|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  15th Meeting: Geneva, CH 23 October – 1 Nov. 2013 | Document: JCTVC-O0198-v2 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Additional experiments and software for frame packing arrangement SEI message for 4:4:4 content in 4:2:0 bitstreams** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Srinath Reddy  Sandeep Kanumuri  Yongjun Wu  Shyam Sadhwani  Gary J. Sullivan  Henrique S. Malvar  One Microsoft Way Redmond WA 98052-6399  USA | Tel: Email: | |  | | --- | | +1-425-703-5308 [srinath.reddy@microsoft.com](mailto:srinath.reddy@microsoft.com)  [skanumu@microsoft.com](mailto:skanumu@microsoft.com)  [yongjunw@microsoft.com](mailto:yongjunw@microsoft.com)  [shyams@microsoft.com](mailto:shyams@microsoft.com)  [garysull@microsoft.com](mailto:garysull@microsoft.com)  [malvar@microsoft.com](mailto:malvar@microsoft.com) | |
| *Source:* | Microsoft Corporation | | |

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

# Abstract

This contribution proposes the use of a frame packing arrangement SEI message to represent 4:4:4 content in nominally 4:2:0 bitstreams. This contribution is an update of the prior contributions JCTVC-K0240, JCTVC-L0316, JCTVC-M0281 and JCTVC-N0270 that provides new experimental results for anticipated use cases where both 4:2:0 and 4:4:4 representations undergo a similar level of compression. Further work is reported with a focus on the investigation of the newer "band separation" variation first proposed in JCTVC-L0316. Substantially improved software for testing the scheme is provided. It is reported that the additional results indicate a substantial coding-efficiency benefit for the band-separation frame-packing modes over the “direct” frame-packing mode – averaging 45% benefit as tested in terms of chroma BD bit rate. It is asserted that the “direct” frame-packing modes of this contribution were adopted in spirit at the last meeting (July 2013) and the planned action was deferred to this meeting. It is suggested that since the new experiments indicate a significant benefit for the band separation mode, this mode should also be included.

For any of the frame-packing modes using the proposed method, it is reported that one constituent frame (e.g. in a top-bottom packing or alternating-frame coding scheme) can be decoded compatibly as an ordinary 4:2:0 image, or can be supplemented with the data from another constituent frame to form a complete 4:4:4 image representation. It is proposed to include support for the additional scheme into the frame packing arrangement SEI message in both AVC and HEVC, to facilitate deployment of systems using this method. Since 4:2:0 is the most widely supported format in products, it is asserted that having an effective way of conveying 4:4:4 content through such decoders can provide the substantial benefit of enabling widespread near-term deployment of 4:4:4 capabilities (especially for screen content coding). The proposed method operates by packing the samples of a 4:4:4 frame into two 4:2:0 frames (main and auxiliary) and encoding the two 4:2:0 frames as the constituent frames of a frame packing arrangement. The semantics of 'content\_interpretation\_type' are extended to signal this packing arrangement. The proposed scheme is asserted to be of high practical value for applications involving screen content. Relative to native 4:4:4 encoding, the proposed scheme can provide the advantage of compatibility with the ordinary 4:2:0 decoding process that is expected to be more widely supported in decoding products. It is reported that the attached software is capable of handling the frame-packing and frame-unpacking processes and can be used in conjunction with any 4:2:0 codec.

# Introduction

Most video codecs that are commercially available today support only the 4:2:0 format, which sub-samples the chroma resolution, as opposed to using a 4:4:4 format, in which the chroma information is represented at the same resolution used for the luma. The YUV 4:2:0 format is considered sufficient and efficient for "mainstream" content (i.e. most natural and animated video content) since users do not ordinarily see a perceptible difference between the two formats for such content. However, there are a variety of existing and emerging applications (e.g., as discussed in [1] and [2]) that operate with screen content, and for such content, the difference between these two formats can be easily perceived (see Appendix I of [4]). In such applications, a 4:4:4 format is strongly preferred over the YUV 4:2:0 format. However, the lack of wide-spread support for video codecs supporting 4:4:4 formats is a hindrance for these applications. Moreover, for certain critical uses such as scrolling titles and hard-edged graphics, there may sometimes be a significant benefit for 4:4:4 use in other scenarios.

