SPECIFICATION OF BITSTREAM SUBSETS

JCTVC-H0568

JONATAN SAMUELSSON
RICKARD SJÖBERG

THE FOLLOWING IS PROPOSED:

1. Copy AVC/H.264 bitstream conformance section into the HEVC WD and split it into two conformance points
   a. Output timing conforming bitstreams (copied from H.264/AVC)
   b. Output order conforming bitstreams (as H.264/AVC but all parts related to timing removed)
2. Copy SVC specification of bitstream subsets for temporal_id with the following changes
   a. Separate output timing conformance from output order conformance
   b. Add a sub-bitstream extraction process for removal of individual access units
3. Make all information needed for sub-bitstream extraction available in fixed length without dependency on PPS or SPS
   a. Temporal Layer Access pictures for defining temporal layer switching points (JCTVC-H0566)
   b. Add a fixed length copy of picture order count into the access unit delimiters

THE FOLLOWING IS PROPOSED:

1. Copy AVC/H.264 bitstream conformance section into the HEVC WD and split it into two conformance points
    a. Output timing conforming bitstreams (copied from H.264/AVC)
    b. Output order conforming bitstreams (as H.264/AVC but all parts related to timing removed)
2. Copy SVC specification of bitstream subsets for temporal_id with the following changes
    a. Separate output timing conformance from output order conformance
    b. Add a sub-bitstream extraction process for removal of individual access units
3. Make all information needed for sub-bitstream extraction available in fixed length without dependency on PPS or SPS
    a. Temporal Layer Access pictures for defining temporal layer switching points (JCTVC-H0566)
    b. Add a fixed length copy of picture order count into the access unit delimiters

OUTPUT TIMING CONFORMING BITSTREAM SPECIFICATION COPIED FROM AVC/H.264

For Type I bitstreams, the number of tests carried out is equal to cpb_cnt_minus1 + 1 where cpb_cnt_minus1 is either the syntax element of hrd_parameters( ) following the vcl_hrd_parameters_present_flag or is determined by the application by other means not specified in this Recommendation | International Standard. One test is carried out for each bit rate and CPB size combination specified by hrd_parameters( ) following the vcl_hrd_parameters_present_flag. Each of these tests is conducted at the Type I conformance point shown in Figure C-1.

For Type II bitstreams there are two sets of tests. The number of tests of the first set is equal to cpb_cnt_minus1 + 1 where cpb_cnt_minus1 is either the syntax element of hrd_parameters( ) following the vcl_hrd_parameters_present_flag or is determined by the application by other means not specified in this Recommendation | International Standard. One test is carried out for each bit rate and CPB size combination. Each of these tests is conducted at the Type I conformance point shown in Figure C-1. For these tests, only VCL and filler data NAL units are counted for the input bit rate and CPB storage.

The number of tests of the second set, for Type II bitstreams, is equal to cpb_cnt_minus1 + 1 where cpb_cnt_minus1 is either the syntax element of hrd_parameters( ) following the nal_hrd_parameters_present_flag or is determined by the application by other means not specified in this Recommendation | International Standard. One test is carried out for each bit rate and CPB size combination specified by hrd_parameters( ) following the nal_hrd_parameters_present_flag. Each of these tests is conducted at the Type II conformance point shown in Figure C-1. For these tests, all NAL units (of a Type II NAL unit stream) or all bytes (of a byte stream) are counted for the input bit rate and CPB storage.

NOTE 2~~1~~ – NAL HRD parameters established by a value of SchedSelIdx for the Type II conformance point shown in Figure C-1 are sufficient to also establish VCL HRD conformance for the Type I conformance point shown in Figure C-1 for the same values of initial_cpb_removal_delay[SchedSelIdx], BitRate[SchedSelIdx], and CpbSize[SchedSelIdx] for the VBR case (cbr_flag[SchedSelIdx] equal to 0). This is because the data flow into the Type I conformance point is a subset of the data flow into the Type II conformance point and because, for the VBR case, the CPB is allowed to become empty and stay empty until the time a next picture is scheduled to begin to arrive. For example, when decoding a coded video sequence conforming to one or more of the profiles specified in Annex A using the decoding process specified in clauses 2-9, when NAL HRD parameters are provided for the Type II conformance point that not only fall within the bounds set for NAL HRD parameters for profile conformance in item j) of subclause A.3.1 or item h) of subclause A.3.3 (depending on the profile in use) but also fall within the bounds set for VCL HRD parameters for profile conformance in item i) of subclause A.3.1 or item g) of subclause A.3.3 (depending on the profile in use), conformance of the VCL HRD for the Type I conformance point is also assured to fall within the bounds of item i) of subclause A.3.1.

