An Improved Hypothetical Reference Decoder For HEVC

Sachin Deshpande[a], Miska M. Hannuksela[b], Kimihiko Kazui[c], Thomas Schierl[d]
[a]Sharp Laboratories of America, [b] Nokia Research Center,
[c] FUJITSU LABORATORIES LTD., [d] Fraunhofer HHI

[a] sdeshpande@sharplabs.com, [b] miska.hannuksela@nokia.com,
[c] kazui.kimihiko@jp.fujitsu.com, [d] thomas.schierl@hhi.fraunhofer.de

ABSTRACT

Hypothetical Reference Decoder is a hypothetical decoder model that specifies constraints on the variability of conforming network abstraction layer unit streams or conforming byte streams that an encoding process may produce. High Efficiency Video Coding (HEVC) builds upon and improves the design of the generalized hypothetical reference decoder of H.264/ AVC. This paper describes some of the main improvements of hypothetical reference decoder of HEVC.

***Index Terms*—** Hypothetical Reference Decoder, HRD, HEVC

1. INTRDUCTION

Video coding standards utilize intra and inter coding techniques for compressing video frames. Typically this results in a variation in the bits required to compress each frame. The compressed video may be transmitted over channels at a nearly constant bitrate. To handle fluctuation in the bitrate variation of the compressed video when transmitting at constant or nearly constant bitrate, buffering is used at the encoder and decoder side. Hypothetical Reference Decoder (HRD) is a hypothetical decoder model that specifies constraints on the variability of conforming network abstraction layer unit streams or conforming byte streams that an encoding process may produce. In a video coding standard the defined HRD behavior is used to check bitstream and decoder conformance.

Principle behind HRD operation is based on the leaky bucket model [1]. A leaky bucket model is analogous to a physical bucket which has a hole in it. The water is added to the bucket in discrete chunks and the water leaks out of the bucket from the hole. A leaky bucket model is characterized by a set of leaky bucket model parameters: transmission bit rate R, buffer size B, and initial decoder buffer fullness F. Initial buffering latency/ delay can be derived from the parameters R and F as F/R. AVC and HEVC HRD build up on the concept of leaky bucket model with constrained arrival time requirement for data into the decoder buffer.

Various past and current video coding standards including MPEG-2 [2], H.263 [3], MPEG-4 Part2 [4], H.264/ AVC [5] have included concept and definition of the behavior of a HRD. In case of MPEG-2 [2] the HRD behavior was defined by Video Buffering Verifier (VBV). MPEG2 VBV can operate either in Constant Bit Rate (CBR) mode or in Variable Bitrate (VBR) mode. MPEG-4 part2 [4] VBV can operate only in CBR mode. H.263 [3] HRD behaves similar to MPEG-2 VBV in CBR mode with some modifications. H.264/ AVC significantly extended the original concept of MPEG VBV with the design of a *generalized* hypothetical reference decoder [5-6]. One of the main novel aspects of the generalized HRD of H.264/ AVC was the definition of *multiple* leaky bucket model parameters, each for a different bitrate schedule. Furthermore a mechanism to interpolate between the defined set of leaky bucket model parameters was defined. This allowed HRD to operate at any desired peak transmission bitrate, buffer size or delay.

Visual Information Processing and Communication IV, edited by Amir Said, Onur G. Guleryuz, Robert L. Stevenson,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8666, 866608 · © 2013 SPIE-IS&T
CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2009907

High Efficiency Video Coding (HEVC) builds upon and improves the design of the hypothetical reference decoder of H.264/ AVC. This paper describes some of the main improvements of hypothetical reference decoder of HEVC.

The rest of this paper is organized as follows. In section 2 we describe various terms used in this paper and then list some of the main improvements of the HEVC HRD compared to the H.264/ AVC HRD. These various improvements are then described in detail in subsequent sections. Section 3 describes the sub-picture based HRD of HEVC for ultra-low latency operation. In section 4 we describe definition of alternate sets of initial buffering parameters at random access points for reduction in the initial buffering latency. Section 5 describes the HRD parameter selection for bitstream conformance. Section 6 provides conclusions.

