Subband Decomposition for High-Resolution Color in HEVC and AVC 4:2:0 Video Coding Systems

Srinath Reddy, Sandeep Kanumuri, Yongjun Wu, Shyam Sadhwani,
Gary J. Sullivan, and Henrique S. Malvar

Microsoft Corporation
One Microsoft Way, Redmond, WA 98052, USA

Abstract: We present a frame packing arrangement (FPA) scheme that enables an encoder to convey high-resolution color video (4:4:4) through a video coding system designed for subsampled color video (4:2:0). Relative to native 4:4:4 encoding, the proposed scheme provides the advantage of compatibility with the 4:2:0 video decoding process that is likely to be available in a variety of products. We present different techniques for frame packing and evaluate their coding efficiencies for cases where the 4:4:4 and 4:2:0 encodings have similar compression levels. The use of this scheme can be indicated by a metadata tag such as the FPA supplemental enhancement information (SEI) message defined in the HEVC (Rec. ITU-T H.265 | ISO/IEC 23008-2) and AVC (Rec. ITU-T H.264 | ISO/IEC 14496-10) video coding standards, in a similar manner as has previously been used to represent the two views of stereoscopic 3D video for compatible encoding. The packing and unpacking procedures are of relatively low complexity, and can easily be implemented in conjunction with any 4:2:0 video codec.

1. Introduction

Most video codecs that are commercially available today support only the 4:2:0 chroma format [1], in which the chroma resolution is half that of the luma resolution both vertically and horizontally, as contrasted with using a 4:4:4 format, in which the chroma information is represented at the same resolution used for the luma [1]. The reason is that for videos containing natural scenes, such as in professional or amateur movies, YC_BC_R (a.k.a. YUV) 4:2:0 format is good enough, as we cannot usually see a noticeable difference between the two formats. However, in applications such as virtual / remote desktop computing, wireless displays, and others, the video to be encoded often contains "screen content" [2], [3] with hard-edged / high-resolution text and graphics. For such applications, the 4:2:0 color format leads to noticeable blurring of the video content [4].

In this work, 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. This paper is an update of prior work [2] that provides new experimental on the screen content and range extensions test set of the HEVC coding standard [14], over a variety of QP ranges. In addition to the "direct" and the band-separation techniques presented in [6], 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 paper also provides

specification text along with a figure illustrating the frame packing of 4:4:4 content into 4:2:0 format. In addition, the corresponding JCT-VC contribution [8] 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.

2. Packing a YUV 4:4:4 frame into main and auxiliary views

As described in [4] a frame in YUV (i.e., YC_BC_R, YCoCg, GBR, etc.) 4:4:4 format [1] can be represented as indicated in the top part of Fig. 1, where Y_{444}, U_{444}, and V_{444} are the Y, U, and V planes comprising the YUV 4:4:4 frame.
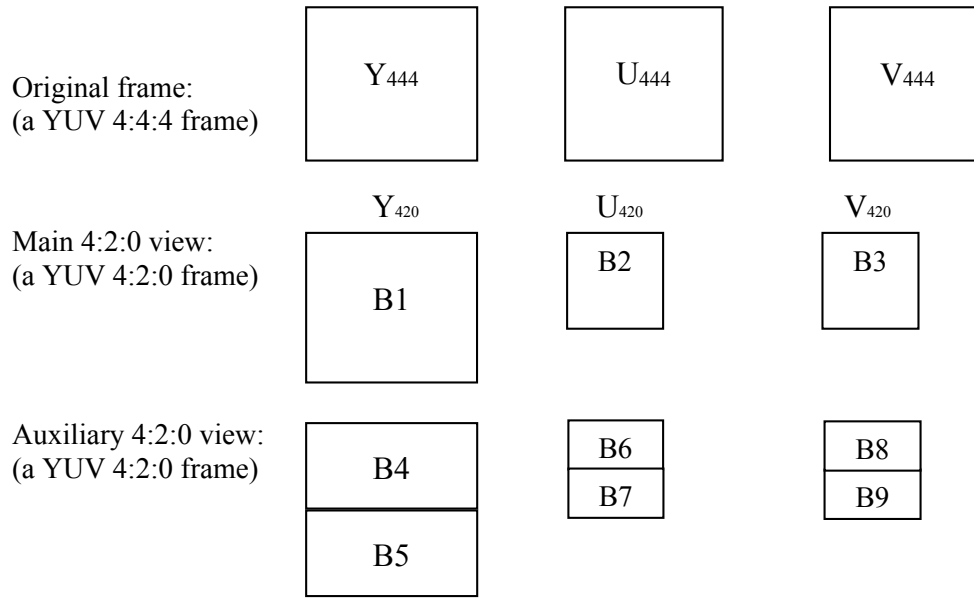

