% | -0.07% | -0.07% | 102% | 101% | -9.43% | -8.94% | -8.95% | 104% | 109% |
| **CE-3.4** | -0.03% | -0.02% | -0.03% | 88% | 89% | 0.03% | 0.03% | 0.03% | 86% | 93% | -8.97% | -8.59% | -8.60% | 105% | 96% |
| **CE-3.5** | -0.03% | -0.05% | -0.06% | 101% | 100% | 0.05% | 0.06% | 0.06% | 103% | 105% | -9.53% | -9.14% | -9.15% | 102% | 107% |
| **CE3.5.b** | -0.08% | -0.10% | -0.11% | 101% | 102% | -0.02% | 0.00% | 0.00% | 107% | 106% | -9.52% | -9.13% | -9.14% | 121% | 115% |
| **CE-3.6** | 0.10% | 0.12% | 0.12% | 99% | 97% | 0.01% | 0.05% | 0.05% | 108% | 110% | -9.46% | -9.10% | -9.12% | 127% | 115% |

Table 3. Simulation results for RA configuration

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **RA** | **HDR PQ** | | | | | **HDR HLG** | | | | | **SVT RGB** | | | | |
| wY | wU | wV | Enc | Dec | Y | U | V | Enc | Dec | R | G | B | Enc | Dec |
| **CE-1.1** | -0.12% | -0.14% | -0.16% | 89% | 95% |  |  |  |  |  | -7.90% | -7.98% | -8.00% | 91% | 87% |
| **CE-1.2a** | -0.13% | -0.14% | -0.14% | 77% | 78% | 0.00% | 0.00% | 0.00% | 96% | 103% | -8.41% | -8.49% | -8.47% | 85% | 85% |
| **CE-1.2b** | -0.13% | -0.14% | -0.14% | 77% | 78% | 0.00% | 0.00% | 0.00% | 96% | 103% | -8.56% | -8.62% | -8.60% | 91% | 90% |
| **CE-1.3** | -0.05% | -0.01% | 0.01% | 100% | 102% | -0.04% | -0.01% | 0.00% | 109% | 108% | -8.30% | -8.33% | -8.37% | 98% | 88% |
| **CE-1.4** | -0.20% | -0.19% | -0.21% | 103% | 102% | -0.09% | -0.08% | -0.07% | 103% | 101% | -9.00% | -8.99% | -8.97% | 107% | 108% |
| **CE-1.5** | -0.19% | -0.17% | -0.20% | 102% | 101% | -0.10% | -0.07% | -0.07% | 101% | 101% | -9.06% | -9.01% | -8.98% | 103% | 107% |
| **CE-2.1** | 0.00% | -0.01% | 0.00% | 99% | 100% | 0.00% | 0.00% | 0.00% | 100% | 101% | -6.75% | -7.04% | -6.99% | 100% | 100% |
| **CE-2.2** | -0.05% | -0.04% | -0.05% | 102% | 101% | 0.00% | 0.00% | 0.00% | 101% | 101% | -6.67% | -6.95% | -6.90% | 101% | 103% |
| **CE-2.3** | -0.02% | -0.02% | -0.02% | 101% | 100% | 0.00% | -0.01% | 0.00% | 100% | 100% | -4.10% | -4.37% | -4.34% | 100% | 99% |
| **CE-3.1** | -0.03% | 0.00% | 0.01% | 105% | 125% | -0.04% | -0.01% | 0.00% | 100% | 89% | -8.57% | -8.57% | -8.56% | 106% | 94% |
| **CE-3.2** | -0.18% | -0.17% | -0.19% | 102% | 104% | -0.10% | -0.07% | -0.07% | 102% | 100% | -9.04% | -9.00% | -8.96% | 104% | 106% |
| **CE-3.3** | -0.19% | -0.18% | -0.19% | 102% | 101% | -0.09% | -0.07% | -0.07% | 102% | 101% | -9.09% | -9.00% | -8.97% | 104% | 107% |
| **CE-3.4** | -0.14% | -0.12% | -0.14% | 86% | 93% | 0.00% | 0.00% | 0.01% | 105% | 109% | -8.56% | -8.62% | -8.59% | 88% | 88% |
| **CE-3.5** | -0.17% | -0.16% | -0.16% | 104% | 103% | 0.02% | 0.03% | 0.03% | 110% | 113% | -9.15% | -9.18% | -9.15% | 95% | 98% |
| **CE3.5.b** | -0.20% | -0.18% | -0.19% | 105% | 105% |  |  |  |  |  |  |  |  |  |  |
| **CE-3.6** | -0.01% | 0.03% | 0.04% | 107% | 107% |  |  |  |  |  | -9.05% | -9.08% | -9.06% | 110% | 97% |

It is noted that the gain for TSRC might be relatively low, as it is less frequently used in case of camera captured content.

Input contribution JVET-U0070 provided additional simulation results for CE-3.5 test. The proposal describes a content adaptive encoder algorithm and simplified high level syntax for design in CE-3.5. Software for the test is available in CE-3.5 GIT repository. Reported simulation results were produced by following CE tests description CTC. Cross-check for this test is available in [JVET-U0113](https://jvet-experts.org/doc_end_user/current_document.php?id=10640).

Tables 4 -6 show simulation results conducted with CTC for HBD/HBR lossy coding (JVET-T2018).

Table 4. Simulation results for AI configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **AI** | **HDR PQ** | | | | | **HDR HLG** | | | | | **SVT RGB** | | | | |
| wY | wU | wV | Enc | Dec | Y | U | V | Enc | Dec | R | G | B | Enc | Dec |
| **CE-3.5.b** | -0.99% | -0.90% | -0.92% | 104% | 107% | -0.55% | -0.41% | -0.41% | 105% | 105% | -15.86% | -14.74% | -14.74% | 113% | 102% |

Table 5. Simulation results for LDB configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **LDB** | **HDR PQ** | | | | | **HDR HLG** | | | | | **SVT RGB** | | | | |
| wY | wU | wV | Enc | Dec | Y | U | V | Enc | Dec | R | G | B | Enc | Dec |
| **CE-3.5.b** | -0.08% | -0.10% | -0.11% | 101% | 102% | -0.02% | 0.00% | 0.00% | 107% | 106% | -9.52% | -9.13% | -9.14% | 121% | 115% |

Table 6. Simulation results for RA configuration.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **RA** | **HDR PQ** | | | | | **HDR HLG** | | | | | **SVT RGB** | | | | |
| wY | wU | wV | Enc | Dec | Y | U | V | Enc | Dec | R | G | B | Enc | Dec |
| **CE-3.5.b** | -0.20% | -0.18% | -0.19% | 105% | 105% |  |  |  |  |  |  |  |  |  |  |

In addition to the high bit depth CTC, supplementary tests have also been conducted for experiments targeting TSRC using the TGM RGB 4:4:4 sequences within the non-420 CTC (JVET-Q2013). These tests have been conducted using the non-420 test conditions but with RExt\_\_HIGH\_BIT\_DEPTH\_SUPPORT = 1 and InternalBitDepth = 12 at QPs -8, -3, 2, 7, 12.

For comparison, CE-1.4, which is the RRC method with the best coding efficiency in CE-1, is also tested using the same test conditions and compared with TSRC methods, CE-2.1, CE-2.2 and CE-2.3, and combined methods, CE-3.1, CE-3.2, CE-3.3 and CE-3.4.

Table 8. Simulation results for TGM test sequences

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| TGM 1080p | **AI** | | | | |
| Test | Y | U | V | EncT | DecT |
| CE-1.4 | -0.48% | -0.51% | -0.48% | 104% | 101% |
| CE-2.1 | -4.89% | -4.67% | -4.72% | 100% | 99% |
| CE-2.2 | -5.52% | -5.28% | -5.32% | 102% | 100% |
| CE-2.3 | -5.02% | -4.78% | -4.84% | 101% | 99% |
| CE-3.1 | -4.71% | -4.51% | -4.54% | 100% | 97% |
| CE-3.2 | -5.05% | -4.83% | -4.87% | 103% | 99% |
| CE-3.3 | -5.15% | -4.93% | -4.97% | 103% | 100% |
| CE-3.4 | -5.07% | -4.86% | -4.90% | 102% | 99% |

It is noted that the TGM test sequences were artificially converted to 12 bit by adding 2 zero bits as LSBs.

Summary of cross-check reports

|  |  |  |
| --- | --- | --- |
| **Test** | **Crosschecker** | **Crosscheck report** |
| *CE-1.1* | Sony | **results:**  Cross-check results match proponent’s with neglible discrepancies at QP=-13 and below  **code verification:**  Consistent with CE description.  **draft text verification:**    **further comments:**  Encoder / decoder run-times did not match |
| *CE-1.2* | Sony | **results:**  Three small discrepancies in SVT LDB (2 in subtest (a), 1 in subtest (b))  **code verification:**  Consistent with CE description.  **draft text verification:**  Subtest (a) draft text provided – matches source code  **further comments:**  Encoder / decoder run-times did not match |
| *CE-1.3* | Kwai | **results:**  Cross-check results completely match proponent’s.  **code verification:**  Consistent with CE description.  **draft text verification:**  Same as the software  **further comments:**  Encoder / decoder run-times did not match for PQ in LDB |
| *CE-1.4* | Qualcomm | **results:**  Simulation results for CTC and completed lossless (Intra and LDB) tests match to results reported by the proponent’s in term of coding performance.  **code verification:**  Consistent with CE description.  **draft text verification:**    **further comments:** |
| *CE-1.5* | Qualcomm | **results:**  Simulation results for CTC and completed lossless (Intra and LDB) tests match to results reported by the proponent’s in term of coding performance.  **code verification:**  Consistent with CE description.  **draft text verification:**    **further comments:**  Reporting CTC excel provided by the proponent featured a typo for PQ classes (PSNR is reported instead of wPSNR). Summary results provided in Tables 1-3 of this document (CE report) are correct. |
| *CE-2.1* | Sharp | **results:**  Cross-check results completely match proponent’s.  **code verification:**  Consistent with CE description.  **draft text verification:**  same as the software  **further comments:** |
| *CE-2.2* | Kwai | **results:**  Cross-check results completely match proponent’s.  **code verification:**  Consistent with CE description.  **draft text verification:**  Same as the software  **further comments:** |
| *CE-2.3* | Kwai | **results:**  Cross-check results completely match proponent’s.  **code verification:**  Consistent with CE description.  **draft text verification:**  Same as the software  **further comments:** |
| *CE-3.1* | Sony | **results:**  Cross-check results completely match proponent’s  **code verification:**  Consistent with CE description.  **draft text verification:**    **further comments:** |
| *CE-3.2* | Sharp | **results:**  Partially cross-check results match proponent’s  **code verification:**  Consistent with CE description.  **draft text verification:**    **further comments:** |
| *CE-3.3* | Qualcomm | **results:**  Simulation results for completed QP points of CTC tests match to results reported by the proponent’s in term of coding performance.  **code verification:**  Consistent with CE description.  **draft text verification:**    **further comments:** |
| *CE-3.4* | Sony | **results:**  Cross-check results completely match proponent’s  **code verification:**  Consistent with CE description.  **draft text verification:**  **further comments:**  Encoder / decoder run-times did not match |
| *CE-3.5* | Kwai | **results:**  Simulation results for completed QP points of CTC tests match to results reported by the proponent’s in term of coding performance.  **code verification:**  Consistent with CE description.  **draft text verification:**    **further comments:** |
| *CE-3.6* | Kwai | **results:**  Simulation results for completed QP points of CTC tests match to results reported by the proponent’s in term of coding performance.  **code verification:**  Consistent with CE description.  **draft text verification:**    **further comments:**  Encoder / decoder run-times did not match for SVT RGB |

CE-1 Summary of complexity estimates per test.

|  |  |  |
| --- | --- | --- |
| **Test** | **Complexity** | **Memory requirement** |
| *CE-1.1* | 2 conditions, 1 shifts, 1 adds per Rice call  Number of operations per 4x4 TU , condition/shift/adds/MSB find:  32/16/16, 64 in total. | 4 storages, 32 bits in ROM/RAM. |
| *CE-1.2.a* | 1 condition, 2 shifts, 1 add, 1 MSB find.  Number of operations per 4x4 TU , condition/shift/adds/MSB find:  16/32/16/16, 80 in total. | No memory increase. |
| *CE-1.2.b* | 10 conditions, 31 shifts, 12 adds, 16 MSB finds, 69 in total. | 12 bits of storage for history counters. |
| *CE-1.3* | 1 condition, 1 mult (1shift+1add), 2 adds, 1 Most Significant Bit (MSB) find per Rice call  Number of operations per 4x4 TU , condition/shift/adds/MSB find:  16/16/48/16, 96 in total. | No memory increase. |
| *CE-1.4* | For coding a value (per coeff):  1 variable shift, 1 add  For updating a counter (per coeff):  3 conditionals, 2 adds, 1 fixed shift  For selecting a parameter set (per coeff):  2 conditional, 1 add  For updating associated parameters (per coeff):  6 adds, 3 fixed shifts, 2 variable shifts  Total per 4x4 TU: 352  80 conditionals, 48 variable shifts, 64 fixed shifts, 160 adds | Eight stores each comprising 1\*10 bit, 2 \* 24 bit, 2\*6 bit unsigned values |
| *CE-1.5* | For coding a value (per coeff):  1 variable shift, 1 add  For updating a counter (4 per block):  3 conditionals, 2 adds, 1 fixed shift  For selecting a parameter set (per coeff):  2 conditionals, 1 add  For updating associated parameters (3 per block):  6 adds, 3 fixed shifts, 2 variable shifts  Total per 4x4 TU: 133  44 conditionals, 22 variable shifts,9 fixed shifts, 58 adds | Eight stores, each comprising a single 10 bit unsigned value. |

CE-2 Summary of complexity estimates per test.

|  |  |  |
| --- | --- | --- |
| **Test** | **Complexity** | **Memory requirement** |
| *CE-2.1* | 1 condition and 1 add per slice. | No memory increase. |
| *CE-2.2* | For coding a value (per coeff):  1 variable shift, 1 add  For updating a counter (per coeff):  3 conditionals, 3 adds, 1 fixed shift  For updating associated parameters (per coeff):  6 adds, 3 fixed shifts, 2 variable shifts  Total per 4x4 TU: 320  48 conditionals, 48 variable shifts, 64 fixed shifts, 160 adds | Two stores each comprising 1\*10 bit, 2 \* 24 bit, 2\*6 bit unsigned values |
| *CE-2.3* | For coding a value (per coeff):  1 variable shift, 1 add  For updating a counter (4 per block):  3 conditionals, 3 adds, 1 fixed shift  For updating associated parameters (1 per block):  6 adds, 3 fixed shifts, 2 variable shifts  Total per 4x4 TU: 71  12 conditionals, 18 variable shifts, 7 fixed shifts, 34 adds | Two stores, each comprising a single 10 bit unsigned value. |

CE-3 Summary of complexity estimates per test.

|  |  |  |
| --- | --- | --- |
| **Test** | **Complexity** | **Memory requirement** |
| *CE-3.1* | 1 condition, 1 mult (1shift+1add), 2 adds, 1 Most Significant Bit (MSB) find per Rice call.  Number of operations per 4x4 TU , condition/shift/adds/MSB find:  16/16/48/16, 96 in total. | No memory increase. |
| *CE-3.2* | For coding a value (per coeff):  1 variable shift, 1 add  For updating a counter (4 per block):  3 conditionals, 2 adds, 1 fixed shift  For selecting a parameter set (per coeff):  2 conditionals, 1 add  For updating associated parameters (3 per block):  6 adds, 3 fixed shifts, 2 variable shifts  Total per 4x4 TU: 133  44 conditionals, 22 variable shifts,9 fixed shifts, 58 adds | Eight stores each comprising 1\*10 bit, 2 \* 24 bit, 2\*6 bit unsigned values |
| *CE-3.3* | For coding a value (per coeff):  1 variable shift, 1 add  For updating a counter (4 per block):  3 conditionals, 2 adds, 1 fixed shift  For selecting a parameter set (per coeff):  2 conditionals, 1 add  For updating associated parameters (3 per block):  6 adds, 3 fixed shifts, 2 variable shifts  Total per 4x4 TU: 133  44 conditionals, 22 variable shifts,9 fixed shifts, 58 adds | Eight stores, each comprising a single 10 bit unsigned value. |
| *CE-3.4* | 1 condition, 2 shift, 1 add, 1 MSB find.  16/32/16/16, total 80. | No memory increase. |
| *CE-3.5* | Number of operations per 4x4 TU , condition/shift/adds/MSB find:  CE-1.2.a: 16/32/16/16  History update: 1/1/2/1  History usage: 0/1/19/0  Total number: 17/34/37/17Total 105 per 4x4 TU. | 12 bits of storage for history counters. |
| *CE-3.6* | Number of operations per 4x4 TU , condition/shift/adds/MSB find:  CE-1.3: 16/16/48/16  History update: 1/1/2/1  History usage: 0/1/19/0  Total number: 17/18/71/17  Total 121 per 4x4 TU. | 12 bits of storage for history counters. |

To visualize identified gain vs. complexity ratios, total number of operations per TU size of 4x4 estimated in Section 7, have been agreated in Table along with coding results (average over 3 color components), presented in section 3.

