Source	**Telecom ParisTech, Canon Research Centre France**
Status	**For consideration at the 109th MPEG Meeting**
Title	**Clarifications on WD3 of Layered HEVC File Format**
Author	Jean Le Feuvre, Cyril Concolato, Franck Denoual, Frédéric Mazé, Eric Nassor

1   Introduction

At the 108th MPEG meeting, the systems subgroup has issued WD3 of 14496-15 for the storage of HEVC layered video in ISOBMFF (w14328). This contribution reviews the current text and proposes some fixes and clarifications on layered HEVC description.

As a general comment, we believe the text should be rewritten in amendment style to avoid any integration issue with any pending COR to 14496-15.

2   Editorial Fixes

- In sections 8.4.1.1.1 and 8.4.1.1.3, there are wrong references to the HEVCDecoderConfigurationRecord definition pointing respectively to sections 5.3.3.1 and 5.3.3. We suggest the editors to use the correct sections: respectively 8.3.3.1 and 8.3.3.

- In section 8.5.4.2, the font style is not correct in the syntax of TileSetGroupEntry. We suggest the editors to fix this.

- In section 8.5.4.3, the semantics of the groupID fields are a bit hard to read, we propose to rephrase them as follows:

Replace:

`tileGroupID` indicates the tile groupID value (as defined by a TileRegionGroupEntry) or of the a tile or tile set groupID value (as defined by a TileSetGroupEntry) that belongs to this tile set.

`topLeftTileGroupId` indicates the top left tile groupID value of the top-left tile or tile set that belongs to this tile set, when this tile set is rectangular.

`bottomRightTileGroupId` indicates the bottom right tile groupID value of the bottom-right tile or tile set that belongs to this tile set, when this tile set is rectangular.

`dependencyTileGroupID` gives the groupId value of the identifier of a Tile tile (as defined by a TileRegionGroupEntry) or a Tile Set tile set (as defined by a TileSetGroupEntry) that this tile set depends on.

`topLeftDependencyTileGroupId` indicates the top left tile groupID value of the top-left tile or tile set in of the rectangular area on which this tile set depends.

`bottomRightDependencyTileGroupId` indicates the bottom right tile groupID value of the bottom-right tile or tile set in the rectangular area on which this tile set depends.

with:

`tileGroupID` indicates the groupID value of a tile or tile set that belongs to this tile set.

`topLeftTileGroupId` indicates groupID value of the top-left tile or tile set that belongs to this tile set, when this tile set is rectangular.

`bottomRightTileGroupId` indicates the groupID value of the bottom-right tile or tile set that belongs to this tile set, when this tile set is rectangular.

`dependencyTileGroupID` gives the groupID value of a tile or tile set that this tile set depends on.

`topLeftDependencyTileGroupId` indicates the groupID of the top-left tile or tile set in the rectangular area on which this tile set depends.

`bottomRightDependencyTileGroupId` indicates the groupID of the bottom-right tile or tile set in the rectangular area on which this tile set depends.

- Still in section 8.5.4.3, the `dependency_list` definition duplicates the semantics of "idr_" fields, and is error-prone, we suggest to remove the last sentence as follows:

`dependency_list` indicates that a list of dependent tiles for this tile set is given. If set to zero, this implies that there is no temporal coding dependencies between tiles from this set and tiles outside of this set. This ensures that decoding only the tiles from the set and from the dependency list is possible without decoding other tiles. ~~If set to 3, an additional list of tile dependencies is given for the case where the sample is a sync sample as defined for this HEVC layer.~~

- Still in section 8.5.4.3, the text refers to "idr" while it really means RAP, as indicated in the note, so propose to do the following change:

Replace:
`idr_dependency_tile_count` and `idr_dependencyTileGroupID`: specify the list of Tile Region or Tile Sets this tile set depends on when the sample this tile belongs to is a sync sample as defined for this HEVC layer.
with:
`irap_dependency_tile_count` and `irap_dependencyTileGroupID`: specify the list of Tile Region or Tile Sets this tile set depends on when the sample this tile belongs to is a IRAP as defined for this HEVC layer.

