Joint Collaborative Team on Video Coding (JCT-VC)
of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 Document: JCTVC-O0043
15th Meeting: Geneva, CH, 23 October – 1 November 2013

Title:	Best-effort decoding of 10-bit sequences: use-cases, requirements and specification methods
Status:	Input Document to JCT-VC
Purpose:	Proposal

Author(s):	David Flynn Gaëlle Martin-Cocher Dake He	*dflynn@blackberry.com* *gmartincocher@blackberry.com* *dhe@blackberry.com*
Source:	BlackBerry Ltd.	

Abstract

HEVC, like other well-known codecs, contains profiles that accommodate the decoding of non-8-bit video sequences. In particular, HEVC contains a 10-bit profile that permits the carriage of 10-bit video sequences. In M0255, it was demonstrated that it is possible to re-purpose an existing 8-bit decoder design to decode 10-bit bitstreams. Moreover, through choice of rounding method, it is possible to reduce the accumulation of decoder error than might otherwise be expected.

This contribution summarises the requirements basis, use case information, the previous work, and provides additional information regarding the decoding of 12-bit video sequences with a 10-bit decoder. Specification text is proposed to allow such techniques to be applied in the context of HEVC.

1 Use-cases and requirements

The guidance document [1] providing the vision, applications and requirements for the development of HEVC suggests that “complexity scalability in the encoder and decoder is desirable”.

Currently the scope of decoder complexity scalability is extremely limited – there exist two profiles with identical tool sets with the only difference being the bit-depth at which processing occurs. In effect, the system scalability exists wholly with the encoder.

Adding a best-effort decoding solution that permits the decoding of Main-10 bitstreams with the resources of an 8-bit decoder would go some way to providing some balance to the receiving application.

In [2], use cases are described for maintaining compatibility between devices for user generated content, better managing the user experience when confronted with complex streams, generating “browse quality” representations of higher quality masters, and hybrid transmission schemes that permit the migration from a lower bitdepth to a higher bitdepth.

One notable effect of best-effort decoding techniques defined to apply to previously defined profiles is the departure of bit-exact decoding. Referring back to the guidance document, we do not believe that this is currently precluded, since, “the full decoding process should be specified, preferably with no mismatch”.

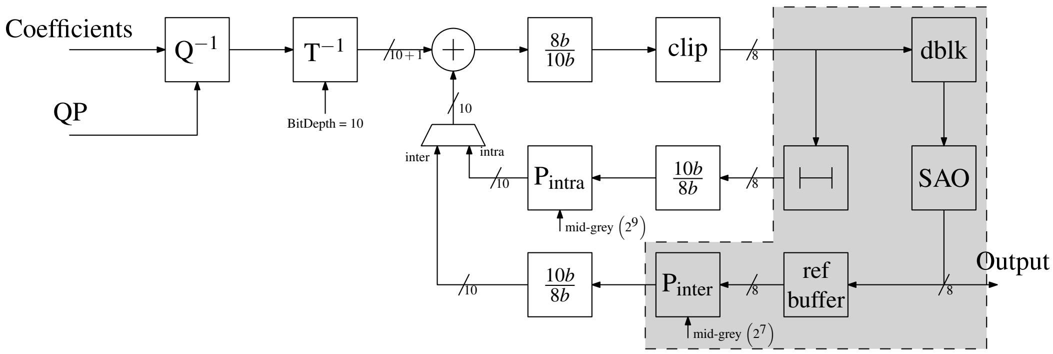
2 Summary of previous work using rounding control

The contribution JCTVC-M0255 [3] provides two methods for decoding a 10-bit sequence using a modified 8-bit decoder. The first method is a straight forward adjustment that re-interprets the system bitdepth in the transform and subsequent stages to be 8-bit. The only extra burden that the decoder must handle is to decode any signalled SAO offsets using a truncated unary code configured for 10-bit operation. This method, while straightforward, is characterized by DC drift in intra pictures that accumulates towards the bottom right of the picture, and by drift in colour and/or saturation that evolves over many frames.

The second method proposed (see Figure 1) does not modify the transform behaviour, rather, it introduces a rounding stage after the residual is combined with the prediction and a consequential bit-depth increase during the prediction process. The rounding stage is implemented using round-to-even to maintain the mean expected signal value and thereby reduce the drift accumulation through the removal of biases.

This design is particularly effective, even for 8-bit software decoders, since many calculations in the transform, prediction and combining processes must be calculated at a bit-depth higher than 8-bit. Ie, capabilities of the current design are exploited.

