|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  10th Meeting: Stockholm, SE, 11 – 20 July 2012 | Document: JCTVC-J0107 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **AHG9: On RAP pictures** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Ye-Kui Wang Ying Chen Rajan L. Joshi Adarsh K. Ramasubramonian  5775 Morehouse Drive San Diego, CA 92121 USA | Tel: Email: | 1-858-651-8345 [yekuiw@qualcomm.com](mailto:yekuiw@qualcomm.com)  1-858-845-6589 [cheny@qualcomm.com](mailto:cheny@qualcomm.com)  1-858-658-4511 [rajanj@qualcomm.com](mailto:rajanj@qualcomm.com)  1-858-658-5804 [aramasub@qualcomm.com](mailto:aramasub@qualcomm.com) |
| *Source:* | Qualcomm Incorporated | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

This document includes the following proposals related to RAP pictures (i.e., IDR, CRA and BLA pictures):

1. To include the support for handling a CRA picture as a BLA picture based on an indication through external means
2. To enable prediction from decodable leading pictures (DLPs, i.e. non-TFD leading pictures) associated with a RAP picture by normal pictures associated with the same RAP picture, and by leading pictures associated with the next RAP picture (wherein leading pictures associated with a RAP picture are those pictures following the RAP picture in decoding order but preceding the RAP picture in output order, and normal pictures associated with a RAP picture are those pictures following a RAP picture in both decoding order and output order and preceding, in decoding order, the next RAP picture)
3. To fix the definition of RAP picture
4. To mandate the activation of VPS, SPS, PPS and APS at each BLA picture
5. To include a constraint to disallow output-order interleaving of DLP pictures with TFD pictures or pictures earlier than the same associated CRA or BLA picture in decoding order, and a constraint to disallow decoding-order interleaving of TFD pictures and normal pictures associated with a RAP picture
6. To fix two bugs related to the inference of no\_output\_of\_prior\_pics\_flag equal to 1
7. To use one more NAL unit type to differentiate TFD & TLA pictures and non-TLA TFD pictures

# Introduction

## Random access

Random access refers to a decoding of a video bitstream starting from a coded picture that is not the first coded picture in the bitstream. Random access to a bitstream is needed in many video applications, such as broadcasting and streaming, e.g., for users to tune-in to a program anytime, to switch between different channels, to jump to specific parts of the video, or to switching to a different bitstream for stream adaptation (of the bit rate, frame rate, spatial resolution, and so on). This feature is enabled by inserting random access pictures or random access points, many times in regular intervals, into the video bitstream.

## Bitstream splicing

Bitstream splicing refers to the concatenation of two or more bitstreams or parts thereof. For example, a first bitstream may be appended by a second bitstream, possibly with some modifications to either one or both of the bitstreams to generate a spliced bitstream. The first coded picture in the second bitstream is also referred to as the splicing point. Therefore, pictures since the splicing point in the spliced bitstream were originated from the second bitstream while pictures preceding the splicing point in the spliced bitstream were originated from the first bitstream.

Splicing of bitstreams is performed by bitstream splicers. Bitstream splicers are often lightweight and much less intelligent than encoders. For example, they may not be equipped with entropy decoding and encoding capabilities.

Bitstream switching may be used in adaptive streaming environments. A bitstream switching operation at certain picture in the switch-to bitstream is effectively a bitstream splicing operation wherein the splicing point is the bitstream switching point, i.e., the first picture from the switch-to bitstream.

## Random access point (RAP) pictures

Instantaneous decoding refresh (IDR) pictures as specified in AVC or HEVC can be used for random access. However, since pictures following an IDR picture in decoding order cannot use pictures decoded prior to the IDR picture as reference, bitstreams relying on IDR pictures for random access can have significantly lower coding efficiency.

