Joint Video Experts Team (JVET)
of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11
12th Meeting: Macao, CN, 3–12 Oct. 2018

Document: JVET-L0523-v2

| | | | |
|---|---|---|---|
| *Title:* | **CE11-related: Very strong deblocking filtering with conditional activation signaling** | | |
| *Status:* | Input document to JVET | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Christian Helmrich<br>Benjamin Bross | Email: | christian.helmrich@hhi.fraunhofer.de<br>benjamin.bross@hhi.fraunhofer.de |
| *Source:* | Fraunhofer HHI, Einsteinufer 37, 10587 Berlin, Germany | | |

# Abstract

Improved luma and chroma deblocking for the Versatile Video Coding specification is studied extensively in Core Experiment (CE) 11, described in JVET-L0031. While these designs generally succeed in increasing the video reconstruction quality for many SDR sequences, we noticed that, for some SDR and, particularly, HDR material, even more aggressive deblocking filters (using even longer filter kernels) may be beneficial. Usage of such very strong filtering, however, increases the risk of adverse artifacts in some image regions.

As a solution to the issue noted above, this contribution proposes the use of a simple very strong deblocking algorithm, whose activation can be controlled by the encoder. To minimize the signaling overhead required for this “superstrong” deblocking mode, a conditional transmission of the activation flag is employed on a CTU-wise level based on the partitioning and residual coding information already present in the bit-stream.

Due to the fact that an activation flag is not signaled in every CTU, the Bjøntegaard delta (BD) rate increase caused by said flag (a 1-bit element) is less than 0.04 % on average. In CTUs for which no flag is transmitted, very strong deblocking is disabled, and the deblocking resorts to the conventional filtering as in, e. g., CE11. Given the achievable subjective benefit, as exemplified in this document, and the absence of any significant BD-PSNR degradations, also as reported here (AI: 0.04 %), we recommend to study the proposed conditional activation together with (very) strong deblocking filters towards the next ($13^{th}$) JVET meeting in 2019.

# 1 Introduction to Deblocking In-Loop Filtering

Contemporary perceptual (i. e., lossy) block transform image and video codecs can achieve very good visual reconstruction quality even at relatively low bit-rates. At very low rates, however, artifacts such as blurring or discontinuities around the block boundaries, also known as “blocking”, appear. To reduce these typically annoying artifacts, deblocking post-processing methods are used in modern codecs such as HEVC or VVC. In video coding, a typical deblocking post-processor operates as an *in-loop* filter on each decoded image or frame, i. e., on each inter-picture prediction (also called motion compensation) source image or frame before coding of the next image/frame in the coding loop. The deblocking post-filter analyzes the boundary pixel values of each reconstructed sub-block of the decoded image in terms of potential discontinuities. If a weak discontinuity is detected, it is assumed to be caused by the low-rate coding itself and not to be part of the original image. Hence, this discontinuity is minimized by means of smoothing of the boundary pixel values (e. g., the addition of adaptive pixel value offsets). A detailed discussion of this process is published in [1].

In VVC, the maximum transform block size has doubled compared to the size allowed in HEVC [2], which was found to necessitate the use of stronger deblocking filters (i. e., deblocking post-processors modifying a wider range of pixels) especially around the larger-block boundaries, as studied in CE 11 [3]. Such more aggressive deblocking, however, increases the risk of smoothing out – and thus, potentially erasing – original image content which was *not* caused by the low-rate coding. An example is provided in Fig. 1 (c) where, undesirably, some parts of the thin clouds are largely removed by strong deblocking filtering (see the arrow). At the same time, the bright part of the sky around the sun is not fully smoothened, not even by the strong deblocking, so blocking artifacts remain. The bit-rate in this example is a reasonable 4.83 Mbit/s on average.

