ISO/IEC JTC 1/SC 29/WG 03 N0764

**ISO/IEC JTC 1/SC 29/WG 03  
MPEG Systems   
Convenorship: KATS (Korea, Republic of)**

**Document type:** Output Document

**Title:** Technologies under Consideration for ISO/IEC 14496-15

**Status:** Approved

**Date of document:** 2023-03-16

**Source:** ISO/IEC JTC 1/SC 29/WG 03

**No. of pages:** 1 (with cover page)

**Email of Convenor:** young.L @ samsung . com

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg3>

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 03 MPEG SYSTEMS**

**ISO/IEC JTC 1/SC 29/WG 03 N0764**

**March 2023, Virtual**

|  |  |
| --- | --- |
| **Title** | **Technologies under Consideration for ISO/IEC 14496-15** |
| **Source** | **WG 03, MPEG Systems** |
| **Status** | **Approved** |
| **Serial Number** | **22288** |

**Abstract**

This document proposes three approaches for carriage of a Base bitstream (e.g. AVC, HEVC, VVC) and an LCEVC enhancement bitstream in a “Single-Track”, that is without the need to use two separate PIDs or Tracks, but rather carrying the base NALUs and the LCEVC NALUs as if they were a single bitstream.

Three possible solutions to carry a complete set composed by Base and LCEVC bitstreams are:

(1) Interleaved NALUs approach: insert LCEVC NALUs as interleaved NALUs within the Base NALUs.

(2) SEI approach: insert LCEVC NALUs as SEI messages in the Base NALUs.

(3) Aggregators to aggregate the Base bitstream with the LCEVC bitstream.

The following sections describe the three possible Single-Track solutions making explicit reference to four existing MPEG Video Coding specification:

• ISO/IEC 14496-10 Advanced Video Coding (AVC/H264)

• ISO/IEC 23008-2 High Efficiency Video Coding (HEVC/H265)

• ISO/IEC 23090-3 Versatile Video Coding (VVC/H266)

• ISO/IEC 23094-1 Essential Video Coding (EVC)

Contents

[**1.** **NALU aproach** 4](#_Toc126880204)

**[1.1](#_Toc126880205)****[AVC/H264 NALU header format](#_Toc126880205)** [4](#_Toc126880205)

**[1.2](#_Toc126880206)****[HEVC/H265 NALU header format](#_Toc126880206)** [5](#_Toc126880206)

**[1.3](#_Toc126880207)****[VVC/H266 NALU header format](#_Toc126880207)** [5](#_Toc126880207)

**[1.4](#_Toc126880208)****[LCEVC NALU header format](#_Toc126880208)** [6](#_Toc126880208)

**[1.5](#_Toc126880209)****[Compatibility of interleaving LCEVC NAL units with AVC/HEVC/VVC NAL units](#_Toc126880209)** [6](#_Toc126880209)

[**2.** **SEI aproach** 8](#_Toc126880210)

**[2.1](#_Toc126880211)****[Carriage of LCEVC NALUs in SEI messages](#_Toc126880211)** [8](#_Toc126880211)

**[2.2](#_Toc126880212)****[Suggested solution for SEI carriage](#_Toc126880212)** [8](#_Toc126880212)

[**3.** **Aggregators approach** 10](#_Toc126880213)

[This approache proposes the following ISOBMFF based codec-agnostic solution for the single-track carriage and storage of MPEG-5 LCEVC bitstream together with the base bitstream. 10](#_Toc126880214)

[**13** **LCEVC elementary streams and sample definitions** 10](#_Toc126880215)

**[13.1](#_Toc126880216)****[Overview](#_Toc126880216)** [10](#_Toc126880216)

**[13.4](#_Toc126880217)****[Derivation from ISO base media file format](#_Toc126880217)** [10](#_Toc126880217)

[13.4.1 LCEVC video stream definition: sample entry name and format 10](#_Toc126880218)

[13.4.2 LCEVC mixed sample entry 11](#_Toc126880219)