In this contribution, we propose an approach to use codecs designed for YUV 4:2:0 content to compress and represent 4:4:4 content through the use of frame packing. The contribution is an update of the prior contributions JCTVC-K0240 [2], JCTVC-L0316 [3], JCTVC-M0281 [4] and JCTVC-N0270 [5] that provides new experimental results covering a wide-range of useful operational points wherein both main and auxiliary frames undergo a similar level of compression thereby obtaining a similar level of compression for both 4:2:0 (main frame) and 4:4:4 (main frame + auxiliary frame) representations. We believe that the typical use-case scenario for frame-packing 4:4:4 content in 4:2:0 format involves the compression of main frame at a level (measured in terms of noise introduced by compression) similar or lower than the level at which the auxiliary frame is compressed and our experimental results are representative of the typical use-case scenario. This contribution also provides specification text along with a figure illustrating the frame packing of 4:4:4 content into 4:2:0 format. In addition, this contribution provides an updated software capable of handling the frame-packing and frame-unpacking processes.

This proposal is similar to the frame packing of stereo (3D) content into 2D images, and builds on the framework established for that by extending the semantics of the frame packing arrangement SEI message. Unlike the frame packing of stereo content wherein there is a left and right view, the frame packing of the 4:4:4 content is done using a main view and an auxiliary view. Both the main and auxiliary views are in an equivalent of a 4:2:0 format. The main view may be independently useful, while the auxiliary view is useful when interpreted appropriately together with the main view.

# Proposed modifications

## Packing a YUV 4:4:4 frame into main and auxiliary views

A YUV (YCbCr, YCoCg, GBR, etc.) 4:4:4 frame can be represented as follows, wherein Y444, U444, and V444 are the Y, U, and V planes comprising the YUV 4:4:4 frame. Let the resolution of these planes be represented by width and height .

Y444

U444

V444

The YUV 4:4:4 frame represented above can be packed into two YUV 4:2:0 frames (as main and auxiliary frames) as follows:

Y420 U420 V420

Main view

B1

B2

B3

(YUV 4:2:0 frame)

B6

B7

B8

B9

B4

B5

Auxiliary view

(YUV 4:2:0 frame)

**Figure 1:** Illustration of main and auxiliary frame regions formed by proposed frame packing scheme.

The areas marked as B1 to B9 make up the Y, U and V planes of the two YUV 4:2:0 frames representing the main and auxiliary views. These areas are related to Y444, U444, and V444 as follows:

**Main view**

Area B1, , where the range of is .

Area B2, , where the range of is .

Area B3, , where the range of is .

**Auxiliary view**

Area B4, , where the range of is .

Area B5, , where the range of is .

Area B6, , where the range of is .

Area B7, , where the range of is .

Area B8, , where the range of is .

Area B9, , where the range of is .

In the above equations, and are either the same as or represent filtered versions of and respectively, where the range of is . This choice is explained in more detail in section 2.4.

### Advantages

The proposed packing method is designed such that:

* The main view is a YUV 4:2:0 equivalent of the original YUV 4:4:4 frame
  + Systems can optionally just display just the main view if YUV 4:2:0 output is needed
* The auxiliary view fits the content model of a YUV 4:2:0 frame and is well suited for compression in this manner, in terms of
  + Geometric consistency across its Y, U and V components
  + Motion is highly correlated across its Y, U and V components

The packing method is illustrated by the following example wherein a YUV 4:4:4 frame contains a circle represented using gray color (checkerboard pattern) for the Y plane, blue color (horizontal lines) for the U plane and red color (vertical lines) for the V plane and how the resultant main and auxiliary views are formed in YUV 4:2:0 format.