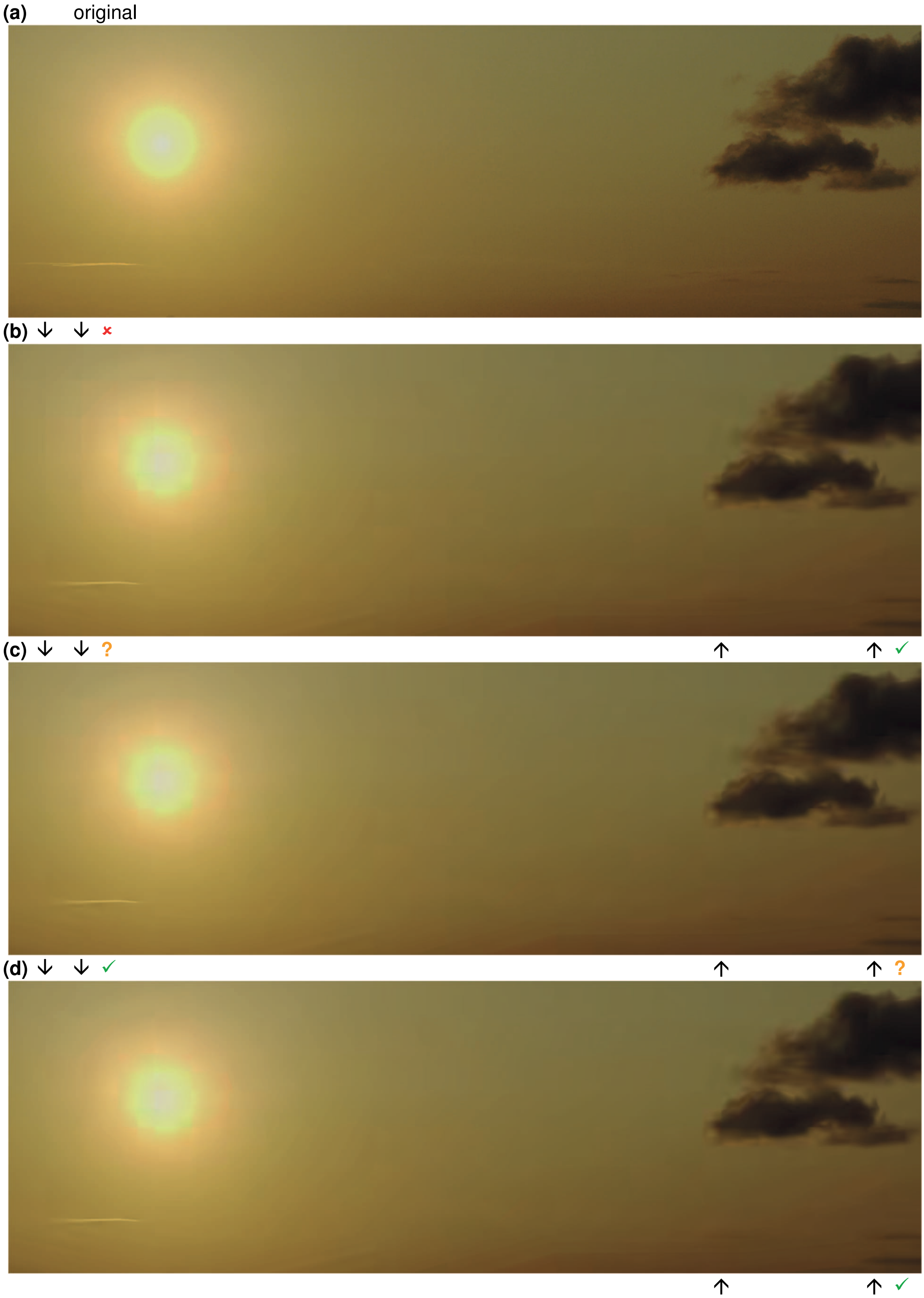


*Fig. 1. Effect of deblocking on first frame of SunsetBeach sequence. (a) Input image, (b) HEVC deblocking only, with bugfixes as in VTM 2.0.1, (c) additional strong deblocking equivalent to CE 11, (d) our proposal. Arrows mark the locations of visible advantages and disadvantages of the deblocking methods.* $QP_{base} = 37$.