To visualize the results, CE tests are color marked. Example of such table for AI case, and Figures for AI/LDB and RA coding cases are shown below.

Table 11. Summary of coding gain and number of operations per 4x4 TU per test, AI case.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **AI** | **HDR PQ** | | | | **HDR HLG** | | | | **SVT RGB** | | | |  |
| wY | wU | wV | Average | Y | U | V | Average | R | G | B | Average | num Ops |
| **CE-1.1** | -0.60% | -0.67% | -0.67% | -0.65% | -0.22% | -0.24% | -0.24% | -0.23% | -13.86% | -13.23% | -13.23% | -13.44% | 64 |
| **CE-1.2a** | -0.76% | -0.79% | -0.79% | -0.78% | -0.31% | -0.31% | -0.32% | -0.31% | -14.54% | -13.82% | -13.82% | -14.06% | 80 |
| **CE-1.2b** | -0.76% | -0.79% | -0.79% | -0.78% | -0.31% | -0.31% | -0.32% | -0.31% | -14.79% | -13.97% | -13.98% | -14.25% | 69 |
| **CE-1.3** | -0.92% | -0.91% | -0.90% | -0.91% | -0.50% | -0.41% | -0.41% | -0.44% | -14.33% | -13.76% | -13.78% | -13.96% | 96 |
| CE-1.4 | -0.94% | -0.89% | -0.91% | -0.91% | -0.60% | -0.45% | -0.45% | -0.50% | -16.00% | -14.70% | -14.71% | -15.14% | 352 |
| CE-1.5 | -0.92% | -0.85% | -0.85% | -0.87% | -0.63% | -0.47% | -0.46% | -0.52% | -15.96% | -14.57% | -14.60% | -15.04% | 133 |
| **CE-2.1** | -0.12% | -0.09% | -0.08% | -0.10% | -0.10% | -0.05% | -0.05% | -0.07% | -11.84% | -11.48% | -11.43% | -11.58% | 0 |
| **CE-2.2** | -0.18% | -0.20% | -0.19% | -0.19% | -0.07% | -0.05% | -0.05% | -0.06% | -11.99% | -11.58% | -11.53% | -11.70% | 320 |
| **CE-2.3** | -0.15% | -0.14% | -0.13% | -0.14% | -0.17% | -0.12% | -0.12% | -0.14% | -8.74% | -8.30% | -8.26% | -8.43% | 71 |
| **CE-3.1** | -0.92% | -0.87% | -0.86% | -0.88% | -0.54% | -0.43% | -0.43% | -0.47% | -14.62% | -13.95% | -13.94% | -14.17% | 96 |
| **CE-3.2** | -0.92% | -0.81% | -0.81% | -0.85% | -0.63% | -0.47% | -0.46% | -0.52% | -15.54% | -14.39% | -14.37% | -14.77% | 133 |
| **CE-3.3** | -0.92% | -0.84% | -0.84% | -0.87% | -0.67% | -0.51% | -0.51% | -0.56% | -15.96% | -14.60% | -14.61% | -15.06% | 133 |
| **CE-3.4** | -0.75% | -0.74% | -0.75% | -0.75% | -0.36% | -0.33% | -0.33% | -0.34% | -14.66% | -13.95% | -13.93% | -14.18% | 80 |
| **CE-3.5** | -0.95% | -0.87% | -0.88% | -0.90% | -0.46% | -0.32% | -0.32% | -0.37% | -15.88% | -14.77% | -14.77% | -15.14% | 105 |
| **CE-3.5.b** | -0.99% | -0.90% | -0.92% | -0.94% | -0.55% | -0.41% | -0.41% | -0.46% | -15.86% | -14.74% | -14.74% | -15.11% | 105 |
| **CE-3.6** | -1.05% | -0.92% | -0.91% | -0.96% | -0.56% | -0.37% | -0.37% | -0.43% | -15.76% | -14.79% | -14.80% | -15.12% | 121 |

AI case:

Note: horizontal axis is clipped at 150 ops.

Note: horizontal axis is clipped at 150 ops.

Note: horizontal axis is clipped at 150 ops.

LDB case

RA case

The method with 0 additional operations (CE2-1) is using a slice-level signalling (based on statistics of past slices). This is conceptually similar to locally history based approaches of some other proposals, but less complex, and avoiding dependencies. This method has only been investigated for TSRC, which is typically less used in camera captured content.

Generally, more significant gain is observed for SVT. The results above are however averaging coding of 16 bit content (which is the original content) and a version where only the upper 12 bits were input to the encoder. For the latter, it is reported that the gain is significantly lower.

Furthermore, it is asked if the higher gain of SVT is due to the fact that it is coded in RGB format?

The results might indicate that different methods might be useful for 12 and 16 bit cases. This could also be attractive under the assumption that 12 bit (in particular for HDR) might come to consumer applications. However, the results reported are with low QP, not a typical range for consumer.

In terms of complexity, the simplest methods with local adaptation have around 64 additional operations per 4x4 subblock, which increases the number of operations per Rice call by approximately 25%.

Some of the methods have impact by introducing additional dependencies (and some even introducing dependencies across CTUs).

A similar method like CE-2.1 could probably be applied for RRC, where currently the Rice parameter can only be varied between 0 and 3.

Some experts mentioned preference for a simple slice-level approach.

On the other hand, also HEVC included a history based approach with cross-block dependency in RExt.

BoG (A. Browne) to perform more analysis on the impact of different proposals in terms of introducing dependencies between blocks, and possible impact on parallel operation, pipelining and throughput. Also include proposals from 5.1.3 in this analysis.

It is also requested to provide a new version of the CE summary, separating the results SVT for 12 and 16 bit.

Revisit.

[JVET-U0050](https://jvet-experts.org/doc_end_user/current_document.php?id=10562) CE-1.3 and CE-3.1: Rice parameter derivation for high bit-depth coding [T. Hashimoto, T. Ikai (Sharp)]

[JVET-U0109](https://jvet-experts.org/doc_end_user/current_document.php?id=10636) Crosscheck of JVET-U0050 (CE-1.3: Rice parameter derivation for high bit-depth coding) [H.-J. Jhu (Kwai)] [late]

[JVET-U0057](https://jvet-experts.org/doc_end_user/current_document.php?id=10571) CE-1.4, CE-1.5, CE-2.2 and CE-2.3: Rice parameter selection for high bit depths [A. Browne, S. Keating, K. Sharman (Sony)]

[JVET-U0106](https://jvet-experts.org/doc_end_user/current_document.php?id=10621) Crosscheck of JVET-U0057: CE-1.4, CE-1.5: Rice parameter selection for high bit depths [D. Rusanovskyy (Qualcomm)] [late]

[JVET-U0110](https://jvet-experts.org/doc_end_user/current_document.php?id=10637) Crosscheck of JVET-U0057 (CE-2.2 and CE-2.3: Rice parameter selection for high bit depths) [H.-J. Jhu (Kwai)] [late]

[JVET-U0058](https://jvet-experts.org/doc_end_user/current_document.php?id=10572) CE-3.2: Combination of CE-1.5 and CE-2.1 [A. Browne, S. Keating, K. Sharman (Sony)]

[JVET-U0126](https://jvet-experts.org/doc_end_user/current_document.php?id=10653) Crosscheck of JVET-U0058 (CE-3.2: Combination of CE-1.5 and CE-2.1) [T. Hashimoto (Sharp)]

[JVET-U0059](https://jvet-experts.org/doc_end_user/current_document.php?id=10573) CE-3.3: Combination of CE-1.5 and CE-2.3 [A. Browne, S. Keating, K. Sharman (Sony)]

[JVET-U0107](https://jvet-experts.org/doc_end_user/current_document.php?id=10622) Cross-check of JVET-U0059: Combination of CE-1.5 and CE-2.3 [D. Rusanovskyy (Qualcomm)] [late]

[JVET-U0064](https://jvet-experts.org/doc_end_user/current_document.php?id=10578) CE-1.1 and CE-1.2: On the Rice parameter derivation for high bit-depth coding [D. Rusanovskyy, L. P. Van, M. Coban, M. Karczewicz (Qualcomm)]

[JVET-U0090](https://jvet-experts.org/doc_end_user/current_document.php?id=10604) Crosscheck of JVET-U0064 (CE-1.1 and CE-1.2: On the Rice parameter derivation for high bit-depth coding) [A. Browne (Sony)]

[JVET-U0065](https://jvet-experts.org/doc_end_user/current_document.php?id=10579) CE-3.4: Combination of CE-1.2 and CE-2.1 [D. Rusanovskyy, L. P. Van, M. Coban, M. Karczewicz (Qualcomm)]

[JVET-U0095](https://jvet-experts.org/doc_end_user/current_document.php?id=10610) Crosscheck of JVET-U0065 (CE-3.4: Combination of CE-1.2 and CE-2.1) [A. Browne (Sony)]

[JVET-U0066](https://jvet-experts.org/doc_end_user/current_document.php?id=10580) CE-3.5 and CE-3.6: Combination of CE-1.2, CE-1.4/1.5 and CE-1.3 [D. Rusanovskyy, L. P. Van, M. Coban, M. Karczewicz (Qualcomm)]

[JVET-U0111](https://jvet-experts.org/doc_end_user/current_document.php?id=10638) Crosscheck of JVET-U0066 (CE-3.5 and CE-3.6: Combination of CE-1.2, CE-1.4/1.5 and CE-1.3) [H.-J. Jhu (Kwai)] [late]

[JVET-U0117](https://jvet-experts.org/doc_end_user/current_document.php?id=10644) Crosscheck of JVET-U0066 (CE-3.6) [T. Zhou (Sharp)] [late]

[JVET-U0075](https://jvet-experts.org/doc_end_user/current_document.php?id=10589) 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-U0125](https://jvet-experts.org/doc_end_user/current_document.php?id=10652) Crosscheck of JVET-U0075 (CE-2.1: Slice based Rice parameter selection for transform skip residual coding) [T. Hashimoto (Sharp)]

### CE related contributions: Entropy Coding for High Bit Depth and High Bit Rate Coding (3)

Contributions in this area were discussed in Session X at XXXX UTC on XXday X January 2021 (chaired by XXX).

[JVET-U0051](https://jvet-experts.org/doc_end_user/current_document.php?id=10563) Non-CE: Rice parameter derivation for high bit-depth coding with state value [T. Hashimoto, T. Ikai (Sharp)]

[JVET-U0112](https://jvet-experts.org/doc_end_user/current_document.php?id=10639) Crosscheck of JVET-U0051 (Non-CE: Rice parameter derivation for high bit-depth coding with state value) [H.-J. Jhu (Kwai)] [late]

[JVET-U0062](https://jvet-experts.org/doc_end_user/current_document.php?id=10576) CE related: On Rice Parameter Derivation with Content Adaptation [K. Kawamura, K. Unno (KDDI)]

[JVET-U0070](https://jvet-experts.org/doc_end_user/current_document.php?id=10584) CE Related: On signalling and encoder optimization for Rice parameter derivation [D. Rusanovskyy, M. Coban, M. Karczewicz (Qualcomm)]

[JVET-U0113](https://jvet-experts.org/doc_end_user/current_document.php?id=10640) Crosscheck of JVET-U0070 (CE Related: On signalling and encoder optimization for Rice parameter derivation) [H.-J. Jhu (Kwai)] [late]

### Adaptation of transforms and other tools for high bit rate and high bit depth (5)

Contributions in this area were discussed in Session 15 at 1300 UTC on Tuesday 12 January 2021 (chaired by JRO and GJS).

[JVET-U0052](https://jvet-experts.org/doc_end_user/current_document.php?id=10564) AHG8: Transform coefficients range extension for high bit-depth coding [T. Zhou, T. Chujoh, E. Sasaki, T. Ikai (Sharp)]

This contribution proposes an extended coefficient precision for high bit-depth coding. Compared to the VVC version1, where the transform coefficient range is always equal to 15 regardless of bit-depth, this proposal uses Min(20, BitDepth + 6) for the coefficient range. An enabled flag is signaled in SPS extension but the method is only enabled when the flag is equal to 1 and the bit-depth is greater than 10 to avoid 10 bit case changes.

The experimental results are reported (anchor: extended precision = 0, test: extended precision = 1):

Method 1 (sps\_extention\_flag, Min(BitDepth + 6, 20) ):

* 0.64 %, 1.43 % and 5.89 % in bdrateY in PQ, HLG and SVT in AI
* 0.58 %, 1.71 % and 5.22 % in bdrateY in PQ, HLG and SVT in LDB
* 0.54 %, xxx % and 4.62 % in bdrateY in PQ, HLG and SVT in RA

Method 2 (no flag, Max(Min(BitDepth + 5, 20),15) ):

* 0.61 %, 1.33 % and 5.86 % in bdrateY in PQ, HLG and SVT in AI
* 0.54 %, 1.49 % and 5.17 % in bdrateY in PQ, HLG and SVT in LDB
* 0.56 %, 1.40 % and 4.58 % in bdrateY in PQ, HLG and SVT in RA

v2 added no ExtendedPrecisionFlag solution (Method2) where extended range is always used when the bit-depth is greater than 10. There is no change compared to VVC version 1 since the extended range is disabled when bit-depth is equal to or less than 10 bit. It is asserted that no flag solution could significantly reduce verification cost.

Method 1 is equivalent to HEVC and also implemented in VTM (where the extended precision flag from HM was retained).

More detailed results:

Software: VTM11.0 with CE-1.3 in JVET-U0050

Anchor: ExtendedPrecision = 0, Log2TransformRange =15

Test: ExtendedPrecision = 1, Log2TransformRange =Min (20, BitDepth + 6)

Table 3. Results of proposal method 1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | **AI** |  |  |
|  |  | **Over VTM11.0** |  |  |
| psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -0.68% | -1.18% | -1.19% | 104% | 96% |
| PQ422 | -0.60% | -0.79% | -0.79% | 102% | 96% |
| **Overall** | -0.64% | -0.98% | -0.99% | 103% | 96% |
| HLG444 | -1.74% | -3.03% | -3.31% | 106% | 92% |
| HLG422 | -1.12% | -1.94% | -2.06% | 103% | 90% |
| **Overall** | -1.43% | -2.48% | -2.69% | 104% | 91% |
| SVT16 | -11.23% | -9.76% | -9.96% | 97% | 81% |
| SVT12 | -0.55% | -1.16% | -1.17% | 100% | 93% |
| **Overall** | -5.89% | -5.46% | -5.56% | 98% | 87% |
|  |  |  | **RA** |  |  |
|  |  | **Over VTM11.0** |  |  |
| psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -0.58% | -0.70% | -0.77% | 104% | 125% |
| PQ422 | -0.51% | -0.51% | -0.56% | 105% | 126% |
| **Overall** | -0.54% | -0.61% | -0.66% | 105% | 125% |
| HLG444 | -2.16% | -2.86% | -3.00% | 98% | 99% |
| HLG422 |  |  |  |  |  |
| **Overall** |  |  |  |  |  |
| SVT16 | -8.50% | -8.76% | -8.79% | 102% | 83% |
| SVT12 | -0.74% | -0.78% | -0.79% | 101% | 93% |
| **Overall** | -4.62% | -4.77% | -4.79% | 102% | 88% |
|  |  |  | **LDB** |  |  |
|  |  |  | **Over VTM11.0** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -0.61% | -0.71% | -0.76% | 90% | 82% |
| PQ422 | -0.55% | -0.48% | -0.50% | 98% | 94% |
| **Overall** | -0.58% | -0.60% | -0.63% | 94% | 88% |
| HLG444 | -2.30% | -2.92% | -3.07% | 99% | 99% |
| HLG422 | -1.11% | -1.79% | -1.84% | 98% | 93% |
| **Overall** | -1.71% | -2.36% | -2.46% | 98% | 96% |
| SVT16 | -9.55% | -8.85% | -8.87% | 103% | 82% |
| SVT12 | -0.89% | -0.81% | -0.82% | 100% | 100% |
| **Overall** | -5.22% | -4.83% | -4.84% | 102% | 91% |

Table 4. Results of proposal method 2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | **AI** |  |  |
|  |  | **Over VTM11.0** |  |  |
| psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -0.65% | -1.12% | -1.13% | 121% | 116% |
| PQ422 | -0.57% | -0.74% | -0.75% | 119% | 114% |
| **Overall** | -0.61% | -0.93% | -0.94% | 120% | 115% |
| HLG444 | -1.62% | -2.73% | -2.99% | 124% | 110% |
| HLG422 | -1.03% | -1.76% | -1.87% | 120% | 105% |
| **Overall** | -1.33% | -2.25% | -2.43% | 122% | 108% |
| SVT16 | -11.23% | -9.76% | -9.96% | 97% | 81% |
| SVT12 | -0.49% | -1.02% | -1.02% | 118% | 111% |
| **Overall** | -5.86% | -5.39% | -5.49% | 107% | 96% |
|  |  |  | **RA** |  |  |
|  |  | **Over VTM11.0** |  |  |
| psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -0.56% | -0.68% | -0.73% | 104% | 127% |
| PQ422 | -0.49% | -0.48% | -0.51% | 105% | 134% |
| **Overall** | -0.53% | -0.58% | -0.62% | 105% | 131% |
| HLG444 | -1.91% | -2.62% | -2.75% | 98% | 101% |
| HLG422 | -0.84% | -1.57% | -1.64% | 95% | 95% |
| **Overall** | -1.37% | -2.09% | -2.19% | 97% | 98% |
| SVT16 | -8.50% | -8.76% | -8.79% | 102% | 83% |
| SVT12 | -0.65% | -0.71% | -0.71% | 101% | 99% |
| **Overall** | -4.58% | -4.74% | -4.75% | 102% | 91% |
|  |  |  | **LDB** |  |  |
|  |  |  | **Over VTM11.0** |  |  |
|  | psnrY | psnrU | psnrV | EncT | DecT |
| PQ444 | -0.57% | -0.68% | -0.72% | 91% | 83% |
| PQ422 | -0.53% | -0.46% | -0.47% | 102% | 98% |
| **Overall** | -0.55% | -0.57% | -0.59% | 96% | 90% |
| HLG444 | -2.02% | -2.65% | -2.79% | 102% | 99% |
| HLG422 | -0.96% | -1.65% | -1.66% | 102% | 95% |
| **Overall** | -1.49% | -2.15% | -2.23% | 102% | 97% |
| SVT16 | -9.55% | -8.85% | -8.87% | 103% | 82% |
| SVT12 | -0.78% | -0.73% | -0.73% | 106% | 102% |
| **Overall** | -5.17% | -4.79% | -4.80% | 105% | 92% |

Part of the gain may come from the fact that the precision is always clipped to 20 bits (as otherwise the influence on ExpGolomb codeword length could lead to negative impact of several percents of loss in 16 bit case). It is noted that such a dependency was not there in HEVC, and its interaction with the increased precision flag was only recently recognized.