To improve coding efficiency, the concept of clean random access (CRA) pictures was introduced in HEVC to allow pictures that follow a CRA picture in decoding order but precede it in output order to use pictures decoded before the CRA picture as reference. Pictures that follow a CRA picture in decoding order but precede the CRA picture in output order are referred to as leading pictures associated with the CRA picture (or leading pictures of the CRA picture). The leading pictures of a CRA picture are correctly decodable if the decoding starts from an IDR or CRA picture before the current CRA picture. However, the leading pictures of a CRA picture may be non-decodable when random access from the CRA picture occurs; hence the leading pictures are typically discarded during random access decoding. To prevent error propagation from reference pictures that may not be available depending on where the decoding starts, all pictures that follow a CRA picture both in decoding order and output order shall not use any picture that precedes the CRA picture either in decoding order or output order (which includes the leading pictures) as reference.

The concept of broken link access (BLA) picture was further introduced in HEVC after the introduction of CRA pictures and is based on the concept of CRA pictures. A BLA picture typically originates from bitstream splicing at the position of a CRA picture, and in the spliced bitstream the splicing point CRA picture is changed to a BLA picture.

IDR pictures, CRA pictures and BLA pictures are collectively referred to as random access point (RAP) pictures.

The most essential difference between BLA pictures and CRA pictures is as follows. For a CRA picture, the associated leading pictures are correctly decodable if the decoding starts from a RAP picture before the CRA picture in decoding order, and may be non-correctly-decodable when random access from the CRA picture occurs (i.e., when the decoding starts from the CRA picture, or in other words, when the CRA picture is the first picture in the bitstream). For a BLA picture, the associated leading pictures may be non-decodable in all cases, even when the decoding starts from a RAP picture before the BLA picture in decoding order.

For a particular CRA or BLA picture, some of the associated leading pictures are correctly decodable even when the CRA or BLA picture is the first picture in the bitstream. These leading pictures are referred to as decodable leading pictures (DLPs), and other leading pictures are referred to as non-decodable leading pictures (NLPs). NLPs are also referred to as tagged for discard (TFD) pictures in the latest HEVC draft specification.

# Handling of a CRA picture as a BLA picture

## Problem

In streaming adaptation based on CRA pictures, changing of a CRA picture to a BLA picture typically needs to be performed by a media server or an intermediate network element, e.g. a media-aware network element (MANE) or even a media-unaware network element such as an HTTP cache or web proxy, which is typically preferable to be lightweight and may not be able to change the bitstream at all.

## Proposal

Instead of relying on a server or an intermediate network element to change a BLA picture to a CRA picture, a server or an intermediate network element may generate a message to be sent to the decoder side (i.e. the client), notifying that a bitstream switching operation has occurred at certain CRA picture and that CRA picture should be handled as a BLA picture. In the context of dynamic adaptive streaming over HTTP (DASH), the decoder side may also infer such a message by itself through the change of the URL it used for requesting stream data and the reception of the media data associated with the changed URL.

To work with the above approach, the HEVC specification needs to be changed, such that a CRA picture is handled as a BLA picture when it is indicated so by an external means. Such an external indication can be passed to the decoder, by a function of the decoder side, through inference or reception from a server or an intermediate network element.

The proposed decoding process change is as follows.

The variable HandleCraAsBlaFlag is associated with each CRA picture. When the value of HandleCraAsBlaFlag for a particular CRA picture is not specified by external means, it is set to 0.

When decoding (including parsing) each coded slice NAL unit, if HandleCraAsBlaFlag is equal to 1 and nal\_unit\_type is equal to 4 or 5, the following applies:

1. The value of nal\_unit\_type is increased by 2.
2. The value of no\_output\_of\_prior\_pics\_flag is set to 1.
3. If the previous picture in decoding order is a RAP picture and the rap\_pic\_id of the current slice is equal to the rap\_pic\_id of the previous picture, the following applies:
   * If the next picture in decoding order is not a RAP picture, the value of rap\_pic\_id of the current slice is changed to be different than the rap\_pic\_id of the previous picture in decoding order.
   * Otherwise (the next picture in decoding order is a RAP picture), the value of rap\_pic\_id of the current picture is changed to be a value that is different than the rap\_pic\_id of both the previous picture and the next picture in decoding order.

