|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  16th Meeting: San Jose, US 9 January – 17 Jan. 2014 | Document: JCTVC-P0216 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Additional content interpretation type and experiments 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, JCTVC-N0270 and JCTVC-O0198 that provides new experimental results for the SC and RExt test sets and QP ranges. A content interpretation type that uses “lifting-based band separation” (to remove rounding error effects, with clipping to eliminate the bit-depth expansion) is also discussed and evaluated. The lifting-based concept was mentioned in prior contributions but had not previously been well tested. It is reported that the additional results indicate a coding-efficiency benefit for the lifting-based scheme over both the ordinary “band-separation” and “direct” frame-packing modes. It is suggested that the additional content interpretation type should be supported as well as the others.

For any of the interpretation types using the proposed frame packing approach, 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 (or a similar new SEI message) in both AVC and HEVC, to facilitate deployment of systems using this method. Relative to native 4:4:4 encoding, the proposed scheme has the advantage of compatibility with the ordinary 4:2:0 decoding process.

This feature of conveying 4:4:4 through conventional 4:2:0 decoders is the main motivation for this proposal, to enable more widespread deployment of 4:4:4 content usage by avoiding the need for decoders to support a different decoding process for it. It is reported that the attached software (now updated with 10 bit capability) is capable of handling the frame-packing and frame-unpacking processes and can be used in conjunction with any 4:2:0 codec.

# Introduction

There are a variety of existing and emerging applications (e.g., as discussed in [1] and [2]) that operate with screen content where 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] and JCTVC-O0198 [6] that provides new experimental results on the screen content and range extensions test set, over the main-tier and high-tier QP ranges. Also, a variation of the band-separated approach, a lifting-based band-separation technique is proposed and evaluated. It is shown that the lifting-based band-separation approach is superior to the prior band-separation approach for most screen content and eliminates rounding errors, which is a frequent issue with band-separation approach. 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 updated software capable of handling the frame-packing and frame-unpacking processes.

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. This ability to transmit 4:4:4 signal through conventional 4:2:0 decoders is the motivating factor for this proposal, and it is expected to enable quicker and widespread deployment of 4:4:4 content.

# 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.

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. Thus systems can display only 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.

## 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 [7] and HEVC [8] specifications. The text for the proposed extension is made on top of the HEVC Range Extensions Draft Specification [9] and attached with this contribution.

The content interpretation types proposed herein are shown in Table 1.

Table 1: Proposed content interpretation type codes

|  |  |
| --- | --- |
| 3 | Indicates that the two constituent frames form the main and auxiliary views (4:2:0 frames) representing a 4:4:4 frame, with frame 0 being associated with the main view and frame 1 being associated with the auxiliary view. Indicates that the chroma samples of frame 0 should be interpreted as unfiltered samples of the 4:4:4 frame (without anti-alias filtering). |
| 4 | Indicates that the two constituent frames form the main and auxiliary views (4:2:0 frames) representing a 4:4:4 frame, with frame 0 being associated with the main view and frame 1 being associated with the auxiliary view. Indicates that the chroma samples of frame 0 should be interpreted as having been anti-alias filtered prior to frame packing. |
| 5 | Indicates that the two constituent frames form the main and auxiliary views (4:2:0 frames) representing a 4:4:4 frame, with frame 0 being associated with the main view and frame 1 being associated with the auxiliary view. Indicates that the chroma samples of frame 0 should be interpreted as having been band-separation filtered prior to frame packing. |
| 6 | Indicates that the two constituent frames form the main and auxiliary views (4:2:0 frames) representing a 4:4:4 frame, with frame 0 being associated with the main view and frame 1 being associated with the auxiliary view. Indicates that the chroma samples of frame 0 should be interpreted as having been lifting-based band-separation filtered prior to frame packing. |

## 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 for "direct" packing types

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 anti-alias 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).

## Lifting-based Band separation filtering for the auxiliary frame

As discussed in section 2.5, the band-separation frame packing approach incurs rounding error. In this section, another variation of the band-separation approach, referred to as lifting-based band separation, is discussed which can mitigate the rounding error issue. The underlying feature of this approach is the same as in the band-separation approach, and a three-band wavelet decomposition is applied to the and signals prior to the encoding process.

Mathematically, for an array , where *=* or , define the following:

, for .

, for .

, for .

, for .

, for

, for

, for

, for

), for

), for

, for

), for

), for

As with the band-separation approach described in section 2.5, the “HL” and “HH” bands are not created; instead, the vertical high-pass signal is kept at full horizontal resolution. The decoder applies the corresponding inverse wavelet operations after decoding the main and auxiliary frames to obtain and samples.

The notable feature of this approach is that rounding error is removed, although the clipping step potentially introducing clipping errors during the generation of the chroma samples in the "auxiliary frame". However clipping errors occur only when the adjacent chroma samples differ by a very high margin. It is asserted that this is very rare even for screen content, and hence clipping artifacts are very infrequent. Even in cases where clipping occurred, we did not observe any significant visual artifacts introduced by clipping during informal viewing. In comparison, rounding errors prevalent in band-separated approach are more frequent for most screen content. Thus it is expected that the lifting-based band separation will ordinarily provide superior performance compared to the band-separation approach for most types of screen content.