2. TERMINOLOGY AND LIST OF MAIN IMPROVEMENTS IN HEVC HRD

In this section we describe various terms used in this paper and then list some of the main improvements of the HEVC HRD compared to the H.264/ AVC HRD.

The HRD contains a coded picture buffer (CPB), which holds the coded access units (AU), i.e. coded video frames, in a first-in first-out fashion. An access unit holds a set of network abstraction layer (NAL) units, containing video coding layer (VCL) data of one or more coded slices associated with the video frame. The slices are conveyed in decoding order within the access unit. A slice consists of one or more slice fragments, i.e. a preceding independent slice segment and zero or more dependent slice segments. Each slice segment is contained in its own NAL unit. A subset of the access unit is defined as a decoding unit, which may consist of one or more slices as well as parts of a slice, i.e. slice fragments, as shown in Figure 1. An access unit may further comprise NAL units carrying non-VCL data such a parameter sets or supplemental enhancement information (SEI) messages.

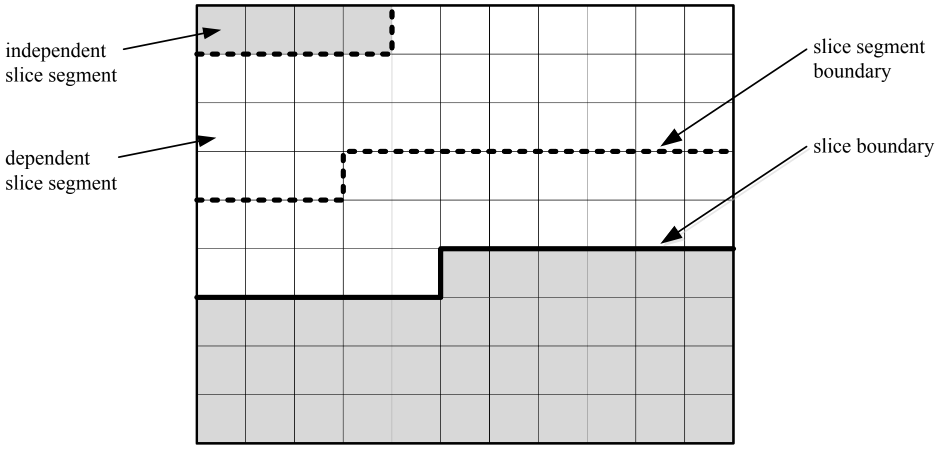

Figure 1: A picture divided into two slices, the first of which is partitioned into three slice segments.

At high level the HEVC HRD operates as follows. Data associated with decoding units that flow into the CPB according to a specified arrival schedule are delivered by the Hypothetical Stream Scheduler (HSS). There are two modes the HRD may operate in, i.e. either at access unit level or a sub-picture level based on decoding units, where the latter one defines a new timing on sub-picture level. In the sub-picture mode, the data associated with each decoding unit are removed and decoded instantaneously by the instantaneous decoding process at CPB removal time of the decoding unit. In the access unit level mode, all data associated with an access unit is removed and decoded instantaneously by the instantaneous decoding process at CPB removal time of the access unit.

Each decoded picture is placed in the Decoded Picture Buffer (DPB) for being referenced by the decoding process as well as for output and cropping. A decoded picture is removed from the DPB at the later of the DPB output time or the time that it becomes no longer needed for inter-prediction reference.

The operation of the HEVC CPB is specified in sub-clause C.2 of [7]. The operation of the HEVC DPB is specified in sub-clause C.3 [7]. We do not describe each of these steps in detail in this paper. Instead we focus on the main improvements of HEVC HRD compared to the prior art.

The main improvements of HEVC HRD compared to H.264/ AVC HRD include:

- Sub-picture based HRD operation for ultra-low latency
- Support for defining alternate sets of initial buffering parameters at random access points
- HRD parameter selection for bitstream conformance

These will be described in the next sections.

3. SUB-PICTURE BASED HRD OPERATION FOR ULTRA-LOW LATENCY

Sub-picture based HRD operation is newly introduced in HEVC for ultra-low latency operation. In section 3.1 we provide background concept behind the sub-picture based operation. Then the specification of sub-picture based operation in HEVC HRD is summarized in section 3.2.