# Prediction from decodable leading pictures and a bug fix

Currently, it is disallowed to have inter prediction from decodable leading pictures (DLPs, i.e. non-TFD leading pictures) associated with a CRA or BLA picture by normal pictures associated with the same RAP picture or by leading pictures associated with the next CRA or BLA picture. However, with the current BLA or CRA definition, the DLP pictures can be correctly decoded when the decoding starts from the associated BLA or CRA picture (e.g. in random access or streaming switching) or from the previous RAP picture (e.g. in normal playback). Therefore, it is not necessary to disallow the DLP pictures to be used to predict the normal pictures.

Note that leading pictures associated with an IDR picture are all DLP pictures, which can be used to predict the normal pictures.

Therefore, it is proposed to allow for inter prediction of the normal pictures from the DLP pictures associated with the same BLA or CRA pictures.

This way, the behavior of DLP pictures is unified among all types of RAP pictures.

Similarly, it is also proposed to allow for inter prediction from DLP pictures associated with a CRA or BLA picture by leading pictures associated with the next CRA or BLA picture.

The required text changes are as follows (additions are highlighted, removals are strikethrough):

**8.3.2 Decoding process for reference picture set**

**...**

It is a requirement of bitstream conformance that the reference picture set is restricted as follows:

* There shall be no reference picture with temporal\_id greater than that of the current picture included in RefPicSetStCurrBefore, RefPicSetStCurrAfter ~~and~~ or RefPicSetLtCurr.
* When the current picture is a TLA picture, there shall be no reference picture included in the reference picture set with temporal\_id greater than or equal to the temporal\_id of the current picture.
* Let prevRapPic be the previous RAP picture in decoding order, if any. Let prevRapPic2 be the second previous RAP picture in decoding order, if any.
  + If the current picture is a RAP picture, ~~T~~there shall be no reference picture included in the reference picture set that precedes prevRapPic in decoding order.
  + Otherwise if the current picture precedes prevRapPic in output order, there shall be no reference picture included in the reference picture set that precedes prevRapPic2 in decoding order or that are TFD pictures associated with prevRapPic2.
  + Otherwise (the current picture is not RAP picture and follows prevRapPic in output order), there shall be no reference picture included in the reference picture set that precedes prevRapPic in decoding order or that are TFD pictures associated with prevRapPic.
* Unless either of the following conditions is true, there shall be no entry in RefPicSetStCurrBefore, RefPicSetStCurrAfter or RefPicSetLtCurr that is equal to "no reference picture": a) the first coded picture in the bitstream is a CRA picture and the current coded picture is a TFD picture associated with the first coded picture in the bitstream; b) the previous RAP picture preceding the current coded picture in decoding order is a BLA picture and the current coded picture is a TFD picture associated with the BLA picture.

NOTE 5 – A reference picture cannot be included in more than one of the five reference picture set lists. [Ed. (YK): Why is this expressed as a note, not as a normative requirement?]

# Definition of RAP picture

The definition of RAP picture is as follows:

**random access point (RAP) picture**: A coded *picture* containing only *I slices* and for which each slice has nal\_unit\_type in the range of 4 to 8, inclusive; a RAP picture may be a *BLA picture*, a *CRA picture* or an *IDR picture*; all *coded pictures* that follow the RAP picture both in *decoding order* and *output order* do not use *inter prediction* from any *picture* that precedes the RAP picture either in *decoding order* or *output order*; and any *picture* that precedes the RAP picture in *decoding order* also precedes the RAP picture in *output order*.

However, a normal picture (following in both decoding order and output order) associated with an IDR picture may be predicted from any of the IDR picture and all the pictures between the IDR picture and the normal picture in decoding order, including the leading pictures (all the leading pictures associated with an IDR picture are DLP pictures, following in decoding order but preceding in output order).

