|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  9th Meeting: Geneva, CH, 27 April – 7 May 2012 | Document: JCTVC-I0349 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Sub-picture-level low-delay CPB behavior** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Ye-Kui Wang Ying Chen  5775 Morehouse Dr 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) |
| *Source:* | Qualcomm Incorporated | | |

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

# Abstract

This document proposes a sub-picture based coded picture buffer (CPB) behavior to achieve reduced codec delay in an interoperable manner. In the proposal, CPB removal may be carried out either at access unit level or sub-picture level. When at sub-picture level, each slice and the associated non-VCL NAL units are defined as a decoding unit, and the CPB removal delay for each decoding unit is signaled.

# Background

## Video applications and end-to-end delay

Video applications include local playback, streaming, broadcast/multicast and conversational applications. Conversational applications include video telephony and video conferencing and are also referred to as low-delay applications. Conversational applications require a relatively low end-to-end delay of the entire systems, i.e., the delay between the time when a video frame is captured and the time when the video frame is displayed. Typically acceptable end-to-end delay for conversational applications should be less than 400 ms, and an end-to-end delay of around 150 ms is very good. Each processing step may contribute to the overall end-to-end delay, e.g., capturing delay, pre-processing delay, encoding delay, transmission delay, reception buffering delay (for de-jittering), decoding delay, decoded picture output delay, post-processing delay, and display delay. Thus, typically, the codec delay (encoding delay, decoding delay and decoded picture output delay) should be minimized in conversational applications. In particular, the coding structure should ensure that the pictures' decoding order and output order are identical such that the decoded picture output delay is equal to zero.

## Hypothetical reference decoder (HRD)

Each video coding standard should include a specification of video buffering model. In AVC and HEVC, the buffering model is referred to as hypothetical reference decoder (HRD), which includes a buffering model of both the coded picture buffer (CPB) and the decoded picture buffer (DPB), and the CPB and DPB behaviors are mathematically specified. The HRD directly imposes constraints on different timing, buffer sizes and bit rate, and indirectly imposes constraints on bitstream characteristics and statistics. A complete set of HRD parameters include five basic parameters, initial CPB removal delay, CPB size, bit rate, initial DPB output delay, and DPB size.

In AVC and HEVC, bitstream conformance and decoder conformance are specified as parts of the HRD specification. Though it is named some kind of decoder, HRD is typically needed at the encoder side to guarantee bitstream conformance, while typically not needed at the decoder side. Two type of bitstream or HRD conformance, namely Type I and Type II, are specified. Also, two types of decoder conformance, output timing decoder conformance and output order decoder conformance are specified.

In the AVC and HEVC HRD models, decoding or CPB removal is access unit based, and it is assumed that picture decoding is instantaneous. In practical applications, if a conforming decoder strictly follows the decoding times signaled, e.g., in the picture timing SEI messages, to start decoding of access units, then the earliest possible time to output a particular decoded picture is equal to the decoding time of that particular picture plus the time needed for decoding that particular picture. Apparently, the time needed for decoding a picture in the real-world cannot be equal to zero.

## Sub-picture based coded picture buffer (CPB) behavior

A sub-picture based CPB behavior was proposed in JCTVC-G188 in order to achieve coding delay of less than one picture period in an interoperable way. The method is summarized as follows. A picture is evenly divided into M groups of treeblocks, i.e., the first M treeblocks in treeblock raster scan of the picture belong to the first group of treeblocks, the second M treeblocks in treeblock raster scan of the picture belong to the second group of treeblocks, and so on. The value M is signaled in buffering period SEI messages. This value is used to derive the CPB removal time (i.e., the decoding time) of each group of treeblocks. In this sense, the CPB behavior is sub-picture based, wherein each sub-picture is a group of treeblocks. It is assumed that access unit level CPB removal times are signalled as usual (using picture timing SEI messages), and within each access unit, the CPB removal times for the treeblock groups are assumed to linearly or evenly divide the interval from the CPB removal time of the previous access unit to the CPB removal time of the current access unit. This further implies the following assumptions or bitstream requirements:

1. Within each picture each treeblock group is encoded in a way that requires the same amount of decoding time (not just in the HRD model but for real-world decoders), wherein the coded data of the first treeblock group is considered to include all non-VCL NAL units in the same access unit and before the first VCL NAL unit.
2. Within each picture, the number of bits for each treeblock group is identical, wherein the coded data of the first treeblock group is considered to include all non-VCL NAL units in the same access unit and before the first VCL NAL unit.

The sub-picture based CPB behavior in JCTVC-G188 has the following problems:

1. The requirement that the amount of coded data for each treeblock group in a coded picture is identical is hard to achieve with a balanced coding performance (wherein areas with more detailed texture or motion activity in a picture use more bits).
2. When more than one treeblock group is included in a slice, there is no easy way to split the coded bits of treeblocks belonging to different treeblock groups and separately send them at the encoder side and separately remove them from the CPB (i.e., separately decode them).

# Proposal

To solve the above problems, a generic design for support of sub-picture based CPB behavior is proposed in this document.

When the signaling indicates that the sub-picture based CPB behavior is in use, CPB removal or decoding is based on sub-picture, or equivalently, decoding unit, that consists of a coded slice and the associated non-VCL NAL units. In other words, each time a decoding unit, instead of an access unit, is removed from the CPB for decoding. The removal time of a decoding unit from the CPB is derived from a signaled initial CPB removal delay and the CPB removal delay signaled for the decoding unit.

The DPB output and removal processes still operates at the picture level or access unit level, i.e., each time an entire decoded picture is output or removed from the DPB. Removal of decoded pictures from the DPB happens instantaneously at the CPB removal time of the first decoding unit of the current access unit.

With the proposal, coding of sub-pictures and allocation of bits to different sub-pictures in a picture can be performed as usual, without assuming or requiring that each sub-picture in one picture is coded with the same amount of bits.

The proposed changes are in the attachment of this document.

# 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).**