For output timing conforming bitstreams, all of the following conditions shall be fulfilled for each of the tests:

1. For each access unit n, with n>0, associated with a buffering period SEI message, with $\Delta t_{g,90}(n)$ specified by

   $$\Delta t_{g,90}(n) = 90000 * (t_{r,n}(n) - t_{af}(n - 1)) \qquad \text{(C-14)}$$

   the value of initial_cpb_removal_delay[SchedSelIdx] shall be constrained as follows.

   – If cbr_flag[SchedSelIdx] is equal to 0,

   $$\text{initial\_cpb\_removal\_delay[SchedSelIdx]} <= \text{Ceil}(\Delta t_{g,90}(n)) \qquad \text{(C-15)}$$

   – Otherwise (cbr_flag[SchedSelIdx] is equal to 1),

   $$\text{Floor}(\Delta t_{g,90}(n)) <= \text{initial\_cpb\_removal\_delay[SchedSelIdx]} <= \text{Ceil}(\Delta t_{g,90}(n)) \qquad \text{(C-16)}$$

   NOTE 3~~2~~ – The exact number of bits in the CPB at the removal time of each picture may depend on which buffering period SEI message is selected to initialise the HRD. Encoders must take this into account to ensure that all specified constraints must be obeyed regardless of which buffering period SEI message is selected to initialise the HRD, as the HRD may be initialised at any one of the buffering period SEI messages.

2. A CPB overflow is specified as the condition in which the total number of bits in the CPB is larger than the CPB size. The CPB shall never overflow.

3. A CPB underflow is specified as the condition in which $t_{r,n}(n)$ is less than $t_{af}(n)$. When low_delay_hrd_flag is equal to 0, the CPB shall never underflow.

4. The nominal removal times of pictures from the CPB (starting from the second picture in decoding order), shall satisfy the constraints on $t_{r,n}(n)$ and $t_r(n)$ expressed in subclauses A.3.1 through A.3.3 for the profile and level specified in the bitstream when decoding a coded video sequence conforming to one or more of the profiles specified in Annex A using the decoding process specified in clauses 2-9.

5. ~~Immediately after any decoded picture is added to the DPB, the fullness of the DPB shall be less than or equal to the DPB size as constrained by Annexes A, D, and E and subclauses G.10, G.13, G.14, H.10, H.13, and H.14 for the profile and level specified in the bitstream.~~ At any time instant, there shall be at most max_dec_frame_buffering pictures in the DPB.

6. All reference pictures shall be present in the DPB when needed for prediction. Each picture shall be present in the DPB at its DPB output time unless it is not stored in the DPB at all, or is removed from the DPB before its output time by one of the processes specified in subclause C.2.

7. The value of $\Delta_{to,dpb}(n)$ as given by Equation C-13, which is the difference between the output time of a picture and that of the first picture following it in output order and having OutputFlag equal to 1, shall satisfy the constraint expressed in subclause A.3.1 for the profile and level specified in the bitstream when decoding a coded video sequence conforming to one or more of the profiles specified in Annex A using the decoding process specified in clauses 2-9.

WHY DO WE NEED OUTPUT ORDER CONFORMING BITSTREAMS?

- In many applications timing is handled by the systems layer instead of the bitstream layer
  - The timing model might not be compatible with the AVC/H.264 timing model (based on buffering_period and pic_timing SEI)

THE FOLLOWING IS PROPOSED:

1. Copy AVC/H.264 bitstream conformance section into the HEVC WD and split it into two conformance points
   a. Output timing conforming bitstreams (copied from H.264/AVC)
   b. Output order conforming bitstreams (as H.264/AVC but all parts related to timing removed)
2. Copy SVC specification of bitstream subsets for temporal_id with the following changes
   a. Separate output timing conformance from output order conformance
   b. Add a sub-bitstream extraction process for removal of individual access units
3. Make all information needed for sub-bitstream extraction available in fixed length without dependency on PPS or SPS
   a. Temporal Layer Access pictures for defining temporal layer switching points (JCTVC-H0566)
   b. Add a fixed length copy of picture order count into the access unit delimiters

SPECIFICATION OF BITSTREAM SUBSETS

- SVC specifies bitstream subsets in section G.8.8
  - Definition of sub-bitstream extraction process for priority_id, temporal_id, dependency_id and quality_id.