Similarly, if the proposal piece in Section 3 is adopted to allow for inter prediction from decodable leading pictures (DLPs, i.e. non-TFD leading pictures) associated with a CRA or BLA picture by normal pictures associated with the same RAP picture, the current definition of RAP picture disallows the same case for CRA and BLA pictures.

To solve the IDR picture issue, the following change to the definition of RAP picture is proposed (just remove the four strikethrough words):

**random access point (RAP) picture**: A coded *picture* containing only *I slices* and for which each slice has nal\_unit\_type in the range of 4 to 8, inclusive; a RAP picture may be a *BLA picture*, a *CRA picture* or an *IDR picture*; all *coded pictures* that follow the RAP picture both in *decoding order* and *output order* do not use *inter prediction* from any *picture* that precedes the RAP picture ~~either~~ in *decoding order* ~~or~~ *~~output order~~*; and any *picture* that precedes the RAP picture in *decoding order* also precedes the RAP picture in *output order*.

The fix also covers the CRA and BLA case mentioned above. Even if the above-mentioned proposal piece in Section 3 is not adopted, the above change to the definition of RAP picture is still needed to fix the IDR picture issue.

# Mandating parameter set activation at BLA pictures

In both AVC and the current HEVC draft, a parameter set is activated when it is not already active, its ID is referred, and the ID is different than the ID of the parameter set (of the same type) that is currently active. A parameter set is also activated when it is not already active, its ID is referred, its ID is the same as that of the active parameter set of the same type, and the content of these two parameter sets are different.

BLA pictures come from splicing of bitstreams at CRA pictures, when the splicing-point CRA pictures should be changed to be BLA pictures. Thus, each parameter set in a BLA access unit is most likely different from a parameter set of the same type with the same ID in an earlier access unit in decoding order. For this reason, it makes sense just to activate a parameter set referred to by a slice of a BLA picture without checking whether the activation conditions are satisfied.

Therefore, we propose that a VPS/SPS/PPS/APS referred to by a slice of a BLA picture is always activated, to avoid the need of checking whether the VPS/SPS/PPS/APS is has the same content (including parameter set ID) as the same as the current active VPS/SPS/PPS/APS, if any.

# Two constraints on DLP pictures

## Proposal

We propose the following constraints to be included in the specification:

1. Any picture preceding a CRA or BLA picture in decoding order shall precede any DLP picture associated with the CRA or BLA picture in output order. Any TFD picture associated with the CRA or BLA picture shall precede any DLP picture associated with the CRA or BLA picture in output order.
2. A TFD picture associated with a CRA or BLA picture shall precede, in decoding order, any picture that follows the CRA or BLA picture in both decoding and output order.

The first constraint excludes cases wherein DLP pictures are interleaved, in output order, with TFD pictures or pictures earlier, in decoding order, than the associated CRA or BLA picture. In those cases, when random accessing or performing stream switching from the CRA or BLA picture, the beginning picture rate would be smaller than after the DLP pictures are output. In adaptation streaming with frequent stream switching, e.g. in DASH context, such fluctuation in picture rate would cause undesirable user experience.

Currently, for a CRA or BLA picture, there shall be no picture preceding it in decoding order and following it in output order. The inclusion for that restriction was for a similar reason.

The second constraint disallows decoding-order interleaving of TFD pictures and normal pictures associated with a CRA or BLA picture. This helps checking which pictures are TFD pictures when needed, e.g., for discarding TFD pictures by a MANE or a decoder, because it would be known that there is no more TFD pictures when seeing the first normal picture.

## Discussion

We concluded that there is no need for a restriction to disallow interleaving, in decoding order, of the TFD pictures and the DLP pictures associated with a same CRA or BLA picture. This is because that such interleaving may be beneficial, as in hierarchical coding structures, those pictures closer to the CRA or BLA picture in output order may be encoded as DLP pictures, while those closer to the previous key picture of the same GOP in output order may be encoded as TFD pictures, and these pictures are often interleaved in decoding order.