The complexity of transform implementation increases with both methods. Method 2 has the advantage that it only needs implementing one mode. It would however increase the internal bit depth of the transform would also increase from 15 to 18 in method 2, which is not the case with method 1 when the flag is off for 12 bit (as in CTC).

Question: How do the gains relate to the gains reported in the CE? Some of the results are not directly comparable, as different anchors are used.

Though the gains are lower than those in the CE, CE results might significantly change if a relatively simple approach such as 20 bit precision clipping would be applied.

It is suggested to perform further investigation of the relationship between transform precision and entropy coding modifications.

Further study (CE)

[JVET-U0123](https://jvet-experts.org/doc_end_user/current_document.php?id=10650) Cross-check of JVET-U0052 [D. Rusanovskyy (Qualcomm)]

[JVET-U0063](https://jvet-experts.org/doc_end_user/current_document.php?id=10577) A constraint of max transform size for high bit depth [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 know the impact of coding efficiency, experiments to restrict transform size 64 is carried out. For AI test case, it is reported that the simulation results are {0.02%, 0.08%, 0.09%} for PQ contents, {0.05%, -0.03%, 0.03%} for HLG contents and {0.00%, 0.00%, 0.00%} for SVT RGB contents.

* Change history
  + v2: Update simulation results.
  + v3: Add additional test results for SVT RGB contents.

Results might be different in higher QP range.

It is commented that the currently investigated QP range might not reflect all relevant application (particular for 12 bit), and that also the set of test sequences is not necessarily representative.

Complexity benefit not too obvious – it is clear that more implementation cost is higher for higher bit depth (16 bit mainly used in professional devices).

Further study (after re-defining the CTC at least for higher QP).

[JVET-U0130](https://jvet-experts.org/doc_end_user/current_document.php?id=10657) Crosscheck of JVET-U0063 (A constraint of max transform size for high bit depth) [T. Hashimoto (Sharp)] [late]

[JVET-U0067](https://jvet-experts.org/doc_end_user/current_document.php?id=10581) AHG8: On ALF clipping of high bit-depth coding [M. G. Sarwer, Y. Ye (Alibaba)]

It was observed that the coding gain of ALF is significantly reduced at high bit-depth coding. At the very high bit-rate, the reconstructed frame consists of lot of details and high frequency components. Applying ALF may remove the details/high frequencies which ultimately impact the coding performance. In order to avoid the over filtering, this contribution proposes to adjust the ALF clipping values in high bit-depth coding.

Following results are reported:

* PQ sequences: -0.20% (wpsnrY), -0.84% (wpsnrU), -0.81% (wpsnrV)
* HLG sequences: -0.02% (psnrY), -0.02% (psnrU), -0.11% (psnrV)
* SVT sequences: 0.00%(psnrG), 0.00% (psnrB), 0.00% (psnrR)

V2 of this contribution includes the additional results of non-linear ALF.

V3 of this contribution includes complete results of non-linear ALF.

The benefit of ALF in the low QP range might not be too large, anyway.

One expert points out that in this range the gain reported may be within the expected error range of BD rate.

No action.

[JVET-U0124](https://jvet-experts.org/doc_end_user/current_document.php?id=10651) Cross-check on JVET-U0067 [D. Rusanovskyy (Qualcomm)] [late]

[JVET-U0069](https://jvet-experts.org/doc_end_user/current_document.php?id=10583) AHG8: CABAC-bypass alignment for high bit-depth coding [M. G. Sarwer, J. Chen, Y. Ye, R. -L. Liao (Alibaba)]

HEVC range extension supports CABAC bypass alignment flag to improve the throughput of the CABAC engine in high bit-rate applications. This contribution proposes to add the same CABAC-bypass alignment concept to VVC, where the value of the ivlCurrRange of the CABAC engine is set to 256 right before starting of the bypass coding of a coefficient group. Two alignment options are proposed. In the first option, only 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. In the proposed methods, the bypass coded bins can be decoded just by looking at the most-significant-but-one bits of ivlOffset which ultimately allows decoding of multiple bypass coded bins in parallel. In order to allow the flexibility of enable/disable the proposed tool, both sequence and picture level flags are also proposed.

Following results are reported

As compared to VTM-11.0:

* Option1:
  + SVT sequences: 0.50%(G), 0.52% (B), 0.53% (R)
  + HLG sequences: 1.31%(Y), 1.30% (U), 1.33% (V)
  + PQ sequences: 1.21%(Y), 1.51% (U), 1.48% (V)
* Option2:
  + SVT sequences: 0.50%(G), 0.502% (B), 0.53% (R)
  + HLG sequences: 1.31%(Y), 1.30% (U), 1.33% (V)
  + PQ sequences: 1.21%(Y), 1.50% (U), 1.48% (V)

As compared to CE-1.2.a:

* Option1:
  + SVT sequences: 0.58%(G), 0.60% (B), 0.60% (R)
  + HLG sequences: 1.30%(Y), 1.31% (U), 1.33% (V)
  + PQ sequences: 1.21% (Y), 1.52% (U), 1.48% (V)
* Option2:
  + SVT sequences: 0.58%(G), 0.60% (B), 0.60% (R)
  + HLG sequences: 1.31%(Y), 1.31% (U), 1.33% (V)
  + PQ sequences: 1.21% (Y), 1.52% (U), 1.48% (V)

V2 of this contribution includes additional simulation results with enabling the proposed methods only for the frames with temporal id is equal to 0.

V3 of this contribution includes complete results.

This is an encoder option signalled by a flag. Syntax with one flag at SPS and one at PH.

Results when only applying to TId=0 only have minor loss (0.1%)

Option 1 is similar to HEVC. Option 2 is similar combined with the limit mechanism for context coded bins that is defined in VVC. Crosschecker’s opinion is that option 2 is more consistent.

Several other experts also indicated preference for option 2, and that this is relevant at high bit rate.

One expert suggests that it might be more appropriate making the option mandatory (to avoid that decoders have to implement both flag on and off cases, and decoder might also face a throughput problem). On the other hand, it has some impact on compression performance, and would not be useful at higher QP. Would this only apply to certain profiles, such as the high throughput profile in HEVC?

Revisit – come back to this when discussing next steps in high bit rate.

[JVET-U0114](https://jvet-experts.org/doc_end_user/current_document.php?id=10641) Crosscheck of JVET-U0069 (AHG8: CABAC-bypass alignment for high bit-depth coding) [H.-J. Jhu (Kwai)] [late]

[JVET-U0134](https://jvet-experts.org/doc_end_user/current_document.php?id=10661) Crosscheck of JVET-U0069 (AHG8: CABAC-bypass alignment for high bit-depth coding) [A. Browne (Sony)] [late]

[JVET-U0121](https://jvet-experts.org/doc_end_user/current_document.php?id=10648) AHG8: Combination of JVET-U0069 and CE-2.1 [M. G. Sarwer, J. Chen, Y. Ye, R.-L. Liao (Alibaba), H.-J. Jhu, X. Xiu, Y.-W. Chen, W. Chen, C.-W. Kuo, X. Wang (Kwai Inc.)] [late]

(include abstract)

No need for detailed presentation – indicates consistent behaviour of JVET-U0069 with modified entropy coding.

[JVET-U0131](https://jvet-experts.org/doc_end_user/current_document.php?id=10658) Crosscheck of JVET-U0121 (AHG8: Combination of JVET-U0069 and CE-2.1) [T. Hashimoto (Sharp)] [late]

## AHG11: Neural network-based video coding (22)

### General (1)

[JVET-U0116](https://jvet-experts.org/doc_end_user/current_document.php?id=10643) A video dataset for training in neural network based video coding [X. Xu, S. Liu, R. Yao, L. Wang, S. Tian (Tencent), D. Wu (Shenzhen Boyan Technology Ltd.), Y. Hu, J. Li, J. Xia, W. Qi, J. Zhang, J. Wen (Tsinghua University)] [late]

To be presented and discussed in BoG.

### EE contributions: Neural network-based video coding (7)

Contributions in this area were discussed in Session 4 at 1530-1730 UTC on Wednesday 6 January 2021, in Session 8 at 1535-1730 UTC on Thursday 7 January 2021, and in Session 11 at 1300-1500 UTC on Friday 8 January 2021 (chaired by JRO and GJS).

[JVET-U0023](https://jvet-experts.org/doc_end_user/current_document.php?id=10634) EE Summary Report: Neural Network-based Video Coding [E. Alshina, S. Liu, W. Chen, Y. Li, R.-L. Liao, Z. Ma, H. Wang (EE coordinators)]

In the 20th JVET meeting it was decided [1] to setup an exploration experiment on NN-based video coding to learn how to assess NN technology, which is the major purpose of the EE activity. The important goal of Exploration Experiment is to exercise test conditions and complexity assessment methodology.

Anchor for this EE supposed to be VTM10.0 [1]. Reason for this is relatively short meeting circle and later for start EE code work after it release of VTM11.0.

Proponent were encouraged to use Results Reporting Template for performance and complexity assessment methodology [2]. Most of proponents follow this recommendation. None of the proponents reported MS-SSIM. Mostly because HDRTools SW, recommended for MS-SSIM computation was not very convenient (reconstructed YUV storage is needed). VTM11.0 comes with MS-SSIM computation integrated, so in a future this metric computation should be easier. One proponent reported VMAF (and it was observed that PSNR gain corresponds in some cases in drop in VMAF metric).

5 EE participants studied CNN based filters as addition to VVC. All used float-point implementation, 32bits float representation of NN parameters. PyTorch or Tensorflow packages needed additionally to VTM SW. Most of proponents provided detailed instructions for SW usage (which far less trivial than just running VTM). Originally build instruction were provided for Linux only (later updated for Windows).

All CNN filters proponents used results reporting template and reported key parameters for complexity assessment.

It looks fair to compare run time between different variants tested by same proponent, but run-time comparison between 2 different proponents looks tricky at current stage, since NN implementations and levels of optimization vary in wide range.

Among CNN filters proposals compression gain up to 5% was demonstrated. In all cases this gain comes with significant decoder complexity (decoder run time increment is 33 … 579).

One EE test can be seen as combination of super-resolution and enchantment filters with almost any base codec (both HEVC and VVC as base codec were tested by proponents). NN VC test conditions do not suit well to assessment such kind of technologies (for getting comparable bit-rate proponents used QP > 42). How to asses technologies of this kind need to be further discussed during meeting.

Due to the short meeting circle several important aspects (such as limited precision of computations inside NN) study was not conducted.

High level summary of complexity aspects and compression performance demonstrated in EE tests is summarised in Tables below. More details can be found in excel sheet attached to this report.

Using performed experiments following questions regarding CNN-based in-loop filters were to be answered and useful information obtained.

1. Regular convolution gives ~0.3% compression gain compared to DSC (Depth wise Separable convolution), but leads to decoder run time increment ×1.25, also the number of NN parameters by ~ ×4. Thanks to Tencent for this study in JVET-U0060.
2. SE (squeeze-and-excitation) block contributed ~1% compression gain for models trained with SSIM as part of loss-function. Inclusion SSIM as part of loss-function during training results in PSNR drop (>1% in RA cfg), but shows gain in VMAF metric. Visual quality needs to be studied to explain this phenomenon. Thanks to Tencent for this information in JVET-U0074.
3. Nearly 5% compression gain was demonstrated by Qualcomm in JVET-U0094. Default number of hidden convolution layers is 12. If reduced to 6 then performance degrades roughly 0.6…0.8% but decoder run time is reduced by half. Residual scale in NN filter contributes ~0.5% compression gain, w/o decoder run-time implication. Increment number of models (from 4 to 12) with content adaptive selection between those actually leads to performance *degradation* 0.1…0.4% (possible reasons include signalling overhead, adaptivity introduced by using QP as input, etc.). Same NN based filter placed prior to SAO performs ~1% better compared after ALF. Many thanks to Qualcomm for this extensive study in JVET-U0094.
4. 2 ways if Luma and Chroma sizes equalization (Chroma up-sampling and Luma split to 4 planes) have been compared in JVET-U0101. Chroma up-sampling appears to perform ~1% better than Luma split. But in terms of decoding run time (also memory usage) Luma split is roughly twice simpler. With 10 residual blocks 3.4% compression gain in RA cfg can be achieved by NN based filter. With twice more residual blocks (20) complexity doubles and 0.7...0.8% additional compression gain can be achieved. Thanks to Kwai for conducting this study in JVET-U0101.