We, thus, conclude that very strong deblocking filtering is desirable for some low-rate coded high-resolution image and video content and that it is essential to allow for highly selective control of the application of the very strong deblocking filtering. Naturally, a bit-flag may be signaled for each sub-block (e. g., each coding tree unit, CTU) to indicate to the receiver (i. e., decoder) whether to allow the application of a "superstrong" deblocking filter. This approach, however, would lead to many additional signaling bits being included in the bit-stream, thereby increasing the coding bit-rate to unacceptable levels especially at very low bit-rates.

# 2 Description of the Proposed Conditional Deblocking

In view of the abovementioned issues and the need for a more efficient solution, the following is proposed:

1. selective signaling of an in-loop filter control parameter (FCP) per CTU to *allow* the application of very strong deblocking filtering; this FCP shall only be signaled in a CTU if that CTU is partitioned into fewer than *A* sub-units (TUs) and if residual coefficient coding is used in at least *B* of the TUs.
2. use of a very strong deblocking design, with a filter kernel size of 16 taps, for large transform units. This "superstrong" deblocking is *disallowed* in CTUs for which no FCP is signaled according to 1.

With regard to the particular example of Fig. 1, this solution results in the decoded image shown in Fig. 1 (d). The following two subsections provide a more detailed explanation of the two aspects of the proposal.

## *2.1 Selective Signaling of Filter Control Parameter*

In HEVC and VVC, a CTU can be subject to sub-partitioning into multiple square or rectangular sub-blocks (called CUs). Examples are depicted in Fig. 2, where Fig. 2 (a) also implies the absence of sub-partitioning.

CTUs with low visual activity, i. e., few image details [4], are usually not sub-partitioned or sub-partitioned into only a few relatively large CUs. Moreover, for these low-activity CTUs, the spatial intra-picture prediction (and temporal inter-picture prediction, if applicable) typically operates very efficiently. As a result, the prediction residual sub-units (TUs) in said CTUs comprise very little signal energy and, thus, can often be fully quantized to zero. In doing so, these TUs are exempt from transmission, which is indicated by a coded block flag (CBF) value of zero in the bit-stream. Sometimes, however, at least one CU in such a low-activity CTU exhibits relatively high signal variance in its prediction residual, thus requiring the transmission of at least one (coarsely) quantized TU which is not fully zero and likely to cause visible blocking upon decoding.

We observed that, in particular, CTUs sub-partitioned into fewer than $A = 9$ CUs with, at the same time, the non-zero (i. e., $\text{CBF} = 1$) coding and transmission of at least $B = 1$ TUs in the CTU, benefit most from very strong deblocking filtering. Examples for such CTU segmentations, including rectangular CUs, are shown in Figs. 2 (b)–(e). Note that, at both the encoder and decoder side, the case of partitioning into fewer than *A* CUs can be identified by way of the CTU's coding tree, whereas the presence of at least *B* non-zero TUs can be detected by counting the number of TUs in the CTU for which $\text{CBF} = 1$. Given that, for every CTU, the coding tree as well as the CBF data are available in both the encoder and decoder, it can be stated that

**Condition 1**: very strong deblocking filtering shall be *allowed* in a CTU if

- the signaled coding tree indicates a partitioning of said CTU into fewer that *A* CUs and
- the number of non-zero quantized TUs with $\text{CBF} = 1$ signaled in said CTU is at least *B*.

In other words, if condition 1 is not met, very strong deblocking shall be *disallowed* and shall, therefore, be *disabled* for each CU/TU in the considered CTU at both the encoder and decoder side, i. e., a conventional deblocking solution such as one of those in CE 11 [3] is applied in that CTU. If, however, condition 1 is met in a CTU, very strong deblocking is allowed, but this does not necessarily mean that it is also *enabled*. In fact, as discussed above, it is highly desirable to provide a means for realizing highly selective control over the application of "superstrong" deblocking filtering. Our proposed approach to provide such a means is to

**Condition 2**: signal a deblocking filter control parameter (FCP) in a CTU if

- condition 1 evaluates to true in said CTU; otherwise, assume $\text{FCP} = 0$.