# On inference of no\_output\_of\_prior\_pics\_flag to 1

The semantics of no\_output\_of\_prior\_pics\_flag is as follows:

**no\_output\_of\_prior\_pics\_flag** specifies how the previously-decoded pictures in the decoded picture buffer are treated after decoding of an IDR or a BLA picture. See Annex C. When the IDR or BLA picture is the first IDR or BLA picture in the bitstream, the value of no\_output\_of\_prior\_pics\_flag has no effect on the decoding process. When the IDR or BLA picture is not the first IDR or BLA picture in the bitstream and the value of pic\_width\_in\_luma\_samples or pic\_height\_in\_luma\_samples or sps\_max\_dec\_pic\_buffering[ sps\_max\_temporal\_layers\_minus1 ] derived from the active sequence parameter set is different from the value of pic\_width\_in\_luma\_samples or pic\_height\_in\_luma\_samples or sps\_max\_dec\_pic\_buffering[ sps\_max\_temporal\_layers\_minus1 ] derived from the sequence parameter set active for the preceding picture, no\_output\_of\_prior\_pics\_flag equal to 1 may (but should not) be inferred by the decoder, regardless of the actual value of no\_output\_of\_prior\_pics\_flag.

There are two issues with the semantics:

1. The wording "may (but should not)" in the semantics is conflicting with the HRD behavior in both in Annexes C.3.1 and C.5.2, wherein the wording is as follows:

When the IDR or BLA picture is not the first IDR or BLA picture decoded and the value of pic\_width\_in\_luma\_samples or pic\_height\_in\_luma\_samples or sps\_max\_dec\_pic\_buffering[ i ] for any possible value of i derived from the active sequence parameter set is different from the value of pic\_width\_in\_luma\_samples or pic\_height\_in\_luma\_samples or sps\_max\_dec\_pic\_buffering[ i ] derived from the sequence parameter set that was active for the preceding picture, respectively, no\_output\_of\_prior\_pics\_flag is inferred to be equal to 1 by the HRD, regardless of the actual value of no\_output\_of\_prior\_pics\_flag.

To be consistent with Annex C, "may (but should not)" should be changed to "shall" or "is".

1. The two instances of "the first IDR or BLA picture" should be changed to "the first picture", as the inference of no\_output\_of\_prior\_pics\_flag equal to 1 also applies even when the IDR or BLA picture is the first IDR or BLA picture in the bitstream but not the first picture in the bitstream, e.g., when the first picture in the bitstream is a CRA picture.

Similar as in the second bullet item above, the instances of "the first IDR or BLA picture" in Annexes C.3.1 and C.5.2 related to the inference of no\_output\_of\_prior\_pics\_flag equal to 1 should be changed to "the first picture".

# NAL unit types for TFD and TLA pictures

Currently, NAL unit types 2 and 3 are specified as follows:

|  |  |  |
| --- | --- | --- |
| 2 | Coded slice of a TFD picture  slice\_layer\_rbsp( ) | VCL |
| 3 | Coded slice of a non-TFD TLA picture  slice\_layer\_rbsp( ) | VCL |

There are three types of TLA and TFD pictures:

* A: TLA pictures that are not TFD pictures
* B: TFD pictures that are TLA pictures (or TLA pictures that are TFD pictures)
* C: TFD pictures that are not TLA pictures

B and C are currently using the same NAL unit type 2. Thus, it is not possible to know whether a TFD picture is also a TLA picture or not.

To fully utilize the feature of TLA pictures, we propose to burn one more NAL unit type, such that B and C have their own NAL unit types, as follows ():

|  |  |  |
| --- | --- | --- |
| 2 | Coded slice of a TFD and TLA picture  slice\_layer\_rbsp( ) | VCL |
| 3 | Coded slice of a non-TFD TLA picture  slice\_layer\_rbsp( ) | VCL |
| 9 | Coded slice of a non-TLA TFD picture  slice\_layer\_rbsp( ) | VCL |

# Patent rights declaration(s)

**Qualcomm Incorporated 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).**