Figure 1. Top: Representation of an original frame in YUV 4:4:4 format. Bottom: Decomposition of the frame into two YUV 4:2:0 views: "main" and "auxiliary"

Let the resolution of these planes be represented by width W and height H. The YUV 4:4:4 frame represented above are packed into two YUV 4:2:0 frames (as main and auxiliary view frames) as shown in the bottom part of Fig. 1. 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 Y_{444}, U_{444}, and V_{444} as follows:

Main view:

Area B1: $Y_{420}^{main}(x,y) = Y_{444}(x,y)$, where the range of (x,y) is $[0, W-1] \times [0, H-1]$.

Area B2: $U_{420}^{main}(x,y) = \tilde{U}_{444}(2x, 2y)$, with (x,y) in $\left[0, \frac{W}{2} - 1\right] \times [0, \frac{H}{2} - 1]$.

Area B3: $V_{420}^{main}(x,y) = \tilde{V}_{444}(2x, 2y)$, with (x,y) in $\left[0, \frac{W}{2} - 1\right] \times [0, \frac{H}{2} - 1]$.

Auxiliary view:

Area B4: $Y_{420}^{aux}(x,y) = U_{444}(x, 2y+1)$, with (x,y) in $[0, W-1] \times [0, \frac{H}{2} - 1]$.

Area B5: $Y_{420}^{aux}\left(x, \frac{H}{2} + y\right) = V_{444}(x, 2y+1)$, with (x,y) in $[0, W-1] \times [0, \frac{H}{2} - 1]$.

Area B6: $U_{420}^{aux}(x,y) = U_{444}(2x+1, 4y)$, with (x,y) in $\left[0, \frac{W}{2} - 1\right] \times [0, \frac{H}{4} - 1]$.

Area B7: $U_{420}^{aux}\left(x, \frac{H}{4} + y\right) = V_{444}(2x+1, 4y)$, with (x,y) in $\left[0, \frac{W}{2} - 1\right] \times [0, \frac{H}{4} - 1]$.

Area B8: $V_{420}^{aux}(x,y) = U_{444}(2x+1, 4y+2)$, with (x,y) in $\left[0, \frac{W}{2} - 1\right] \times [0, \frac{H}{4} - 1]$.

Area B9: $V_{420}^{aux}\left(x, \frac{H}{4} + y\right) = V_{444}(2x+1, 4y+2)$, with (x,y) in $\left[0, \frac{W}{2} - 1\right] \times [0, \frac{H}{4} - 1]$.

In the above equations, $\tilde{U}_{444}(2x, 2y)$ and $\tilde{V}_{444}(2x, 2y)$ are either the same as or represent anti-alias filtered values corresponding to $U_{444}(2x, 2y)$ and $V_{444}(2x, 2y)$, respectively, where the range of (x,y) is $[0, W/2 - 1] \times [0, H/2 - 1]$. This choice is explained in more detail in Section 3. The packing method is designed such that the main view is the YUV 4:2:0 equivalent of the original YUV 4:4:4 frame. Systems can just display the main view if YUV 4:4:4 is either not supported or is considered not necessary for the decoder.

2.1. Advantages

The proposed packing method has the following key characteristics:

- The main view is a YUV 4:2:0 frame equivalent of the original YUV 4:4:4 frame
    - Systems can opt to display just the main view if only YUV 4:2:0 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 in Figure 2, 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.