# Experiments

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 [11] and the HEVC HM 10.0 with 4:4:4: range extensions [12] 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 [13]. In October, we further provided test results on common 4:4:4 screen content test sequences for the band-separation and the direct-mode frame packing approaches [6].

In this contribution, we present results for SC and RExt test sequences for the main and high tier QP ranges. The same test setup is used, as described in contribution JCTVC-O0198 [6]. 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 for the reconstruction with the anti-aliased "direct" packing type to simplify the initial testing).

In each experiment, the main frame luma QP parameter was varied from 17 to 37 (for so-called "main tier" and "high tier" QP scenarios). The main frame Chroma QP was kept at the same value as the luma QP. The auxiliary frame 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.

We compared the Chroma BD-rate performance using the HEVC HM12.1 with Range extensions (Encoder Version "[12.1\_RExt5.1][Windows][VS 1700][64 bit]") [12] on the 4:4:4 screen content and range extensions test sequences [13]. The BD-rate comparison result is shown in Table 2, and Table 3, and the detailed PSNR curves are provided in the attached spreadsheet. It can be seen that band-separation out-performs “direct” frame packing approach by a significant margin for all the sequences. Furthermore, the lifting-based band-separation approach outperforms band-separation by a slight margin, except with a couple of sequences. One sequence (BirdsInCage) has an abnormal BD-rate delta values presumably due the operating region for the BD-plots, as seen in Table 3.

We also compared the performance of native 4:4:4 encoding using the HM12.1 (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 lifting-based band-separation or 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, mode-decision and segmentation information that needs to be sent for the auxiliary frame if frame-packing approaches are used. Also the original signal is separated into different components and transforms applied to non-ordinary image data, which may result in compression inefficiencies. At lower bitrates, both band-separation and lifting-based band-separation techniques become more competitive in terms of compression performance.

# Discussion

Some discussion been raised in previous JCT-VC meetings about optimal bit-allocation for the chroma samples in the main and auxiliary frames. For the frame-packing approaches, we have previously 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) to achieve optimal bit-allocation for the chroma information in the auxiliary frame. It was observed that using a more optimal bit-allocation does provide a slight improvement in the BD rate performance, however it does not approach the native HM444 coding quality, due to additional overhead for frame-packing approaches as described above. The precise values for QPdelta do not really seem to affect the overall outcome very much. 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. QP-delta of 0) for both band-separation and “direct” frame packing approaches to simplify testing. Any encoder utilizing any of the frame-packing is, of course, free to select a more optimal bit-allocation suitable to the packing approach used.

On the other hand, there was also some reflector discussion on changing the luma/chroma balance for the sake of simplifying BD-rate evaluation with conventional 4:4:4 encoding using HM. One of the ideas discussed was to adjust the chroma QP for the main frame (4:2:0 representation) in addition to adjusting the auxiliary frame QPs (luma and chroma) so that the 4:4:4 chroma PSNR obtained by this method matches the corresponding value obtained by HM 4:4:4 software. It should be noted that such adjustments need sequence-specific values. This is evident from another contribution [15] to the same meeting that attempts to match the PSNRs using a fixed differential of 12 between luma and chroma QPs. While this adjustment helps to bring the PSNRs closer for some sequences, the same adjustment does an over-correction to the other side for a few other sequences (notably the Rext sequences) and does not necessarily help BD-rate evaluations. For example, the sequence BirdsInCage with LB-MT (Low Delay B – Main Tier) setting gets a luma BD-rate of -57%, U BD-rate of 77% and V BD-rate of 245% when compared to HM 4:4:4. With the QP adjustment, the luma BD-rate changes to 383%, U BD-rate changes to 33% and V BD-rate changes to 87%. Neither of these operational points are useful in evaluating compression efficiency. Hence, our recommendation is to use the BD-graphs (and not BD-rates) to come up with a rough understanding of the compression efficiency loss compared to HM 4:4:4. We would like to also draw attention that the goal of this method is not to provide compression efficiency gains but to provide a method for 4:2:0 codecs to compress 4:4:4 content.

It was suggested in the previous meeting that the performance analysis of this contribution be evaluated against other mechanisms for 4:4:4 encoding, e.g. use of "auxiliary pictures" (not to be confused with a different use of the term "auxiliary" in this contribution) [14]. Testing the approach described in [14] would be feasible (although it would take some work, as we did not have software for it readily at hand).

**We emphasize that the different approaches to conveying 4:4:4 video provide quite different functionality, and we thus consider the mere comparison of coding efficiency between the approaches as not especially relevant. The same is true, of course, for comparison with an ordinary 4:4:4 single-layer codec design.**

Both a single-layer 4:4:4 codec and the scheme proposed in [14] require the internal operation of the decoder to support some capabilities beyond those of an ordinary 4:2:0 decoder – whereas the approach presented here has a more simple architecture – it requires only a small pre-processing stage for an encoder and corresponding post-processing stage for a decoder, while these functional units themselves can be used without alteration.