3.1 Background

The HRD models of previous standards define the decoding time of a picture (i.e. an AU). When both an encoder and a decoder operate according to previous standards, minimal end-to-end delay in an actual system could reach up to 4*t* seconds (e.g. *t* is equal to 0.033 for 29.97 Hz video signal) as shown in Figure 2 (A). The reason is the following. As shown in Figure 2 (A), a typical encoder captures and encodes a picture both in *t* seconds. Then a coded picture is transmitted from the encoder to a decoder in the fixed duration, which could equal *t* seconds in the CBR case. Finally the decoder decodes and displays a coded picture both in *t* seconds. Those five operations should be carried in a non-overlapped manner because of the variation in the amount of bits of each macroblock and in the coding order of each macroblock. Regarding the former variation, the worst case is that the last macroblock or the first macroblock in a coded picture consumes whole amount of bits allocated to each picture. In those cases, the encoder cannot start sending a coded picture before the encoding is done or the decoder cannot start decoding a coded picture before the last bit of the coded picture arrives at the CPB of the decoder. Regarding the latter variation, for example, the scanning orders of each macroblock in capturing (or displaying) and in encoding (or decoding) could be different when e.g. Flexible Macroblock Ordering (FMO) of H.264/ AVC is used. In this case, the encoder cannot start encoding a picture before the capturing is done and the decoder cannot start displaying a reconstructed picture before the decoding is done.

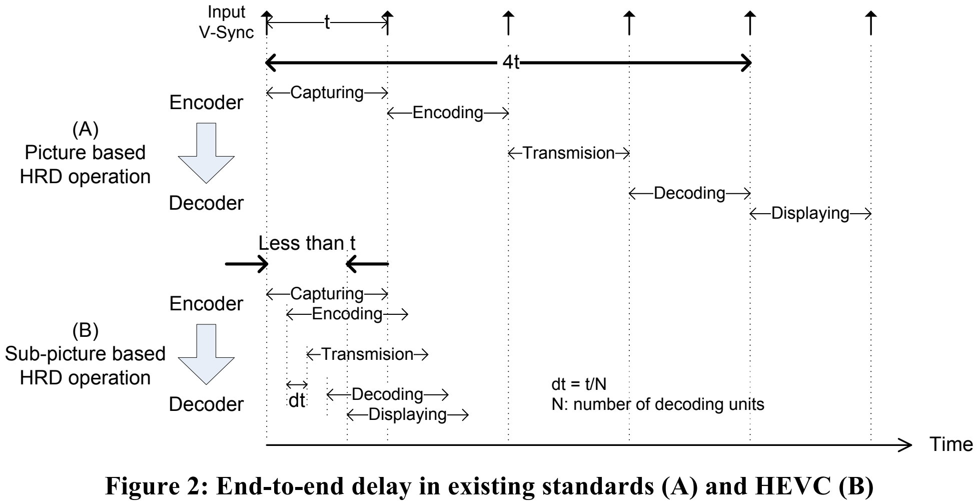

Figure 2: End-to-end delay in existing standards (A) and HEVC (B)

Some bi-directional video transmission applications, such as remote camera control, require ultra-low latency less than 0.1 seconds. In order to achieve this requirement, several video codec products in the broadcasting market realize ultra-low latency by minimizing the fluctuation of the amount of bits of each set of macroblocks (i.e. sub-picture) and by aligning the scanning order of each macroblock both in capturing and encoding. The former minimization can be achieved by encoding all macroblocks in intra mode with proper rate control, for example. However, existing standards do not specify the means to signal that a picture is encoded in such a manner. Thus the interoperable ultra-low latency is not possible or difficult with previous standards.

In order to realize interoperable ultra-low latency, HEVC newly defines and signals the decoding time of a sub-picture (i.e. a DU). When an encoder encodes a picture in such a way that each DU contains an equal amount of Coding Tree Unit (CTU of HEVC consists of up to 64x64 pixels and corresponds to a macroblock in previous standards) and an equivalent amount of bits, then the decoding time of each DU is uniformly distributed between the CPB removal times of two succeeding AUs. End-to-end delay is reduced as shown in Figure 2 (B) since the transmission of each DU can be started when the encoding of the DU is done, and the decoding of each DU can be started before the last DU of the AU. The degree of ultra-low latency depends on the number of DUs. For example, if each DU consists of one CTU row and the input video is 1920x1080 /59.94 Hz frame sequence, then the number of DUs is 17 and end-to-end delay of about 0.004 seconds could be achieved. It is noted that end-to-end delay becomes large if a decoder cannot start displaying a picture before the decoding of the picture is done.