2.2. Extension to frame packing arrangement SEI message

To use the frame packing of YUV 4:4:4 content approach in AVC and HEVC involves extending the semantics of the syntax element "`content_interpretation_type`", which is part of the frame packing arrangement for supplemental enhancement information (SEI) message as defined in the AVC [8] and HEVC [12] specifications. The text for the proposed extension is described in detail in [7].
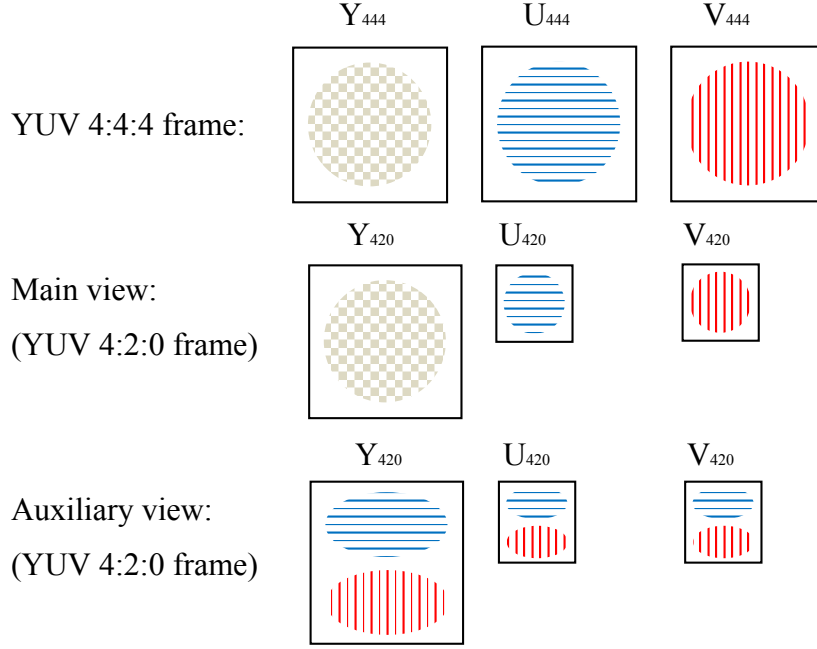
Figure 2. Illustration of the spatial correspondence relationships among the components of the proposed frame packing scheme.

3. Pre-processing and post-processing considerations

For a particular value of `content_interpretation_type`, the indication would be that none of the chroma samples underwent an anti-alias filtering operation during the process of frame packing i.e. $\widetilde{U}_{444}(2x, 2y) = U_{444}(2x, 2y)$ and $\widetilde{V}_{444}(2x, 2y) = V_{444}(2x, 2y)$. 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 [4], 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 main view can be generated using filtered/pre-processed versions of the 4:4:4 chroma planes. A different value of `content_interpretation_type` could be used to indicate that the anti-aliasing has been applied. In that case, it is recommended that the filter be 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 $\tilde{U}_{444}(2x, 2y)$ and $\tilde{V}_{444}(2x, 2y)$ are obtained as follows:

$$\tilde{U}_{444}(2x, 2y) =$$

$$\frac{[U_{444}(2x, 2y) + U_{444}(2x + 1, 2y) + U_{444}(2x, 2y + 1) + U_{444}(2x + 1, 2y + 1) + 2]}{4}$$

$$\tilde{V}_{444}(2x, 2y) =$$

$$\frac{[V_{444}(2x, 2y) + V_{444}(2x + 1, 2y) + V_{444}(2x, 2y + 1) + V_{444}(2x + 1, 2y + 1) + 2]}{4}$$