**Table 1.** Summary of CNN-based in-loop filters experiments.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Proposal | Anchor, VTM, GOP size | Total Parameter Num. | Total Conv. Layers | Position | Variations | | Y-PSNR | U-PSNR | V-PSNR | EncT, × | DecT, × | Y-PSNR | U-PSNR | V-PSNR | EncT, × | DecT, × |
| [JVET-U0060](https://jvet-experts.org/doc_end_user/current_document.php?id=10574) | 10.0 | 2.2E+04 | 13 | prior  SAO | **DSC** |  | -1.0% | -3.8% | -3.8% | 1.1 | 35 | -1.3% | -2.8% | -3.6% | 1.1 | 33 |
| (JVET-T0057) | 16 | 7.7E+04 | 13 | **Regular Conv** | | -1.3% | -5.8% | -5.9% | 1.1 | 44 | -1.6% | -4.5% | -5.4% | 1.1 | 41 |
| [JVET-U0074](https://jvet-experts.org/doc_end_user/current_document.php?id=10588) | 10.0 | 1.3E+05 | 45 | prior  SAO | **SE block** | **0.8 SSIM; 0.2 L1** | -0.8% | -7.9% | -5.9% | 1.4 | 243 | -2.6% | -10.1% | -9.2% | 1.4 | 303 |
| (JVET-T0069) | 16 | 1.3E+05 | 45 | SE block | 0.0 SSIM; **1.0 L1** | -2.2% | -10.4% | -8.6% | 1.4 | 288 | -2.9% | -9.3% | -9.1% | 1.4 | 319 |
|  |  | 1.3E+05 | 45 | **no** | 0.8 SSIM; 0.2 L1 | 0.1% | -3.7% | -1.7% | 1.4 | 215 | -1.4% | -5.7% | -5.7% | 1.4 | 267 |
|  |  | 1.3E+05 | 45 | **no** | 0.0 SSIM; **1.0 L1** | -2.0% | -8.3% | -7.6% | 1.4 | 246 | -2.5% | -7.3% | -7.7% | 1.3 | 218 |
| [JVET-U0094](https://jvet-experts.org/doc_end_user/current_document.php?id=10609) | 10.0 | 4.0E+06 | 14 | prior  SAO | **N=12,** | **X=4, res. scale** | -5.3% | -13.1% | -12.9% | 1.4 | 149 | -4.8% | -10.1% | -10.6% | 1.3 | 85 |
| (JVET-T0079) | 32 | 2.0E+06 | 8 | **N=6,** | X=4, res. scale | -4.6% | -11.8% | -11.6% | 1.2 | 78 | -4.2% | -8.8% | -9.9% | 1.1 | 45 |
|  |  | 4.0E+06 | 14 | N=12, | X=4, **NO** res. scale | -4.8% | -11.7% | -11.9% | 1.4 | 148 | -4.6% | -9.4% | -10.4% | 1.2 | 85 |
|  |  | 1.2E+07 | 14 | N=12, | X=12, res. scale | -4.9% | -12.7% | -12.6% | 2.1 | 148 | -4.4% | -9.9% | -10.7% | 1.6 | 85 |
|  |  | 1.2E+07 | 14 | **a.ALF** | N=12, | X=12, res. scale | -4.0% | -12.1% | -12.3% | 2.1 | 148 | -3.3% | -8.3% | -9.3% | 1.6 | 85 |
| [JVET-U0101](https://jvet-experts.org/doc_end_user/current_document.php?id=10616) | 10.0 | 1.6E+06 | 40 | prior  SAO | **Ch.Up** | **ResBlocks 20** | -4.1% | -18.5% | -17.8% | 1.5 | 410 | -5.3% | -16.7% | -17.7% | 1.5 | 404 |
| (JVET-T0094) | 32 | 8.9E+05 | 21 | Ch.Up | **ResBlocks 10** | -3.4% | -14.2% | -13.9% | 1.2 | 187 | -4.5% | -12.6% | -13.8% | 1.3 | 193 |
|  |  | 1.5E+06 | 40 | **L.split** | ResBlocks 20 | -3.3% | -15.9% | -15.1% | 1.3 | 212 | -4.2% | -14.1% | -14.8% | 1.3 | 224 |
| [JVET-U0054](https://jvet-experts.org/doc_end_user/current_document.php?id=10568) | 10.0 | 9.0E+06 | 66 | prior  SAO | **Ch.Up** | **ResBlocks 32** |  |  |  |  |  |  |  |  |  |  |
| (JVET-T0128) | 16 |  |  | (\*)only classes BCD | | -5.0% | -15.9% | -17.1% |  |  |  |  |  |  |  |

In the table above, “X” refers to number of models

Useful information group gets out of last EE test is

1. Compression gain demonstrated is 8.7% (if VVC is used as base codec) and 24.3% (if HEVC is used as base codec). Thanks to Nanjing University and OPPO for this information in JVET-U0096.

Changes relatively to VTM much more significant for this test, so only Low-delay-P results were provided, no run-time, no Chroma quality metrics.

**Table 2.** Summary experiment on Decomposition, Compression, Synthesis technology.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Proposal | Anchor, VTM, GOP size | Total Parameter Num. | Total Conv. Layers | Variations | | Y-PSNR | U-PSNR | V-PSNR | EncT,  | DecT,  |
| [JVET-U0096](https://jvet-experts.org/doc_end_user/current_document.php?id=10611) | 10.0 |  |  | vs HEVC |  | -24.3% |  |  | 0.2 |  |
| (JVET-T0125) | 16 |  |  | vs VVC (QP=32,37,42,47) | | -8.7% |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |

Larger number of NN parameters (more memory for storage and higher computational complexity) results in higher compression gain (as shown on Fig. 1-3).

Fig. 1. Larger number of NN parameters results in higher compression gain.

Fig. 2. RA gain vs. number of conv. layers

Fig. 3. RA gain vs. decoder runtime increment (measured on CPU)

4 Training and test conditions

For training proponents used BVI-DVC (JVET-U0054, JVET-U-0101, JVET-U-0094), DIV2K (JVET-U0060, JVET-U0074), SJTU 4K (JVET-U0096).

It is commented that methods which used BVI for training had somewhat higher gain (which is probably due to the fact that it is video whereas DIV2K is still images). It would be useful to use same sets for training in future investigation.

5 Conclusion

Results reporting template and complexity assessment methodology was used in this EE. It appears to provide useful information and recommended to further refined.

One aspect hat seems to be missing is the area of operation of a certain method: whole picture, block-wise, etc.

Several technical questions which can result in practical recommendation for NN technologies development were answered during this EE round. It is recommended to continue EE practice.

Different trends in PSNR and VMAF (gain in one metric, but drop in another) were demonstrated during this EE. Inclusion and active report of more than just PSNR is highly encouraged for future EE.

It is recommended to review all EE contributions for further details

Another expert mentions that it is known that VMAF has sometimes strange behaviour when reporting BD results. MS-SSIM might be more stable (exact method still to be discussed).

It is also mentioned that if MS-SSIM would be used for training, it can be expected that a proposal performs better in that. This might however not be critical if it is known, and would be needed for interpretation of results. Another expert mentions that, when using SSIM for comparison, it would be fair to use SSIM in the RDO of the anchor as well.

It is mentioned that the measurement of decoder runtimes between different proposals may be misleading, as it is highly dependent on software implementation and degree of code optimization.

In a follow-up, a short discussion on subjective viewing was conducted by the end of session 7. It is agreed that subjective viewing is important in the NNVC exploration. Some of the established expert viewing methods could be used. However, also informal viewing for understanding the visual effects of NNVC technology is necessary to understand the nature of artifacts that are occurring (which could be quite different in particular in case of end-to-end technology), or see if traditional artifacts are removed in case of hybrid approaches.

It is suggested to define a certain subset of sequences / (at approximate) rate points for which YUV files should be provided, such that interested experts could perform individual viewing, or these could be used in a more organized expert viewing. These could also be high bit rate mp4 encodings.

One expert confirms that according to his experience visual artifacts are quite different (and sometimes surprising) in particular for end-to-end methods.

As discussed before, PSNR is not appropriate when comparing very different technology.

Exchanging thoughts about this could also be included in the JPEG liaision. A more automated solution than we currently have in our expert viewing would be desirable.

A synchronized viewing as we had it is probably desirable, as it allows to discuss the impression among the participating experts.

Investigate if it is possible to organize a trial of such viewing during the meeting (as a demonstration). A. Segall, E. Alshina to coordinate with M. Wien on organizing such a session with appropriate cases from EE.

[JVET-U0054](https://jvet-experts.org/doc_end_user/current_document.php?id=10568) EE: Neural network based in-loop filtering [Z. Wang, R.-L. Liao, C. Y. Ma, Y. Ye (Alibaba)] [late]

This contribution reports the results of Exploration Experiment 1.6 on neural network (NN) based in-loop filtering. It is a similar version of the prior contribution JVET-T0128 while is retrained on the BVI-DVC dataset and tested conforming the common test conditions. The experimental results show that our method can achieve -4.26%, -15.14% and -14.88% coding gain over VVC for Y, U and V components under RA configuration.

Presented Wed. 6 Jan. 17:00

Presentation deck to be provided.

Training was done with compressed data with four QP ranges (four models). Number of parameters 2.25 M/model. Network has 60 layers with 64 filters each

Training uses 2000 epochs. It is used that some other proposals used much less.

It is noted that currently it is only optional to report the training results such as training error convergence. Should this be made mandatory?

The proposal is not inputting QP into the network, but switches model depending on base QP. The assumption is that the QP is not frequently changed, so only one model is used. It is noted that other proposals are inputting local QP.

14 M iterations were used in training from BVI dataset. Only a subset of 200 sequences was used (full resolution except 4K), randomly selected

There is on/off signalling at CTU level (separately for Y,U,V), but all CTU related flags are put into PH.

Statistics how often it is turned on? No

It is recommended that also the peak number of parameters per layer should be reported. If several models are used, the total number of parameters, and the largest possible model size should be reported.

[JVET-U0060](https://jvet-experts.org/doc_end_user/current_document.php?id=10574) EE-1.1: A comparison of depthwise separable convolution and regular convolution with the JVET-T0057 neural network based in-loop filter [C. Auyeung, W. Wang, W. Jiang, X. Li, S. Liu (Tencent)]

This contribution provides description of the NN-based in-loop filter in EE1.1 from JVET-T2023. The EE1.1 compares the complexity and coding performance of the in-loop filter when the depthwise separable convolution (DSC) from JVET-0057 is replaced by regular convolution (RC). When the DSC in JVET-T0057 is replaced by RC, the memory usage and the number of GMAC of the NN with RC increases to 3.5 times of the NN with DSC. In an average of the RA/LB/LP/AI configurations, the luma BD-Rate changed from ‑0.99% with DSC to ‑1.30% with RC. The decoding time changed from 3837% with DSC to 4831% with RC. It this EE, RC and DSC is about the same in %YBDR per %DecTime for RA and AI. However, RC is better than DSC for LB and LP.

Weights are the same in RC and DSC, but the latter saves computation, reduction to roughly 1/3. This is however not reflected similarly in terms of decoding time reduction.

A mean shift (with global mean) is performed such that the mean is approximately zero at network input.

It is pointed out that DSC might also have advantage for integerization.

[JVET-U0074](https://jvet-experts.org/doc_end_user/current_document.php?id=10588) EE: SSIM based CNN model for in-loop filtering [T. Ouyang, H. Zhu, Z. Chen (Wuhan Univ.), X. Xu, S. Liu (Tencent)]

This contribution reports the experimental results of Exploration Experiment 1 on Neural Network based filtering for video coding of JVET-T0069. In the test, the impact of Squeeze-and-excitation network (SE block) and SSIM based loss functions are investigated. More specifically, several sets of experiments are set up to explore their impacts separately.

Due to the limitation of computer resources and time, four sets of experiments are set up, which are implemented on VTM-10.0. The experimental results under All Intra (AI) and Random Access (RA) configuration with 5 QPs recommended by AHG11 are reported in this contribution. From the results, the SE-block is proven to have more BD-Rate savings over PSNR as well as VMAF (Video Multimethod Assessment Fusion). SSIM based loss function shows superiority in preserving subjective quality in compressed video (through VMAF evaluation), while the model trained with L1 loss performs better in saving more BD-Rate over PSNR.

The contribution exercises training on L1 loss, on SSIM loss, and indicates that based on that the performance either in PSNR or in VMAF becomes better. The original proposal from the previous meeting used a weighted combination of both metrics as a compromise.

In another “Test 3”, the “SE box” (Squeeze-and-excitation network) is removed. Some experts noticed that this does not deliver consistent conclusion regarding improvements (sometimes better, sometimes worse VMAF).

Would be interesting to see if the switch L1/SSIM or using SE has any impact on visual quality.

It is noted that the decoder runtime is much higher than in some other proposals.

Was MSE used for optimization? No.

SSIM results are not presented.

[JVET-U0094](https://jvet-experts.org/doc_end_user/current_document.php?id=10609) EE: Tests on Neural Network-based In-Loop Filter [H. Wang, M. Karczewicz, J. Chen, A. M. Kotra (Qualcomm)]

This contribution reports the EE test results of JVET-T0079. In the proposed Neural Network-based method, one NN-based filter can be selected out of X candidates as an in-loop filter for each picture. Scaling factors are derived and signaled at picture level for each color component to further improve the quality of the filter output. In this EE, several aspects of the proposed NN filter, including different numbers of candidate filters (X=4 or 12), model sizes (1M or 510K), residual scaling on and off, and positions of the filter (placed after ALF or after deblocking ) are tested.

Presentation deck to be provided.

Training uses BVI-DVC. Training data with QPs 20 … 45

Input is 128x128 block (CTU, can be turned on/off per CTU

5 different models were tested, three using QP map as input, otherwise varying in number of layers, position in LF chain (after ALF / prior to SAO). All models were trained with same conditions. Models different for RA and AI. Roughly 1 M parameters per model. Selection between 4 models possible in case of QP input, 12 in case of no QP input.

The output is scaled by a factor and added to the input. This factor is constant per picture and signalled in picture header (separate for the three components).

QP as input tends to give somewhat better results.

Currently based on fixed QP assumption, but could be modified for variable QP.

[JVET-U0096](https://jvet-experts.org/doc_end_user/current_document.php?id=10611) EE: Tests on Decomposition, Compression and Synthesis (DCS)-based Technology [M. Lu, Z. Ma (Nanjing Univ.), L. Xu, D. Wang (OPPO)]

Inspired by the facts that retinal cells actually segregate the visual scene into different attributes (e.g., spatial details, temporal motion) for respective neuronal processing, it is proposed to first decompose the input video into respective spatial texture frames (STF) at its native spatial resolution that preserve the rich spatial details, and the other temporal motion frames (TMF) at a lower spatial resolution that retain the motion smoothness; then compress them together using standardized video coder; and finally synthesize decoded STFs and TMFs for high-fidelity video reconstruction at the same resolution as its native input. For this resolution-adaptive synthesis at decoder, a motion compensation network (MCN) is devised on TMFs to efficiently align and aggregate temporal motion features that will be jointly processed with corresponding STFs using a nonlocal texture transfer network (NL-TTN) to better augment spatial details, by which the compression and resolution resampling noises can be effectively alleviated with better rate-distortion efficiency. Such DCS based scheme is codec agnostic, exemplifying ≈1 dB PSNR gain or ≈25% BD-Rate saving on average, against the HEVC LDP anchor using common test conditions. A recent extension to VVC has also promised the similar gains. Given that this preliminary studies have offered very encouraging results, it is suggested to further study the DCS techniques deeply for future standardization exploration.

Presentation deck to be provided.

4 QPs 32 … 47, 4 models per QP

At lower QPs, the method might not be showing good performance, as the synthesis might no longer be competitive if the data rate is high enough.

Significant loss for BQ Terrace, and slight loss for Basketball

It is pointed out that the method might benefit from using the downsampling from RPR.

Was the QP properly adjusted for the downsampled pictures? Was not investigated in detail.

Only BD rates for Y PSNR are shown. Chroma is also downsampled but conventionally upsampled. The method is currently not applied to chroma.

Does the synthesis sometimes generate artifacts? Not known – should be included in the viewing session.

[JVET-U0101](https://jvet-experts.org/doc_end_user/current_document.php?id=10616) EE-2.1.5: In-loop filtering based on neural network [W. Chen, X. Xiu, Y.-W. Chen, H.-J. Jhu, C.-W. Kuo, X. Wang (Kwai)]

This contribution reports the results of Exploration Experiment 2.1.5 on in-loop filtering based on neural network (NN). There are three tests in this EE test. In Test #1, the original NN-based in-loop filter scheme as proposed in JVET-T0094 is tested. In Test #2, the performance of a simplified NN model with reduced number of residual blocks and reduced size of feature map is tested. In Test 3, the performance of the proposed NN model with removing the convolution-based chroma up/down sampling is investigated. The simulation results reportedly show that under RA configuration, the corresponding average {Y, U, V} BD-rate savings of the three tests are {-4.15%, -18.49%, -17.77%}, {-3.43%, -14.15%, -13.87%} and {-3.25%, -15.87%, -15.05%}, respectively.

Training performed with 4 QPs as in CTC. Only one model, QP information is fed into the network (also QP scaling factor). 2 pass encoding to determine the best scaling factor

Number of parameters roughly 1.5 M in tests 1 and 3, 900 K in test 2.

It is commented that the number of (Mega)MAC may be better to assess the complexity rather than number of parameters. Test 1 has 1.5 MMAC, Test 2 0.19 MMAC, Test 3 0.38 MMAC, even though 1 and 3 have same number of parameters. Overall, Test seems best in terms of tradeoff complexity /performance.

It is asked whether the frequency response of chroma upsampling filter (which is trained) was investigated. This was not the case.

The concept of chroma upsampling is similar to JVET-U0094.

It is reported that convergence in training is similar in all three tests.

The network is operated on CTU basis, can be enabled/disabled per CTU.

### EE related contributions: Neural network-based video coding (5)

Contributions in this area were discussed in Session 11/12 at 1405-1530/1550-1710 UTC on Friday 8 January 2021 (chaired by JRO).

[JVET-U0055](https://jvet-experts.org/doc_end_user/current_document.php?id=10569) AHG11: Multi-density network for in-loop filtering [Z. Wang, R.-L. Liao, C. Y. Ma, Y. Ye (Alibaba)] [late]

This contribution is an improved version of the contribution JVET-U0054 where a neural network based in-loop filter (NNLF) is proposed. This contribution proposes a multi-density network to be used in NNLF and the other information is similar to JVET-U0054. The experimental results show that our method can achieve 5.60%, 14.92% and 15.24% coding gain on VTM-10.0 for Y, U and V components under RA configuration.

Does the weight branch have the same number of channels as feature branch? Yes.

Roughly 1% better than JVET-U0054, most of this gain comes by the multi-branch method. Only 2 seconds were encoded in the results of JVET-U0055, so it might not be exactly comparable.

Reported decoding time is with GPU.

Model size is 3.44 M parameters (JVET-U0054 was approx. 2.25)

Motivation is an approximation of receptive field – one expert points out that this looks similar to “inception network”.