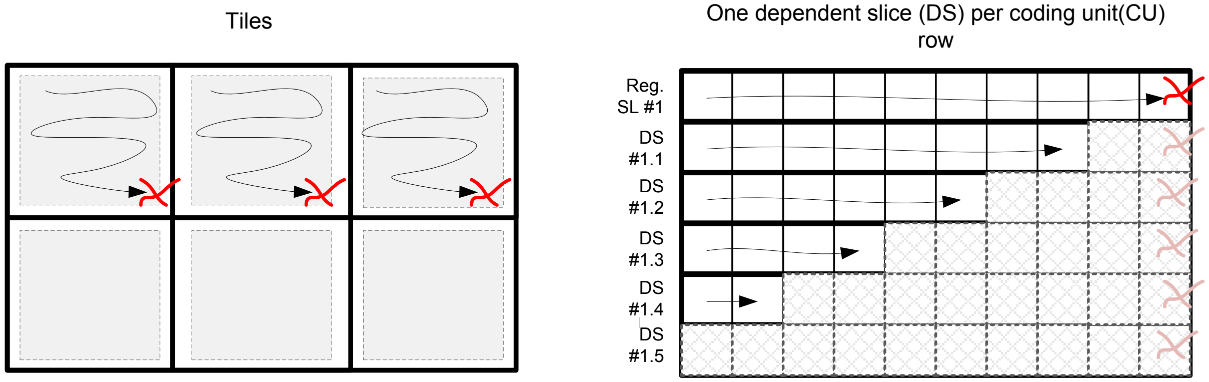

Figure 3: Ultra-low delay with high-level parallelism, tiles (left) and WPP (right) with end of DU indicated by "red marker"

HEVC further provides the ability of using high-level parallelization techniques [11] in order to reduce processing demands in multi-processing unit environments. Therefore, HEVC allows to sub-divide the picture into Wavefront Parallel Processing (WPP) substreams or tiles. Both methods may be used with ultra-low latency operations, where the WPP case can be only achieved using dependent slice fragments, as defined earlier.

The ultra-low latency mode for high-level parallelism is shown in Figure 3. The left part of the figure shows the coding process of tiles, where the first three tiles (marked red) are bound to the same decoding unit. In the right part of the figure, six WPP substreams are coded, where each substream belongs to a single decoding unit, each consisting of a single slice fragment.

3.2 Specification

The decoding time of an AU (or a DU) is defined as the time when the AU (or the DU) is removed from the CPB. It is required that the last bit of an AU (or a DU) should arrive at the CPB before its removal time.

In order to maintain backward compatibility, HEVC specifies and signals both the removal times of an AU and that of all DUs. It is up to a decoder to choose picture based HRD operation or sub-picture based HRD operation.

Figure 4 shows various parameters relating to the removal times of an AU and that of DUs. Those parameters are signaled in Video Usability Information (VUI) and various Supplemental Enhancement Information (SEI) messages in HEVC.

Similar to H.264/AVC, the removal time of an AU is derived from nal_initial_cpb_removal_delay/ vcl_initial_cpb_removal_delay in the Buffering Picture SEI message (BPSEI) available for the AU and au_cpb_removal_delay_delta_minus1 in the Picture Timing SEI message (PTSEI) of the AU.