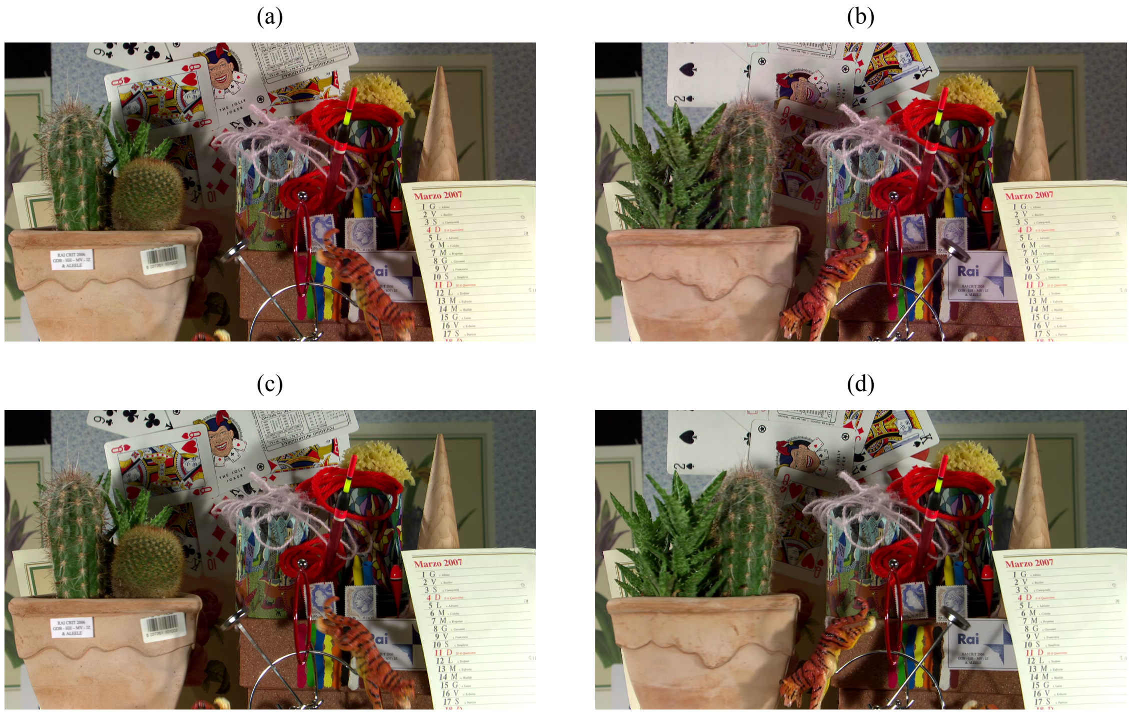
Figure 1 – Illustrative signal coding path of hybrid 8-bit–10-bit decoder



To examine the modified decoder behaviour in unfavourable conditions, the HM-10 low-delay-B main-10 sequences [4] were decoded and the PSNR measured against the original input sequences. Table 1 illustrates the method 2 per-sequence losses when compared against the PSNR of the normal decoder[1]. Figure 3 shows how the PSNR loss evolves as the sequence progresses. Visual inspection of the decoded video shows that the gross DC drifts observed using the first method are avoided, to the extent that under single stimulus conditions it can be difficult to observe a degradation.

Figure 2 illustrates the typical drift effect observed when using the second method compared to the vanilla HM-10 decoder after decoding the ten-second 500-frame sequence Cactus encoded using the low-delay main-10 configuration at QP=27. The figure shows the first intra and the last inter frame of the sequence. Video sequences using both methods are available at the meeting for viewing.

Figure 2 – Typical distortion introduced by 8-bit decoding of 10-bit sequences. Sequence cactus QP=27, first frame (left), last frame (right). (a,b) Method–2, (c,d) HM-10.



[1]The Cactus sequence used in figure 2 is marked in black

Figure 3 – Plot of frame-number versus ΔPSNR for all sequences at particular QPs using Method–2.

