Joint Video Experts Team (JVET)
of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11
14th Meeting: Geneva, CH, 19–27 March 2019

Document: JVET-N0282-v3

Title:	**CE7-related: Joint chroma residual coding with multiple modes**
Status:	Input document to JVET
Purpose:	Proposal

Author(s) or Contact(s): Christian Helmrich, Christian Rudat, Tung Nguyen, Heiko Schwarz, Detlev Marpe, Thomas Wiegand

Fraunhofer HHI
Einsteinufer 37
10587 Berlin, Germany

Email: christian.helmrich@hhi.fraunhofer.de
christian.rudat@hhi.fraunhofer.de
tung.nguyen@hhi.fraunhofer.de
heiko.schwarz@hhi.fraunhofer.de
detlev.marpe@hhi.fraunhofer.de
thomas.wiegand@hhi.fraunhofer.de

Source: Fraunhofer HHI

Abstract

In this contribution, an extension of the joint chroma residual coding suggested in JVET-M0305 is proposed. In contrast to JVET-M0305, in which the addition of one joint chroma residual coding mode (given by Cr = −Cb) was suggested, this contribution proposes three modes for joint chroma residual coding with different mixing factors (given by Cr = ±Cb/2, Cr = ±Cb, Cb = ±Cr/2). The sign used for deriving the second chroma residual is coded in the tile group header. The usage of a joint chroma coding mode is indicated by a TU-level flag and the selected mode is implicitly indicated by the chroma coded block flags.

The following average results are reported for the common test conditions relative to VTM 4.0:

AI: −0.58% (Y), −0.76% (Cb), −1.13% (Cr) at 103% enc. and approximately 100% dec. time;
RA: −0.34% (Y), −2.53% (Cb), −1.74% (Cr) at 102% enc. and approximately 100% dec. time;
LD: −0.03% (Y), −1.47% (Cb), −4.95% (Cr) at 101% enc. and approximately 100% dec. time.

In addition, results for an alternative design of the chroma residual coding modes are reported. In this version, one of the three modes is replaced by a mode in which two chroma channels are coded that represent the transform coefficients of a Hadamard transform across the chroma components. For this version, the following average results are reported:

AI: −0.67% (Y), −0.67% (Cb), −1.12% (Cr) at 107% enc. and approximately 98% dec. time;
RA: −0.37% (Y), −2.28% (Cb), −1.68% (Cr) at 104% enc. and approximately 100% dec. time;
LD: −0.03% (Y), −1.02% (Cb), −4.34% (Cr) at 104% enc. and approximately 100% dec. time.

1 Introduction

In contribution JVET-M0305 [1], the addition of a joint chroma coding mode is proposed, in which only the Cb residual is coded and the Cr residual is derived according to Cr = −Cb. In order to improve coding efficiency for transform units in which the chroma residuals have different signal energies or different levels of cross-component correlation, we propose to support three different modes for joint chroma residual coding, which differ in the mixing factors for Cb and Cr. Similarly to JVET-M0305 [1], in all joint chroma coding modes, only one chroma transform block is transmitted, and the other block of chroma residuals is derived based on the samples of the coded chroma transform block.

2 Proposed Extension of Joint Chroma Residual Coding

It is proposed to support three modes for joint chroma residual coding. In all of these three joint chroma residual coding modes, a single chroma transform block is coded (using the residual coding of VTM 4 [2]) and the other block of chroma residual samples is derived using simple arithmetic operations.

The following three joint chroma coding modes are supported:

Mode 1: Cb is coded and Cr is derived according to Cr = CSign * Cb/2;

Mode 2: Cb is coded and Cr is derived according to Cr = CSign * Cb;

Mode 3: Cr is coded and Cb is derived according to Cb = CSign * Cr/2,

where CSign represents the sign used for deriving the second chroma residual block. CSign is indicated using a tile group header syntax element; it is either −1 or 1. Note that if CSign is equal to −1, mode 2 is the same as the joint chroma coding mode suggested in JVET-M0305 [1].