The removal time of a DU is further derived from the removal time of the AU that contains the DU and corresponding set of du_cpb_removal_delay_increment_minus1 in the PTSEI of the AU or the Decoding Unit Information SEI (DISEI) message. The nominal removal time of the last DU in an AU and that of the AU is aligned. Instead of using du_cpb_removal_delay_increment_minus1, the removal time of a DU can be also derived from du_common_cpb_removal_delay_increment_minus1 in the PTSEI of the AU, or from du_spt_cpb_removal_delay_increment in the Decoding Unit Information SEI (DISEI) message which is associated with each DU. The former method is applied when du_cpb_removal_delay_incremement_minus1 of all DUs are identical. The latter method is applied when an encoder decides the removal time of a DU after encoding the DU. It is noted that when DISEI is not used, an encoder has to decide the removal times of all DUs in an AU before encoding the DUs since the BPSEI and the PTSEI are required to always be transmitted before the DUs and the encoding of the BPSEI and the PTSEI after encoding the DUs introduces one picture delay. A further point to note is that using the DISEI allows for generating timing information after a DU is encoded. In a typical ultra-low delay system, the bits of the preceding DU are already on the wire, i.e. they are already transmitted or still being transmitted. The case not using DISEIs implies that an encoder might occasionally generate a bitstream in which an underflow of the CPB at a DU occurs (i.e. the last bit of the DU is not delivered to the CPB before its CPB removal time). This situation may be improved by using the low_delay_hrd_flag in VUI similarly as in H.264/ AVC. In order to maintain the additional delay caused by underflow equally both in AU and DU based HRD operation, it is required that an encoder should generate an AU which causes overflow at AU level when a DU inside the AU causes overflow.

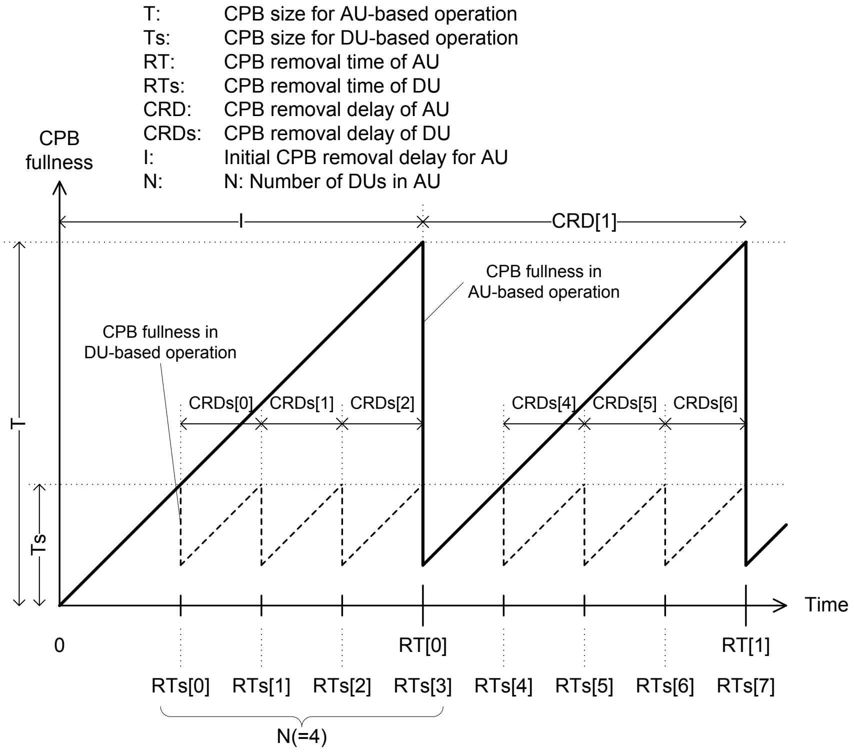
T: CPB size for AU-based operation
Ts: CPB size for DU-based operation
RT: CPB removal time of AU
RTs: CPB removal time of DU
CRD: CPB removal delay of AU
CRDs: CPB removal delay of DU
I: Initial CPB removal delay for AU
N: N: Number of DUs in AU
CPB fullness
I
CRD[1]
CPB fullness in AU-based operation
CPB fullness in DU-based operation
CRDs[0]
CRDs[1]
CRDs[2]
CRDs[4]
CRDs[5]
CRDs[6]
T
Ts
Time
0
RT[0]
RT[1]
RTs[0]
RTs[1]
RTs[2]
RTs[3]
RTs[4]
RTs[5]
RTs[6]
RTs[7]
N(=4)

Figure 4: Parameters for describing removal time of AU and DU

Table 1 shows a summary of various HEVC HRD parameters, their location, the functionality they support, whether they are applicable to access unit level/ sub-picture level operation and their comparison with H.264/ AVC HRD.