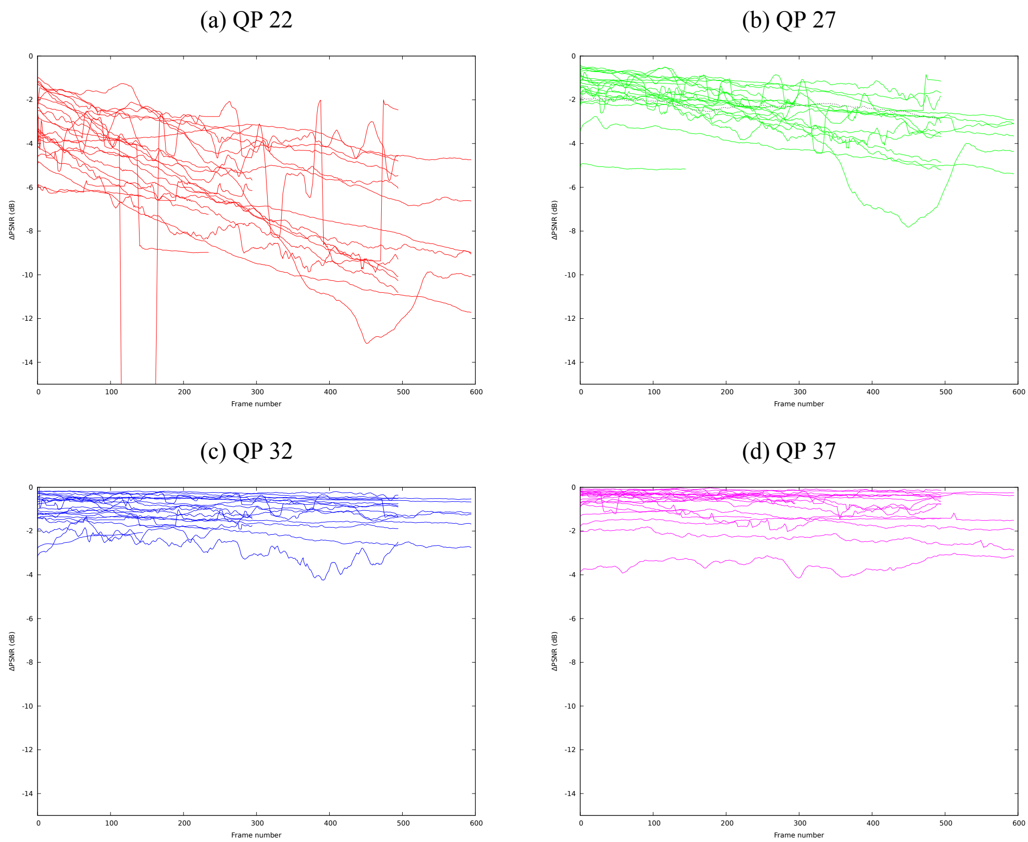


Table 1 – PSNR loss associated with method–2 of JCTVC-M0255, averaged over 10s sequences at QPs 22/27/32/37 using low-delay main-10 configuration

Sequence	Luma PSNR	Chroma PSNR
BQMall	−7.1/−3.5/−1.1/−0.6	−3.7/−2.7/−1.5/−1.2
BQSquare	−3.9/−1.7/−0.3/−0.2	−4.7/−2.0/−0.8/−0.6
BQTerrace	−6.3/−2.6/−0.8/−1.3	−4.2/−3.2/−1.7/−1.0
BasketballDrillText	−7.2/−2.8/−0.6/−0.4	−5.4/−2.9/−1.0/−0.4
BasketballDrill	−6.5/−3.1/−0.5/−0.5	−5.3/−4.1/−1.5/−0.3
BasketballDrive	−3.9/−2.3/−1.7/−0.7	−5.8/−2.6/−2.3/−0.9
BasketballPass	−4.0/−1.3/−0.3/−0.2	−2.8/−1.0/−0.4/−0.3
BlowingBubbles	−3.5/−1.3/−0.5/−0.3	−4.6/−1.7/−1.4/−0.8
Cactus	−6.2/−2.3/−1.5/−1.5	−5.1/−4.1/−1.5/−0.8
ChinaSpeed	−7.7/−3.0/−2.8/−0.6	−8.5/−4.0/−2.7/−1.2
FourPeople	−8.9/−4.2/−1.5/−3.5	−4.3/−4.3/−2.3/−0.8
Johnny	−7.2/−2.4/−2.1/−1.4	−7.8/−4.1/−2.3/−1.2
Kimono1	−5.6/−1.7/−1.4/−1.1	−2.7/−1.8/−2.6/−1.0
KristenAndSara	−5.4/−2.6/−0.9/−2.3	−7.8/−3.4/−4.0/−1.4
ParkScene	−5.6/−2.0/−1.0/−0.5	−4.6/−2.8/−0.8/−0.7
PartyScene	−3.3/−2.9/−0.6/−0.4	−3.4/−1.7/−1.0/−0.9
PeopleOnStreet	−3.6/−1.2/−1.2/−0.7	−4.4/−3.2/−3.5/−1.3
RaceHorses	−3.9/−1.4/−0.8/−0.3	−5.1/−0.9/−0.5/−0.7
RaceHorsesC	−2.4/−1.3/−0.8/−0.4	−4.2/−2.9/−1.9/−0.7
SlideEditing	−6.1/−2.0/−1.5/−0.3	−1.9/−0.7/−0.3/−0.2
SlideShow	−6.5/−1.9/−0.6/−0.3	−4.4/−2.4/−1.1/−0.5
Traffic	−6.2/−5.1/−2.3/−1.1	−4.5/−3.0/−2.4/−0.8

3 Further results using 12 bit source video

The current range extensions test set contains the 12-bit Traffic sequence in RGB. Additional decoding results are presented for the case when the 12-bit sequence is decoded using method 2 configured to decode at 10-bit against native 12-bit decoding. Figure 4 illustrates the losses associated in this process. All PSNR measurements are performed using 12-bit data (lower-bitdepth data is padded to 12-bit prior to measurement).