The usage of joint chroma residual coding is indicated by a TU-level flag tu_joint_chroma_residual_flag. This flag is present if either or both of the two chroma coded block flag (CBF) syntax elements are equal to 1. If tu_joint_chroma_residual_flag is equal to 1, one of the joint chroma residual coding modes is used. The mode used is indicated by the chroma CBFs, as specified in the following table:

tu_cbf_cb	tu_cbf_cr	joint chroma coding mode
1	0	mode 1
1	1	mode 2
0	1	mode 3

If a joint chroma coding mode is chosen, the QP for coding the joint chroma component is decreased by 1 (for modes 1 and 3) or 2 (for mode 2).

At the encoder side, the joint chroma residual is derived by a corresponding down-mixing of the Cb and Cr residuals. One of the three supported chroma coding modes is pre-selected based on a minimization of the mixing distortion (i.e., the distortion obtained by first down-mixing the Cb and Cr residuals and then reconstructing, or up-mixing, the Cb and Cr residuals from the joint chroma residual, without quantization). Only the pre-selected mode is tested as an additional mode in the mode decision process (i.e., using transform, quantization, and entropy coding). Due to the low-complexity pre-selection of one candidate mode for each TU, the encoding time is virtually not changed relative to JVET-M0305.

The tile group header syntax element that indicates the sign (CSign) for deriving the second chroma component is determined by analyzing the correlation between high-pass filtered versions of the original Cb and Cr components for the tile group.

Further details are provided in the following subsections. The suggested normative changes are specified relative to VTM 4 [2].

2.1 Syntax and Semantics

The proposed changes for syntax and semantics are summarized in the following. All changes are marked in red.

7.3.4.1 General tile group header syntax

tile_group_header( ) {	**Descriptor**
…	
if(ChromaArrayType != 0)	
joint_cb_cr_sign_flag	u(1)
…	
}	

7.3.4.1 General tile group header semantics

…

joint_cb_cr_sign_flag specifies whether, in joint chroma residual coding, the signs for the residual samples of the second chroma component are inverted. When joint chroma residual coding is used for a transform unit (as indicated by the transform unit syntax element tu_joint_chroma_residual_flag[ ][ ]), joint_cb_cr_sign_flag equal to 0 specifies that the signs for the residual samples of the second chroma component are identical to the signs of the corresponding residual samples of the first chroma component and joint_cb_cr_sign_flag equal to 1 specifies that the signs for the residual samples of the second chroma component are inverted.

…

7.3.7.10 Transform unit syntax

transform_unit(x0, y0, tbWidth, tbHeight, treeType, subTuIndex) {	**Descriptor**
…	
if((treeType = = SINGLE_TREE || treeType = = DUAL_TREE_CHROMA) {	
if((IntraSubPartitionsSplitType = = ISP_NO_SPLIT && !(cu_sbt_flag && ((subTuIndex = = 0 && cu_sbt_pos_flag) || (subTuIndex = = 1 && !cu_sbt_pos_flag)))) || (IntraSubPartitionsSplitType != ISP_NO_SPLIT && (subTuIndex = = NumIntraSubPartitions − 1))) {	
tu_cbf_cb[x0][y0]	ae(v)
tu_cbf_cr[x0][y0]	ae(v)
}	
}	
if(tu_cbf_cb[x0][y0] || tu_cbf_cr[x0][y0])	
tu_joint_chroma_residual_flag[x0][y0]	ae(v)
…	
if(tu_cbf_cb[x0][y0] || tu_joint_chroma_residual_flag[x0][y0])	
residual_coding(xC, yC, Log2(wC), Log2(hC), 1)	
if(tu_cbf_cr[x0][y0] && !tu_joint_chroma_residual_flag[x0][y0])	
residual_coding(xC, yC, Log2(wC), Log2(hC), 2)	
}	

7.4.7.10 Transform unit semantics

…

tu_joint_chroma_residual_flag[x0][y0] specifies whether the residual samples for both chroma components Cb and Cr are coded as a single transform block. The array indices x0, y0 specify the location (x0, y0) of the top-left luma sample of the considered transform unit relative to the top-left luma sample of the picture.

tu_joint_chroma_residual_flag[x0][y0] equal to 1 specifies that the transform unit syntax includes the transform coefficient levels for a single transform block from which the residual samples for both Cb and Cr are derived. tu_joint_chroma_residual_flag[x0][y0] equal to 0 specifies that the transform coefficient levels of either or both chroma components are coded as indicated by the syntax elements tu_cbf_cb[x0][y0] and tu_cbf_cr[x0][y0]. When tu_joint_chroma_residual_flag[x0][y0] is not present, it is inferred to be equal to 0.