When pre-processing is used, the main view does not contain samples $U_{444}(2x, 2y)$ and $V_{444}(2x, 2y)$ but contains their filtered counterparts $\tilde{U}_{444}(2x, 2y)$ and $\tilde{V}_{444}(2x, 2y)$. 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 $U_{444}(2x, 2y)$, $V_{444}(2x, 2y)$ as $U'_{444}(2x, 2y)$, $V'_{444}(2x, 2y)$ from the encoded packed frame. For example, a simple suggested estimation of $U'_{444}(2x, 2y)$ and $V'_{444}(2x, 2y)$ would be as follows:

$$U'_{444}(2x, 2y) = (1 + \alpha + \beta + \gamma)\,\tilde{U}_{444}(2x, 2y) - \alpha\,U_{444}(2x + 1, 2y) - \beta\,U_{444}(2x, 2y + 1)$$
$$- \gamma\,U_{444}(2x + 1, 2y + 1)$$

$$V'_{444}(2x, 2y) = (1 + \alpha + \beta + \gamma)\,\tilde{U}_{444}(2x, 2y) - \alpha\,V_{444}(2x + 1, 2y) - \beta\,V_{444}(2x, 2y + 1)$$
$$- \gamma\,V_{444}(2x + 1, 2y + 1)]$$

In the proposed form, with the SEI descriptor `content_interpretation_type` indicating anti-aliasing filtering and with "`chroma_sample_loc_type_top_field`" and "`chroma_sample_loc_type_bottom_field`" set equal to 1, with the suggested anti-alias filter of [0.5 0.5], the value $\alpha = \beta = \gamma = 1$ 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).

3.1. Band separation filtering for the auxiliary frame

In the frame packing scheme illustrated in Section 2, sample values of U_{444} and V_{444} 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 U_{444} and V_{444} before the actual encoding process. Mathematically, for an array A, where $A=U_{444}$ or V_{444}, define the following:

$$C_A(x,y) = A(x,2y) + A(x,2y+1), \text{ for } x = 0, \dots, W-1, y = 0, \dots, \frac{H}{2}-1.$$

$$D_A(x,y) = A(x,2y) - A(x,2y+1), \text{ for } x = 0, \dots, W-1, y = 0, \dots, \frac{H}{2}-1.$$

$$E_A(x,y) = C_A(2x,y) + C_A(2x+1,y), \text{ for } x = 0, \dots, \frac{W}{2}-1, y = 0, \dots, \frac{H}{2}-1.$$

$$F_A(x,y) = C_A(2x,y) - C_A(2x+1,y), \text{ for } x = 0, \dots, \frac{W}{2}-1, y = 0, \dots, \frac{H}{2}-1.$$

$$Y_{420}^{\text{main}}(x,y) = Y(x,y), \text{ for } x = 0, \dots, W-1, y = 0, \dots, H-1.$$

$$U_{420}^{\text{main}}(x,y) = \frac{E_U(x,y)}{4}, \text{ for } x = 0, \dots, \frac{W}{2}-1, y = 0, \dots, \frac{H}{2}-1.$$

$$V_{420}^{\text{main}}(x,y) = \frac{E_V(x,y)}{4}, \text{ for } x = 0, \dots, \frac{W}{2}-1, y = 0, \dots, \frac{H}{2}-1.$$

$$Y_{420}^{\text{aux}}(x,y) = \frac{D_U(x,y)}{2} + 2^{B-1}, \text{ for } x = 0, \dots, W-1, y = 0, \dots, \frac{H}{2}-1.$$

$$Y_{420}^{\text{aux}}\left(x,\frac{H}{2}+y\right) = \frac{D_V(x,y)}{2} + 2^{B-1}, \text{ for } x = 0, \dots, W-1, y = 0, \dots, \frac{H}{2}-1.$$

$$U_{420}^{\text{aux}}(x,y) = \frac{F_U(x,2y)}{4} + 2^{B-1}, \text{ for } x = 0, \dots, \frac{W}{2}-1, y = 0, \dots, \frac{H}{4}-1.$$

$$U_{420}^{\text{aux}}\left(x,\frac{H}{4}+y\right) = \frac{F_V(x,2y)}{4} + 2^{B-1}, \text{ for } x = 0, \dots, \frac{W}{2}-1, y = 0, \dots, \frac{H}{4}-1.$$

$$V_{420}^{\text{aux}}(x,y) = \frac{F_U(x,2y+1)}{4} + 2^{B-1}, \text{ for } x = 0, \dots, \frac{W}{2}-1, y = 0, \dots, \frac{H}{4}-1.$$

$$V_{420}^{\text{aux}}\left(x,\frac{H}{4}+y\right) = \frac{F_V(x,2y+1)}{4} + 2^{B-1}, \text{ for } x = 0, \dots, \frac{W}{2}-1, y = 0, \dots, \frac{H}{4}-1.$$