In other words, if condition 1 is not met, no FCP is signaled for the considered CTU. If, however, condition 1 is met, a FCP is written to, and read from, the bit-stream at the end of the CTU (de)coding process. A more detailed description of the envisioned change to the VVC syntax specification will be provided in Annex A.

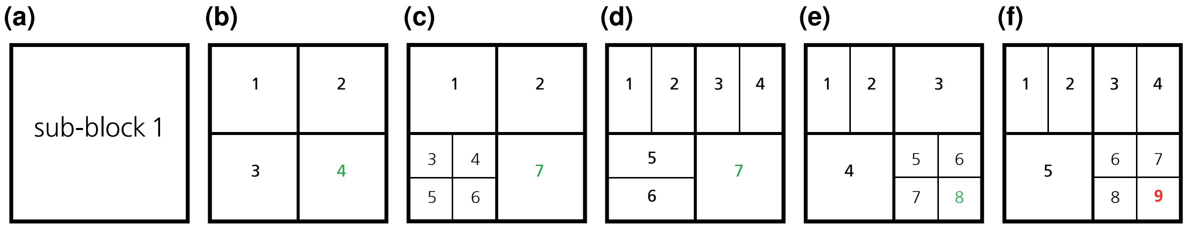


***Fig. 2***. *Examples of sub-block partitioning of a CTU into CUs. (a) Single sub-block, (b–c) square sub-blocks (quad-tree), (d–f) generalized rectangular sub-blocks (QTBT). Condition A < 9 is met in (a)–(e) but not (f).*

## *2.2 Application of Very Strong Deblocking Filtering*

If the FCP is present in the bit-stream (i. e., condition 1 is met) for a given CTU, then the FCP value determines whether the decoder is to enable the very strong deblocking mode (value 1) or to disable it (value 0).

If FCP = 0 in a CTU, traditional deblocking with decoder-side filter strength detection on the reconstructed image component [1], [3] is applied for each TU sub-area of the CTU. If FCP = 1 in a CTU, the very strong deblocking filter described below is used on each TU whose width and height both equal **32** or more pixels. It can be extended by a decoder-side filter strength detection. Note that the specific details of the superstrong deblocking filter do not appear to be critical and that the proposed filter could be replaced by a more efficient third-party solution, as long as the filter remains strong enough (i. e., employs between 8 and 16 filter taps).

The first step in our superstrong deblocking algorithm is the derivation of left, right, top, and bottom boundary offsets for each TU satisfying the above size constraint. Specifically, given $d = \{\text{left}, \text{right}, \text{top}, \text{bottom}\}$,

$$\text{offset}_d = \text{Clip3}\,(-127, 127, (P_d - Q_d + N_d) >> (\log_2(N_d) + 1)\,),$$

where $P_d$ and $Q_d$ are the sums of the outer and inner boundary reconstructed samples of the TU, respectively (excluding the outer four corner samples), along the direction $d$ of length $N_d$ (i. e., the TU width resp. height). Unavailable $P_d$ sums at slice or image borders are replaced by the adjacent $Q_d$ sums. Then, upon deblocking in case of FCP = 1 in a CTU (see also Section 2.3), weighted additions of $\text{offset}_d$ are performed, for each $d$, along the **16** inner boundary sample columns or rows perpendicular to $d$. This adds a linear ramp with slope $\text{offset}_d \div \mathbf{16}$ towards the $d$ TU boundary, reaching $\text{offset}_d$ at the boundary, with blend-overs at the TU corners.

In our VTM software modification, the very strong deblocking is executed before the traditional deblocking on the luma as well as chroma channels (note that, for 4:2:0 chroma, the three **bold** values above are halved).