[13.4.3 LCEVC track structure 12](#_Toc126880220)

[13.4.4 Parameter sets 12](#_Toc126880221)

[13.4.5 'sync' sample 12](#_Toc126880222)

[**Annex A** (normative) **In-stream structures** 13](#_Toc126880223)

**[A.1](#_Toc126880224)****[General](#_Toc126880224)** [13](#_Toc126880224)

**[A.10 Base aggregators for LCEVC](#_Toc126880225)** [13](#_Toc126880225)

**[A.10.1 Definition](#_Toc126880226)** [13](#_Toc126880226)

**[A.10.2 Syntax](#_Toc126880227)** [14](#_Toc126880227)

**[A.10.3 Semantics](#_Toc126880228)** [14](#_Toc126880228)

1. **NALU aproach**

LCEVC encoded data units are Network Abstraction Layer units (NALU) as defined in ISO/IEC 23094-2, Sec. 7.3.2.

Each of the Base video coding standards under considerations here (AVC, HEVC, VVC) defines its own NALU syntax.

The format of the LCEVC NALU, in fact, allows for their unambiguous detection even when parsed according to the NALU syntax of the AVC, HEVC or VVC Base bitstream. This property allows for an “interleaved” single stream base plus enhancement, which means a single stream where LCEVC NALUs are inserted among base NALUs, within the same NALU sequence.

In the resulting interleaved single stream, each Access Unit, defined as the set of NALUS that result in each decoded picture, will contain both the LCEVC NALUs (an LCEVC Access Unit contains only one LCEVC NAL unit) and the Base NALUs relevant for the specific Access Unit.



***Figure 1 – Diagram of Interleaved NALU “single track” for LCEVC.***

* 1. **AVC/H264 NALU header format**

The AVC NALU header is defined in ISO/IEC 14496-10, Sec. 7.3.1, with the following syntax:

| ***Syntax*** | **Descriptor** |
| --- | --- |
| nal\_unit\_header( ) { |  |
| forbidden\_zero\_bit | u(1) |
| nal\_ref\_idc | u(2) |
| nal\_unit\_type | u(5) |
| } |  |

Table 6 – AVC NALU header syntax

The NALU type values and semantics for AVC are specified in Table 7-1 of the specification (IS 14496-10). Table 7 summarizes the usage of the AVC NALU types. Since the AVC NALU type is a field of 5 bits, the possible values are from 0 to 31.

|  |  |  |  |
| --- | --- | --- | --- |
| nal\_unit\_type | Name of nal\_unit\_type | Content of NAL unit and RBSP syntax structure | NAL unit type class |
| 0 - 5 | (…) | (…) | VCL |
| 6 - 20 | (…) | (…) | Non VCL |
| 21 - 23 | RSV | Reserved | Non VCL |
| 24 - 31 | UNSPEC | Unspecified | Non VCL |

Table 7 – AVC NALU types

* 1. **HEVC/H265 NALU header format**

The HEVC NALU header is defined in ISO/IEC 23008-2, Sec. 7.3.1.2, with the following syntax:

| ***Syntax*** | **Descriptor** |
| --- | --- |
| nal\_unit\_header( ) { |  |
| forbidden\_zero\_bit | f(1) |
| nal\_unit\_type | u(6) |
| nuh\_layer\_id | u(6) |
| nuh\_temporal\_id\_plus1 | u(3) |
| } |  |

Table 8 – HEVC NALU header syntax

The NALU type values and semantics for HEVC are specified in Table 7-1 of the specification (IS 23008-2). Table 9 summarizes the usage of the HEVC NALU types. Since the HEVC NALU type is a field of 6 bits, the possible values are from 0 to 63.

|  |  |  |  |
| --- | --- | --- | --- |
| nal\_unit\_type | Name of nal\_unit\_type | Content of NAL unit and RBSP syntax structure | NAL unit type class |
| 0 - 21 | (…) | (…) | VCL |
| 22 - 31 | RSV | Reserved | VCL |
| 32 - 40 | (…) | (…) | Non VCL |
| 41 - 47 | RSV | Reserved | Non VCL |
| 48 - 63 | UNSPEC | Unspecified | Non VCL |