Y444 U444 V444

YUV 4:4:4 frame

Y420 U420 V420

Main view

(YUV 4:2:0 frame)

Y420 U420 V420

Auxiliary view

(YUV 4:2:0 frame)

**Figure 2:** Illustration Spatial correspondence relationships in proposed frame packing scheme.

## Extension to frame packing arrangement SEI message

The proposed method to signal the frame packing of YUV 4:4:4 content involves extending the semantics of the syntax element 'content\_interpretation\_type' which is part of the frame packing arrangement SEI message as defined in the AVC [6] and HEVC [7] specifications. The text for the proposed extension is made on top of the HEVC Range Extensions Draft Specification [8] and attached with this contribution.

## System operation at the decoding end

The proposed extension to the frame packing arrangement SEI message informs the decoder that the decoded pictures contain main and auxiliary views of a 4:4:4 frame as the constituent frames of the frame packing arrangement. This information can be used to process the main and auxiliary views appropriately for display or other purposes.

When the system at the decoding end desires the video in 4:4:4 format and is capable of reconstructing the 4:4:4 frames from the main and auxiliary views, it should do so and the output format should be 4:4:4. Otherwise, only the main view should be given as output and the output format will then be 4:2:0.

## Pre-processing and post-processing

When content\_interpretation\_type is set to 3, the indication is that none of the chroma samples underwent an anti-alias filtering operation during the process of frame packing i.e. and =. In such a case, the chroma samples comprising the main view are a result of a direct sub-sampling of the chroma planes representing the 4:4:4 frame. As shown in the appendix, direct sub-sampling without filtering can create aliasing artifacts for certain types of screen content when only the main view is used to generate a 4:2:0 output.

In order to reduce the aliasing artifacts and improve the visual quality for the case where only the main view is used, the content\_interpretation\_type can be set to 4 and the main view can be generated using filtered/pre-processed versions of the 4:4:4 chroma planes. In such a case, it is recommended that the filter choice be made based on the chroma sample grid alignment with luma sample grid (inferred from chroma\_sample\_loc\_type\_top\_field and chroma\_sample\_loc\_type\_bottom\_field). For simplicity, in the case when the chroma sample grid aligns with the luma sample grid for each particular direction (horizontal/vertical), it is suggested that that the 3-tap filter [0.25 0.5 0.25] be used in that direction. If the chroma sample grid positions are centered between the luma sample positions for a particular direction (horizontal/vertical), then it is suggested that the 2-tap filter [0.5 0.5] be used in that direction. Another possible filter choice for the latter case is [0.125 0.375 0.375 0.125].

For example, if we consider the case where the chroma sample grid is not aligned with the luma sample grid in both horizontal and vertical directions (i.e. when chroma\_sample\_loc\_type\_top\_field and chroma\_sample\_loc\_type\_bottom\_field are equal to 1), the 2-tap filter [0.5 0.5] would be applied in both directions, such that and are obtained as follows:

When pre-processing is used (content\_interpretation\_type set to 4), the main view does not contain samples and but contains their filtered counterparts and . The auxiliary view contains the other chroma samples.

If the decoding system decides to output a 4:4:4 frame, a post-processing step should be applied to estimate the samples , as , from the encoded packed frame. For example, a simple suggested estimation of and would be as follows:

In the proposed form, with content\_interpretation\_type equal to 4 and chroma\_sample\_loc\_type\_top\_field and chroma\_sample\_loc\_type\_bottom\_field equal to 1, with the suggested anti-alias filter of [0.5 0.5], the value would perfectly reconstruct the input values in the absence of quantization error and rounding error. When considering quantization error, using somewhat different values would be advised (e.g., as determined by quantization step-size-dependent cross-correlation analysis).