- In section 8.5.6.2.3, the current WD does not indicate a precise reference in the HEVC standard to the temporal Motion-constrained tile SEI Message.

- In section 9.7.1, the following statement is made:

"If a layer represented by a track uses another layer represented by another track for inter-layer prediction reference, a track reference of type 'scal' shall be included in the track referring to the source track for inter-layer prediction."

This statement is not correct. The 'scal' dependency shall be set only because extractors are present in the layered track. We suggest removing the statement.

- In section 9.7.2 the following statement is made:

"Tracks that carry the relevant layers for an operation point can be obtained by following the 'scal' track references and information in the track contents box."

We believe there is no need to mention 'scal' references here, the set of operation points can be found following 'oref' to the 'oinf' of the base HEVC, and if needed using 'tcon' of the track.

- In section 9.7.5.1, the title "Random accessible sample sample group" can let reader think that there is a mistake (due to the duplication of "sample").
  We suggest renaming into: "Sample group for Random Accessible Samples
  As well as in the Definition section 9.7.5.1.1

- In section 9.7.6 there is a wrong reference to 14496-2, it should be 14496-12.

- In section 9.6.2.2 (Decoder Configuration Record) there is a field called hevc_baselayer_flag. This information can be found from track references: if the track referred to by an 'sbas' reference is not an HEVC track, then the base layer is not HEVC. We suggest removing this flag.

3 Clarifications on 'oinf' and 'tcon'

There are simple yet quite common use cases where only one enhancement layer is used and stored as a separate track in a file. In such a case, operating point information is not useful, as all tier/profile/level information for the enhancement layer would be included in the track carrying the layer, in particular in the DecoderConfigurationRecord, as depicted in the figure below. There is therefore no reason to make the 'oinf' and 'tcon' boxes mandatory.
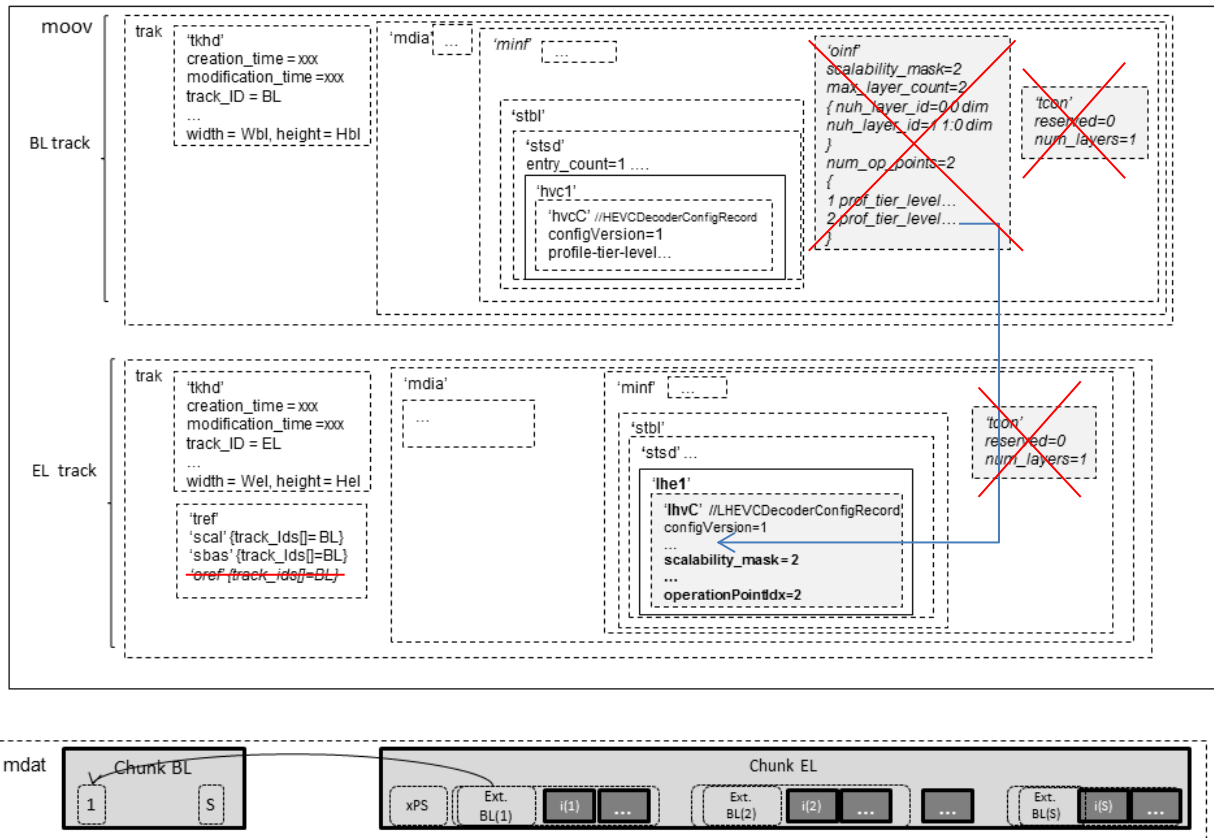
Figure 1: Example of layered HEVC encapsulation in separate tracks (in gray, boxes discussed in this contribution)