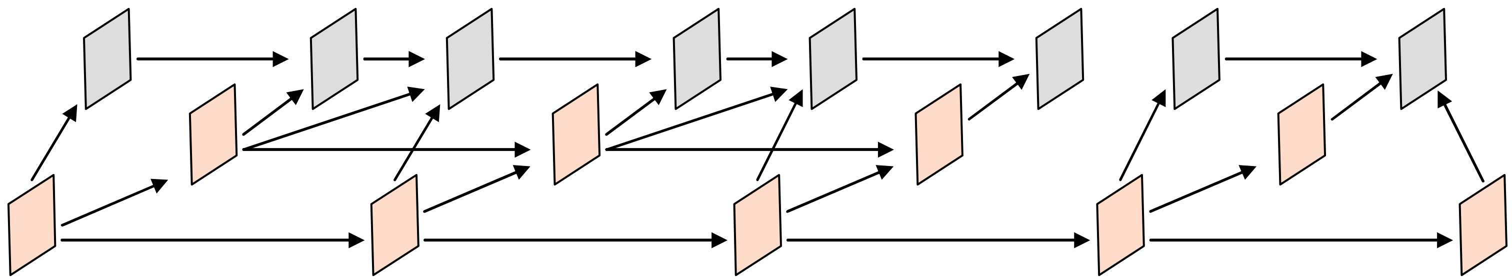
- We propose to add a similar sub-bitstream extraction process to HEVC for temporal_id only

- Two problems with the SVC sub-bitstream extraction process:
  a. The encoder is responsible of guaranteeing bitstream conformance for each possible sub-bitstream
  b. Only specified for scaling an entire bitstream to a specific layer

ENTIRE BITSTREAM THINNING

SVC:
"It is requirement of bitstream conformance that any sub-bitstream that is the output of the process specified in this subclause … shall be conforming to this Recommendation | International Standard."

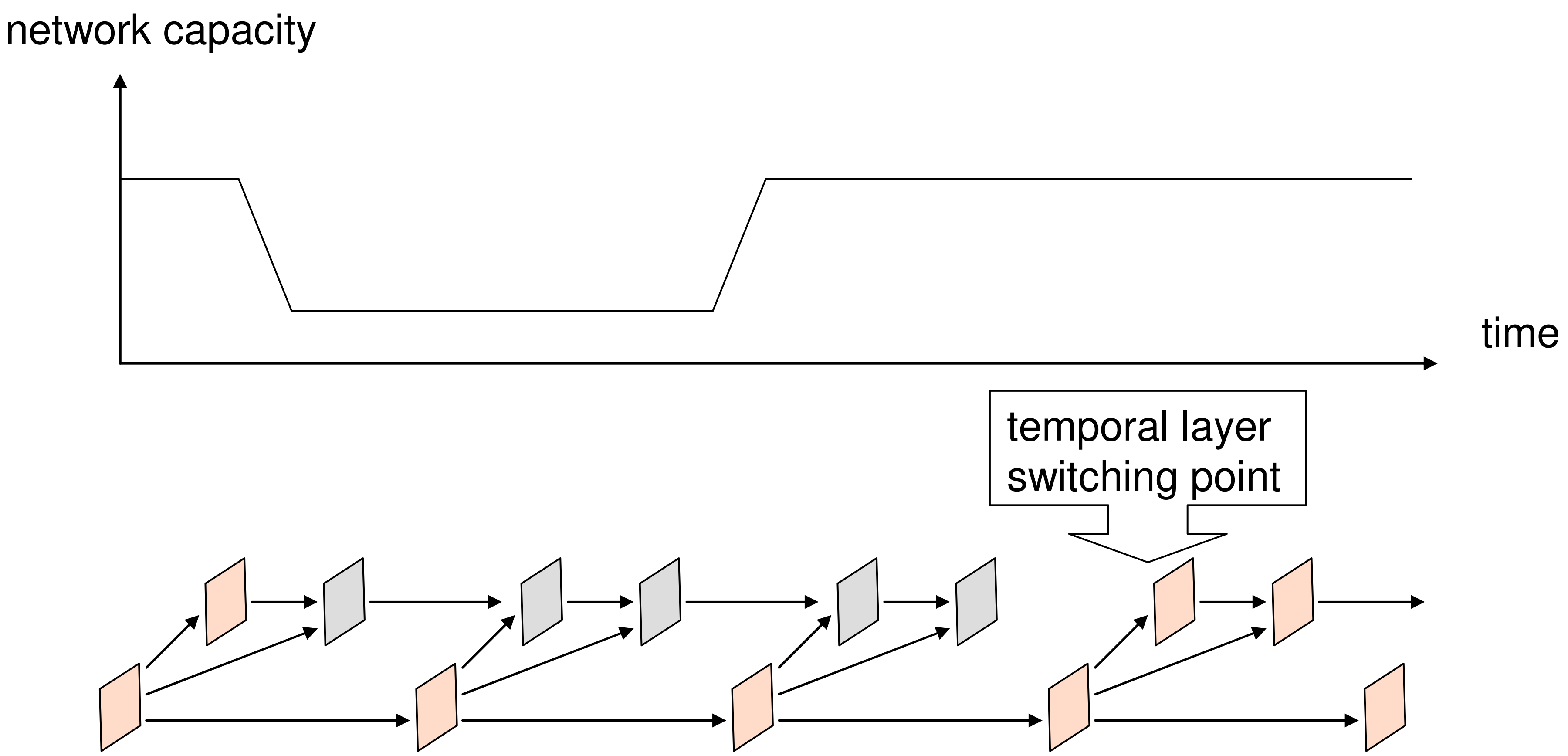
This becomes a burden for the encoder which has to evaluate output timing conformance for each temporal layer.