Depending on tu_joint_chroma_residual_flag[x0][y0], tu_cbf_cb[x0][y0], and tu_cbf_cr[x0][y0], the variable TuCResMode[x0][y0] is derived as follows:

– If tu_joint_chroma_residual_flag[x0][y0] is equal to 0, the variable TuCResMode[x0][y0] is set equal to 0;

– Otherwise, if tu_cbf_cb[x0][y0] is equal to 1 and tu_cbf_cr[x0][y0] is equal to 0, the variable TuCResMode[x0][y0] is set equal to 1;

– Otherwise, if tu_cbf_cb[x0][y0] is equal to 1, the variable TuCResMode[x0][y0] is set equal to 2;

– Otherwise, the variable TuCResMode[x0][y0] is set equal to 3.

…

2.2 *Decoding Process*

The proposed changes for the decoding process are summarized in the following. All changes are marked in red.

8.7.2 Scaling and transformation process

…

The variable cCodedIdx is derived as follows:

– If cIdx is not equal to 0 and TuCResMode[xTbY][yTbY] is not equal to 0, cCodedIdx is set equal to 1;

– Otherwise, cCodedIdx is set equal to cIdx.

The (nTbW)x(nTbH) array of residual samples resSamples is derived as follows:

1. The scaling process for transform coefficients as specified in clause 8.7.3 is invoked with the transform block location (xTbY, yTbY), the transform block width nTbW and the transform block height nTbH, the colour component variable cCodedIdx and the bit depth of the current colour component bitDepth as inputs, and the output is an (nTbW)x(nTbH) array of scaled transform coefficients d.

2. The (nTbW)x(nTbH) array of residual samples r is derived as follows:

   – If transform_skip_flag[xTbY][yTbY] is equal to 1 and cIdx is equal to 0, the residual sample array values r[x][y] with x = 0..nTbW − 1, y = 0..nTbH − 1 are derived as follows:

     r[x][y] = d[x][y] << tsShift

   – Otherwise (transform_skip_flag[xTbY][yTbY] is equal to 0 or and cIdx is not equal to 0), the transformation process for scaled transform coefficients as specified in clause 8.7.4.1 is invoked with the transform block location (xTbY, yTbY), the transform block width nTbW and the transform block height nTbH, the colour component variable cCodedIdx and the (nTbW)x(nTbH) array of scaled transform coefficients d as inputs, and the output is an (nTbW)x(nTbH) array of residual samples r.

3. The intermediate samples res[x][y] with x = 0..nTbW − 1, y = 0..nTbH − 1 are derived as follows:

   res[x][y] = (r[x][y] + (1 << (bdShift − 1))) >> bdShift

4. The residual samples resSamples[x][y] with x = 0..nTbW − 1, y = 0..nTbH − 1 are derived as follows:
   – If TuCResMode[xTbY][yTbY] is equal to 2 and cIdx is equal to 2,

     resSamples[x][y] = (joint_cb_cr_sign_flag ? −1 : 1) * res[x][y]

   – Otherwise, if (TuCResMode[xTbY][yTbY] is equal to 1 and cIdx is equal to 2) or (TuCResMode[xTbY][yTbY] is equal to 3 and cIdx is equal to 1),

     resSamples[x][y] = ((joint_cb_cr_sign_flag ? −1 : 1) * res[x][y] + 1) >> 1

   – Otherwise,

     resSamples[x][y] = res[x][y]

8.7.3 Scaling process for transform coefficients

…

The variable dQPC is derived as follows:

– If TuCResMode[xTbY][yTbY] is equal to 2, dQPC is set equal to 2;

– Otherwise, if TuCResMode[xTbY][yTbY] is equal to 1 or 3, dQPC is set equal to 1;

– Otherwise, dQPC is set equal to 0.