Table 9 – HEVC NALU types

* 1. **VVC/H266 NALU header format**

The VVC NALU header is defined in ISO/IEC 23090-3, Sec. 7.3.1.2, with the following syntax:

| ***Syntax*** | **Descriptor** |
| --- | --- |
| nal\_unit\_header( ) { |  |
| forbidden\_zero\_bit | f(1) |
| nuh\_reserved\_zero\_bit | u(1) |
| nuh\_layer\_id | u(6) |
| nal\_unit\_type | u(5) |
| nuh\_temporal\_id\_plus1 | u(3) |
| } |  |

Table 10 – VVC NAL unit header syntax

The NALU type values and semantics for VVC are specified in Table 5 of the specification (IS 23090-3). Table 11 summarizes the usage of the VVC NALU types. Since the VVC NALU type is a field of 5 bits, the possible values are from 0 to 31.

|  |  |  |  |
| --- | --- | --- | --- |
| nal\_unit\_type | Name of nal\_unit\_type | Content of NAL unit and RBSP syntax structure | NAL unit type class |
| 0 - 11 | (…) | (…) | VCL |
| 12 - 25 | (…) | (…) | Non VCL |
| 26 - 27 | RSV | Reserved | Non VCL |
| 28 - 31 | UNSPEC | Unspecified | Non VCL |

Table 11 – VVC NALU types

* 1. **LCEVC NALU header format**

The LCEVC NALU header is defined in ISO/IEC 23094-2, Sec. 7.3.2, with the following syntax:

| ***Syntax*** | **Descriptor** |
| --- | --- |
| nal\_unit\_header( ) { |  |
| forbidden\_zero\_bit | u(1) |
| forbidden\_one\_bit | u(1) |
| nal\_unit\_type | u(5) |
| reserved\_flag | u(9) |
| } |  |

Table 12 - LCEVC NAL unit header

The NALU type values and semantics for LCEVC are specified in Table 17 of the specification (IS 23094-2). Table 13 summarizes the usage of the LCEVC NALU types. Since the LCEVC NALU type is a field of 5 bits, the possible values are from 0 to 31.

|  |  |  |  |
| --- | --- | --- | --- |
| nal\_unit\_type | Name of nal\_unit\_type | Content of NAL unit and RBSP syntax structure | NAL unit type class |
| 0 - 27 | UNSPEC | Unspecified | Non VCL |
| 28 - 29 | (…) | (…) | VCL |
| 30 | RSV | Reserved | Non VCL |
| 31 | UNSPEC | Unspecified | Non VCL |

Table 13 – LCEVC NALU types

The two NALU type values used to identify VCL NALUs are:

28 = 0x1C = 0b1.1100 (LCEVC NALU type 28Non IDR)

29 = 0x1D = 0b1.1101 (LCEVC NALU type 29IDR)

The two NALU header bytes for the two VCL NALU types of LCEVC are as follows:

0111.1001:1111.1111 (LCEVC NALU type 28 Non-IDR)

0111.1011:1111.1111 (LCEVC NALU type 29 IDR)

* 1. **Compatibility of interleaving LCEVC NAL units with AVC/HEVC/VVC NAL units**

The following table summarizes the position and semantic of the main fields of the NALU headers of the four MPEG specifications: AVC, HEVC, VVC, LCEVC:

* in yellow the NALU Type field
* in blue the Layer ID field
* in green the Temporal ID field

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | | 14 | 15 |
| AVC | 0 |  |  | NALU Type | | | | |  |  |  |  |  |  | |  |  |
| HEVC | 0 | NALU Type | | | | | | Layer ID (6) | | | | | | | Temp ID (3) | | | |
| VVC | 0 | 0 | Layer ID (6) | | | | | | NALU type | | | | | | Temp ID (3) | | | |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | | 14 | 15 |
| LCEVC | 0 | 1 | 1 | 1 | 1 | 0/1 | 0/1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 |
| LCEVC 28 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 |
| LCEVC 29 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 |