## Band separation filtering for the auxiliary frame

In the frame packing scheme illustrated in section 2.1, sample values of and frames are placed directly into (and are directly unpacked from) the auxiliary frames. We thus refer to these schemes as “direct” packing approaches. Alternatively, we can consider the auxiliary frame samples as an enhancement layer signal to be combined with the main frame (or base layer frame) data. The main and auxiliary frame data can be formed using low-pass and high-pass band separation filtering, instead of “direct” sample packing. With this variation, the primary signal energy can be concentrated into the main frame, and arbitrarily low bit rates can be allocated to the supplemental auxiliary frame data that forms the enhancement signal.

Instead of encoding auxiliary frame samples directly, a two-dimensional, three-band wavelet decomposition can first be applied to and before the actual encoding process. Mathematically, for an array , where *=* or , define the following:

, for .

, for .

, for .

, for .

, for

, for

, for

, for

, for

, for

, for

, for

, for

A typical four-band wavelet decomposition breaks the frame into “LL”, “LH”, “HL” and “HH” subbands (“LL” = low-pass in both vertical and horizontal directions, “LH” = low-pass vertical, high-pass horizontal, and so forth). However, in our wavelet packing scheme as defined by the above equations, the “HL” and “HH” bands are not created; instead, the vertical high-pass signal is kept at full horizontal resolution, i.e., B2 and B3 are the “LL” bands of and respectively, B4 and B5 are vertical high-pass signals, i.e. a vertical “H” band of and , respectively, B6 and B8 consist of even-numbered rows of the “LH” band of , and B7 and B9 consist of odd-numbered rows of the “LH” band of . That way, the decoder would apply the corresponding inverse wavelet operations after decoding the main and auxiliary frames to obtain and samples. Moreover, an additional vertical band separation can be performed, such that B6 and B8 are an “LHL” and “LHH” band of , and B7 and B9 are an “LHL” and “LHH” band of .

For the scenario where the auxiliary frame is transmitted at lower bit rates (lower quality relative to the main frame), the chroma information from the main frame ( and ) sets the minimum level of quality for the and reconstruction, and any information from the auxiliary frame is used to improve beyond that minimum quality level. In the case of the “direct” frame packing method however, samples from the auxiliary frame are directly unpacked into and frames. This approach would cause the chroma samples obtained from the auxiliary frame (3 out of 4) to have a lower quality compared to the chroma samples obtained from the main frame. However, the band-separation frame packing approach potentially incurs a larger rounding error in the pre-processing steps than the “direct” frame packing approach because of the additional filtering operations involved (in the absence of bit-depth expansion).

We have also experimented with the use of lifting-based band-separation filtering to mitigate the rounding error effects, with clipping to eliminate the bit-depth expansion. This topic is further discussed in our recent conference publication at the *IEEE Data Compression Conference* (DCC 2013) [9].

# Experiment results

We initially tested an end-to-end system for packing a 4:4:4 frame into two 4:2:0 frames, based on Microsoft’s implementation of an AVC software encoder and decoder with a simple “IPPP” (forward-predictive) coding structure [2]. We have since conducted similar tests using the HEVC HM 9.0 encoder [10] and the HEVC HM 10.0 with 4:4:4: range extensions [11] using the Low Delay Main configuration. In January, we presented some test results using our own screen content test sequences [3]. In April and July, we provided results using the common 4:4:4 screen content test sequences that were used in other committee contributions [12].

Previously, we have run tests by keeping the main-frame (that yields the 4:2:0 representation of the scene) QP to be constant and varying the auxiliary frame QP. However, the typical use-case for 4:4:4 frame-packing scenarios is one where the main-frame undergoes a similar or lower compression-level compared to the auxiliary frame. Hence it is more instructive to analyze the behavior of the different frame packing approaches in this region.

In the current contribution, we tested both the “direct” frame packing approach and the band-separation frame packing approach using HEVC codec. We have modified our test conditions keeping in line with the above principle for operating region. We also fixed a bug in the code for band-separated frame packing which makes it perform better than previously reported.