The quantization parameter qP is derived as follows:

– If cIdx is equal to 0, the following applies:

$$qP = Qp'_{Y}$$

– Otherwise, if cIdx is equal to 1, the following applies:

$$qP = Qp'_{Cb} - dQPC$$

– Otherwise (cIdx is equal to 2), the following applies:

$$qP = Qp'_{Cr} - dQPC$$

…

2.3 Entropy Coding

The binary syntax element tu_joint_chroma_residual_flag[x0][y0] is coded using CABAC. One of three context models are used. The context index ctIdx indicating the context model used is derived according to

ctIdx = (2 * tu_cbf_cb[x0][y0] + tu_cbf_cr[x0][y0] – 1).

2.4 *Encoding*

The tile group header syntax element that indicates the sign for deriving the second chroma component is determined as follows. The original chroma components Cb and Cr for the tile group are high-pass filtered using the FIR filter

$$\begin{pmatrix} -1 & -2 & -1 \\ -2 & 12 & -2 \\ -1 & -2 & -1 \end{pmatrix}.$$

Given the high-pass filtered components, the correlation is determined. The sign of the correlation factor is used as sign CSign for deriving the second chroma component.

On a TU level, the encoding proceeds as follows. First, the three down-mix components C_1, C_2, and C_3 (corresponding to joint coding modes 1, 2, and 3, respectively) are determined. Then, for each of the joint coding modes, the mixing distortion is derived as follows:

$$D_1 = \sum_k (Cb[k] - C_1[k])^2 + (Cr[k] - 0.5 * CSign * C_1[k])^2$$
$$D_2 = \sum_k (Cb[k] - C_2[k])^2 + (Cr[k] - CSign * C_2[k])^2$$
$$D_3 = \sum_k (Cr[k] - C_3[k])^2 + (Cb[k] - 0.5 * CSign * C_3[k])^2$$

The mode k with minimum distortion D_k is selected as candidate mode for the TU. Finally, the conventional chroma coding mode (with separate Cb and Cr components) and the selected candidate mode are tested using conventional Lagrangian mode decision and the mode minimizing the rate-distortion cost is selected.

3 Experimental Results

The proposed joint chroma coding was tested on top of VTM 4.0 [3] according to the JVET common test conditions [4] (CTC QPs) and the low-QP configuration requested by CE7 (incl. restriction to 100 coded frames). The average coding results are given in the following tables. Detailed results are provided in the accompanying Excel documents. Note that the decoding runtime (**dec. time**) for all configurations lies between 98% and 101% and, due to short execution times, could not be measured more accurately.

	All Intra					Random Access			
Class	**Gain Y**	**Gain Cb**	**Gain Cr**	**Enc. time**		**Gain Y**	**Gain Cb**	**Gain Cr**	**Enc. time**
A1	–0.71%	–0.32%	2.72%	103%		–0.48%	**–3.73%**	1.38%	103%
A2	–1.24%	–2.47%	1.77%	106%		–0.69%	**–5.12%**	0.54%	102%
B	–0.39%	0.57%	–1.70%	103%		–0.16%	–1.02%	**–3.28%**	102%
C	–0.39%	–1.42%	**–3.03%**	104%		–0.18%	–1.59%	**–3.85%**	102%
E	–0.34%	–0.85%	**–4.39%**	102%					
Overall	**–0.58%**	**–0.76%**	**–1.13%**	103%		**–0.34%**	**–2.53%**	**–1.74%**	102%
D	–0.36%	–1.58%	–2.11%	105%		–0.15%	–1.99%	**–3.08%**	102%
F	–0.44%	–2.57%	**–4.38%**	103%		–0.46%	–2.42%	**–4.37%**	103%

Table 1: Bjøntegaard delta-rate results of proposal relative to VTM 4.0 [3], SDR common test conditions.

