

**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/M36560  
June 2015, Warsaw, Poland**

**Source**    **Telecom ParisTech, Canon Research Centre France**  
**Status**    **For consideration at the 112th MPEG Meeting**  
**Title**      **Simplifications of L-HEVC File Format**  
**Author**    Cyril Concolato, Jean Le Feuvre, Franck Denoual, Frédéric Mazé, Naël Ouedraogo,  
Eric Nassor

## **1 Introduction**

During MPEG 111, question was raised whether further simplification of the storage of L-HEVC could be made, more specifically the restriction to use a single layer per track, the usage of implicit versus explicit reconstruction of AUs, and the usage of a single code point (lhc1/...) instead of indicating the layer type (shv1/...) as was the case before the DIS text. This contribution reviews some use cases and proposes a modification of the existing draft text.

## **2 AU reconstruction**

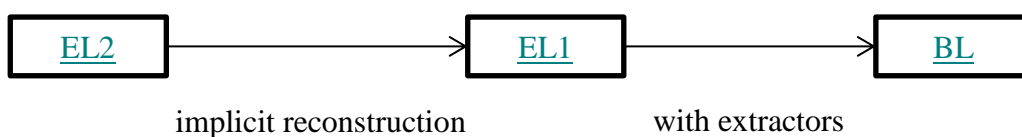
### **2.1 Tile tracks**

During our experiments with tile tracks, we found out that the most common operation expressed by the extractors is a simple concatenation of tile VCL NALUs with base track non-VCL NALUs (possibly with appending of suffix SEI messages at the end of the reconstructed AU). With the current approach based on extractors, each extractor costs 12 bytes, and the overhead can be quite significant for large tiling configurations: a 6x6 tiling at 50 hz costs 172 kbps for a pattern which is always the same. We believe that this situation should be avoided, and simple rules for AU reconstruction of tile tracks should be defined when extractors are not needed.

It was identified that explicit versus implicit AU reconstruction should be indicated in L-HEVC file format, and we propose to rely on AU reconstruction for tile tracks as well. This may be also applicable with fixed patterns of extractors for instance in SHVC or MV-HEVC.

### **2.2 AU reconstruction and extractors**

Having a mix of tracks using extractors and tracks not using extractors makes things very complicated. Consider a 3 layer configuration BL, EL1, EL2. EL1 uses extractors to get AUs from BL, while EL2 does not use extractors at all but depends on EL1.



The processing of EL1 is therefore, depending on whether EL2 is used:

- EL1 is the top layer of the current operating point (EL2 is not used): extractors can and shall be followed
- EL1 is NOT the top layer, EL2 is: in this case EL2 reconstruction rules state that we need to get all NALUs from BL, EL1 and EL2 and reorder them, which implies that extractors have to be skipped when getting EL1 sample.

We believe allowing both options mixed in the same file is too complex. We propose a solution allowing either usage of extractors everywhere (i.e. for all tracks with no implicit reconstruction), or nowhere (implicit reconstruction only, no extractor used in none of the tracks).

Additionally, when extractors are not used (implicit AU reconstruction), usage of the ‘scal’ track reference is not very useful, as pointed by several NBs during the last ballot. Indeed, the oinf box gives a complete list of layer IDs for each operating point, and the layer IDs are indicated in the ‘tcon’ box. We propose to get rid of ‘scal’ track references when no extractors are used.

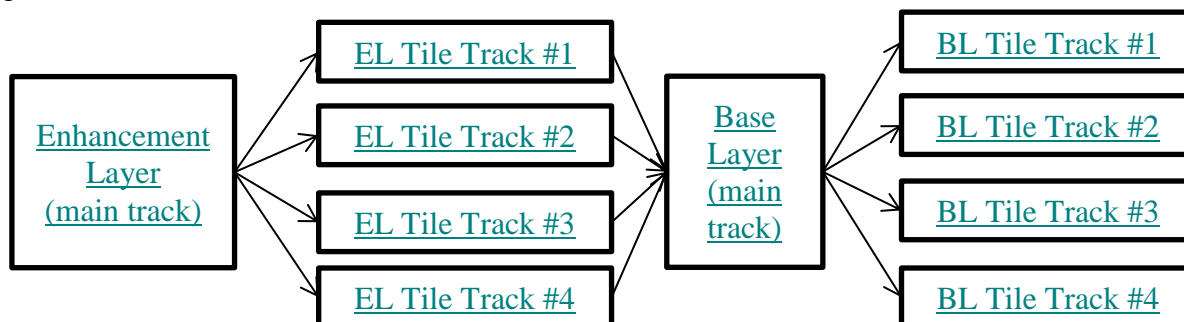
### 2.2.1 Example 1

Consider a base (BL) and its enhancement layer (EL1). The following storage options would be possible:

- 1) EL1 uses extractors, ‘scal’ references from EL1 to BL are present, ‘sbas’ reference from EL1 to BL
- 2) EL1 does not use extractors, no ‘scal’ references are present on EL1, but ‘sbas’ reference from EL1 to BL remains

#### 2.2.1.1 Example 2

Consider a base layer, tiled with a 2x2 grid, and its enhancement layer, also tiled with the same grid.