For each frame packing approach, the encoder starts with a 4:4:4 input frame, generates the main and auxiliary views for that frame packing approach, constructs a 4:2:0 frame with twice the height of the 4:4:4 frame, places the main view in the top half and the auxiliary view in the bottom half of the 4:2:0 frame, and encodes the 4:2:0 frame. This corresponds to the use of the frame packing arrangement SEI message with frame\_packing\_arrangement\_type equal to 4. The decoder decodes the 4:2:0 frame, extracts the main and auxiliary views and reassembles the 4:4:4 frame for output (using to simplify the initial testing).

In each experiment, the main frame luma QP parameter was varied from 10 to 34 (high-bitrate to low-bitrate scenarios). The main frame chroma QP was kept at the same value as the luma QP. The luma and chroma QPs are maintained at the same value by using appropriate cb\_qp\_offset / cr\_qp\_offset values at either the picture and/or slice-level. For both packing approaches (“direct” and band-separation), we experimented with different values of QPdelta ( QP of auxiliary frame – QP of main frame) for the auxiliary frame (representing 3/4th of the chroma samples) and the final 4:4:4 chroma BD-rate performance was compared for the different approaches. We observed that using a negative value for QPdelta (-6) for the auxiliary frame resulted in a very slight performance gain while a positive QPdelta (+6) caused a significant drop in performance for both approaches. Based on the results, we kept the auxiliary frame (luma and chroma) QP at the same value as the main frame chroma QP (i.e. QPdelta of 0) for both band-separation and “direct” frame packing approaches to simplify testing. Owing to the nature of the experiments, the chroma signal in the main frame will have equal or better quality at the 4:2:0 representation-level than the chroma quality at the resultant 4:4:4 representation-level.

We compared the BD-rate performance using the HEVC HM12.0 (Encoder Version "[12.0][Windows][VS 1700][64 bit]") [10] on widely-used 4:4:4 screen content test sequences [12]. The BD-rate comparison result is shown in Table 1 and the detailed PSNR curves are attached in the spread sheet. It can be seen that band-separation out-performs “direct” frame packing approach by a significant margin for all the sequences.

It can also be seen in the spreadsheet that the two schemes have pretty much the same coding-efficiency (luma within 0.1% and chroma within 0.8% in terms of BD bit rate) on the main frame (4:2:0 representation), which is expected since the main frame is identical for both schemes (any minor difference is due to differences in encoder decisions triggered due to the auxiliary frame). This indicates that all of the difference in the coding-efficiency performance for the 4:4:4 representation is due to the chroma signal carried in the auxiliary frame.

We also compared the performance of native 4:4:4 encoding using the HM12.0 (with Range Extension software for 4:4:4 coding), for current operating scenario. It can be seen that encoding the 4:4:4 test sequences with the 4:4:4 HM directly in the native 4:4:4 format yields a superior compression performance compared to encoding them using the band-separation frame packing, and substantially better compression performance than “direct” frame packing. To some extent, this is due to the fact that there is additional overhead for frame-packing approaches; for example, there is redundant motion and segmentation information that needs to be sent for the auxiliary frame if frame-packing approaches are used. At low-bitrates, band-separation becomes more competitive in terms of compression performance. Also, both frame packing schemes are useful when the 4:4:4 format is not supported by the decoder.

Table 1: BD-rate comparison of band-separation scheme relative to “direct” packing  
on common 4:4:4 screen content test sequences, for typical use-case scenarios

|  |  |  |
| --- | --- | --- |
| **Test Sequence** | **Chroma BD-rate** | **Luma BD-rate** |
| *web\_browsing* | -26.5% | -12.8% |
| *word\_editing* | -32.4% | -10.8% |
| *programming* | -50.5% | -25.5% |
| *map* | -52.2% | -22.0% |
| *viking* | -73.4% | -11.8% |
| *robot* | -70.3% | -22.3% |
| *twist\_tunnel* | -12.8% | -0.1% |
| *Average* | -45.4% | -15.0% |

**Figure 3: S**ample rate-distortion plot of different packing schemes on a common screen content test sequence *programming*. Also shown is the rate-distortion plot of native HM 4:4:4 encoder. The resolution of the 4:4:4 test sequence is 1280x720 with 150 frames at a frame rate 60 fps.