The following table reports how the base parsers for AVC, HEVC, VVC would interpret the two bytes of NALU headers for the significant LCEVC NALUs.

|  |  |  |  |
| --- | --- | --- | --- |
|  | LCEVC 28|29 | Layer ID | Temp ID |
| AVC parser | AVC 25|27 |  |  |
| HEVC parser | HEVC 60|61 | 63 | 7 |
| VVC parser | VVC 31|31 | 57|59 | 7 |

Editor’s Note:

*For the AVC parser, the LCEVC NALU types (interpreted as 25 or 27) fall in the Unspecified range from 24 to 31.*

*For the HEVC parser, the LCEVC NALU types (interpreted as 60 or 61) fall in the Unspecified range from 48 to 63.*

*For the VVC parser, the LCEVC NALU types (interpreted as 31) fall in the Unspecified range from 28 to 31.*

1. **SEI aproach**

LCEVC encoded data units are Network Abstraction Layer (NAL) units as defined in ISO/IEC 23094-2, Sec. 7.3.2.

All the MPEG base video coding standards considered (e.g., AVC, HEVC, VVC) provide metadata messages that can be used for the carriage of LCEVC. AVC, HEVC and VVC, employ NAL units as basic data units, and additionally the type of NAL unit identified as Supplemental Enhancement Information (SEI) that can be used to embed the LCEVC NAL unit stream.



***Figure 2 – Diagram of Embedded SEI “single track” for LCEVC.***

**2.1 Carriage of LCEVC NALUs in SEI messages**

When the base encoding for LCEVC is an MPEG standard, the elementary stream is a NALU stream. In this case, the encapsulation of LCEVC Access Units as metadata is implemented using the SEI messages specific for each Base codec. AVC, HEVC and VVC have each a different NALU format (i.e., with different NALU headers) breakdown of NALU types and payloads. However, all of them comprise SEI messages, identified with a nal\_unit\_type field as in the following table, where RBSP stands for raw byte sequence payload (see Table 1 below).

|  |  |  |
| --- | --- | --- |
| MPEG Standard | SEI nal\_unit\_type | Corresponding NAL unit payload |
| AVC | 6 | sei\_rbsp() |
| HEVC | 39\* | sei\_rbsp() |
| VVC | 23\* | sei\_rbsp() |
| \* prefix SEI | | |

Table 1 - SEI NALU type

**2.2 Suggested solution for SEI carriage**

The proposed solution for using SEI encapsulation of an LCEVC bitstream consists in defining a new SEI message for LCEVC and referencing it as a new Payload Type in each of the base layer video coding specifications.

In AVC (IS 14496-10), the SEI message payloadType is a field of 8 bits (with an escape mechanism). The Payload Types are specified in Section D.1.1, with values ranging from 0 to 56, and from 137 to 201, with the other values up to 255 allocated as reserved\_sei\_message.

In HEVC (IS 23008-2), as well, the SEI message payloadType is a field of 8 bits (with an escape mechanism). The Payload Types are specified in Section D.2.1, with values ranging from 0 to 56, and from 128 to 201, with other values up to 255 allocated as reserved\_sei\_message.

In VVC (IS 23090-3), as well, the SEI message payloadType is a field of 8 bits (with an escape mechanism). The Payload Types are specified in Section D.2.1, with values ranging from 0 to 45, and from 129 to 204, with other values up to 255 allocated as reserved\_sei\_message.

The suggested value of payloadType to be allocated to LCEVC in the three specifications (AVC, HEVC, VVC) is payloadType value 57, that falls in the “reserved range” of all the three spec.

1. **Aggregators approach**

This approach proposes the following ISOBMFF based codec-agnostic solution for the single-track carriage and storage of MPEG-5 LCEVC bitstream together with the base bitstream.

* A LCEVC mixed track is defined for carriage of LCEVC and the base streams
* An aggregator is defined for carrying base stream NAL units in LCEVC mixed track.