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 U_{444} and V_{444} respectively, B4 and B5 are vertical high-pass signals, i.e. a vertical "H" band of U_{444} and V_{444}, respectively, B6 and B8 consist of even-numbered rows of the "LH" band of U_{444}, and B7 and B9 consist of odd-numbered rows of the "LH" band of V_{444}. That way, the decoder would apply the corresponding inverse wavelet operations after decoding the main and auxiliary frames to obtain U_{444} and V_{444} samples. Moreover, an additional vertical band separation can be performed, such that B6 and B8 are an "LHL" and "LHH" band of U_{444}, and B7 and B9 are an "LHL" and "LHH" band of V_{444}.

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 (U_{420}^{main} and V_{420}^{main}) sets the minimum level of quality for the U_{444} and V_{444} 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 U_{444} and V_{444} 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).

3.2.   Lifting-based Band separation filtering for the auxiliary frame

As discussed in Section 3.1, 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 U_{444} and V_{444} signals prior to the encoding process.

Mathematically, for an array A, where $A=U_{444}$ or V_{444}, define the following:

$C_A(x,y) = A(x,2y) - A(x,2y+1)$, for $x = 0, \ldots, W-1, y = 0, \ldots, \frac{H}{2}-1$.

$D_A(x,y) = A(x,2y+1) + C_A(x,y)/2$, for $x = 0, \ldots, W-1, y = 0, \ldots, \frac{H}{2}-1$.

$E_A(x,y) = D_A(2x,y) - D_A(2x+1,y)$, for $x = 0, \ldots, \frac{W}{2}-1, y = 0, \ldots, \frac{H}{2}-1$.

$F_A(x,y) = D_A(2x+1,y) + E_A(x,y)/2$, for $x = 0, \ldots, \frac{W}{2}-1, y = 0, \ldots, \frac{H}{2}-1$.

$Y_{420}^{main}(x,y) = Y(x,y)$, for $x = 0, \ldots, W-1, y = 0, \ldots, H-1$.

$U_{420}^{main}(x,y) = F_U(x,y)$, for $x = 0, \ldots, \frac{W}{2}-1, y = 0, \ldots, \frac{H}{2}-1$.

$V_{420}^{main}(x,y) = F_V(x,y)$, for $x = 0, \ldots, \frac{W}{2}-1, y = 0, \ldots, \frac{H}{2}-1$.

$Y_{420}^{aux}(x,y) = Clip(C_U(x,y) + 2^{B-1})$, for $x = 0, \ldots, W-1, y = 0, \ldots, \frac{H}{2}-1$.

$Y_{420}^{aux}\left(x,\frac{H}{2}+y\right) = Clip(C_V(x,y) + 2^{B-1})$, for $x = 0, \ldots, W-1, y = 0, \ldots, \frac{H}{2}-1$.

$U_{420}^{aux}(x,y) = Clip(E_U(x,2y) + 2^{B-1})$, for $x = 0, \ldots, \frac{W}{2}-1, y = 0, \ldots, \frac{H}{4}-1$.

$U_{420}^{aux}\left(x,\frac{H}{4}+y\right) = Clip(E_V(x,2y) + 2^{B-1})$, for $x = 0, \ldots, \frac{W}{2}-1, y = 0, \ldots, \frac{H}{4}-1$.

$V_{420}^{aux}(x,y) = Clip(E_U(x,2y+1) + 2^{B-1})$, for $x = 0, \ldots, \frac{W}{2}-1, y = 0, \ldots, \frac{H}{4}-1$.

$V_{420}^{aux}\left(x,\frac{H}{4}+y\right) = Clip(E_V(x,2y+1) + 2^{B-1})$, for $x = 0, \ldots, \frac{W}{2}-1, y = 0, \ldots, \frac{H}{4}-1$.

As with the band-separation approach described in Section 3.1, 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 U_{444} and V_{444} 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.