[JVET-U0061](https://jvet-experts.org/doc_end_user/current_document.php?id=10575) EE-1.1-related: BD-Rate improvements to JVET-T0057 neural network based in-loop filter using depthwise separable convolution and regular convolution [C. Auyeung, X. Li, S. Liu (Tencent)]

(include abstract)

Main goal is improvement of performance, not simplification

One of the advantages where the new method gives gain is by performing a clipping at the output of the upsampling.

[JVET-U0068](https://jvet-experts.org/doc_end_user/current_document.php?id=10582) AHG11: Convolutional Neural Network-based 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. It is an extended version of the prior contribution JVET-T0088. The proposed CNN-architecture features a convolutional layer with a stride of two and several residual blocks to increase the valid receptive field and enable a smooth optimization. To better capture local characteristics of an image, the proposed technique embraces adaptive model selection at CTU level and slice level. Compared with VTM-9.0, the proposed method reportedly shows on average {8.33%, 23.11%, 23.55%}, {10.28%, 28.22%, 27.97%}, and {9.46%, 23.74%, 23.09%} BD-rate reductions for {Y, Cb, Cr}, under AI, RA, and LDB configurations, respectively.

Deblocking and SAO disabled, ALF/CCALF after the NN

Models for intra were trained using Div2K, RA/LB trained with UB set. Three models are used for luma, and three for chroma.

Model selection can be done at CTU level. This could imply that a frequent loading of model parameters is necessary, and imposes some more burden on the encoder.

It is commented that a tendency can be observed from different proposals (including this one) that local adaptation improves the performance.

For luma, only luma is used. For chroma, all three components are input.

It would be interesting to provide more analysis about percentage usage of models.

One expert points out that there seems to be a discrepancy between encoder and decoder time usage of models. Proponents say there may be some inaccuracy in computing run times.

It is claimed by proponents that probably encoder run times could be further reduced by fast decision, or reducing the model size which should be possible due to CTU adaptive selection.

How much gain due to CTU based model selection? Not known.

An additional input to the model is the prediction. What is the additional benefit? Not known.

Did anybody try also feeding in motion vectors?

Subjective quality problem due to removing deblocking and SAO? This should be checked. The proponents report that they observed some quality improvement at least for lower rates.

Training loss seems to vary (expressed in PSNR)? It is asserted that this is due to varying characteristics of sequences per batch.

[JVET-U0077](https://jvet-experts.org/doc_end_user/current_document.php?id=10591) AHG11: Revisiting SAO in-loop filter with Neural Networks [P. Bordes, F. Galpin, T. Dumas, P. Nikitin (InterDigital)]

This contribution proposes a Neural Network-based In-loop filter (NN filter). The proposed NN filter replaces the regular SAO filter while re-using some of its features.

It is reported that averagely 1.76%, 4.96%, and 4.52% BD rate saving for RA, 1.62%, 6.62%, and 5.65% BD rate saving for LDB and 1.91%, 2.91%, and 3.03% BD rate saving for AI, for Y, Cb, and Cr components respectively, are achieved.

7 models for luma (selected per CTU), 1 model for chroma. Models can also be combined (weighted linear). Were trained separately.

(relatively) low encoding and decoding times (compared to other proposals)

This is due to the fact that relatively small models are used, and also computation in integer precision (16 bit, MAC in 32 bits).

Why replacing SAO? It is reported that the similar syntax is used, which otherwise would duplicate. Furthermore, SAO is conceptually similar but simpler in terms of classifying pixels.

What is coding performance difference by converting float to integer precision? There is no floating point implementation in VTM.

It is commented that it would be interesting also seeing results as post filter.

Integerization could be an interesting for the next round of EE, and this is the first example of a proposal doing that.

Visual difference? The proponents did not do extensive viewing, but dis not notice differences.

[JVET-U0104](https://jvet-experts.org/doc_end_user/current_document.php?id=10619) AHG11: In-loop filtering with convolutional neural network and large activation [J. Chen, H. Wang, A. M. Kotra, M. Karczewicz (Qualcomm)]

This contribution proposes a Neural Network-based in-loop filter (NN filter). A CNN with large activation layer is employed and the filter is applied after deblocking filter. The proposed filters achieve:

3.85%, 8.75% and 8.11% BD rate saving for RA and 3.45%, 6.54%, 6.84% BD rate saving for AI, for Y, Cb and Cr components respectively, with ~138K model parameters.

4.53%, 12.38% and 12.25% BD rate saving for RA and 4.02%, 9.00% and 9.77% BD rate saving for AI, for Y, Cb and Cr components respectively, with ~270K model parameters.

x.xx%, xx % and xx% BD rate saving for RA and 4.50%, 10.40%, 11.23% BD rate saving for AI, for Y, Cb and Cr components respectively, with ~510K model parameters.

Extension of JVET-U0094 subtest 2 with same model size (but some layers have larger size, and also number of layers increases). It cannot be identified which of the changes has most impact. M is chosen approx.. 3 times larger than K.

Run time for same model size (e.g. 510K) is slightly larger than in EE.

Compared to EE, in AI it is reducing BD rate by another 0.3% (RA not finished yet).

[JVET-U0115](https://jvet-experts.org/doc_end_user/current_document.php?id=10642) AHG11: Neural Network-based In-Loop Filter Performance with No Deblocking Filtering stage [H. Wang, J. Chen, A. M. Kotra, M. Karczewicz (Qualcomm)] [late]

This contribution proposes a Neural Network-based in-loop filter (NN filter). The NN model design is based on the subtest 1 of JVET-U0094. Additionally, 8 lines of neighboring pixels to each side of the current CTU are included into the input reconstructed samples to the NN model. The deblocking boundary strength information of VVC is also used as additional NN model input information but the actual deblocking filtering process is bypassed. Simulation results reportedly show 6.95%, 13.92 % and 14.97% BD rate saving for RA and 6.13%, 11.87%, 13.46% BD rate saving for AI, for Y, Cb and Cr components respectively.

Input is 144x144 block instead of 128x128. Uses additional samples from neighboring CTUs, output is still 128x128. Output is scaled residual added to the input. Furthermore, BS from deblocking is input to the network. Also in training only inner 128x128 used to compute the loss.

It is suggested that alternatively blending could be used with overlapping blocks.

BS values are put on boundary positions such that the network knows where the boundaries are.

Subjective quality? Not thoroughly investigated. General observation that edges are sharper and colour is cleaner.

Is deblocking also disabled when the network is off? Yes. However apparently, the network is used in most cases.

The scaling factor can be changed frame by frame. Could this also be changed on CTU level?

### Tools in “hybrid” architectures (2)

Contributions in this area were discussed in Session 13 at 1300-1355 UTC on Monday 11 January 2021 (chaired by Y. Ye).

[JVET-U0087](https://jvet-experts.org/doc_end_user/current_document.php?id=10601) AHG11: Updated information on inter-prediction coding tool with deep neural network [Z. Li, B. Choi, W. Wang, W. Jiang, X. Xu, S. Liu (Tencent)]

This informative contribution reports the updated results of deep neural network (DNN) utilization for inter-prediction with newly designed and trained models. The basic framework is the same as JVET-T0058 whose idea is inserting of a new (virtual) reference picture, which is generated by inference processes of trained networks, into reference picture list (RPL). The new network consists of several sub-network models for flow estimation/compensation and detail enhancement. The new models show improved coding gains 1.87%, 3.26%, 4.99% respectively, for classes B, C & D in luma for RA with high four QPs.

One virtual reference pictures was generated for pictures using both forward and one backward reference pictures (in display order) that are of the same temporal distance to the current picture, and the temporal distance can be up to 8.

Only one model (regardless of QP and regardless of temporal level) was trained and used for all applicable pictures.

It was asked why the coding gain was higher for smaller resolution video sequences. The proponent said this could be because objects in smaller resolution video are more likely to occupy a larger portion of the picture content, and potentially better fit the patch size of 480x272x3.

It was asked why the decoding run time increase was so large. The proponent said this was because more complicated model was used in this contribution.

It was commented that increasing the number of reference picture by 1 using another “normal” reference pictures (non-CNN generated) could also increase coding performance.

[JVET-U0105](https://jvet-experts.org/doc_end_user/current_document.php?id=10620) AHG11-related: Investigation on CNN-based Intra Prediction [M. Meyer, C. Rohlfing (RWTH Aachen Univ.)

This document reports several findings from an investigation on CNN-based intra prediction. This includes a comparison of different training loss functions, findings related to the integration of coding artifacts into the training data as well as the handling of optionally available parts of the reference area. Further, a comparison of different network architectures with regard to their prediction quality as well as their computational complexity is conducted and the effects of pruning these networks are evaluated.

The proponent was asked to upload a revision with correct document header in v3. This contribution is for information only.

On top of HM-16.9, Y/U/V coding gain was reported to be -2.61%/-1.70%/-1.87% for Class B and -2.39%/-2.14%/-2.31% for Class C.

Training and inference are QP “agnostic” – models are trained for luma and chroma separately but do not rely on QP.

### “End to end” architecture concepts (3)

Contributions in this area were discussed in Session 17 at 1525-1625 UTC on Tuesday 12 January 2021 (chaired by Y. Ye).

[JVET-U0079](https://jvet-experts.org/doc_end_user/current_document.php?id=10593) A DNN Architecture for Intra-Frame Coding in YUV 4:2:0 format with Cross-Component Prediction [H. E. Egilmez, A. K. Singh, M. Coban, M. Karczewicz (Qualcomm)]

Most existing deep neural network (DNN) based video coding architectures are designed to operate in non-subsampled input formats such as RGB or YUV 4:4:4. However, video coding standards such as HEVC and VVC are designed primarily for the YUV 4:2:0 colour format. In the 20th JVET meeting, JVET-T0123 introduced three distinct network architectures including a solution jointly coding luma and chroma components. This contribution document proposes two alternative joint coding designs for intra-frame coding in YUV 4:2:0 colour format. The experimental results demonstrate that the proposed methods outperform HM-16.20 about 12.5% in luma coding. Over the joint coding scheme in JVET-T0123, Method 1 provides 6.43% Y, 5.56% U, 6.76% V, and Method 2 achieves 6.64% Y, 6.97% U, 8.34% V coding improvements.

The Y/U/V performance compared to HM-16.20 is reported to be -12.41%/46.07%/13.65% and -12.61%/37.48%/10.91%, i.e., luma has coding gain but chroma has coding loss.

Compared to JVET-T0123, this contribution has higher coding gain in all 3 colour components, with somewhat reduced number of parameters.

It was noted that the ratio of luma weight and chroma weight in the loss function is 4x.

Regarding the chroma coding performance loss, it was asked if the loss is reflected in visual quality. It was reported that the average loss is mainly due to only a few sequences, ParkingRunning and Campfire. And the proponent reported that the visual quality of these sequences appear to be OK.

Regarding the observation that GDN can be removed without performance loss (method 2 vs. method 1), it was commented that the additional 1x1 convolution layer in this contribution on top of JVET-T0123 seems to make GDN unnecessary.

It was commented that bit allocation between different colour components may be needed to address the large chroma loss in some sequences.

Further study of YUV 4:2:0 coding in the context of E2E framework is encouraged.

[JVET-U0080](https://jvet-experts.org/doc_end_user/current_document.php?id=10594) Balancing Luma-Chroma Channel Coding for DNN-based Intra-Frame Coding in JVET-U0079 [A. K. Singh, H. E. Egilmez, M. Coban, M. Karczewicz (Qualcomm)]

This contribution document presents a reformulation of the loss function used to train end-to-end DNN for YUV 4:2:0 intra-frame compression. It is asserted that by adjusting the weight between luma and chroma distortion (in mean-square error), a more balanced RD trade-off can be achieved.

Based on JVET-U0079 method 1, this contribution experimented with different weights for luma and chroma in loss function. In U0079, a ratio of Y:U:V weights of 4:1:1 was used, and this contribution changes the ratio to 2:1:1.

It was reported that, with this 2:1:1 ratio, (Y, U, V) performance over HM-16.20 becomes (-9.53%, 9.97%, -13.60%), compared to (-12.41%, 46.07%, 13.65%) as reported in JVET-U0079 method 1.

ParkRunning and Campfire chroma performance was also significantly improved, though chroma performance still shows a loss.

It was commented that non-equal weights between U and U components may help to reduce the imbalance as well.

Further study of YUV 4:2:0 coding in the context of E2E framework is encouraged.

[JVET-U0102](https://jvet-experts.org/doc_end_user/current_document.php?id=10617) AHG11: Variable rate end-to-end image compression [C. Lin, F. Chen, L. Wang (Hikvision)] [late]

In order to yield compressed images with different quality, most existing end-to-end (E2E) image compression models require training separate networks for different compression rates. This contribution proposes to use conditional convolution for image compression, which can adjust the compression rates by input parameter with only one model. We test this conditional convolution on existing E2E image compression model. Experimental results that the proposed model is comparable with the BPG 4:4:4 while it performs a little worse than mbt2018 model.

The reported performance is about 0.3 – 0.4 dB loss compared to the original mbt2018 method, which uses 7 models, whereas this contribution uses only 1 model.

In training stage, the loss function is modified to use randomized lambda values covering a range of bpp values.

In inference stage, the lambda value is used as an additional network input.

In the proposed network, conditional convolution replaces regular convolution in a few places of the mbt2018 network. It was noted that this increases the number of parameters by about 296K (or approximately 2% of the total number of parameters).

Training and test were performed on RGB images, which are different from the JVET NN CTC.

### Further study based on JVET NN CTC is encouraged. Super resolution and post filtering (3)

Contributions in this area were discussed in Session 17 at 1630-1745 UTC on Tuesday 12 January 2021 (chaired by Y. Ye).

[JVET-U0053](https://jvet-experts.org/doc_end_user/current_document.php?id=10565) AHG9/AHG11: Level information for super-resolution neural network [T. Chujoh, E. Sasaki, T. Suzuki, T. Ikai (Sharp)]

In this contribution, level information for a super-resolution neural network is proposed. In VVC, RPR (Reference Picture Re-sampling) has been introduced. As a result of several experiments to change the resolution of the whole sequence, in several 4K sequences, there are coding gains. Also, instead of the up-sampling filter, applying a super-resolution using neural network, the visual quality is more improved. At the previous meeting, JVT-T0092 proposed an SEI which sent the neural network parameters of super-resolution post-filter. Using MPEG NNR which is standardized in SC29/WG4 can describe neural network parameters efficiently. In this contribution, to realize arbitrary super-resolution processing, an SEI which defines level information for the complexity of neural network processing is proposed.

This was discussed earlier in session 10, see notes in 6.2.

Compared to coding in original resolution using VTM-11.0, the following table shows performance coding in half resolution in each dimension and performing upsampling using RPR filters.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Random access Main10 | | | | |
|  |  | Y | U | V | EncT | DecT |
| Cass A1 4K | Tango2 | -9.78% | -7.64% | -0.21% | 55% | 170% |
|  | FoodMarket4 | -10.77% | 2.04% | 0.55% |
|  | Campfire | -15.34% | 13.60% | 3.29% |
| Class A2 4K | CatRobot1 | -4.63% | 18.21% | 23.62% | 48% | 154% |
|  | DaylightRoad2 | -0.57% | -0.03% | 1.09% |
|  | ParkRunning3 | -11.17% | 114.24% | 51.05% |

Compared to coding in half resolution in each dimension using VTM-11.0 and performing upsampling using RPR filters, the following table shows performance also coding in half resolution using VTM-11.0 but upsampling using trained super-resolution NN.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Random access Main10 | | | | |
|  |  | Y | U | V | EncT | DecT |
| Cass A1 4K | Tango2 | -0.37% | -4.72% | -2.67% | 100% | 581% |
|  | FoodMarket4 | -0.98% | -0.62% | -2.36% |
|  | Campfire | -2.65% | -1.78% | -4.27% |
| Class A2 4K | CatRobot1 | -4.96% | -4.09% | -4.63% | 100% | 528% |
|  | DaylightRoad2 | -5.99% | -3.13% | -2.48% |
|  | ParkRunning3 | -2.44% | -4.07% | 4.11% |

NN parameters are fixed in this contribution, and signalling of NN parameters is not included in rate calculation.

Downsampling is performed using RPR filters in VVC, not trained NN filters, for both cases.

It was noted that QP used in this contribution were 32, 37, 42, and 47, which were lower than the JVET CTC QP.

It was commented that JVET-U0099 was closely related. See notes under JVET-U0099.

[JVET-U0091](https://jvet-experts.org/doc_end_user/current_document.php?id=10605) AHG9/AHG11: SEI message for carriage of neural network information for post filtering [B. Choi, Z. Li, W. Wang, W. Jiang, X. Xu, S. Wenger, S. Liu (Tencent)] [late]

This contribution proposes an SEI message design for carriage of a neural network topology and parameters that are utilized for post filtering with neural network models. The proposed straw-man design of high level syntax structure and elements includes both approaches; 1) internal carriage of description of topology information and compressed network parameters in SEI message, for simple convolutional neural network (CNN) use cases, and 2) external linkage information to provide where the network topology information and corresponding parameters are present in external files or remote locations, in case that the data size of neural network information is too huge to be contained in SEI messages. The goal of this proposal initiates the discussion on how to specify neural-network based post/in-loop filtering in VSEI, because the development of CNN-based post/in-loop filtering shows the good progress. The design of the proposed syntax structure is aimed to be specified in VSEI as codec-agnostic approach, but potentially the similar mechanism can be specified in parameter sets targeting VVC/HEVC-extensions.