Our suggestion:

- All sub-bitstreams are output order conforming.
- Optionally, sub-bitstreams can be output timing conforming.

ADAPTIVE BITSTREAM THINNING



We propose to add a bitstream extraction process for this

THE FOLLOWING IS PROPOSED:

1. Copy AVC/H.264 bitstream conformance section into the HEVC WD and split it into two conformance points
   a. Output timing conforming bitstreams (copied from H.264/AVC)
   b. Output order conforming bitstreams (as H.264/AVC but all parts related to timing removed)
2. Copy SVC specification of bitstream subsets for temporal_id with the following changes
   a. Separate output timing conformance from output order conformance
   b. Add a sub-bitstream extraction process for removal of individual access units
3. Make all information needed for sub-bitstream extraction available in fixed length without dependency on PPS or SPS
   a. Temporal Layer Access pictures for defining temporal layer switching points (JCTVC-H0566)
   b. Add a fixed length copy of picture order count into the access unit delimiters

EXAMPLE OF NETWORK NODE'S MODE OF OPERATION

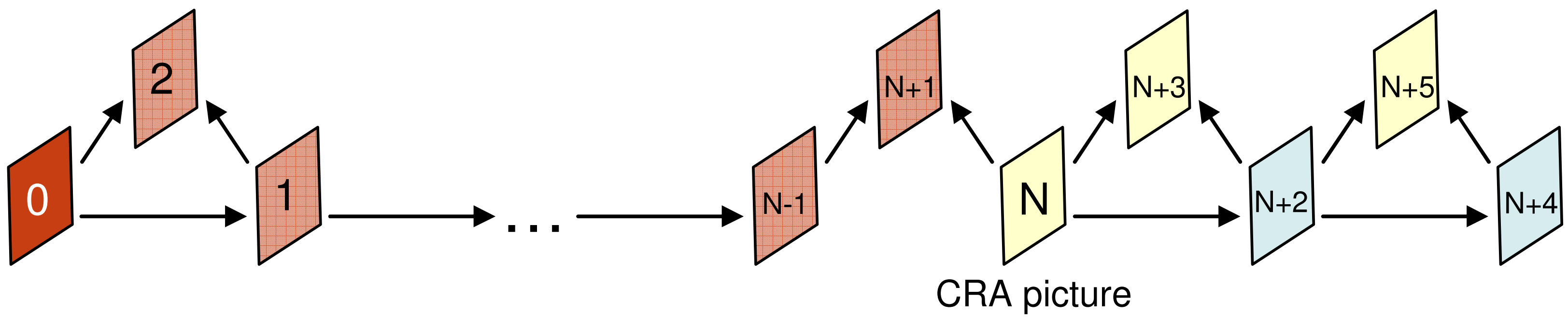
1. A network node relays access units for a stream with multiple temporal layers
2. At a time where the network conditions requires the bitstream to be thinned the network node stops relaying access units of the highest temporal layer(s)
3. When the network conditions are better the network node can relay a TLA* picture from the higher temporal layer and all access units with picture order count value higher than the TLA picture. **The network node will know that everything that is relayed is output order conforming.**

With temporal_id in the NAL unit header, TLA picture as a NAL unit type and a fixed length copy of picture order count in the access unit delimiter the network node can perform this without parsing any slice headers, parameter sets or variable length codes.

*Temporal layer access picture – a temporal layer switching point with "switch-to" definition

ALL BITSTREAMS STARTING WITH A CRA PICTURE ARE OUTPUT ORDER CONFORMING

- CRA pictures are a special case of TLA pictures
- The sub-bitstream extraction process for removal of individual access units ensures that all bitstreams starting with a CRA picture are output order conforming (if the original bitstream was output order conforming).