The proposed changes, with respect to the current “Dual Track” carriage of LCEVC, are highlighted with yellow text.

The clause numbering refers to the numbering currently used in IS 14496-15, Clause 13.

**13 LCEVC elementary streams and sample definitions**

**13.1 Overview**

…

~~Since the~~ LCEVC elementary streams carry enhancement to a "base" codec such as the ones listed above. A LCEVC elementary stream and a “base” codec elementary stream may be present in the same track. When, ~~the~~ a LCEVC elementary stream is in its own track, it makes a reference to a "base" codec elementary stream in a separate track, so that the LCEVC stream can be decoded in conjunction with the "base" stream, while the "base" stream can be decoded independently of the LCEVC stream.

An aggregator structure is specified to allow LCEVC elementary stream and a “base” codec elementary stream to be present in the same track. The aggregator structure enables grouping of NAL units from the base codec elementary stream into aggregated data units.

This clause defines the carriage of LCEVC elementary streams in the ISO base media file format as defined in this specification.

…

**13.4 Derivation from ISO base media file format**

### 13.4.1 LCEVC video stream definition: sample entry name and format

#### 13.4.1.1 Definition

…

#### 13.4.1.2 Syntax

…

#### 13.4.1.3 Semantics

…

### 13.4.2 LCEVC mixed sample entry

#### 13.4.2.1 Definition

Sample Entry and Box Types: 'lvms'

Container: Sample Description Box ('stsd')

Mandatory: The 'lvms' sample entry is mandatory

Quantity: One or more sample entries may be present

An LCEVC mixed sample entry shall contain a LCEVCConfigurationBox and a BaseConfigurationBox, as defined below. The BaseConfigurationBox contains the sample entry type and the decoder configuration box of the base stream (e.g. AVCConfigurationBox, HEVCConfigurationBox).

An optional BitRateBox may be present in the LCEVC mixed sample entry to signal the bit rate information of the LCEVC and the the base stream. Extension descriptors that should be inserted into the Elementary Stream Descriptor, when used in MPEG-4, may also be present.

The sample entry name 'lvms' specifies that the track to which this sample entry applies contains both a LCEVC stream and the base stream.

Base aggregators, as specified in subclause A.10, shall be used for aggregating the base stream in 'lvms' tracks. The order of all coded data included in a base aggregator is exactly the decoding order as if the coded data were present in a sample not containing aggregators or LCEVC NAL units.

If the sample of an 'lvms' track contains unspecified NAL unit types as defined in ISO/IEC 23094-2, the NAL units or NAL-unit-like structures having unspecified NAL unit types shall be discarded from the sample before providing the sample to the LCEVC decoder.

#### 13.4.2.2 Syntax

class LCEVCMixedSampleEntry() extends VisualSampleEntry('lvms'){

LCEVCConfigurationBox config1;

BaseConfigurationBox config2;

MPEG4ExtensionDescriptorsBox(); // optional

}

class BaseConfigurationBox() extends Box('blcf'){

unsigned int(32) base\_4cc;

Box config; // E.g., AVCConfigurationBox

Box other\_boxes[]; // optional boxes that are allowed for base\_4cc

}

#### 13.4.2.3 Semantics

Compressorname in the base class VisualSampleEntry indicates the name of the compressor used with the value "\014LCEVC Coding" being recommended (\014 is 10, the length of the string in bytes).

BaseConfigurationBox contains the decoder configuration box of the base stream (e.g. AVCConfigurationBox, HEVCConfigurationBox).

base\_4cc is the sample entry type that the base stream conforms to.

config is the decoder configuration box of the base stream (e.g. AVCConfigurationBox, HEVCConfigurationBox).

Editor’s Note:

*The Compressorname field “LCEVC Coding” is actually 12 bytes, so the text should read:*

*\014 is 12, the length of the string in bytes.*

### 13.4.3 LCEVC track structure

…

When the base track is coded using EVC, the base track shall be constructed according to clause 12.