This was discussed previously in session 10, see notes in 6.2.

During the discussion of session 10, it was asked for us to discuss the following:

- identifying which types of networks are required e.g. for post processing, superresolution, and what their benefit is in terms of compression performance

- identifying if it hypothetically requires normative specification or could be considered as non-normative post processing

- investigate which mechanisms for adaptation to specific content would be beneficial in terms of compression performance (e.g. enabling/disabling locally, or transmitting parameters)

It was commented that some sort of high-level signalling could be needed, although we should perhaps first develop a deeper understanding of NN architectures for super-resolution/upsampling and other post-processing purposes.

Should super-resolution/upsampling NN architectures, e.g. those used in JVET-U0053 and JVET-U0099, be investigated in EE? Proponents were encouraged to consider this for EE, and propose EE tests if there is interest in doing so.

It was commented that such NN can be used for different purposes, in addition to super-resolution/upsamling, but also post processing of other purposes (normative or non-normative). Further study is encouraged.

[JVET-U0099](https://jvet-experts.org/doc_end_user/current_document.php?id=10614) AHG11: Neural Network-based Super Resolution [A. M. Kotra, K. Reuzé, J. Chen, H. Wang, M. Karczewicz, J. Li (Qualcomm)]

This contribution 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 up-sampled version of the decoded picture. The upsampled picture is generated by the Neural Network-based up-sampling filter instead of the existing VTM up-sampling filter. On average, for Class A1 sequences, Luma BD-Rate gains of 5.74% and 8.79% for AI and RA configurations respectively were observed.

Number of parameters per model was ~1.3 million, and two models were used in this contribution for different bit rate ranges.

Training was performed in two ways: sequences coded in the original resolution, and sequences coded in the downsampled resolution.

NN training takes YUV 4:4:4 as input by repeating every other chroma sample in YUV 4:2:0 domain.

CTC QP 22 to 42 were used in testing. For all cases, VVC RPR filters were used in downsampling, and the reported decoding time does not include upsampling.

The following table shows the performance of half resolution coding compared to full-resolution coding with VTM-11.0, with upsampling performed using RPR filters.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Random access Main10 | | | | |
|  |  | Over VTM-11.0 (QP 22,27,32,37,42) | | | | |
|  |  | Y | U | V | EncT | DecT |
| Cass A1 4K | Tango2 | -4.41% | 0.19% | 10.52% | 50% | 33% |
|  | FoodMarket4 | -0.57% | 10.37% | 9.06% |
|  | Campfire | -2.56% | 95.98% | 35.38% |
| Class A2 4K | CatRobot1 | 24.07% | 44.74% | 68.09% | 40% | 30% |
|  | DaylightRoad2 | 42.46% | 18.66% | 41.19% |
|  | ParkRunning3 | -0.06% | 297.20% | 146.30% |

The following table shows coding performance of half-resolution coding compared to full-resolution coding with VTM-11.0, with upsampling performed using NN filters trained using method 1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Random access Main10 | | | | |
|  |  | Over VTM-11.0 (QP 22,27,32,37,42) | | | | |
|  |  | Y | U | V | EncT | DecT |
| Cass A1 4K | Tango2 | -7.48% | -11.02% | -1.39% | 92% | 32% |
|  | FoodMarket4 | -4.03% | 7.27% | 5.77% |
|  | Campfire | -12.68% | 79.71% | -3.04% |
| Class A2 4K | CatRobot1 | 11.86% | 21.67% | 32.83% | 99% | 37% |
|  | DaylightRoad2 | 25.36% | 5.73% | 21.00% |
|  | ParkRunning3 | -4.87% | 194.75% | 89.13% |

The following table shows coding performance of half-resolution coding compared to full-resolution coding with VTM-11.0, with upsampling performed using NN filters trained using method 2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Random access Main10 | | | | |
|  |  | Over VTM-11.0 (QP 22,27,32,37,42) | | | | |
|  |  | Y | U | V | EncT | DecT |
| Cass A1 4K | Tango2 | -7.71% | -9.67% | 4.70% | 90% | 31% |
|  | FoodMarket4 | -5.20% | 4.88% | 4.81% |
|  | Campfire | -13.44% | 85.12% | 6.07% |
| Class A2 4K | CatRobot1 | 7.85% | 23.33% | 40.51% | 74% | 31% |
|  | DaylightRoad2 | 21.33% | 5.31% | 23.25 |
|  | ParkRunning3 | -7.44% | 197.45% | 93.99% |

It was commented that training method 2 seems to be better than training method 1.

It was noted that the NN in JVET-U0053 had fewer parameters.

It was commented that visual quality of video coded at QP 47 (as done in JVET-U0053) should be checked.

It was commented that for some of the sequences where BD rate shows a gain, e.g. Campfire, crossing of the luma RD curves between the reference and the tested method can be observed, where the tested method is better at lower rates but worse at higher rates.

Further study of NN upsampling filters was encouraged, perhaps in the context of EE. See notes under JVET-U0091.

## Other compression technologies (3)

Contributions in this area were discussed in Session 15 at 0815 UTC on Tuesday 12 January 2021 (chaired by JRO and GJS).

[JVET-U0048](https://jvet-experts.org/doc_end_user/current_document.php?id=10560) Evaluation of Template Matching Prediction for VVC [K. Naser, F. Le Léannec, T. Poirier, F. Galpin (InterDigital)]

This document studies the coding efficiency of the classical coding tool of template matching. It is reported that this tool provides AI coding gain of 0.2% to 0.5% and up to 16% for TGM class.

It is asked if this gain is additional to IBC. It is confirmed that IBC is enabled in TGM, where significant additional bit rate reduction is observed. It is noted that encoder and decoder run times are significantly increased, and that much less complex versions of TM prediction were proposed earlier.

[JVET-U0093](https://jvet-experts.org/doc_end_user/current_document.php?id=10608) YCgCo-R: Additional Experiments on the new representation [D. Buitenhuis (VideoLAN), A. Tourapis (Apple Inc)] [late]

TBP

[JVET-U0100](https://jvet-experts.org/doc_end_user/current_document.php?id=10615) Compression efficiency methods beyond VVC [Y.-J. Chang, C.-C. Chen, J. Chen, J. Dong, H. E. Egilmez, N. Hu, H. Huang, M. Karczewicz, J. Li, B. Ray, K. Reuze, V. Seregin, N. Shlyakhov, L. Pham Van, H. Wang, Y. Zhang, Z. Zhang (Qualcomm)]

This contribution presents changes and tools in the areas of intra prediction, inter prediction, transform and coefficient coding, in-loop filtering, and entropy coding added on top of VVC standard. The implementation was done based on VTM-10.0 with added fixes related to MCTF and it demonstrates -11.50%, -12.91%, -13.19% BD rate for luma and chroma components, respectively with 243% encoder and 392% decoder run time in RA configuration using VTM CTC.

Intra prediction:

* Multi-model LM (JVET-D0110)
  + Neighboring samples are classified into 2 classes
  + Linear model is derived per class
  + Linear least square method is used
* Gradient PDPC (JVET-Q0391)
  + Gradient PDPC method of horizontal/vertical modes is extended to other modes
* Secondary MPM (JVET-D0114)
  + Primary MPM consists of 5 entries
  + Secondary MPM list of 16 entries is introduced
  + Shape dependent order of neighboring blocks to construct the list
* Reference sample interpolation (JVET-D0119)
  + 4-tap cubic interpolation is replaced by 6-tap filter
  + 6-tap Gaussian filter is used for larger blocks
  + 4-tap filtering is used instead of the nearest neighbor to derive extended reference
* Decoder side intra mode derivation (JVET-O0449)
  + Intra prediction is derived as a weighted average between Planar and two derived directions
  + The first DIMD mode is stored with a block and is included into the primary MPM list of the neighboring blocks

Inter prediction:

* Local illumination compensation (JVET-O0066)
* Non-adjacent spatial candidate (JVET-L0399)
* Template matching for MV refinement (JVET-J0021)
* Multi-pass MV refinement
* OBMC (JVET-L0101)
  + Top and left CU boundaries are refined using neighbor block motion
  + Top, left, bottom, and right subblock boundaries are refined using neighbor subblock motion
* Sample-based BDOF
  + Instead of the block basis, BDOF refinement is derived per sample
  + 5x5 window around each sample is used to derive Vx and Vy refinement
  + On 8x8 subblock basis, whether to apply BDOF or not is determined by checking the SAD between the two reference subblocks against a threshold
* Multi-hypothesis prediction (JVET-M0425)
  + Applied to bi-prediction in AMVP mode (for non equal BCW weights), affine AMVP, and regular merge mode
  + Up to 2 additional predictors are signalled or inherited for merge mode
* 12-tap interpolation for luma

Transform and coeff. coding:

* CTU and max transform size is extended to 256 (for class A)
* The number of LFNST sets is extended to 35 with 3 candidates
* Full 64x64 LFNST matrix is used (no zeroing out)
* 8-state DQ is utilized (JVET-Q0243)
* Sign prediction is applied for up to 6 transform coefficients (JVET-D0031, JVET-J0021)

Entropy coding:

* The precisions for two probability states are both increased to 15 bits, in comparison to 10 bits and 14 bits in VVC
* LPS range update process is modified
* CABAC context window is defined for each slice type

ALF:

* Luma and chroma filter shapes are extended to 9x9
* Classification is performed on 2x2 basis
* Luma filtering is applied using 1 signaled F2 9x9 filter and 2 fixed F0, F1 13x13 filters
* Weighting coefficients for F0 and F1 fixed filters are signalled
* Different 2x2 based classifiers are used for F0, F1, and F2 filters
  + Directionality is derived by comparing the ratio of horizontal, vertical, and diagonal gradients with a set of thresholds, which supports more edge strengths

Selection of tools was done based on complexity vs. performance assessment. The powerpoint deck includes a table of tool-on test for some of the individual tools.

Why was bilateral filter not included? A version was tested, but was not performing in combination with the modified ALF.

LDB results not available for class B, but should be in similar range of gain as for C/D

Some encoder optimization was done.

Max. BT/TT size is also 256x256.

More detailed analysis of all individual tool performance would be desirable, also in terms of memory bandwidth, etc.

It is suggested to establish an exploration experiment and an AHG on “XXXX”, as an organized activity to investigate also other technology than NN, but also being able to compare benefit of both NN and “conventional” tools, or their combination.

This exploration should not be too restricted by complexity considerations.

Some experts suggest to primarily run tool-on test against VVC in first place, and only based on such results define a kind of “default package”. Others are also preferring tool-off test.

BoG (V. Seregin) to define this EE – meets Wednesday 1520-1720

# High-level syntax (HLS) proposals (XX)

## AHG9: SEI message studies and proposals (6)

Contributions in this area were discussed in Session 6 at 0720-0930 UTC on Thursday 7 January 2021 and in Session 9 at 0500-0710 UTC on Friday 8 January 2021 (chaired by JRO and GJS).

[JVET-U0045](https://jvet-experts.org/doc_end_user/current_document.php?id=10557) AHG9: Picture output suppression SEI message [M. Pettersson, R. Sjöberg, M. Damghanian, J. Ström (Ericsson)]

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 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. Spatial and SNR scalability in this proposal are 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 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 is pointed out that typically in scalability applications it is decided at the decoder which layer to output.

It is pointed out that VVC already has the 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 is also pointed out that this might also be applicable for other use cases such as stereo/multiview.

No action at this point – further study.

[JVET-U0082](https://jvet-experts.org/doc_end_user/current_document.php?id=10596) AHG9: Scalability dimension SEI message and three HEVC SEI messages [Y.-K. Wang, L. Zhang, K. Zhang, Z. Deng, Y. Wang (Bytedance), A. Vetro (MERL), M. Mrak, S. Blasi (BBC)]

This contribution reproposes the scalability dimension SEI message proposed in JVET-T0070, for signalling of scalability dimension information for a VVC bitstream, including 1) whether the bitstream is a multiview bitstream, and if yes, the view ID of each layer, and 2) whether one or more layers in the bitstream represent auxiliary information such as alpha, depth, etc., and if yes, which layers represent what.

Furthermore, this contribution also proposes to include three HEVC SEI messages into VSEI/VVC: the multiview acquisition information SEI message, the depth representation information SEI message, and the alpha channel information SEI message.

For HEVC, equivalent information proposed for the scalability dimension SEI is carried in VPS. Such an information is definititely needed for making use of such kind of data.

It is pointed out that the SEI messages from HEVC are well established, but have been defined for representing camera-acquired depth data. Nowadays, other types of depth data are becoming more common (which might require other SEI messages). It should also be checked that the precision of depth and alpha data is flexible as coming from the bitstream.

Some editing was done on correcting syntax during the session – new version to be uploaded.

Decision: Adopt for VSEI extensions

[JVET-U0083](https://jvet-experts.org/doc_end_user/current_document.php?id=10597) Signalling of decoder initialization information [Y.-K. Wang, K. Zhang, L. Zhang, Y. Wang, J. Xu, Z. Deng (Bytedance)]

This contribution proposes to signal decoder initialization information (DII) in order to avoid or minimize the need of decoder reinitialization when decoding a video bitstream. The DII is proposed to be signalled either in the DCI NAL unit or in a new SEI message, named DII SEI message.

In version 1, DCI information was kept to a minimum (such as max profile/level/tier tha would occur in the bitstream). It is pointed out that including similar information (such as max pic width/height, etc.) was proposed before for DCI but not included. Alternatively, it is proposed defining a new SEI message for this.

It is argued that a decoder could initialize the DPB once for the highest resolution expected, even when the bitstream starts with lower resolution, such that re-initialization could be avoided.

It is asked whether a decoder would not anyway need to be able for quickly re-initializing DPB due to potential presence of RPR. One expert mentions that in a reasonable hardware implementation, the allocation of memory would be done sophisticated enough to be flexible and still re-initialize fast enough. The proposal of knowing the proposed information in advance would not help for re-configuration.

The proponent answers that many decoder implementations are not sophisticated and face problems with re-initialization.

One suggestion is that another option would be to define much DPB size rather than max width and max height. In a case where a bit stream would contain one part with very high and narrow and another with a very wide picture (both using the same max DPB size) this might lead to contradiction.

Revisit (after offline discussion with M. Zhou)

[JVET-U0084](https://jvet-experts.org/doc_end_user/current_document.php?id=10598) AHG9: Cross RAP referencing (CRR) SEI message [Y.-K. Wang, Y. Wang, L. Zhang, K. Zhang, Z. Deng (Bytedance)]

This contribution is a follow-up of JVET-T0071 and proposes a new SEI message, named cross RAP referencing (CRR) indication SEI message. A usage of CRR in adaptive streaming based on DASH is described. Simulation results comparing CRR with DRAP as well as with just using IRAP for providing the random access functionality are also presented.

The relationship with temporal sublayering had been previously discussed, and was brought up again.

A difference was said to be that this indicates a specific number of prior picture dependencies, which is not indicated by the use of temporal sublayering.

This further discussed in session 9 on Friday Jan. 8 at 0500 UTC. An illustration is provided in a slide deck uploaded with v2. It is assumed that in case of a random access, a client requests the initial IDR, and another preceding CRR (which is encoded by only using the initial IDR) in order to decode the current CRR. This way, the coding efficiency is increased compared to DRAP, where each DRAP is only allowed to use the initial IDR.

The following results were presented:

The test results of CRR compared to VTM-11.0 and DRAP are shown in the two tables below.

**Test results of CRR compared to VTM-11.0**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Random Access Main 10** | | |
|  | **Over VTM-11.0** | | |
|  | Y | U | V |
| Class A1 |  |  |  |
| Class A2 |  |  |  |
| Class B | -4.19% | -6.74% | -5.66% |
| Class C | -7.56% | -9.82% | -8.43% |
| Class E |  |  |  |
| **Overall** |  |  |  |
| Class D | -4.99% | -9.06% | -8.48% |
| Class F (optional) | -22.93% | -22.78% | -21.93% |

**Test results of CRR compared to DRAP**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Random Access Main 10** | | |
|  | **Over DRAP** | | |
|  | Y | U | V |
| Class A1 |  |  |  |
| Class A2 |  |  |  |
| Class B | -3.55% | -10.10% | -10.25% |
| Class C | -2.87% | -8.74% | -7.99% |
| Class E |  |  |  |
| **Overall** |  |  |  |
| Class D | -2.47% | -10.89% | -10.37% |
| Class F (optional) | -7.76% | -8.26% | -8.20% |

It is pointed out that this requires more additional information than in DRAP case, as the referred CRR is sent additionally. This additional rate is not considered, as it is assumed that this does not happen frequently.