Table 1: HEVC HRD Parameters

Parameter	**Functionality:** **Used in derivation of**	**AU Level**	**Sub-picture Level**	**Location**	**Compared to H.264/ AVC**
nal_initial_cpb_removal_delay **nal_initial_cpb_removal_offset** **vcl_initial_cpb_removal_delay** **vcl_initial_cpb_removal_offset**	Removal time of first AU	Yes	No	BPSEI	Unchanged
nal_initial_alt_cpb_removal_delay **nal_initial_alt_cpb_removal_offset** **vcl_initial_alt_cpb_removal_delay** **vcl_initial_alt_cpb_removal_offset**	Removal time of first AU of RAP pictures/ sub-picture level operation	Yes	Yes	BPSEI	Alternate sets added (see section 4)
au_cpb_removal_delay_delta_minus1	Removal time of following AU	Yes	No	PTSEI	Unchanged
du_common_cpb_removal_delay_flag **du_common_cpb_removal_delay_increment_minus1**	Removal time of DUs which has the same removal time interval	No	Yes	PTSEI	New
num_nalus_in_du_minus1[i] **du_cpb_removal_delay_increment_minus1**[i]	Number of NAL units in a DU, and removal time of the DU.	No	Yes	PTSEI	New
du_spt_cpb_removal_delay_increment	Removal time of the DU following this DISEI.	No	Yes	DISEI	New
pic_dpb_output_delay	DPB output delay	Yes	No	PTSEI	Unchanged

4. ALTERNATE SETS OF INITIAL BUFFERING PARAMETERS

In this section we describe HEVC support for alternate sets of initial buffering parameters at random access points for reduction in the initial buffering latency.

When random access occurs at a random access point (RAP) picture, there can be leading pictures that follow the RAP picture in decoding order and precede it in output order. It is possible to discard these leading pictures following the RAP picture without affecting the decoding operation, as the leading pictures are not specified for output and have no effect on the decoding process of any other pictures that are specified for output.

HEVC Buffering Period SEI message supports signaling alternate set of initial buffering delay and initial buffering delay offset parameters. These can be signaled at a RAP picture. One set of values can specify the required initial buffering when leading pictures are present in the bitstream. The other set of values can specify the required initial buffering when leading pictures are discarded. The HRD CPB initialization in HEVC at RAP pictures defines using the correct set of initial buffering delay parameters based on if the leading pictures are discarded or not discarded.

Figure 5 illustrates the high level concept behind how the correct set of parameters is selected.
Signaling the alternate sets of initial buffering parameters at RAP pictures:

- Can reduce the initial buffering latency when starting playback at a RAP picture when associated leading pictures are discarded.

- Guarantees that HRD will not underflow/ overflow when associated leading pictures after a RAP picture are discarded.

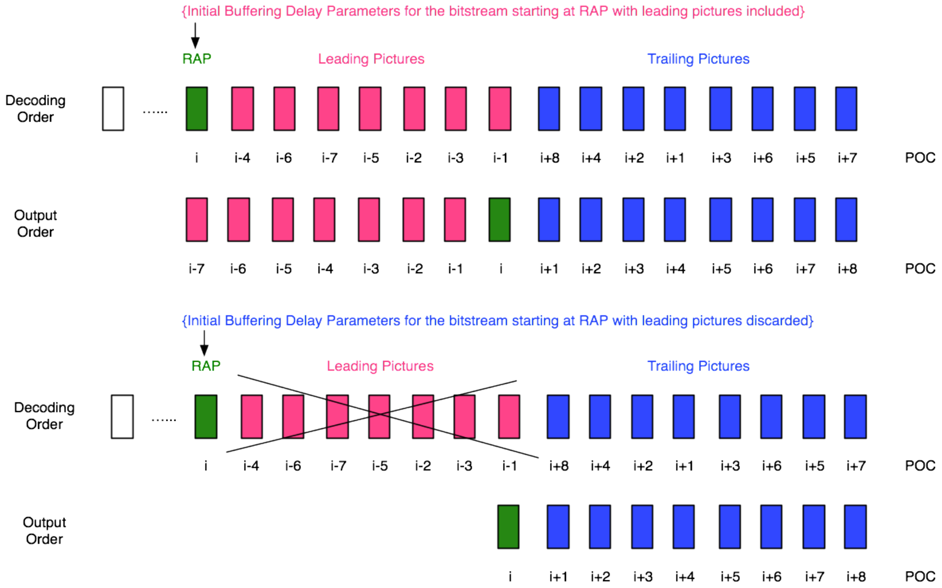
{Initial Buffering Delay Parameters for the bitstream starting at RAP with leading pictures included}
RAP
Leading Pictures
Trailing Pictures
Decoding Order
......
i i-4 i-6 i-7 i-5 i-2 i-3 i-1 i+8 i+4 i+2 i+1 i+3 i+6 i+5 i+7
POC
Output Order
i-7 i-6 i-5 i-4 i-3 i-2 i-1 i i+1 i+2 i+3 i+4 i+5 i+6 i+7 i+8
POC
{Initial Buffering Delay Parameters for the bitstream starting at RAP with leading pictures discarded}
RAP
Leading Pictures
Trailing Pictures
Decoding Order
......
i i-4 i-6 i-7 i-5 i-2 i-3 i-1 i+8 i+4 i+2 i+1 i+3 i+6 i+5 i+7
POC
Output Order
i i+1 i+2 i+3 i+4 i+5 i+6 i+7 i+8
POC