# Conclusion

This proposal enables the creation of a system in which the existing 4:2:0 decoding process becomes the core component of a 4:4:4 decoder. Moreover, a subset of the decoded output can provide compatibility with existing 4:2:0 decoding systems. Since 4:2:0 is the most widely supported format in products, having an effective way of conveying 4:4:4 content through such decoders can provide the substantial benefit of enabling widespread near-term deployment of 4:4:4 capabilities. At the Vienna meeting in July, the “direct” frame packing approach (content\_interpretation\_type equal to 3 and 4) was adopted in spirit. We request the committee to formalize that decision. We further request the adoption of the band-separated frame packing approach as well (content\_interpretation\_type equal to 5).

# References

1. T. Lin *et. al,* “Syntax and semantics of Dual-coder Mixed Chroma-sampling-rate (DMC) coding for 4:4:4 screen content”, JCTVC-J0233, 10th Meeting: Stockholm, SE, 11–20 July 2012.
2. Y. Wu*,* S. Kanumuri, S. Sadhwani, L. Zhu, S. Sankuratri, G. J. Sullivan, and B. A. Kumar, “Frame packing arrangement SEI for 4:4:4 content in 4:2:0 bitstreams”, JCTVC-K0240, 11th Meeting: Shanghai, CN, 10–19 October, 2012.
3. Y. Zhang, Y. Wu*,* S. Kanumuri, S. Sadhwani, G. J. Sullivan, and H. S. Malvar, “Updated proposal for frame packing arrangement SEI for 4:4:4 content in 4:2:0 bitstreams”, JCTVC-L0316, 12th Meeting: Geneva, CH, 14–23 January, 2013.
4. Y. Zhang, Y. Wu*,* S. Kanumuri, S. Sadhwani, G. J. Sullivan, and H. S. Malvar, “Additional experiment results for frame packing arrangement SEI message for 4:4:4 content in 4:2:0 bitstreams”, JCTVC-M0281, 13th Meeting: Incheon, Korea, 18–23 April, 2013.
5. Y. Zhang, Y. Wu*,* S. Kanumuri, S. Sadhwani, G. J. Sullivan, and H. S. Malvar, “Updated proposal with software for frame packing arrangement SEI message for 4:4:4 content in 4:2:0 bitstreams”, JCTVC-N0270, 14th Meeting: Vienna, AT, 25 July –2 August, 2013.
6. Rec. ITU-T H.264 | ISO/IEC 14496-10, *Advanced video coding*.
7. B. Bross *et. al*, “High efficiency video coding (HEVC) Text Specification Draft 9”, JCTVC-L1003, 11th meeting, Shanghai, October 2012.
8. D. Flynn, J. Sole and T. Suzuki, “High Efficiency Video Coding (HEVC) Range Extensions text specification: Draft 4”, JCTVC-N1005, 14th Meeting: Vienna, AT, 25 July –2 August, 2013.
9. Y. Wu, S. Kanumuri, Y. Zhang, S. Sadhwani, G. J. Sullivan, and H. S. Malvar, “Tunneling High-Resolution Color Content through 4:2:0 HEVC and AVC Video Coding Systems”, *Proc. IEEE Data Compression Conf.* (DCC 2013), Snowbird, Utah, pp. 3–12, March 2013.
10. “HEVC software repository”, <https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/>.
11. <http://hevc.kw.bbc.co.uk/trac/browser/branches/HM-range-extensions>
12. <ftp://ftp.tnt.uni-hannover.de/testsequences/FrExt-candidate-sequences/screen_content/>

# Patent rights declaration(s)

**Microsoft Corporation may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**