

June 21, 2023



# Apple HEVC Stereo Video Interoperability Profile (Beta)

---

Version 0.9 (Beta)

June 21, 2023

Note: The information contained within this document is preliminary and is subject to change.

<b>Introduction</b>	<b>3</b>
<b>References</b>	<b>3</b>
<b>Apple HEVC Stereo Video Profile</b>	<b>4</b>
Constraints on the Movie and Track	4
Constraints on Sample Description / Sample Entry	4
Constraints on Video Frame Samples	8
Diagram of decoder configurations and a sample	9
File Type Brand	9
<b>Annex A—Establishing left and right view IDs</b>	<b>10</b>
Overview	10
Algorithm description	10
Diagram of view ID connection	13
<b>Document Revision History</b>	<b>14</b>

# Introduction

The ability to carry stereoscopic video views for the left and right eyes produces a richer user experience. Stereoscopic video—sometimes called “3D video”—presents one image to the user’s left eye and another image (typically correlated) to the user’s right eye to produce the *stereopsis* effect, defined as:

the perception of depth produced by the reception in the brain of visual stimuli from both eyes in combination; binocular vision.

While the production and display of this effect and its implementation and storage of views varies, this document describes use of an HEVC defined technique enabling professional and consumer workflows at more distribution-friendly bit rates.

Annex G of the HEVC Specification (Recommendation ITU-T H.265 or ISO/IEC 23008-2), building on Annex F, defines how multiple layers corresponding to related views may be represented and decoded using HEVC. Syntax allows several related layers within encoded video bit-streams. A use anticipated is multiview HEVC having left-eye and right-eye stereoscopic views. This document describes a highly focused profile for video with stereo views, using the syntax and tools in the HEVC Specification, with the goal of providing a clear target for file writers and file readers that maximizes interoperability.

Video tracks with stereo views encoded in this manner may be stored in QuickTime Movie files and MPEG-4 files.

**Note:** The QuickTime Movie File Format Specification and the ISO Base Media File Format Specification use different terminology for broadly equivalent concepts: atoms and boxes; sample descriptions and sample entries. This document uses the former specification's terminologies without loss of generality.

**Note:** The words "may", "should", and "shall" are used in the conventional specification sense, that is, respectively, to denote permitted, recommended, or required behaviors.

## References

[QTFF] QuickTime File Format Specification (QTFF), 2016

[HEVC] ISO/IEC 23008-2:2020 “High efficiency video coding” or ITU-R H.265:2021 “High efficiency video coding”

[ISOBMFF] ISO/IEC 14496-12:2020(E) ISO Base Media File Format

[ISONALU] ISO/IEC 14496-15:2019(E) “Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format”

# Apple HEVC Stereo Video Profile

In order for a QuickTime Movie file [QTFF] or MPEG-4 file [ISOBMFF, ISONALU] to conform to this profile, the following constraints must be satisfied.

Adhering to the principle that writers of an interoperable data format should be conservative in what they write, while readers should be somewhat flexible in what they accept, this document

- Recommends specific values for writers to use for the left and right stereo layers' nuh\_layer\_id, SPS ID, and PPS ID
- Recommends that readers implement their support in a way that is not fragile in the face of files using alternative ID values, where they are used consistently with the HEVC Specification [HEVC]
- Constraints from Section 9 of [ISONALU] should be followed along with additional restrictions described in this document.
- Other uses of “base layer” and “enhancement layer” within HEVC than specified for multiview carriage in HEVC are not constrained by this profile.

**Note:** Figure 1 below is a diagram relating sample descriptions, decoder configurations, and a representative video-frame sample. It may be useful for reference in reading the following sections.

## Constraints on the Movie and Track

- A single video track shall contain both the corresponding primary (or base) eye's view layer and the secondary eye's view layer. The primary eye view may be the left or right viewer eye. The secondary eye view should be the other viewer eye (i.e., if primary is left then secondary is right, if primary is right then secondary is left).
- A reader supporting HEVC but not multiview HEVC should be able to decode the video track as a single view corresponding to the base layer. This allows Apple HEVC Stereo Video Interoperability Profile tracks to be deployed that offer backwards compatibility with already deployed HEVC playback.