Figure 5: Selection from alternate sets of initial buffering parameters at a RAP picture

5. HRD PARAMETERS SELECTION FOR BITSTREAM CONFORMANCE

In this section we describe HEVC HRD parameter selection for determining bitstream conformance.

Extending H.264/AVC, SVC, and MVC with new scalability types, such as depth views, has been and is complicated due to at least the following reasons:

- The coded slice NAL units of the new scalability types are VCL NAL units according to the new amendment but non-VCL NAL units according to the "old" versions of the standard. As the HRD makes a difference between the VCL and non-VCL NAL units in its operation, different sets of HRD parameters are needed depending on the interpretation of the NAL unit types to either VCL or non-VCL NAL units.
- The sub-bitstream extraction process is specified for the NAL units and scalability types of the "old" versions of the standard, e.g. for dependency_id, quality_id, temporal_id and priority_id in Annex G of H.264/AVC [8] and for temporal_id, priority_id and view_id in Annex H of H.264/AVC [9]. However, new NAL unit types are introduced for new types of scalability, such as NAL unit type 21 for coded depth views in the current specification draft of MVC inclusion for depth maps, also known as MVC+D [10], and the existing sub-bitstream extraction process leaves those new NAL unit types intact even if they would also contain the "old" scalability dimensions, such as temporal_id and view_id in the case of depth views.

The NAL unit header for HEVC NAL units includes a 6-bit reserved field (nuh_reserved_zero_6bits). It is envisioned that this 6-bit field will be used as a generic layer identifier in the upcoming scalable extensions of HEVC. The semantics of layer identifier values are envisioned to be specified in a sequence-level syntax structure, such as a video parameter set (VPS) [7].

The base layer of a bitstream conforming to any of the future scalable extensions of HEVC should also be decodable by a decoder conforming only to HEVC version 1. Moreover, decoders conforming only to HEVC version 1 should be able to obtain appropriate HRD parameters that enable them to use correct initial buffering delays even if the incoming bitstream contained layers that they are not able to decode.

To solve the above drawbacks, the following design decisions were made in HEVC:

- Separate ranges of NAL unit type values for VCL and non-VCL NAL units were specified, including reserved VCL and non-VCL NAL unit type values for future extensions. This enables HEVC version 1 decoders and HRD handle the reserved NAL unit types (taken into use in a future extension) according to their correct categorization to VCL and non-VCL NAL units.

- The sub-bitstream extraction process for HEVC version 1 is specified to use temporal identifier and a set of values for nuh_reserved_zero_6bits (which in the future scalable versions are planned to signal layer information) as inputs. Consequently, it is possible to specify the HRD parameters for future scalable extensions already in HEVC version 1 by referring to the sub-bitstream extraction process with specific inputs.

- For the base layer, the HRD parameters are defined in VUI for each of the individual temporal sub-layers. A decoder can therefore select the correct HRD parameters for temporal subsets of HEVC version 1 bitstreams.