Since many display systems are limited to 8-bit (or worse), an additional set of relative PSNR results are presented in Figure 5 are obtained where all video data is truncated to 8-bit prior to calculating the PSNR.

Figure 4 – Plot of frame-number versus ΔPSNR measured at 12-bit for Traffic_2560x1600_30_12bit_444_-crop.rgb using Method–2.

(a) ld_main_rext (b) i_main_rext

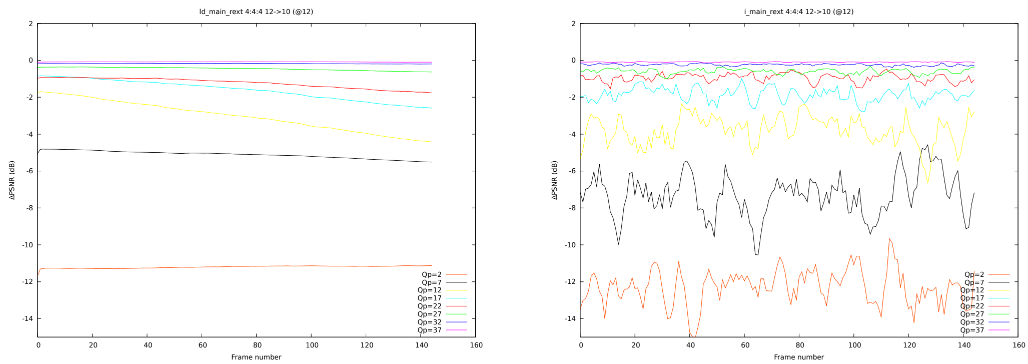
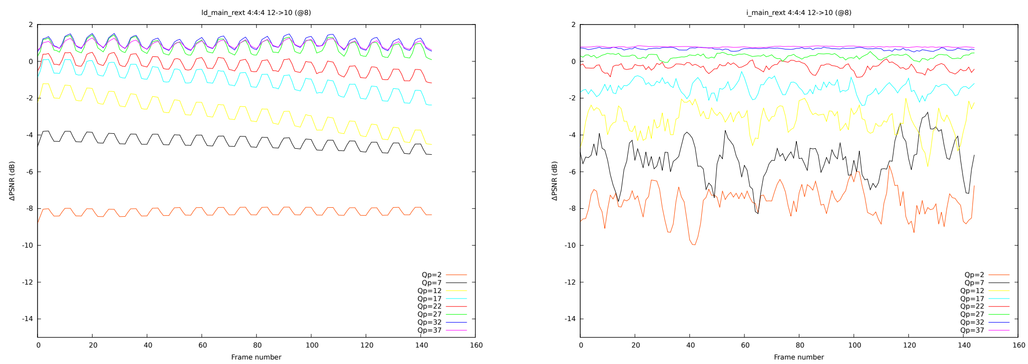


Figure 5 – Plot of frame-number versus ΔPSNR measured at 8-bit for Traffic_2560x1600_30_12bit_444_-crop.rgb using Method–2.

(a) ld_main_rext (b) i_main_rext



4 Proposed specification text

To permit the application of such best-effort techniques in the specification text, the definition of conformance in the general decoding process is extended.

Figure 6 – Modified specification text

8.1 General decoding process

…

The decoding process is specified such that all decoders will produce numerically identical cropped decoded pictures. Any decoding process that produces identical cropped decoded pictures to those produced by the process described herein (with the correct output order or output timing, as specified) conforms to the decoding process requirements of this Specification. For the purpose of best effort decoding, a decoder conforming to one profile at a given tier and level that, and without claiming conformance to, decodes a bitstream conforming to a different tier, level or profile without using a decoding process that produces numerically identical cropped decoded pictures to those produced by the process described herein conforms to the decoding process requirement of this Specification.

BlackBerry Limited 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).

References

[1] "Vision, applications and requirements for high efficiency video coding (hevc)," W11872, ISO/IEC JCT1/SC29/WG11, Jan. 2011.

[2] D. Flynn, G. Martin-Cocher, and D. He, "[ahg21] best-effort decoding of 10-bit sequences: use cases, requirements and specification methods," JCTVC-N0291, JCT-VC, Jul. 2013.

[3] ——, "Decoding a 10-bit hevc sequence using an 8-bit decoder," JCTVC-M0255, JCT-VC, Apr. 2013.

[4] F. Bossen, "Common hm test conditions and software reference configurations," JCTVC-L1100, JCT-VC, Jan. 2013.