We therefore ask to have the 'tcon' and 'oinf' boxes optional, as they are not needed to decode basic files. The following changes need to be made:

- In sections 9.8.1 and 9.8.2, make the 'oinf' and 'tcon' boxes optional.

- In section 9.6.2.2 (Decoder Configuration Record), operationPointIdx shall be the 1-based index of the operating point described in the oinf box, or 0 if unknown or not specified.

- In section 9.8.1.2 (OperatingPointInformationBox), there is an operationPointId field which is not used anywhere else in the spec. If no justification for this field is given, we suggest removing it.

- In section 9.8.1.1, replace

"For a set of tracks with a common 'sbas' track reference, there shall be only one track among this set that carries the 'oinf' box. All tracks with a common 'sbas' referenced track shall be linked to the track that carries the 'oinf' box by using the 'oref' type track reference."

With

"For a set of tracks with a common 'sbas' or 'tbas' track reference, there shall be at most one track among this set that carries the 'oinf' box. If an 'oinf' box is present, all tracks with a common 'sbas' referenced track shall be linked to the track that carries the 'oinf' box by using the 'oref' type track reference."

4 L-HEVC tile tracks

Tile tracks are currently defined for HEVC base layer, but the description information (tile region and especially tile sets) are already defined with inter-layer dependencies in mind. The only current limitation stopping from using layered HEVC tile tracks is the absence of code points for such tracks.

We therefore suggest introducing generic layered-hevc tile tracks using the code point 'lht1' and introducing a new section for clarity.

If desired by the group, this section could be merged with the HEVC tile track section (to avoid too much copy/paste text), in a common HEVC and LHEVC tile tracks, to be moved after the LHEVC definitions.

"Layered HEVC Tile Track

Introduction

In a similar way to HEVC tile tracks, it is possible to define Layered-HEVC tile tracks to allow efficient access of spatial subparts of an enhancement layer. For such cases, tracks may be created using `LHEVCTileSampleEntry` sample description format.

An LHEVC tile track is a video track for which there is a 'tbas' reference to the LHEVC track carrying NALUs of the HEVC layer to which the tile(s) in this track belong. The sample description type shall be 'lht1'. Neither the samples in the tile track nor the sample description box shall contain VPS, SPS or PPS NAL units, these NAL units shall be in the samples or in the sample description box of the track containing the associated layer, as identified by the track reference. Both the LHEVC tile track and the track containing the associated layer, as indicated by the 'tbas' track reference, may use extractors, as defined in Annex B, to indicate how the original bitstream is reconstructed; presence of extractors in these tracks may be constrained in some application domains.

An LHEVC sample stored in a tile track is a complete set of slices for one or more tiles, as defined in ISO/IEC 23008-2. Typically if the track consists of a single tile, only the slice(s) used to code this tile will be found in the sample. A tile track typically includes one `TileRegionGroupEntry` (single-tile track), or one `TileSetGroupEntry` and one or more dependent `TileRegionGroupEntry` this tile set is made of (multi-tile track).