	All Intra				Random Access			
Class	**Gain Y**	**Gain Cb**	**Gain Cr**	**Enc. time**	**Gain Y**	**Gain Cb**	**Gain Cr**	**Enc. time**
A1	–0.04%	0.06%	0.12%	109%	?	?	?	?
A2	–0.10%	0.08%	0.07%	112%	?	?	?	?
B	–0.06%	0.06%	–0.15%	111%	–0.10%	0.22%	–0.42%	106%
C	–0.27%	–0.17%	–0.96%	110%	–0.18%	–0.38%	–2.21%	104%
E	–0.09%	0.17%	–0.27%	108%				
Overall	**–0.12%**	**0.03%**	**–0.27%**	110%	?	?	?	?
D	–0.41%	–0.37%	–0.94%	109%	–0.23%	–1.22%	–2.61%	103%
F	–0.77%	–0.60%	–1.40%	105%	–0.62%	–0.79%	–2.31%	103%

Table 2: Bjøntegaard delta-rate results of proposal relative to VTM 4.0 [3], **low QPs** (100 coded frames).

4 Results for an Alternative Configuration

In the following, we present experimental results for an alternative configuration which realizes a slightly different gain-runtime tradeoff. The main difference to the method described above is that one of the three joint coding modes is replaced with a chroma residual coding mode in which two chroma components C_1 and C_2 are coded, and the decoded Cb and Cr residuals are derived by applying an inverse 2x2 Hadamard transform (without a decoder-side scaling by $\frac{1}{\sqrt{2}}$ in order to minimize the algorithmic complexity).

4.1 Description

A new mode (mode 4) replaces one of the three joint coding modes. If this chroma residual coding mode is selected on a TU level, the reconstructed residual components are derived as follows:

1. The residual blocks C_1 and C_2 are decoded as if they represent conventional Cb and Cr components, respectively. That means, the transform coefficient levels are entropy-decoded and the scaling and inverse transformation are applied.
2. The residual samples for the Cb and Cr components of the TU are derived according to

   $Cb[\,x\,][\,y\,] = C_1[\,x\,][\,y\,] + C_2[\,x\,][\,y\,]$,

   $Cr[\,x\,][\,y\,] = C_1[\,x\,][\,y\,] - C_2[\,x\,][\,y\,]$.

Similarly as in the proposed approach described above, a TU-level flag indicates whether any of the three alternative chroma residual coding modes is used. This TU-level flag is present when either or both chroma CBFs are equal to 1 and, if any of these modes is used, the selected mode is indicated by the chroma CBFs as specified in the following table:

tu_cbf_cb	tu_cbf_cr	joint chroma coding mode
1	0	mode 2 (Cr = ±Cb)
1	1	mode 4 (Hadamard)
0	1	mode 1 or mode 3

Whether mode 1 or mode 3 is used for the last case (tu_cbf_cb = = 0 and tu_cbf_cr = = 1) is indicated by an additional tile group header flag.

4.2 *Experimental Results*

The alternative configuration of the proposed joint chroma coding was also tested on top of VTM 4.0 [3] according to the JVET common test conditions [4] as well as the CE7 low-QP configuration (base QPs 2, 7, 12, 17, limited to 100 coded frames). The average coding results are given in the following tables. Again, detailed results are provided in the accompanying Excel documents, and the decoding runtime (**dec. time**) for all configurations lies between 98% and 101% and could not be measured more accurately.

	All Intra				Random Access			
Class	**Gain Y**	**Gain Cb**	**Gain Cr**	**Enc. time**	**Gain Y**	**Gain Cb**	**Gain Cr**	**Enc. time**
A1	–0.72%	–0.95%	2.64%	107%	–0.46%	**–4.16%**	1.79%	105%
A2	–1.66%	–2.59%	1.65%	111%	–0.82%	**–5.27%**	0.46%	104%
B	–0.42%	0.77%	–1.70%	106%	–0.18%	–0.50%	**–3.38%**	103%
C	–0.44%	–1.03%	–2.85%	108%	–0.20%	–0.84%	**–3.75%**	104%
E	–0.40%	–0.39%	**–4.40%**	104%				
Overall	**–0.67%**	**–0.67%**	**–1.12%**	107%	**–0.37%**	**–2.28%**	**–1.68%**	104%
D	–0.42%	–1.05%	–2.00%	109%	–0.09%	–1.19%	–2.69%	103%
F	–0.49%	–2.39%	**–4.18%**	106%	–0.58%	–1.93%	**–4.23%**	105%

Table 3: Bjøntegaard delta-rate results of alternative proposal relative to VTM 4.0, SDR test conditions.