- Multiple sets of HRD parameters are specified in a VPS for different operation point sets, each identified by a set of nuh_reserved_zero_6bits values. A decoder can therefore select the correct HRD parameters based on the layers (i.e. nuh_reserved_zero_6bits values) present in the bitstream. Moreover, a bitstream extractor can select the correct HRD parameters for its output bitstream from the multiple sets of HRD parameters included in the VPS.

- A scalable nesting SEI message has been specified and serves the following purposes: It enables to indicate that the nested SEI messages pertain to a certain set of layers and/or a certain range of temporal sub-layers rather than all layers and/or sub-layers. Particularly, it enables to indicate the information contained in the buffering period and picture timing SEI messages for any bitstream subset, such as the base layer sub-bitstream.

The bitstream conformance may be verified for any sub-bitstream resulting from the sub-bitstream extraction process with a set of nuh_reserved_zero_6bits values (referred to as OpLayerIdSet) and a maximum temporal identifier value (referred to as OpTid) as inputs. For example, a list of values for all the layer identifier values (i.e. nuh_reserved_zero_6bits values) present in the bitstream may be created and used as input for the sub-bitstream extraction process. The conformance of the sub-bitstream resulting from the sub-bitstream extraction is verified. The HRD parameter sets for bitstream conformance are selected as follows. If OpLayerIdSet contains all the layer identifiers of the bitstream, the HRD parameters in the VUI are used and selected according to OpTid. Otherwise, the HRD parameters in the VPS are used. Correspondingly, the buffering period and picture timing information is selected either from the respective non-nested SEI messages (when OpLayerIdSet contains all the layer identifiers of the bitstream and when OpTid is the highest temporal identifier present in the bitstream) or from the respective SEI messages carried within a scalable nesting SEI message for which the indicated operation point information matches OpLayerIdSet and OpTid.

6. CONCLUSION

Hypothetical Reference Decoder (HRD) is a hypothetical decoder model that specifies constraints on the variability of conforming network abstraction layer unit streams or conforming byte streams that an encoding process may produce.
In this paper we described some of the major improvements of HEVC HRD compared to the HRDs in previous video coding standards. The improvements in HEVC HRD enable ultra-low latency sub-picture based operation, lower initial buffering latency at random access points through the definition of alternate sets of initial buffering parameters, and provide support for future scalable extension design and for sub-bitstream extraction process.

7. REFERENCES

1. A. R. Reibman, B. G. Haskell, “Constraints on variable bit-rate video for ATM networks,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 2, pp. 361–372, Dec. 1992.
2. “Information Technology—Generic Coding of Moving Pictures and Associated Audio Information: Video (MPEG-2/H.262) ISO/IEC 138180-2,” 2000.
3. “Video Coding for Low Bit Rate Communication (ITU-T Recommendation H.263),” 1998.
4. “Coding of Audio-Visual Objects: Video (ISO/IEC 14496-2),” 1998.
5. “Advanced video coding for generic audiovisual services (ITU-T Recommendation H.264),” 2007.
6. J. Ribas-Corbera, P.A. Chou, S. Regunathan, “A generalized hypothetical reference decoder for H.264/AVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 674–687, July 2003.
7. “High Efficiency Video Coding (HEVC) text specification draft 9,” December 2012.
8. Annex G – “Scalable Video Coding” in “Advanced Video Coding for Generic Audiovisual Services (ITU-T Recommendation H.264),” March 2010.
9. Annex H – “Multiview Video Coding” in “Advanced Video Coding for Generic Audiovisual Services (ITU-T Recommendation H.264),” March 2010.
10. “Study Text of ISO/IEC 14496-10:2012/DAM2 MVC extension for inclusion of depth maps,” JCT-3V document JCT3V-B1001, November 2012, Online: http://phenix.int evry.fr/jct3v/doc_end_user/current_document.php?id=462.
11. C. C. Chi, M. Alvarez-Mesa, B. Juurlink, G. Clare, F. Henry, S. Pateux, and T. Schierl, “Parallel Scalability and Efficiency of HEVC Parallelization Approaches,” *IEEE Trans. Circuits Syst. Video Technol.*, *Special Issue on Emerging Research and Standards in Next Generation Video Coding,* vol. 22, no. 12, pp. 1827-1838, December 2012.