An LHEVC sample stored in a tile track is considered as a sync sample if the VCL NAL units in the sample are IRAP nal units and the corresponding access unit has no RASL pictures.

Sub-sample and sample grouping defined for regular LHEVC samples have the same definitions for an LHEVC tile sample.

Note: an implementation may decide to decode only a subset of the complete tiles of an HEVC or Layered HEVC sequence. In this case, it may use the tile dependency information in the `TileRegionGroupEntry` and `TileSetGroupEntry` sample group descriptions to discard un-needed tracks or ignore some extractors while decoding the HEVC sequence.

Sample entry name for LHEVC tile tracks

X.X Definition

Box Types: 'lht1'
Container: Sample Description Box ('stsd')
Mandatory: No
Quantity: Zero or more sample entries may be present

This sample entry describes media samples of an LHEVC tile track. The width and height of VisualSampleEntry for an LHEVC tile track (sample entry type 'lht1') shall be set to the maximum width and height of the Tile(s) or TileSet(s) contained in the track. The layout information in the track header (i.e., layer , matrix , width and height) of a tile track shall be identical to the track header information of the track containing the associated layer as identified by the 'tbas' track reference, and otherwise ignored. Any '*clap*' and '*pasp*' box or boxes derived from other specifications in an 'lht1' sample description shall be ignored.

X.X.1 Syntax

```
class LHEVCTileSampleEntry() extends VisualSampleEntry ('lht1'){

    MPEG4BitRateBox ();                  // optional
    Box extra_boxes[];            // optional
}
```

X.X.2 Semantics

The `LHEVCTileSampleEntry` shall not contain any `LHEVCConfigurationBox`, `HEVCConfigurationBox` or `MPEG4ExtensionDescriptorsBox`; these boxes are found in the base LHEVC track sample description, as indicated by the 'tbas' track reference. Other optional boxes may be included. There are usually as many entries in the `SampleDescriptionBox` of an LHEVC tile track as there are entries in the `SampleDescriptionBox` of the base LHEVC track.

The sub-parameters for the MIME type 'codecs' parameter for LHEVC Tile Track follow the rules defined in annex E.3 of this standard. The decoder configuration record is taken from the base track sample description, as indicated by the 'tbas' track reference. The sub-parameters for the MIME type 'codecs' parameter are then constructed using this decoder configuration record."

5   Identifying scalability type

The WD currently defines the `LHEVCDecoderConfigurationRecord` without indicating the type of scalability used in the bitstream. If a generic 'lhv1' / 'lhe1' is to be used, as currently defined in the WD, a file reader has to parse the VPS extensions to understand the scalability type. This is a complicated task. We therefore believe that the `LHEVCDecoderConfigurationRecord` should include the 16 bits (as currently defined in the SHVC, MV-HEVC annexes) of the scalability mask:

Replace
```
        unsigned int(8) general_level_idc;
        bit(1) complete_representation;
```
with

```
unsigned int(8) general_level_idc;
unsigned int(16) scalability_mask;
bit(1) complete_representation;
```

Similarly, the format for codec subtype parameter for L-HEVC should be extended to include the scalability mask, at least the defined values, in order to help the user agent detect if the stream scalability type is supported. We suggest adding the following sentence at the end of Annex E.3:

"Codec sub-type parameter for L-HEVC are identical to HEVC ones with the following exception: if the codec type identifies a generic L-HEVC media (i.e. 'lhv1' or 'lhe1' code points), the constructed HEVC codec parameter shall have .SXX appended to, 'S' indicating the scalability_mask_flag and XX are the 2 bytes corresponding to the scalability_mask_flag's value for this track; the trailing byte may be omitted if zero."

Annex E.3 should be renamed to "HEVC and derived specifications"

6 Conclusion

We suggest adding the above proposition in the current text of WD of 14496-15 and promote the resulting text to PDAM.