THE FOLLOWING IS PROPOSED:

1. Copy AVC/H.264 bitstream conformance section into the HEVC WD and split it into two conformance points
    a. Output timing conforming bitstreams (copied from H.264/AVC)
    b. Output order conforming bitstreams (as H.264/AVC but all parts related to timing removed)
2. Copy SVC specification of bitstream subsets for temporal_id with the following changes
    a. Separate output timing conformance from output order conformance
    b. Add a sub-bitstream extraction process for removal of individual access units
3. Make all information needed for sub-bitstream extraction available in fixed length without dependency on PPS or SPS
    a. Temporal Layer Access pictures for defining temporal layer switching points (JCTVC-H0566)
    b. Add a fixed length copy of picture order count into the access unit delimiters

How?

PROPOSED SPECIFICATION OF OUTPUT ORDER CONFORMING BITSTREAMS

There are two types of conformance that can be claimed by a bitstream: output timing conformance and output order conformance.

> NOTE – All output timing conforming bitstreams are output order conforming. For each output order conforming bitstream there is at least one configuration of VUI parameters, buffering period SEI messages and picture timing SEI messages for which the bitstream is output timing conforming.

For output order conforming bitstreams, all of the following conditions shall be fulfilled:

1. Each access unit shall be smaller than or equal to the CPB size.
2. At any time instant, there shall be at most max_dec_frame_buffering pictures in the DPB.
3. All reference pictures shall be present in the DPB when needed for prediction.
4. Each picture with output_flag equal to 1 shall be output from the DPB by the processes specified in subclause C.4 in increasing order of PicOrderCntVal.

SUB-BITSTREAM EXTRACTION PROCESS FOR TEMPORAL LAYER REMOVAL

10.1 Sub-bitstream extraction process for temporal layer removal

Input to this process is an output order conforming bitstream and a variable tIdTarget in the range of 0 to max_temporal_layers_minus1, inclusive.

Output of this process is an output order conforming bitstream. When the input to this process is an output timing conforming bitstream and output timing conformance of sub-bitstreams is claimed, the output bitstream shall be an output timing conforming bitstream.

The sub-bitstream is derived by applying the following operations in sequential order:

1. Mark all VCL NAL units for which temporal_id is greater than tIdTarget as "to be removed from the bitstream".
2. Remove all access units for which all VCL NAL units are marked as "to be removed from the bitstream".
3. Remove all VCL NAL units that are marked as "to be removed from the bitstream".

SUB-BITSTREAM EXTRACTION PROCESS FOR REMOVAL OF INDIVIDUAL ACCESS UNITS

10.2 Sub-bitstream extraction process for removal of individual access units

Input to this process is an output order conforming bitstream and a variable X in the range of 0 to N-1, inclusive where N is the number of access units in the bitstream.

Output of this process is an output order conforming bitstream. When the input to this process is an output timing conforming bitstream and output timing conformance of sub-bitstreams is claimed, the output bitstream shall be an output timing conforming bitstream.

Let access unit n be the n-th access unit in decoding order with the first access unit being access unit 0.

The sub-bitstream is derived by applying the following operations in sequential order:

1. When there is an access unit containing a TLA picture or IDR picture with temporal_id lower than or equal to the temporal_id of access unit X that follows access unit X in output order, call that access unit Y.

   NOTE 2 – access unit Y may precede access unit X in decoding order.

   – If there is an access unit Y, then for each access unit Z with temporal_id higher than or equal to the temporal_id of access unit X that follows access unit X in decoding order and precedes access unit Y in output order, mark all NAL units of access unit Z as "to be removed from the bitstream".

   – Otherwise (there is no access unit Y) then for each access unit Z with temporal_id higher than or equal to the temporal_id of access unit X that follows access unit X in decoding order, mark all NAL units of access unit Z as "to be removed from the bitstream".

2. Remove all NAL units marked as "to be removed from the bitstream".

SYNTAX

- TLA pictures described in JCTVC-H0566
- POC in access unit delimiters
  - Side effect: helps identifying reference pictures

7.3.2.5 Access unit delimiter RBSP syntax

access_unit_delimiter_rbsp( ) {	**Descriptor**
primary_pic_type	u(3)
pic_order_cnt	u(32)
rbsp_trailing_bits( )	
}	

pic_order_cnt indicates the picture order count for the primary coded picture contained in the access unit. pic_order_cnt is informative and does not affect the decoding process. pic_order_cnt shall be equal to PicOrderCntVal of the primary coded picture contained in the access unit.

ERICSSON