## *2.3 Encoder-Side Selection of Optimal FCP per CTU*

One major advantage of the present proposal is that the encoder can signal the desired application mode for the strong deblocking by way of the FCP for a specific CTU if condition 1 evaluates to true. Therefore, the encoder can test which mode, FCP = 0 or FCP = 1, leads to better objective and/or subjective coding results in a CTU. We opted for the following straightforward encoder-side R/D search for the optimal set of FCPs:

1. In each frame, run the complete search for an R/D optimal partitioning, prediction, and quantization
2. After this, evaluate condition 1 in CTU coding order (horizontally, then vertically across the picture)
   a. If, in a CTU, condition 1 evaluates to true, determine left/right/top/bottom boundary offsets
3. After this, evaluate condition 1 in CTU coding order again, or reuse the evaluation results of step 2
   a. If, in a CTU, condition 1 evaluates to true, save its reconstructed picture region from step 1
   b. Apply the very strong deblocking filtering of Subsection 2.2 in-place on the CTU's reconstructed picture region, using the left/right/top/bottom boundary offsets obtained in step 2a
   c. If the distortion (i. e., reconstruction error) increased by more than $x$ dB due to step 3b, then set FCP = 0 for the given CTU and restore the reconstructed picture region saved in step 3a
   d. Otherwise, set FCP =1 for the given CTU, keep the picture reconstruction result of step 3b
   e. If, in a CTU, condition 1 evaluates to false, don't execute steps 3a–3d and assume FCP = 0
4. While writing the bit-stream in CTU coding order, if condition 1 is true, write the FCP for that CTU.

We tried $x = 0.0$ and 0.1. Note that, in step 3, other distortion measures are possible but not considered here.

# 3 Experimental Results

Apart from the HDR hybrid log gamma (HLG) sequence evaluated visually as shown in Fig. 1, we measured Bjøntegaard delta (BD) PSNR gains on the SDR-category Common Test Conditions (CTC) set of sequences [5], [6] to verify the absence of significant reductions in coding efficiency due to the present proposal. The underlying software version used for this investigation is VTM 2.0.1 with default settings and enabled ALF.

Table 1 lists the BD-PSNR results for $x = 0$ dB (a more objective encoder tuning) while Table 2 contains the BD-PSNR values for $x = 0.1$ dB (a more subjective encoder tuning, see Subsection 2.3). It can be noted that

- both choices of $x$ lead to very similar overall BD-PSNR numbers,
- no significant BD-PSNR losses are observed for either choice of $x$.

In other words, significant coding losses were successfully avoided while achieving the desired visual gains, especially for $x = 0.1$ dB. Demo bit-streams and YUVs can be made available during the 12th JVET meeting. Note, further, that the encoding run-time does not increase measurably and that the decoding run-time overhead is very moderate. Details on the additional algorithmic decoder complexity can be provided on request.

***Table 1**. BD-PSNR [6] data for VTM 2.0.1 with vs. without proposed conditionally signaled deblocking and $x = \mathbf{0.0}$dB for the SDR category CTC [5]. An Excel sheet with detailed results accompanies this contribution.*

| **All Intra** | **Gain Y** (%) | **Gain Cb** (%) | **Gain Cr** (%) | **Time Enc.** (%) | **Time Dec.** (%) |
|---|---|---|---|---|---|
| Class A1 | 0.04 | **–0.25** | –0.11 | 98 | 101 |
| Class A2 | 0.01 | –0.04 | –0.02 | 99 | 101 |
| Class B | 0.01 | –0.07 | –0.09 | 99 | 101 |
| Class C | 0.00 | 0.00 | –0.01 | 100 | 100 |
| Class E | 0.01 | –0.12 | –0.09 | 99 | 101 |
| **Overall** | 0.01 | –0.09 | –0.06 | 99 | 101 |
| **Random Access** | **Gain Y** (%) | **Gain Cb** (%) | **Gain Cr** (%) | **Time Enc.** (%) | **Time Dec.** (%) |
| Class A1 | 0.05 | 0.06 | 0.10 | 99 | 102 |
| Class A2 | 0.04 | –0.06 | –0.02 | 99 | 100 |
| Class B | 0.01 | 0.01 | 0.04 | 99 | 104 |
| Class C | 0.03 | 0.04 | –0.01 | 100 | 100 |
| Class E | | | | | |
| **Overall** | 0.03 | 0.01 | 0.03 | 99 | 102 |