4. Experimental 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 [6]. We have since conducted similar tests using the HEVC HM 9.0 encoder [13] and the HEVC HM 10.0 encoder with 4:4:4: range extensions [14] using the Low Delay Main configuration.

Previously, we had 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 has a lower compression level than the auxiliary frame. Hence, it is more instructive to analyze the behavior of the different frame packing approaches.

In this paper, we evaluate the "direct", band-separation and lifting-based band-separation techniques for frame packing, and present results for the main and high tier QP ranges for above use-case. The test sequences are the Screen-Content and Range-Extension test-suite used in the range-extensions profile of the HEVC coding standard [14].

The test setup is as follows: 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. The decoder decodes the 4:2:0 frame, extracts the main and auxiliary views and reassembles the 4:4:4 frame for output (using $\alpha = \beta = \gamma = 1$ 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]") [14] on the 4:4:4 screen content and range extensions test sequences used in "common conditions" for experiments in the JCT-VC standardization committee**Error! Reference source not found.**. The BD-rate comparison result is shown in **Table 1** and **Table 2**, 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 Figure 4.

Table 1: 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	−44%	*RainFruits*	−82%
WordEditing	−47%	*BirdsInCage*	−97%
Programming	−56%	*LupoCandlelight*	−78%
Map	−61%	*Kimono*	−93%
Viking	−87%	*VenueVu*	−81
Robot	−87%	*CrowdRun*	−86%
TwistTunnel	−22%	*Traffic*	−87%
SlideShow	−57%		
PCBlayout	−36%		
PPT	−36%		
VideoConferencing	−39%		
Waveform	−40%		

Table 2: 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

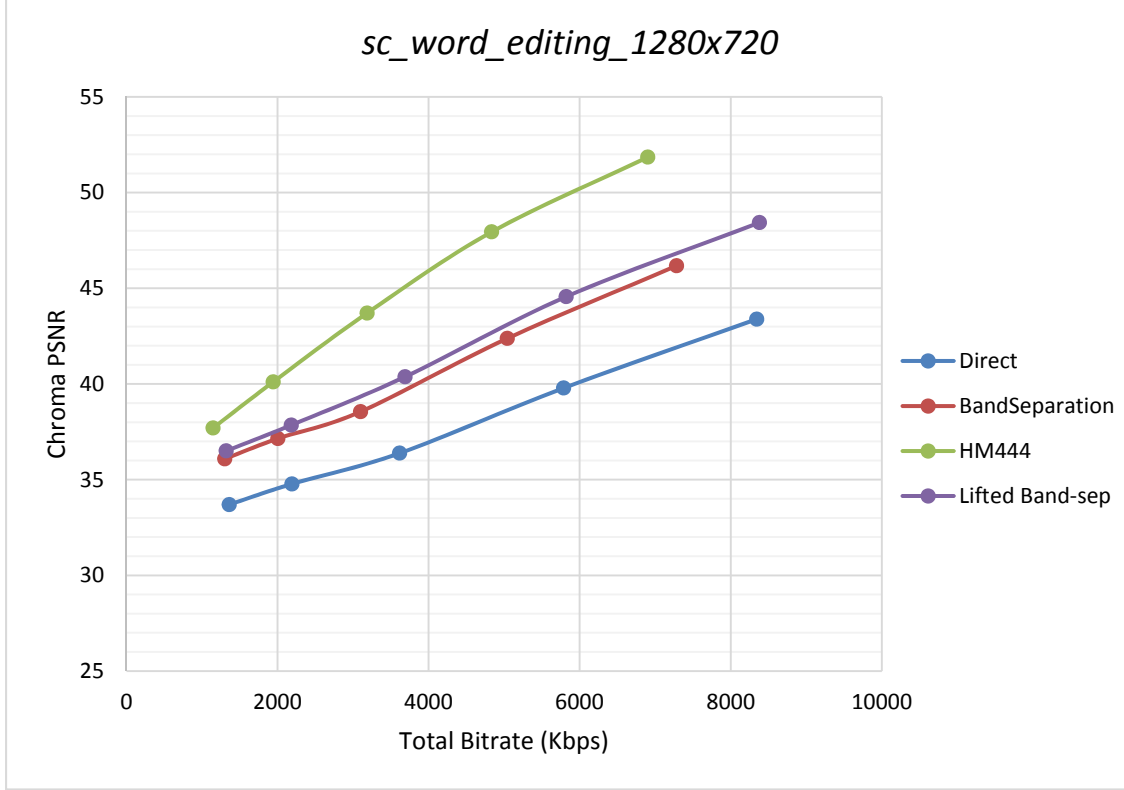| Test Sequences Screen Content	BD-rate delta	Test Sequences Range Extensions	BD-rate delta
WebBrowsing	−1%	*VenueVu*	−6%
WordEditing	−13%	*Rainfruits*	−3%
Programming	−6%	*LupoCandlelight*	−22%
Map	−8%	*BirdsinCage*	56%
Viking	−10%	*Kimono*	10%
Robot	−13%		
TwistTunnel	−14%		
SlideShow	−3%		
PCBlayout	6%		
PPT	−2%		
VideoConferencing	5%		
Waveform	7%		