**This feature of conveying 4:4:4 through conventional 4:2:0 decoders is the main motivation for this proposal, as we suggest that it can enable more widespread deployment of 4:4:4 content.**

We consider both the schemes proposed in [14] and in this contribution to be useful, although we would expect them to be used in somewhat different contexts.

Table 2: Chroma BD-rate comparison of band-separation scheme relative to “direct” packing  
on common 4:4:4 screen content and range extensions test sequences for main-tier Qp ranges

|  |  |  |  |
| --- | --- | --- | --- |
| **Test Sequences**  **Screen Content** | **BD-rate delta** | **Test Sequences**  **Range Extensions** | **BD-rate delta** |
| *WebBrowsing* | −43.5% | *RainFruits* | −81.78% |
| *WordEditing* | −47.11% | *BirdsInCage* | −97.07% |
| *Programming* | −56.24% | *LupoCandlelight* | −78.36% |
| *Map* | −60.7% | *Kimono* | −93.18% |
| *Viking* | −86.80% | *VenueVu* | −80.54% |
| *Robot* | −87.02% | *CrowdRun* | −86.47% |
| *TwistTunnel* | −22.29% | *Traffic* | −86.83% |
| *SlideShow* | −56.97% |  |  |
| *PCBlayout* | −36.15% |  |  |
| *PPT* | −35.84% |  |  |
| *VideoConferencing* | −39.19% |  |  |
| *Waveform* | −39.63% |  |  |

Table 3: Chroma BD-rate comparison of lifting-based band-separation scheme relative to non-lifting band-separation packing on common 4:4:4 screen content and range extensions test sequences for main-tier Qp ranges

|  |  |  |  |
| --- | --- | --- | --- |
| **Test Sequences**  **Screen Content** | **BD-rate delta** | **Test Sequences**  **Range Extensions** | **BD-rate delta** |
| *WebBrowsing* | −0.96% | *VenueVu* | −5.67% |
| *WordEditing* | −13.34% | *Rainfruits* | −2.74% |
| *Programming* | −6% | *LupoCandlelight* | −21.72% |
| *Map* | −8.16% | *BirdsinCage* | 55.52% |
| *Viking* | −9.60% | *Kimono* | 9.87% |
| *Robot* | −13.22% |  |  |
| *TwistTunnel* | −14.20% |  |  |
| *SlideShow* | −3.31% |  |  |
| *PCBlayout* | 5.64% |  |  |
| *PPT* | −1.88% |  |  |
| *VideoConferencing* | 5.27% |  |  |
| *Waveform* | 7.42% |  |  |

**Figure 2**: Sample rate-distortion plot of different packing schemes on a common screen content test sequence *WordEditing* in low delay configuration test. Also shown is the rate-distortion plot of a native HM 4:4:4 encoder. The resolution of the 4:4:4 test sequence is 1280x720 with 50 coded frames at a frame rate of 60 fps.

**Figure 3**: Rate-distortion plot of different packing schemes on *BirdsInCage* sequence inlow delay mode. As can be seen from graphs, the graphs are already mostly “flat”, hence adding a more bits doesn't result in a huge increase the PSNR. Thus the BDrate Delta values show huge differences for different schemes.

# 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. We thus request the addition of the four frame packing content interpretation types (content\_interpretation\_type equal to 3, 4, 5 and 6).

# 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. Srinath *Reddy*, S. Kanumuri, Y. Wu*,* S. Sadhwani, G. J. Sullivan, and H. S. Malvar, “Additional exepriments and software for frame packing arrangement SEI message for 4:4:4 content in 4:2:0 bitstreams”, JCTVC-O0198, 15th Meeting: Geneva, CH, 23 October –1 November, 2013.
7. Rec. ITU-T H.264 | ISO/IEC 14496-10, *Advanced video coding*.
8. B. Bross *et. al*, “High efficiency video coding (HEVC) Text Specification Draft 9”, JCTVC-L1003, 11th meeting, Shanghai, October 2012.
9. 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.
10. 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.
11. “HEVC software repository”, <https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/>.
12. <http://hevc.kw.bbc.co.uk/trac/browser/branches/HM-range-extensions>
13. <ftp://ftp.tnt.uni-hannover.de/testsequences/FrExt-candidate-sequences/screen_content/>
14. M. M. Hannuksela, “REXT/MV-HEVC/SHVC HLS: Auxiliary picture layers”, [JCTVC-O0041](http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=8152) 15th Meeting: Geneva, CH, 23 October – 1 November, 2013.
15. K. Ugur, D. Bugdayci and M. M. Hannuksela, “On frame packing arrangement SEI message for 4:4:4 content in 4:2:0 bitstreams”, JCTVC-P0121, 16th meeting, San Jose, USA, January 2014.

# 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).**