***Table 2**. BD-PSNR [6] data for VTM 2.0.1 with vs. without proposed conditionally signaled deblocking and $x = \mathbf{0.1}$dB for the SDR category CTC [5]. An Excel sheet with detailed results accompanies this contribution.*

| **All Intra** | **Gain Y** (%) | **Gain Cb** (%) | **Gain Cr** (%) | **Time Enc.** (%) | **Time Dec.** (%) |
|---|---|---|---|---|---|
| Class A1 | 0.13 | –0.08 | **0.23** | 99 | 99 |
| Class A2 | 0.03 | 0.10 | 0.11 | 100 | 100 |
| Class B | 0.02 | 0.02 | 0.01 | 99 | 102 |
| Class C | 0.00 | 0.03 | 0.02 | 100 | 100 |
| Class E | 0.03 | –0.05 | –0.04 | 100 | 101 |
| **Overall** | 0.04 | 0.01 | 0.06 | 100 | 100 |
| **Random Access** | **Gain Y** (%) | **Gain Cb** (%) | **Gain Cr** (%) | **Time Enc.** (%) | **Time Dec.** (%) |
| Class A1 | **0.48** | 0.14 | 0.18 | 100 | 101 |
| Class A2 | **0.30** | 0.07 | –0.01 | 100 | 99 |
| Class B | 0.15 | 0.02 | –0.08 | 100 | 101 |
| Class C | 0.03 | 0.11 | 0.03 | 99 | 102 |
| Class E | | | | | |
| **Overall** | **0.21** | 0.08 | 0.02 | 100 | 101 |

# 4 Summary and Conclusion

This contribution proposed a conditionally signaled very strong deblocking mode to address shortcomings in VVC's current deblocking performance, namely, a lack of more aggressive deblocking filtering necessary for certain input video sequences (or still images) and a need for highly efficient encoder-side control over the application of strong deblocking in a picture region such as a CTU. The proposal was shown to increase the visual reconstruction quality especially on HDR content for some input, while Bjøntegaard delta PSNR statistics indicate that the algorithmic extension comes at no significant efficiency loss or run-time increase.

In light of the outcome of this investigation, the authors recommend to study the proposed conditional activation signaling together with (very) strong deblocking filtering towards the 13$^{th}$ JVET meeting in January.

# 5 References


[1] A. Norkin *et al.*, "HEVC Deblocking Filter," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 22, Dec. 2012.

[2] B. Bross, J. Chen, S. Liu, "Versatile Video Coding (Draft 2)," JVET-K1001, ver. 6/7, Sep./Oct. 2018.

[3] A. Norkin, A. M. Kotra, "CE11: Summary report on deblocking," JVET-L0031, Macao/CN, Oct. 2018.

[4] C. Helmrich, H. Schwarz, D. Marpe, and T. Wiegand, "Improved Perceptually Optimized QP Adaptation and Associated Distortion Measure," JVET-K0206, Ljubljana, Slovenia, July 2018.

[5] F. Bossen, J. Boyce, K. Sühring, X. Li, V. Seregin, "JVET Common Test Conditions and Software Reference Configurations for SDR Video," JVET-K1010, Ljubljana, Slovenia, July 2018.

[6] G. Bjøntegaard, "Calculation of Average PSNR Differences between RD-Curves," VCEG-M33, Austin, USA, Mar. 2001.


# 6 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).**