INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 MPEG2015/M35848
February 2015, Geneva, Switzerland

Source **Telecom ParisTech, Canon Research Centre France**
Status **For consideration at the 111th MPEG Meeting**
Title **Clarification on Layered HEVC storage in ISOBMFF**
Author Jean Le Feuvre, Cyril Concolato, Franck Denoual, Frédéric Mazé, Eric Nassor

1 Introduction

During its 110th meeting, MPEG issued the DAM of Carriage of Layered HEVC over MPEG-2 TS [1]. The specification enforces several constraints/options on how HEVC layers may be carried and how operation point signaled that makes the design of the solution simple and efficient. We believe that the current specification for L-HEVC in ISOBMFF could benefit from this design and be simplified for simple, yet most common use cases. This contribution reviews the constraints (hereafter noted *Cx*) and options (hereafter noted *Ox*) proposed in MPEG-2 TS and proposes simplifications of L-HEVC in ISO BMFF.

2 L-HEVC in TS constraints

2.1 Carriage of layers

C1: In MPEG-2 TS, each HEVC layer has to be carried in its own component (PID). Optionally, temporal sub-layers of a layer may be carried in another component.
C2: In MPEG-2 TS, L-HEVC components have dedicated stream types depending on their scalability type. It is therefore straightforward for a terminal to identify SNR/Spatial scalable layers versus multiview layers.
C3: Profile, Tier and Level for each layer is defined in the HEVC Descriptor of the component (PID) carrying it.

2.2 Signaling of Layer Dependencies

In MPEG-2 TS, Hierarchy Descriptor shall be used to describe single dependencies between layers, while Hierarchy Extension Descriptors shall be used to describe multiple dependencies of a given layer. This is similar to the track reference(s) found in the ISOBMFF.

O1: Using implicit dependencies is possible when layers only have a single dependency to another layer, for cases where the layers can non-ambiguously be identified. For example, cases with one base and one scalability layer, or one base and one view from a multi-view stream, etc. While interesting in the TS case because it reduces the size of the PMT (hierarchy descriptors can be dozens of byte for each stream), this option may not be very useful for ISOBMFF.

C4: Concatenation order of components is indicated by the Hierarchy (extension) Descriptor for implicit operation point (*C4#1*), or set through the operation point descriptor for each explicit operation point (*C4#2*).

2.3 Signaling of Operation Points

MPEG-2 TS defines an Operation Point Descriptor that, if present, defines possible operation point, similar to the 'oinf' box.
O2: the Operation Point Descriptor is optional, and only needed for operation points that cannot be guessed from the hierarchy description and PTL info. This reduces the size of the PMT as well, as in most cases such operation points are less than the total set of operation points.

3 Simplification of L-HEVC in ISOBMFF

3.1 Issues on L-HEVC in ISOBMFF

As identified in the FRNB comments on L-HEVC, the extraction process of NALUs is quite confusing in the current document. It is assumed that NALUs are gathered from tracks for the current operation point by following the track references, and then put back in order. As noted in the comments, if extractors are used this is useless. If extractors are not used, the rules should be clearly stated.

3.2 Simplification of L-HEVC in ISOBMFF

We suggest simplifying the storage of L-HEVC in ISOBMFF in the same spirit as the MPEG-2 TS one:

- there shall be VCL NALUs of one and only one layer per track (*matches C1*). In case tile tracks are used for the layer, it is allowed to have the layer spread on different tiled tracks.
- tracks shall use shv1/she1, mhv1/mhe1, as was the case in the WD of L-HEVC in ISOBMFF (*matches C1*). Additional code points ahv1/ahe1 can be defined for auxiliary dimension.
- Simple operation points are known from a given track and its dependencies (*matches 02*). The PTL for such operation points are given in lhvC (*matches C3*).
- Tracks for which it is not desirable to identify an operation points shall be marked as such in the lhvC. For example this could be the case of layers that are never an output layer in any of the defined operation points.
- For simple operation points, the extractors force the order of NALU in the output bitstream. If extractors are not desired, we suggest using another type of track reference, such as 'dpnd', and follow the order of references in that track references (*matches C4#1*)
- For complex (eg signaled in oinf) operation points, extractors to tracks other than hvt1/lht1 shall be ignored and the order of NALU in the output bitstream shall be the same order as the order of layers indicated for this operation point (*matches C4#2*).