Figure 3: 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.

In some cases, the BD-rate delta values have quite large magnitudes. That is partly due to the luma/chroma balance (chroma_qp_delta) for the video sequences, and also in part that we are mostly operating in the flat region of the BD-curve (adding more bits doesn't increase chroma-PSNR). Thus, adjusting the luma-chroma balance brings the chroma BD-rate closer, but results in much higher Luma-BDrate. This is further discussed in [8].

5. 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. The updated analysis presented here shows that frequency band separation and lifting-based band-separation are much more effective than was evident in the previous work on the topic
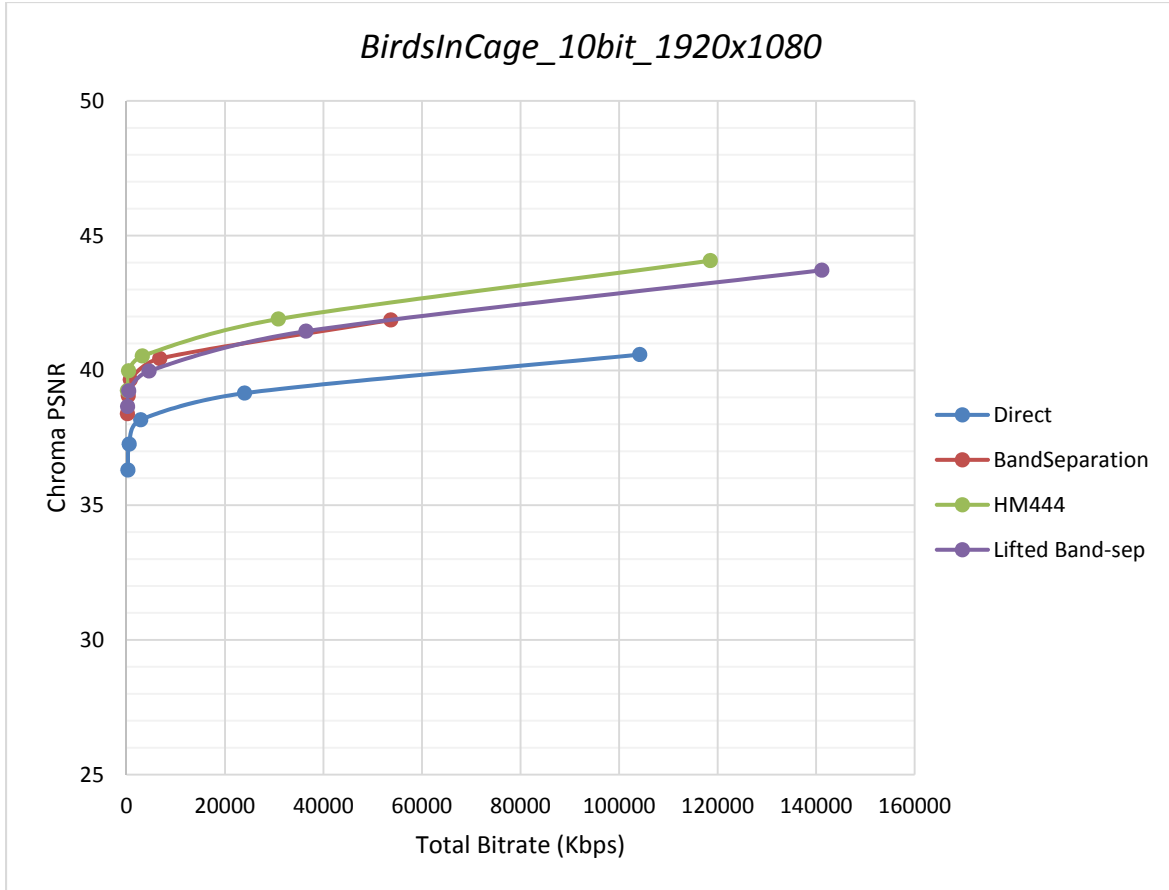
Figure 4: Rate-distortion plot of different packing schemes on *BirdsInCage* sequence in low delay mode. As can be seen from graphs, the graphs are already mostly "flat", hence adding a more bits doesn't result in a huge increase the PSNR. Thus the BDrate Delta values show huge differences for different schemes.