The following storage options would be possible:

- 1) All tracks use extractors:
  - BL uses extractors with ‘scal’ track references on all 4 tile tracks of the base, ‘tbas’ references from BL tile tracks to BL.
  - EL uses extractors with ‘scal’ track references on all 4 tile tracks of the EL and on BL, ‘tbas’ references from EL tile tracks to EL.
- 2) All without extractors:
  - BL does not use extractors, no ‘scal’ track references but ‘sabt’ track references on all 4 BL tile tracks of the base, ‘tbas’ references from BL tile tracks to BL.
  - EL does not use extractors, no ‘scal’ track references but ‘sabt’ track reference on all 4 EL tile tracks of the EL, ‘tbas’ references from EL tile tracks to EL and no ‘scal’ reference (since no extractor), just an ‘sbas’ reference from EL to BL.

### **2.3 Layer data copy in multiple tracks**

Finally, the current text allows the same NALUs of a given layer to be duplicated in multiple tracks ("data sharing by copy").

However, when the AU reconstruction (at least without extractors) is implicit, a file reader reconstructing an AU from two tracks carrying duplicate NALUs from the same layer will have no way to identify the duplication.

We suggest removing the ability to clone data between tracks.

### **2.4 Proposed text**

The attached text “DIS\_4th\_edition-SpecText-Extractors” is a modification of the specification with the following aspects:

- A new track reference of type ‘sabt’, pointing from an HEVC or L-HEVC (tile base) track to its tile tracks (defined in tile section).
- usage of ‘scal’ track references implies usage of extractors (9.7.1, 9.7.2)
- in 9.7.2, the rules for AU reconstruction are clarified

NOTE: We tried to keep the possibility of using extractors, but the whole design could be simplified by not using extractors at all.

## **3 Single layer per track**

### **3.1 Pros and Cons**

Using a single layer per track has the following advantages:

- This simplifies filtering of the layered content, as a track is either used or unused during the filtering. The current multi-layer-per-track approach imposes sample inspection in case the track contains a mix of layers used by the desired OP and layers not used by that OP.
- A single scalability dimension is usually present in the track. This allows using simplified code-points (shcl/...) to indicate the type of track, and simplifies the mime type derivation, without relying on profile and levels to identify the type of scalability.

- It better aligns with MPEG-2 TS carriage of L-HEVC where stream types are defined for each type of scalable content and a single layer is allowed in a track.

Using a single layer per track has the following disadvantages:

- Non-output layers have to be in their own track, and will therefore waste storage, as stbl / truns info will be duplicated. We however consider this use (layer that is never an output layer) case to be quite limited.
- It does not allow for use cases such as base layer stored in one track and all enhancement layers in another track, or all base and enhancement layers stored in one track.

### 3.2 Proposed text

The attached text “DIS\_4th\_edition-SpecText-All” is a modification of the above text (extractors clarification) with the following aspects:

- Specifies the use of a single layer per track;
- Specifies the 'lhc1' and 'lhe1' code points for layers that don't have a single main scalability type (eg both SNR and Multiview dimensions set) or that are non-output layers;
- Re-introduces the 'shv1'/'she1' and 'mhv1'/'mhe1' code points for layers that have a main scalability type identified;
- Removes the 'tcon' box, its information is merged in 'lhvC' as follows:
  - o layer\_id is added to 'lhvC'
  - o min and max temporal sublayer are removed as they cannot be used by 'oinf' to select layers. It is still possible to use the temporal grouping mechanism defined for HEVC.
- Extractors are not allowed in 'hvc1', allowed in 'hvc2'/'hev2' to extract tile data (eg the base track can only contain the base layer);
- Update to the mime type section. Note that it could be nice to allow a more compact form of the mime type for use in DASH, where usually only one OP will be announced with a stream, eg
  - o Make "*BLInternal*," optional
  - o Make "*OlsIdx.MaxTid*." optional

This would allow skipping the *lhevcptl* option and declaring the enhancement layer in the example as:

```
“shv1.A7.1.L120.B0”
```

## 4 Conclusion

We proposed simplification on the L-HEVC design, clarifying AU reconstruction handling and limiting the number of possibilities for multi-layer storage. We strongly suggest the File Format AhG to consider this input and to integrate it in the DIS text to be produced.