	All Intra				Random Access			
Class	**Gain Y**	**Gain Cb**	**Gain Cr**	**Enc. time**	**Gain Y**	**Gain Cb**	**Gain Cr**	**Enc. time**
A1	–0.06%	0.16%	0.15%	117%	?	?	?	?
A2	–0.40%	0.25%	0.13%	123%	?	?	?	?
B	–0.13%	0.18%	–0.23%	123%	–0.16%	0.38%	–0.57%	111%
C	–0.48%	0.11%	–0.94%	121%	–0.31%	–0.19%	–1.86%	109%
E	–0.22%	0.56%	–0.33%	116%				
Overall	**–0.26%**	**0.24%**	**–0.28%**	120%	**?**	**?**	**?**	?
D	–0.60%	–0.16%	–1.17%	120%	–0.35%	–1.02%	–2.31%	107%
F	–1.09%	–0.64%	–1.59%	113%	–0.87%	–0.76%	–2.19%	108%

Table 4: Bjøntegaard delta-rate results of alternative proposal relative to VTM 4.0, **low QPs** (100 frames).

5 References

[1] J. Lainema, "CE7-related: Joint coding of chrominance residuals," Joint Video Experts Team, doc. JVET-M0305, Marrakech, MA, Jan. 2019.
[2] B. Bross, J. Chen, S. Liu (editors), "Versatile Video Coding (Draft 4)," Joint Video Experts Team, doc. JVET-M1001, Marrakech, MA, Jan. 2019.
[3] JVET, "VTM version 4.0," https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/tags/VTM-4.0.
[4] F. Bossen, J. Boyce, X. Li, V. Seregin, K. Sühring, "JVET common test conditions and software reference configurations for SDR video," Joint Video Experts Team, doc. JVET-M1010, Marrakech, MA, Jan. 2019.

6 Appendix: Visual Analysis

Subjective analysis of the proposal was conducted as a supplement to the objective experiments described in sections 3 and 4.2. Two visual experiments were run, one in the context of normal encoder operation and one “visual encoder stress-test” similar to the one illustrated in the CE7-related proposal JVET-N0347.

6.1 Test Employing Normal Encoder Operation

A number of randomly selected sequences were encoded (in random-access configuration at very high and low QPs) and decoded with and without the proposed joint chroma residual coding as well as the described alternative to the proposal. VTM 4.0 was used as the underlying software version, meaning that the version without the proposal(s) is identical to the unchanged VTM 4.0 reference software.

Subjective inspection of the decoded sequences did not reveal any visual quality degradations for any tested sequence or base QP. Moreover, we were under the impression that for some sequences with very high BD-rate gains due to the proposal, such as *CatRobot1* and *BQTerrace* (about 5–6% average chroma-channel gain), a slight visual quality increase due to the proposal could be observed at or above QP37. However, no controlled subjective assessment was performed, and the quality increase may be statistically insignificant.

6.2 Stress-Test Using Forced Encoder Decisions

To assess the perceptual impact of false encoder search-loop decisions with respect to joint chroma residual coding, the proposed encoder was modified to prefer the “best” (in terms of rate-distortion cost) of the given selectable joint chroma coding modes (see sections 2 and 4.1) over conventional separate Cb and Cr coding, even if the latter provided a lower rate-distortion cost. The rest of the encoder search-loop operated as usual.

The figure below depicts the resulting visual artifacts on the first frame of the *RaceHorses* input at a QP of 22, as in N0347. It can be observed that, in the CE7-1-like reference of Fig. 1(b), some chromatic distortion appears at the top-left picture corner and the right edges of the red “X” on the right-hand jockey’s shirt. It is worth noting that these artifacts are much less obvious than those reported in N0347. Our proposal shown in Fig. 1(c) removes the artifacts almost entirely and its alternative in Fig. 1(d) does not exhibit any artifacts.

(a) input sequence (416×240 samples, 8-bit, 4:2:0)

(b) mode = 2, CSign = –1, as in JVET-M0305 [1]

(c) proposed single-channel CE7 extension (sec. 2)

(d) alternative to proposed CE7 extension (sec. 4)

7 Patent rights declaration(s)

Fraunhofer HHI 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).