References

[1]   K. R. Rao and J. J. Hwang, *Techniques and Standards for Image, Video, and Audio Coding*, New Jersey: Prentice-Hall, 1996, Chapter 2.

[2]   T. Lin, P. Zhang, S. Wang, K. Zhou, and X. Chen, "Syntax and semantics of Dual-coder Mixed Chroma-sampling-rate (DMC) coding for 4:4:4 screen content", document JCTVC-J0233, 10th JCT-VC Meeting, Stockholm, Sweden, July 2012.

[3]   Microsoft Corporation, "Microsoft RemoteFX," Available at http://technet.microsoft.com/en-us/library/ff817578(WS.10).aspx, Feb. 2011.

[4]   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.*, Snowbird, UT, pp. 3–12, March 2013.

[5]   S. Reddy, S. Kanumuri, Y. Wu, S. Sadhwani, G. J. Sullivan, and H. S. Malvar, "Subband Decomposition for High-Resolution Color in HEVC and AVC 4:2:0 Video Coding Systems," *Poster, IEEE Data Compression Conf.*, Snowbird, UT, March 2014

[6]   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", document JCTVC-K0240, 11th JCT-VC meeting, Shanghai, China, Oct. 2012.

[7]   S. Reddy, S. Kanumuri, Y. Wu, S. Sadhwani, G. J. Sullivan, and H. S. Malvar, "Additional Experiments and Software for Frame Packing Arrangement SEI message for 4:4:4 Content in 4:2:0 Bitstreams," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, document JCTVC-O0198, 15th JCT-VC meeting, Geneva, Switzerland, Oct. 2013.

[8]   S. Reddy, S. Kanumuri, Y. Wu, S. Sadhwani, G. J. Sullivan, and H. S. Malvar, "Additional content interpretation type and experiments for frame packing arrangement SEI message for 4:4:4 content in 4:2:0 bitstreams", document JCTVC-P0216, 16th JCT-VC meeting, San Jose, USA, Jan. 2014.

[9]   ITU-T and ISO/IEC, *Advanced Video Coding for Generic Audiovisual Services*, Rec. ITU-T H.264 | ISO/IEC 14496-10, Jan. 2012.

[10]   B. Bross *et. al*, "High efficiency video coding (HEVC) Text Specification Draft 9", document JCTVC-P1003, 16th JCT-VC meeting, San Jose, USA, Jan. 2014.

[11]   D. Flynn, J. Sole and T. Suzuki, "High Efficiency Video Coding (HEVC) Range Extensions text specification: Draft 4", document JCTVC-P1005, 16th JCT-VC meeting: San Jose, USA, Jan. 2014

[12]   G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard", *IEEE Trans. on Circuits and Systems for Video Technology*, Dec. 2012.

[13]   "HEVC software repository", https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware.

[14]   "HEVC HM Range Extensions", http://hevc.kw.bbc.co.uk/trac/browser/branches/HM-range-extensions.