It is also pointed out that this causes some additional delay and additional decoder processing resources.

Such “referred CRRs” (which are basically the same as DRAP) are included at every 4th CRR position in the reported results.

The syntax allows up to 8 “referred CRRs” to be used for decoding the current CRR. This may be too much for an application that is described here, but might be useful for flexibility. It is pointed out that as many CRRs might even exceed DPB resources. Furthermore, the additional processing would be excessive and might require an additional decoder. It is asked if this is compliant with HRD (DRAP has some reference to HRD).

Using temporal sublayers and using DRAP at sublayer 0 (e.g.) could achieve a similar functionality. However, CRR is more flexible than DRAP.

The referred CRRs are all carried in one external stream.

A better name would be “extended DRAP”.

Some support is expressed, and no objection is raised against this. It is however pointed out that the usage would require additional design for implementation at the systems side.

Decision: Adopt the proposal to VSEI extensions as “extended DRAP”.

Software implementation is needed.

[JVET-U0092](https://jvet-experts.org/doc_end_user/current_document.php?id=10606) Allocation of SEI message payload type for MPEG-I MIV/V3C carriage [J. Boyce (Intel)] [late]

It is proposed that a new v3c\_metadata( ) SEI message payload type be allocated in the VVC, HEVC, and AVC specifications, with the SEI message payload to be defined by in the ISO/IEC 23090-10 V3C version 2 specification, for carriage of V3C/MIV metadata within a VVC, HEVC, or AVC bitstream.

A similar approach was used for the definition of the HEVC/AVC green metadata SEI message, in which the message payload was defined outside of the HEVC and AVC specifications, with the payload type allocated within the HEVC and AVC specs.

At the time of preparation of the -v1 version of this document, the related MPEG contribution m55800 had not yet been reviewed by the MIV or V-PCC groups in MPEG Video and MPEG 3DG.

It is commented that this would introduce an alternative way of conveying V3C/MIV data, as another approach is currently developed at systems level.

It is commented that carriage in the video stream might have advantages when the metadata information is time dependent.

It is generally agreed that this is an interesting concept that should be further studied. Requires coordination with WGs 3&4&7 (arrange joint meeting next week).

[JVET-U0098](https://jvet-experts.org/doc_end_user/current_document.php?id=10613) AHG9: Composite Picture Information (CPI) SEI Message [Hendry, H. Jang, S. Kim, J. Lim (LGE)]

This contribution proposes a design for an SEI message to support generating composite pictures from pictures in a multi-layer VVC bitstream.

It is asserted that composite picture information (CPI) SEI message provides more advantages when used for implementing the given use-cases (i.e., broadcasting / streaming online game program and video conferencing) compared to using subpictures. The asserted advantages are:

– Better coding efficiency as the approach based on CPI SEI message does not require IRAP subpicture whenever there is change to the display layout.

– Better support for personalization of the layout for each bitstream receiver. Changing the display layout is as simple as replacing the SEI message without having to alter the coded pictures.

It is also asserted that the proposed feature has more benefits when specified as SEI message rather than as descriptor at system level for at least the following reasons:

– The feature would be available to any system that uses VVC, rather than only to system that specified it.

– It allows content creators to provide recommended display layout for their contents without having to worry about which system will be used to transport their contents to receivers.

It is commented that current video conferencing systems are not using centralized picture composition any more. However, for streaming of very large conferences there is a tendency to be used again.

It is further commented that compositing requires some significant additional functionality at the receiver side.

It is also commented that using of subpictures might have a similar functionality with less delay. The proponent highlights as an advantage compared to subpictures would be that a certain video could be moved to another part of the video more easily.

Another participant says that also subpictures could be re-arranged just at the display.

Several experts expressed concern if this SEI message would serve application needs. Some sophisticated processing is required for composition, which may go beyond simple placing of videos in the display side. Seek communication with systems experts to clarify how such mechanisms are invoked, and what is needed. Discuss in joint session (also for other SEI messages).

Revisit: Check software status of SEI messages (Karsten to check).

## HLS signalling for specific tools (3)

Contributions in this area were discussed in Session 10 at 0730-0920 UTC on Friday 8 January 2021 (chaired by JRO and GJS).

[JVET-U0053](https://jvet-experts.org/doc_end_user/current_document.php?id=10565) AHG9/AHG11: Level information for super-resolution neural network [T. Chujoh, E. Sasaki, T. Suzuki, T. Ikai (Sharp)]

In this contribution, level information for a super-resolution neural network is proposed. In VVC, RPR (Reference Picture Re-sampling) has been introduced. As a result of several experiments to change the resolution of the whole sequence, in several 4K sequences, there are coding gains. Also, instead of the up-sampling filter, applying a super-resolution using neural network, the visual quality is more improved. At the previous meeting, JVT-T0092 proposed an SEI which sent the neural network parameters of super-resolution post-filter. Using MPEG NNR which is standardized in SC29/WG4 can describe neural network parameters efficiently. In this contribution, to realize arbitrary super-resolution processing, an SEI which defines level information for the complexity of neural network processing is proposed.

The compressed network parameters require about 0.5 MB (uncompressed 2 MB).

The network parameters are not adaptive. Why is it then necessary to transmit parameters? It is basically post processing, which can be done non-normatively at receiver end and thus might not require standardization.

It is reported that the superresolution network gives better performance (BD rate after upsampling) than RPR filters. The rate for the network parameters is however not counted.

One expert doubts if the size of the compressed file would reflect the network complexity.

It is also pointed out that the topology of the network is not included in MPEG NNR representation.

No action necessary from aspect of HLS – to be further studied and be discussed in AHG11 context.

[JVET-U0078](https://jvet-experts.org/doc_end_user/current_document.php?id=10592) AHG9: Out-of-loop luma mapping with chroma scaling using APS or SEI message parameters signalling [E. François, P. de Lagrange, F. Le Léannec (InterDigital)]

This contribution proposes a variant of LMCS, where it can be applied out of the decoding loop, instead of being applied inside the decoding loop as done in VTM11. Prior to the encoding process, luma mapping and cross-component chroma scaling are applied to the input video. After the decoding process, the decoded video is post-processed by an inverse luma mapping and inverse cross-component chroma scaling. The luma mapping function is built using the same process as in VTM. The cross-component chroma scaling function is a piece-wise linear function built from the VTM chroma scaling LUT values. In a first option, the out-of-loop LMCS parameters are signaled in the APS, as in current VVC specification, and a new SPS flag is added to indicate if LMCS is in-loop or out-of-loop. BD-rate variations compared to the VTM, for the CTCs HDR PQ content, are reported in the table below, for AI, RA and LB\* configurations.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Config. | DE100 | PSNRL100 | wPSNRY | wPSNRU | wPSNRV | PSNRY | PSNRU | PSNRV | EncT | DecT |
| AI | -1.01% | -0.81% | -0.55% | -5.94% | -8.04% | -0.28% | -3.11% | -4.22% | 99% | 94% |
| RA | -2.06% | -1.62% | -1.12% | -5.89% | -8.98% | -0.86% | -2.08% | -3.69% | 96% | 84% |
| LB | -6.97% | -2.61% | -1.54% | -11.45% | -16.60% | -1.30% | -1.85% | -7.17% | 98% | 96% |

*\*LB configuration is not part of the HDR CTCs.*

In another option, it is proposed to signal the out-of-loop LMCS parameters in a new SEI message (“Colour Transform Information”), inspired from the Colour Remapping Information SEI message specified in AVC and HEVC.

Scaling factor for chroma is determined per sample, which is different from the in-loop LMCS. It is asked if the higher gain for chroma would still be appearing if it is done block-wise? Has not been investigated.

The results above are comparing against an anchor with in-loop LMCS enabled. The out-of-loop mapping is performed using a modified version of HEVC’s CRI SEI message, where a cross component mapping (like in chroma scaling) is added.

It is also said that the out-of-loop mapping is not always beneficial (there is loss for Cosmos).

It is argued that defining such an SEI might not enforce everybody to use it. In most application domains such as DVB, ATSC, …, remapping SEI messages are only optionally defined. Several experts expressed opinion that such an SEI message would only be useful if it would be somewhat mandatory (e.g. in a profile).

The chroma mapping is more complex due to sample-wise processing. On the other hand, in-loop LMCS requires one forward and one reverse mapping for luma. The proponent also mentions that perhaps both could be combined (no evidence that this gives benefit), which would definitely be more complex.

Further study on

- complexity aspects (it is mentioned that due to out-of loop processing very likely additional would be necessary. The existing LMCS logic is closely coupled with block-wise processing, whereas out of loop is a post processing).

- potential benefit of combining in-loop and out-loop.

When having this information, it could be decided if the additional coding gain is attractive enough.

To discuss with parent bodies: Possible normative status of SEI messages? Put stuff like this in VVC instead VSEI to make it normative.

(it is mentioned that another SEI that might possibly be desirable to be normative would be film grain)

[JVET-U0118](https://jvet-experts.org/doc_end_user/current_document.php?id=10645) Crosscheck of JVET-U0078 (AHG9: Out-of-loop luma mapping with chroma scaling using APS or SEI message parameters signalling) [Fangjun Pu (Dolby)] [late]

[JVET-U0091](https://jvet-experts.org/doc_end_user/current_document.php?id=10605) AHG9/AHG11: SEI message 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 SEI message design for carriage of a neural network topology and parameters that are utilized for post filtering with neural network models. The proposed straw-man design of high level syntax structure and elements includes both approaches; 1) internal carriage of description of topology information and compressed network parameters in SEI message, for simple convolutional neural network (CNN) use cases, and 2) external linkage information to provide where the network topology information and corresponding parameters are present in external files or remote locations, in case that the data size of neural network information is too huge to be contained in SEI messages. The goal of this proposal initiates the discussion on how to specify neural-network based post/in-loop filtering in VSEI, because the development of CNN-based post/in-loop filtering shows the good progress. The design of the proposed syntax structure is aimed to be specified in VSEI as codec-agnostic approach, but potentially the similar mechanism can be specified in parameter sets targeting VVC/HEVC-extensions.

The proposal also includes elements that describe the network topology.

The intent of the proposal is to describe the neural network, not the association with its operation, e.g. switching on/off at block level.

The next steps should be

- identifying which types of networks are required e.g. for post processing, superresolution, and what their benefit is in terms of compression performance

- identifying if it hypothetically requires normative specification or could be considered as non-normative post processing

- investigate which mechanisms for adaptation to specific content would be beneficial in terms of compression performance (e.g. enabling/disabling locally, or transmitting parameters)

JVET should not define a very general concept of NN representation, rather use existing approaches (if needed)

Further review in AHG11 context

# Plenary meetings, joint meetings, BoG reports, and summary of actions taken

## Plenaries

Monday 0820-0920:

Liaisons: ITU-R WP 6B (VVC profiles) -> Gary, JPEG (NNVC)

White paper on VVC/VSEI (abstract by Friday?) -> Ben, Ye-kui; Gary attends meeting

Joint meetings

Session planning next days

NN EE viewing session being prepared, most likely Wednesday/Thursday; send announcement to reflector for volunteers participating

BoG reports and further tasks

Documents in 4.1 were reviewed

## Information sharing meetings

Beyond the joint meetings listed below, information sharing sessions with other WGs of the MPEG community were held on Monday 11 January 0500–0720, Wednesday 13 January 0500–0610, and Friday 15 January 2100–2230.

## Joint meeting with WGs 3 (Sys), 4(Vid), 7(3DG) & VCEG: SEI for MIV/V3C, Monday 11 January 1520–1550

The following topics were discussed in this joint session.

* JVET-U0092 / m55879 AHG9: Allocation of SEI message payload type for MPEG-I MIV/V3C carriage [J. Boyce (Intel)] Proposing one payload type in AVC, HEVC, VVC, to carry MIV data, referencing (possibly via VSEI) an MPEG standard that would detail the syntax and semantics, similar to Green Metadata handling.
* m55799 Carriage of V3C/MIV data in VVC/HEVC/AVC SEI message (related WG3 input, not discussed in the meeting).
* m56061is another related WG3 input, not discussed in this meeting.
* SEI versus file format carriage was discussed: It was mentionaed that this would be mainly be useful for the single-track case, for multi-track it would be painful digging too deep into the video stream from systems perspective. Multi-track might perhaps be possible with mult-layer streams.
* This would be one way to send the data - not the only way - with a focus on tunneling through previously designed systems - e.g. VLCplayer
* Some concern was expressed about the amount of data. There could be some restrictions imposed. In extreme case, this might require splitting the info carried into multiple parts. The proponent said this was not meant to cover all cases, but to be a restricted case.
* Some concern was expressed regarding fragmentation of approaches.
* There had been prior study of a similar approach but using a multi-layer bitstream; see m49229 and output TuC WG 11 N 18656.
* A registered user data SEI message was mentioned as a possibility.
* If some MPEG WG wants to draft a spec for the SEI message syntax and semantics detail, JVET would allocate a payload type to reference it (as was done with MPEG “green metadata” (although likely via VSEI), but JVET would not write the details of the payload. With such an approach, Systems said they would not provide special functionality for this approach - the data would just be inside the video track as SEI messages in an opaque manner.

## Joint meeting with WG 3 (Sys) & VCEG: SEI for picture composition and decoder initialization, Monday 11 January 1550–1620

The following topics were discussed in this joint session.

* JVET-U0098 AHG9: Composite Picture Information (CPI) SEI Message [Hendry, H. Jang, S. Kim, J. Lim (LGE)] This contribution proposes a design for an SEI message to support generating composite pictures from pictures in a multi-layer VVC bitstream.  
    
  In VCEG discussion, it was noted that such functionality might often be put into an app in some non-standard way. Consulting with ITU-T SG 16 Systems experts (e.g. at the April meeting) could be desirable.  
    
  See also JVET notes.  
    
  The possibility of rearranging subpictures of a decoded picture was also discussed.  
    
  Systems experts said such an SEI message, if present, would not be supported with special functionality at the system level; rather, it would just be carried opaquely within the video bitstream. If a decoder could handle the data, that would be OK from the System perspective.  
    
  Another participant said it would not ordinarily be expected for such a functionality to be supported through video layer signalling.  
    
  From a Systems perspective, the recommended way to handle such functionality was said to be at the system level rather than by sending it through the video bitstream.  
    
  The proponent said that an advantage of the SEI proposed solution is to be able to carry the data regardless of what system environment is being used. Systems experts reiterated that the necessary system infrastructure would need to be established, which is not accomplished by video bitstream SEI signalling alone.  
    
  No action by JVET was recommended on the proposed approach at this time.
* JVET-U0083 Signalling of decoder initialization information [Y.-K. Wang, K. Zhang, L. Zhang, Y. Wang, J. Xu, Z. Deng (Bytedance)] (Aspect 2.2 DII signalling in SEI relates to AHG9): This contribution proposes to signal decoder initialization information (DII) in order to avoid or minimize the need of decoder reinitialization when decoding a video bitstream. The DII is proposed to be signalled either in the DCI NAL unit or in a new SEI message, named DII SEI message.  
  m56019 and m56038 to MPEG Systems also proposed some decoder initialization information. Another related contribution was said to be m56084.  
  There will be further discussion in MPEG Systems on these contributions.  
  No action by JVET was recommended on the proposed approach at this time, pending study in MPEG Systems.

## Joint meeting with WG 2 (Req) & VCEG: VVC profiles, extensions (e.g., norm post-proc), SEI/VSEI, Monday 11 January 1630–1740

The following topics were discussed in this joint session.

* JVET-U0089 8-bit profiles for VVC [Y. Ye, G. Wu, L. Wang, J. Chen (Alibaba), L. Zhang, Y.-K. Wang, K. Zhang (Bytedance), M. Karczewicz (Qualcomm), Y.-W. Huang, S.-M. Lei (MediaTek), X. Wang (Kwai), D. Wang (OPPO), W. Ding (Baidu), Y.-P. Hsiao (Vivo), P. Wu (ZTE), M.-L. Champel (Xiaomi), +More (Twitch, V-Nova)].  
    
  Focused primarily on 4:2:0 (Main & Still). Two options are in the contribution: 1) profile\_idc, and 2) general constraint indication. Proponents prefer option 1 for ease of detection. Another participant commented that option 2 has a desirable advantage of compatibility with existing profiles. It was noted that we are at an early stage of deployment and profiles detection is probably handled in software anyway.  
    
  A proponent commented that a substantial number of applications exclusively use 8 bit content, and that 8-bit profiles of other standards exist and are widely used (particularly AVC).  
    
  For software there would likely be a greater benefit than for hardware, although hardware could also have some power savings.  
    
  Another participant said it seemed unlikely that hardware would be built with only 8 bit capability.  
    
  It was noted that the JVET CTC does not currently test 8 bit content, and the proponents said that coding the 10 bit content as 8 bit content shows similar gains as for 10 bit content. It was commented that Class A UHD SDR shows about 8% compression loss when measured relative to 10 bit original content (not relative to 8 bit input content).  
    