3.3 Proposed Modifications

- Re-introduce specific sample entry type code points used for scalable coding and multiview coding, as in the WD. This simplifies identification of scalability type at the application layer for common operation points.

- Remove the 'tcon' box, since each layer is in its own track. (to match C1)

- Update lhvC syntax as follows:

```
aligned(8) class LHEVCDecoderConfigurationRecord {
   unsigned int(8) configurationVersion = 1;
   unsigned int(2) general_profile_space;
   unsigned int(1) general_tier_flag;
   unsigned int(5) general_profile_idc;
   unsigned int(32) general_profile_compatibility_flags;
   unsigned int(48) general_constraint_indicator_flags;
   unsigned int(8) general_level_idc;
   bit(1) complete_representation;
   bit(1) implicit_operation_point;
   bit(2) reserved = '11'b;
   unsigned int(12) min_spatial_segmentation_idc;
   bit(6) reserved = '111111'b;
   unsigned int(2) parallelismType;
   bit(6) reserved = '111111'b;
   unsigned int(2) chromaFormat;
   bit(5) reserved = '11111'b;
   unsigned int(3) bitDepthLumaMinus8;
   bit(5) reserved = '11111'b;
   unsigned int(3) bitDepthChromaMinus8;
   bit(16) avgFrameRate;
   bit(2) constantFrameRate;
   bit(3) numTemporalLayers;
   bit(1) temporalIdNested;
   unsigned int(2) lengthSizeMinusOne;
   unsigned int(8) numOfArrays;
   for (j=0; j < numOfArrays; j++) {
         bit(1) array_completeness;
         unsigned int(1) reserved = 0;
         unsigned int(6) NAL_unit_type;
         unsigned int(16) numNalus;
         for (i=0; i< numNalus; i++) {
               unsigned int(16) nalUnitLength;
               bit(8*nalUnitLength) nalUnit;
         }
   }
   if (complete_representation) {
        unsigned int(8) layer_ID;
   }
}
```

with the following semantics:

- `implicit_operation_point` : If set to 1, indicates that this track and all its dependencies, as signaled by 'scal' or 'dpnd' track references, form an operation point with a profile, tier and level indications given in this box. Such operation points shall not be listed in 'oinf'. If set to '0', indicates that no operation point is implicitly associated with this track. Operation points using this track may or may not be listed in the oinf box.
- `unsigned int(8) layer_ID`: gives the layer_id of VCL NALUs contained in this track. If `complete_representation` is not set, the track does not contain any VCL NALU.

- Update clause "9.7.2 Data sharing and reconstruction of an access unit" as discussed above (to match C4#2).

More specifically:

"When reconstructing an access unit for an operation point listed in the oinf box, the tracks required for reconstructing the access units are identified using the layers listed in this operation point. If several tracks contain data for the access unit, the alignment of respective samples in tracks is performed on decoding time, i.e. using the time-to-sample table only without considering edit lists. Extractors to tracks other than hvt1/lht1 shall be ignored and the order of NALU in the output bitstream shall be the same as the order of layers indicated in the 'oinf' box for this operation point."

When an operation point is implicitly declared through `implicit_operation_point`, and 'scal' track references are used, Access Units are reconstructed by solving extractors present in the samples.

When an operation point is implicitly declared through `implicit_operation_point`, and 'dpnd' track references are used, the tracks required for reconstructing the access units are listed by the 'dpnd' reference type; if several tracks contain data for the access unit, the alignment of respective samples in tracks is performed on decoding time, i.e. using the time-to-sample table only without considering edit lists. Extractors to tracks other than hvt1/lht1 shall be ignored and the order of NALUs in the output bitstream shall be the same order as the order of references in the 'dpnd' track reference.