A LCEVC mixed track is a track containing both the external base layer stream and the LCEVC enhancement stream, forming a representation of a complete set of encoded information.

The picture dimensions of the base stream and the LCEVC stream, width and height in Luminance samples, are specified by the corresponding relevant DecoderConfigurationRecord(s).

### 13.4.4 Parameter sets

…

### 13.4.5 'sync' sample

…

1. (normative) **In-stream structures**
   1. **General**

Aggregators and Extractors are file format internal structures enabling efficient grouping of NAL units or extraction of NAL units from other tracks.

Aggregators and Extractors use a syntax that is similar to the NAL unit syntax but does not follow the start code emulation prevention mechanism required for the NAL unit syntax as specified in ISO/IEC 14496‑10 or ISO/IEC 23008‑2 or ISO/IEC 23094-2. These NAL-unit-like structures are seen as NAL units in the context of the sample structure. While accessing a sample, Aggregators shall be removed (leaving their contained or referenced NAL units) and Extractors shall be replaced by the data they reference. Aggregators and Extractors shall not be output by file parsers.

These structures use NAL unit types reserved for the application/transport layer by ISO/IEC 14496‑10 or ISO/IEC 23008‑2 or ISO/IEC 23094-2.

See Annex F for more information about use of “reserved”, “unspecified”, “not specified” and “registrant-defined” nal\_unit\_type values.

…

**A.10 Base aggregators for LCEVC**

**A.10.1 Definition**

This subclause specifies base aggregators for LCEVC mixed tracks.

Base aggregators are used to group coded data of the base codec stream belonging to the same sample. When base aggregators are used in an 'lvms' track, the base aggregators shall include base codec stream coded data. A base aggregator is stored within a sample like any NAL unit. All coded data shall remain in decoding order within a base aggregator. A base aggregator shall not be empty, i.e., it shall include a coded picture of the base codec stream.

For storage of ISO/IEC 23094-2 stream together with a NAL-unit-based base codec stream, the following rules apply:

* Aggregators use the same NAL unit header as defined in ISO/IEC 23094-2 but with a different value of NAL unit type.
* If the BaseConfigurationBox in the sample entry contains the AVCConfigurationBox, the aggregator aggregates AVC NAL units.
* Otherwise, if the BaseConfigurationBox in the sample entry contains the HEVCConfigurationBox, the aggregator aggregates HEVC NAL units.
* Otherwise, if the BaseConfigurationBox in the sample entry contains the VVCConfigurationBox, the aggregator aggregates VVC NAL units.
* Otherwise, if the BaseConfigurationBox in the sample entry contains the EVCConfigurationBox, the aggregator aggregates EVC NAL units.

**A.10.2 Syntax**

class aligned(8) BaseAggregator (AggregatorSize) {  
 NALUnitHeader();  
 unsigned int i = sizeof(NALUnitHeader());  
 while (i<AggregatorSize) {  
 unsigned int(8) base\_data\_byte;  
 i++;  
 }  
 }

NOTE: The syntax of base aggregators does not always follow the NAL unit syntax and the NAL unit constraints specified in ISO/IEC 23094-2. For example, there can be three continuous bytes equal to a value in the range of 0x000000 to 0x000010, inclusive. This specifiation disallows the presence of base aggregators in a video bitstream output from parsing a file, therefore formal non-compliance with the video specifications is immaterial as they will never be presented to a video decoder.

**A.10.3 Semantics**

The value of the variable AggregatorSize is equal to the size of the aggregator, and the function sizeof(X) returns the size of the field X in bytes.

NALUnitHeader(): the first two bytes of ISO/IEC 23094-2 NAL units.

nal\_unit\_type shall be set to 31 for ISO/IEC 23094-2 video.

forbidden\_zero\_bit, forbidden\_one\_bit and reserved\_flag shall be set as specified in ISO/IEC 23094-2.

base\_data\_byte: The sequence of base\_data\_byte values shall be a complete base codec sample as defined by theBaseConfigurationBox of the sample entry.