  Coding 8 bit content as 10 bit content with extra LSBs was noted as possible - a concept known as internal bit-depth increase (IBDI). The proponent said that they measured about a 4% coding efficiency difference for this approach.  
    
  The possibility of opportunistically use 8-bit optimization in a decoder for video coded with bit depth 8 was mentioned, especially in software. The proponent said this would impact decoder memory footprint, code size and testing.  
    
  Having an 8 bit profile as a “foot in the door” for transition from AVC to VVC was suggested as a reason to establish such profiles.  
    
  User-generated content was suggested to often be 8 bit only.  
    
  Based on the JVET-U0088 decoder, a memory footprint impact of perhaps 50% increase or more was estimated by a proponent.  
    
  A proponent emphasized the desire to enable use on low-end mobile devices.  
    
  Fragmentation of deployed support was suggested to be an issue that could be created by establishing such profiles, as there could become an installed base that is not capable of 10 bit operation.  
    
  It was noted that an external profile specification could result if such a profile is not specified in the standard.  
    
  It was noted that this is a request by a significant number of companies desiring a conformance point for deployment, including hardware companies, and suggested that such a request from our “customers” should carry weight.  
    
  No consensus for action at this meeting was evident; further study was suggested to clarify and quantify the issues.
* Extensions - e.g., normative post-proc (e.g., U0078), JVET-U0100 - investigation can proceed in JVET
* Other SEI - no particular concerns

## BoGs (2)

The following break-out groups were established at this meeting and produced the below-listed reports.

[JVET-U0133](https://jvet-experts.org/doc_end_user/current_document.php?id=10660) BoG report on CE complexity analysis [A. Browne]

BoG initially reported Tue 0840

Complexity analysis of CE and CE related was conducted (not finished yet)

New results tables with separate results for 12 and 16 bit for SVT have been provided, to be uploaded as new version of JVET-U0022.

Gains are in the range of 2.5..4.3% for 12 bit in RRC (CE-1.x and CE-3.x), around 1.5% for TSRC (CE-2.x) in 12 bit, around 20-28% for 16 bit.

The BoG met again and reported back Wed 13 Jan 0720.

This is a report of activities from the BoG on CE complexity analysis. The BoG held meetings at the following times during the 21st JVET meeting:

• January 8th 13:00 – 15:00

• January 8th 15:20 – 16:40

• January 11th 13:00 – 13:55

(include information from report, detailed analysis of the complexity impact of all proposals)

It was suggested to continue the CE for further studying

* Performance at higher QPs (relevant for consumer applications)
* Interaction between entropy coding modifications and JVET-U0052
* Investigate benefit over HEVC RExt also in low QP range, for 12 and 16 bit

It would be beneficial using more diversified test material. This could also be HD, which would also simplify the amount of simulations.

It should be identified if a simple modification of VTM can be agreed as anchor (at least for the 16 bit case, where something is obviously broken).

BoG (A. Browne) to propose the plan for the next round of CE, identify which of the current proposals should be further investigated, discuss modifications of test conditions (sequences, QP ranges, anchor for 12/16 bit, additional HEVC RExt anchor).

* Meets Wed. 1300-1500 UTC

[JVET-U0139](https://jvet-experts.org/doc_end_user/current_document.php?id=10666) BoG report on high bit rate / high bit depth coding [A. Browne (Sony)]

BoG on CTC/report doc and EE for neural network (A. Segall) – meets Thu. 0500-0700 UTC

CTC update: VTM version, GOP size, MCTF usage, MS-SSIM version & reporting status

Viewing is planned Thursday (see JVET-U0137)

It is noted that at this stage of exploration, it is desirable to collect as much information about technology as possible, and that diversity of methods should be investigated in the EE.

For tools that are primarily targeting utilizing intra-picture compression, it is agreeable investigating only AI conditions.

# Project planning

## Core experiment and exploration experiment planning (update)

A CE on entropy coding for high bit depths and high bit rates was established, as recorded in output document JVET-T2022.

An EE on neural network-based video coding was established, as recorded in output document JVET-T2023.

## 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 Wednesday 30 Dec. 2020.

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-N1010.

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.

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)).

Initial review of AHG plans was conducted in session 21 at 2145 on Thursday 15 October 2020. Further review was conducted in session 22 at 0612 on Friday 16 October 2020.

|  |  |  |
| --- | --- | --- |
| **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-T1003, JVET-T1005, JVET-T1006, JVET-T1008, JVET-T2001, JVET-T2016 JVET-T2017 and JVET-T2019). * 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-T1004 errata output collection. * Produce and finalize JVET-T2002 VVC Test Model 11 (VTM 11) Algorithm and Encoder Description. * Propose improvements to the JCTVC-AN1002 HEVC Test Model 16 (HM 16) Update 14 of Encoder Description * 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, V. Drugeon, S. Kim, S. Liu, J.‑R. Ohm, K. Sharman, 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 model (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. | 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 draft verification test plan JVET-T2009 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. * 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. * 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, particularly including Rep. ITU-R BT.2245. * Prepare availability of viewing equipment and facilities arrangements for future meetings. | V. Baroncini, T. Suzuki, M. Wien (co-chairs), E. François, 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))   * Produce the JVET-T2008 draft conformance testing specification and develop proposed improvements. * Study the requirements of VVC, HEVC, and AVC conformance testing to ensure interoperability. * Maintain and update the conformance bitstream database. * Study additional testing methodologies to fulfil the needs for VVC conformance testing. | J. Boyce and W. Wan (co-chairs), E. Alshina, F. Bossen, I. Moccagatta, K. Kawamura, 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 the 360Lib-12 software version and common test condition configuration files according to JVET-L1012. * Generate CTC anchors and PERP results for the VTM according to JVET-L1012. * Coordinate with AHG4 in preparation for verification testing for 360° video content. * Produce documentation of 360° software usage for distribution with the software. | J. Boyce and Y. He (co-chairs), K. Choi, J.-L. Lin, 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-T2011. * Study the luma/chroma bit allocation in the HDR CTC, especially for HLG content. * Coordinate implementation of HDR anchor aspects in the test model software with AHG3. * 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-T2018 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-T2022. * Identify suitable test material for testing of high bit depth, high bit rate, and high frame rate coding in coordination with AHG 4. | A. Browne and T. Ikai (co-chairs), M. Sarwer, X. Xiu (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 SEI showcase and usage information for SEI messages, including encoder and decoder implementations and bitstreams for demonstration and testing. * Identify potential needs for additional 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 GOP structures, GDR, LMCS and perceptually optimized adaptive quantization for encoder optimization. * Study encoding techniques of optimization for objective quality metrics and their relationship to subjective quality. * Particularly consider neural network-based encoding optimization technologies. * Study the impact of adaptive quantization. * 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. | A. Duenas, A. Tourapis (co-chairs), A. Norkin, R. Sjöberg (vice-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-T2023. * Solicit input contributions on NN based video coding technologies. * Continue to refine the test conditions for neural network-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 on the performance of candidate technology. * Analyse complexity characteristics, perform complexity analysis, and develop complexity reductions of candidate technology. * Identify video test materials, training set materials, and testing methods for assessment of the effectiveness and complexity of considered technology. * Generate and distribute anchor encodings and develop improvements of the JVET-T2006 common test conditions for NNVC technology. * Particularly consider the suitability of sequences from the YouTube UGC data set for future inclusion in the test set. * Coordinate with other relevant groups, including SC29/AG5 on visual quality assessment. | S. Liu, A. Segall, Y. Ye (co‑chairs), E. Alshina, J. Chen, F. Galpin, J. Pfaff, S. S. Wang, M. Wien, P. Wu, J. Xu (vice‑chairs) | Tel.  2 weeks notice |

+ new AHG: Enhanced compression tools

# 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 MPEG output document, a separate version under the MPEG document header should be generated. This version should be sent to GJS and JRO for upload.

[JVET-U1000](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10534) Meeting Report of the 20th JVET Meeting [G. J. Sullivan, J.-R. Ohm] [WG 5 N 4] (2020-11-13)

Initial versions of the meeting notes (d0 … dB) were made available on a daily basis during the meeting.

Remains valid – not updated: [JCTVC-H1001](http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=5095) HEVC software guidelines [K. Sühring, D. Flynn, F. Bossen (software coordinators)]

Remains valid – not updated: [JCTVC-AN1002](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 14 [C. Rosewarne (primary editor), K. Sharman, R. Sjöberg, G. J. Sullivan (co-editors)] [WG 11 N 19473]

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)] (2020-10-30)

WG 5 DoCR N 11.

Ballot comments 54912

* For code value 22 in Table 3, remove the informative reference
* Remove “(historical)” from the reference to SMPTE ST 240
* Reference the 2019 edition of SMPTE 428-1 (assuming that its content is the same in regard to the referenced aspects) and the 2018 edition of ARIB STD-B67 to the latest editions (BT.470-7 and STD-B67. Update other references if identified.
* Include the year in SMPTE references.
* Remove colour from figures 10 and 11.
* Add history commentary about the chroma 4:2:0 grid alignment type.

[JVET-U1004](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10536) Errata report items for HEVC, AVC, Video CICP, and CP usage TR [C. Rosewarne, G. J. Sullivan, Y. Syed, Y.-K. Wang] (2020-12-15, near next meeting)

(C. Rosewarne added.)

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] (2020-10-30)

WG 5 DoCR N 7.

Remove - finished

[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 WD N 16] (2020-12-15)

WG 5 Request N 15. Is a new version needed?

Include errata. The specification will reference VSEI for annotated regions. Target CDAM in April.

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 (editors)] [WG 11 [N 15778](http://phenix.it-sudparis.eu/mpeg/doc_end_user/current_document.php?id=53941&id_meeting=165)]

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 (editors)] [WG 5 TR N 14] (2020-11-13)

WG 5 DoCR N 13.

Remove - finished

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 (editors)]

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 (editors)]

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 (editors)] [WG 11 [N 16049](http://phenix.it-sudparis.eu/mpeg/doc_end_user/current_document.php?id=54889&id_meeting=166)]

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 (editors)]

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 (editors)]

[JVET-U1100](http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=10693) Common Test Conditions for HM Video Coding Experiments [K. Sharman, K. Sühring (editors)]

New version needed. Align intra period, check if MCTF is enabled in config file.

Remains valid – not updated: [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-U2002](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10541) Algorithm description for Versatile Video Coding and Test Model 12 (VTM 12) [J. Chen, Y. Ye, S. Kim] [WG 5 N 23] (2020-12-15, near next meeting)

To include GOP structures (GOP 32 & LD) and MCTF description.

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] (2019-04-01)

[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] (2020-12-15, near next meeting)

Is a new version needed?

This was agreed to include cross-boundary blending for GCMP.

It was noted that this includes some “stale” formats that are no longer subjects of active investigation.

Remains valid for VTM – not updated: [JVET-S2005](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=9678) Methodology and reporting template for coding tool testing [W.-J. Chien and J. Boyce]

Would this be replaced by something for the new EE? Could be combined with CTC and be moved to 10..20 range. -> 2017

[JVET-U2006](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10532) Common Test Conditions and evaluation procedures for neural network-based video coding technology [S. Liu, A. Segall, E. Alshina, R.-L. Liao] (2020-10-30)

Move this to 10..20 range -> 2016

Remains valid – not updated: [JVET-S2007](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=9679) Versatile supplemental enhancement information messages for coded video bitstreams (Draft 5) [J. Boyce, V. Drugeon, G. J. Sullivan, Y.-K. Wang] [WG 11 N 19472]

[JVET-U2008](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10543) Conformance testing for versatile video coding (Draft 5) [J. Boyce, E. Alshina, F. Bossen, K. Kawamura, I. Moccagatta, W. Wan] [WG 5 DIS N XX] (2020-10-30)

[JVET-U2009](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10544) VVC verification test plan (Draft 5) [M. Wien, V. Baroncini, A. Segall, Y. Ye] [WG 5 N XX] (2020-12-15)

The plan is to finalize the VT plan for HD, HDR and 360 in January and conduct the testing and report the results by April. -> move to 2021

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] (2020-10-30)

Updated to reflect the GOP structure change (enable MCTF for RA except category F).

Remains valid – not updated: [JVET-T2011](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] (2020-10-30)

Updated to reflect the GOP structure change (MCTF enabled – may not need to mention that in the document).

Remains valid – not updated: [JVET-L1012](http://phenix.int-evry.fr/jvet/doc_end_user/current_document.php?id=4840) JVET common test conditions and evaluation procedures for 360° video [P. Hanhart, J. Boyce, K. Choi, J.-L. Lin]

No change was needed for the GOP structure issue, and no MCTF is used here.

Is a new version needed?

[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] (2020-10-30)

Updated to reflect the GOP structure change. No change was made for MCTF for these CTCs, although MCTF should operate properly for non-4:2:0 cases – see notes for JVET-T0131.

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]

No change was needed for the GOP structure issue. and no change was made for MCTF for these CTCs,– see notes for JVET-T0131.

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]

Remains valid – not updated: [JVET-T2016](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10547) Summary information on BD-rate experiment evaluation practices [K. Andersson, F. Bossen, J.-R. Ohm, A. Segall, R. Sjöberg, J. Ström, G. J. Sullivan, A. Tourapis] [WG 5 TR N 6] (2020-11-06)

WG 5 DoCR N 5.

Remove – finished?

[JVET-U2017](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10548) Additional SEI messages for VSEI (Draft 1) [J. Boyce, Y.-K. Wang] [WG 5 WD N 18] (2020-11-13)

This has the annotated regions SEI message. It was noted that if we add shutter interval to VSEI, we could reference that in the AVC amendment planned as per above (see JVET-T1006) – no immediate action was taken on that.

WG 5 Request N 17 (target April CDAM).Move this to 1..10 range -> 2006

[JVET-U2018](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10530) Common test conditions for high bit depth and high bit rate video coding [A. Browne, T. Ikai, D. Rusanovskyy, X. Xiu] (2020-10-30)

[JVET-T2019](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10549) New level and additional SEI messages for VVC (Draft 1) [F. Bossen, Y.-K. Wang] [WG 5 WD N 20] (2020-11-13)

WG 5 Request N 19 “Operation range extensions” (target April CDAM).

New level, SEI manifest, SEI prefix, interface to annotated regions.

Is a new version needed? Move this to 1..10 range -> 2005

Remains valid – not updated: [JVET-T2020](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10550) VVC verification test report for UHD SDR video content [V. Baroncini, M. Wien] [WG 5 N 21] (2020-11-13)

[JVET-U2021](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10551) Reference software for versatile video coding (Draft 1) [F. Bossen, K. Suehring, X. Li] [WG 5 DIS N XX] (2020-10-30)

-> move to 2009

[JVET-U2022](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10531) Core Experiment on Entropy Coding for High Bit Depth and High Bit Rate Coding [A. Browne, T. Hashimoto, H.-J. Jhu, D. Rusanovskyy] [WG 5 N XX] (2020-10-30)

If feasible, combinations can be tested, and different encoding algorithms for the same syntax can also be tested. VTM11 (expected 13 November) + 2 wks for T2.

[JVET-U2023](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10529) Exploration Experiment on Neural Network-based Video Coding [E. Alshina, S. Liu, W. Chen, Y. Li, R.-L. Liao, Z. Ma, H. Wang] [WG 5 N XX] (2020-10-30)

[JVET-U2023](http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10529) Exploration Experiment on Enhanced Compression Tools in Video Coding [XXXX] [WG 5 N XX] (2020-10-30)

New output doc: white paper (B. Bross, Y.-K. Wang) (2019 or 2005)

# 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.

Some specific future meeting plans (to be confirmed) were established as follows:

* Wed. 21 – Wed. 28 April 2021, 22nd meeting under ITU-T SG 16 auspices in Geneva, CH.
* Fri. 9 – Fri. 16 July 2021, 23rd meeting under ISO/IEC SC 29 auspices in Prague, CZ.
* Fri. 8 – Fri. 15 October 2021, 24th meeting under ISO/IEC SC 29 auspices in Antalya, TR.
* ?? January 2022 25th meeting under ITU-T SG 16 auspices in Geneva, CH??

The agreed document deadline for the 21st JVET meeting was planned to be Wednesday 30 Dec. 2020.

Vittorio Baroncini and Mathias Wien were thanked for preparing and conducting the VVC verification test for UHD SDR content. Alibaba, ByteDance, GBTech, HHI, Huawei, Mediatek, RWTH Aachen University, Sharp Labs of America, and Tencent were thanked for their great support in this effort.

Alibaba, Ericsson, and Sony were thanked for providing test sequences for usage in video standardization activities.

The 20th JVET meeting was closed at approximately 2350 hours UTC on Friday 16 October 2020.

# 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 XXX people in total), were as follows:

# Annex C to JVET report: Recommendations of the 2nd meeting of ISO/IEC JTC 1/SC 29/WG 5 MPEG Joint Video Coding Team(s) with ITU-T SG 16

**ISO/IEC JTC 1 / SC 29 / WG 5 N XX**