4 Example based on Annex U of MPEG-2 TS

Annex U of MPEG-2 TS shows various configurations for multi-view or scalable L-HEVC. Figure 1 gives an overview of the dependencies on the first example of this annex.

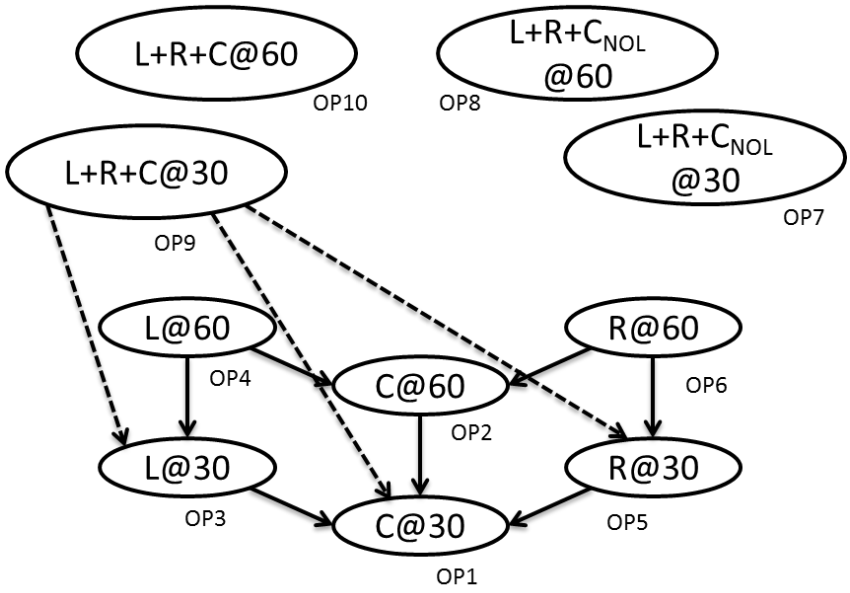

Figure 1 - References and Operation Point for multiview example of Annex U

Such a setup could be stored in ISOBMF as follows:

- Track1: (OP1)
  - Content: Layer 0, Temporal Layer 0 and 1
  - Type 'hvc1'
  - References: none

- Track2: : (OP4)
    - Content: Layer 1 (View1), Temporal Layer 0 and 1
    - Type ‘mvc1’
    - References: scal (or dpnp) to 1, sbas to 1
- Track3: (OP6)
    - Content: Layer 2 (View2), Temporal Layer 0 and 1
    - Type ‘mvc1’
    - References: scal (or dpnp) to 1, sbas to 1

Without oinf box, only the operation point OP1, OP4 and OP6 are described. With ‘oinf’ all operation points can be further described.

Note: it is not clear, from the HEVC specification, which OP should be signaled in the base track. The specification states:

“`general_profile_space`, `general_tier_flag`, `general_profile_idc`, `general_profile_compatibility_flags`, `general_constraint_indicator_flags`, `general_level_idc`, and `min_spatial_segmentation_idc` contain the matching values for the fields general_profile_space, general_tier_flag, general_profile_idc, general_profile_compatibility_flag[i] for i from 0 to 31, inclusive, the 6 bytes starting with the byte containing the general_progressive_source_flag, general_level_idc, and min_spatial_segmentation_idc as defined in ISO/IEC 23008-2, for the stream to which this configuration record applies.”

However, it could be that temporal sub-layers of the base layer have different ptl indication. These PTLs can be indicated in the TemporalLayerEntry sample group, but the specification does not state if indicating the max PTL of all sub-layers in the hvcC is allowed or not.

5 Conclusion

We proposed a simplification of the storage of layers per track and signaling of basic operation points in the current L-HEVC storage in ISOBMFF, mainly:

- One layer per track
- Operating point descriptor only for the operating points involving tracks without any direct dependencies
- Clarification of access unit reconstruction rules

We suggest adding these simplifications while processing the NB comments on [2].

6 References

[1] Text of ISO/IEC 13818-1:201x/DAM 3 Carriage of Layered HEVC
[2] Text of ISO/IEC 14496-15:2013 PDAM 1 Carriage of Layered HEVC