## Constraints on Sample Description / Sample Entry

- The codec type shall be 'hvc1'.
- Both left-eye and right-eye views shall be carried in each video sample of the track (i.e., both views are in the same video frame media sample, so DTS/PTS timings are the same for the two views). One of the two stereo eyes is referred to here as the *primary eye* and the second as the *secondary eye*.
- The primary eye may be either the left or the right eye. The secondary eye is the right eye if the primary is the left. Or, the secondary eye is the left eye if the primary is the right.

- There shall be both an HEVC Decoder Configuration Record ('hvcC') and a L-HEVC Decoder Configuration Record ('lhvC') which together describe the layers for the primary viewer eye and the secondary viewer eye.
  - The HEVC Decoder Configuration Record ('hvcC') carries the configuration of only the primary stereo view layer.
  - The L-HEVC Decoder Configuration Record ('lhvC') carries the configuration of the secondary stereo view layer.
- The configuration of the base and secondary stereo layers should be the same in all characteristics other than the assignment to the left stereo eye versus right stereo eye. This is termed "homogenous" here.
- A reader that does not support multiview video should be able to decode the base layer of the video track for presentation without a stereoscopic treatment. This is to ensure a degree of backwards compatibility. Note that the stereoscopic or monoscopic decoded video may still have other characteristics that make them unsuitable for playback as-is. The mechanism signaling these other characteristics is not described by this document.
- High dynamic range signaling (a.k.a., HDR signaling or HDR metadata) shall be the same for both the left-eye and right-eye views.
- Profiles for the layers:
  - Base layer shall be Main or Main10 profile.
  - Secondary eye layer shall be Multiview Main or Multiview Main10 profile, corresponding to the same bit depth as the base layer (i.e., both 10 bit or not).
- To support left and right stereo eye views, there is a Sequence Parameter Set (SPS) pair and a Picture Parameter Set (PPS) pair that are nearly identical differing only in stereo eye assignment and binding to nuh\_layer\_id values:
  - Two Sequence Parameter Sets (SPS):
    - Both Sequence Parameter Sets shall have sps\_video\_parameter\_set\_id equal to 0.
    - Both Sequence Parameter Sets shall indicate the same characteristics for both stereo eye views such as width, height, bitDepth, chromaFormat and conformanceWindow. This can be accomplished in the second layer by using update\_rep\_format=1 or by explicitly setting the same values for the corresponding seq\_parameter\_set\_ rbsp fields
    - The base layer SPS may have vui\_parameters that indicates video\_full\_range\_flag = 0 or video\_full\_range\_flag = 1. Both layers should have the same flag.
    - Both Sequence Parameter Sets shall indicate the same luma and chroma bit depth, but can be either 8-bit or 10-bit:
      - To use 8-bit luma and chroma, the Sequence Parameter Set shall indicate bit\_depth\_luma\_minus8 equal to 0 and bit\_depth\_chroma\_minus8 equal to 0 (per HEVC Specification G.11.1.1).

- To use 10-bit luma and chroma, the Sequence Parameter Set shall indicate `bit_depth_luma_minus8` equal to 2 and `bit_depth_chroma_minus8` equal to 2.
- Two Picture Parameter Sets (PPS):
  - One for the base layer, with `nuh_layer_id` equal to 0, with `pps_seq_parameter_set_id` equal to 0, and with `pps_pic_parameter_set_id` equal to 0. This corresponds to the primary eye.
  - One for the secondary eye layer, with `nuh_layer_id` equal to the secondary eye `nuh_layer_id`, and with `pps_seq_parameter_set_id` equal to the secondary eye layer SPS ID, and with `pps_pic_parameter_set_id` equal to a non-zero number referred to here as the *secondary eye layer PPS ID*.
  - All reference pictures must have the same CTB size as the picture that refers to them.
    - Specifically, the sum of `log2_min_luma_coding_block_size_minus3` and `log2_diff_max_min_luma_coding_block_size` in the layers' Sequence Parameter Sets must be the same.
    - $sps0.log2\_min\_luma\_coding\_block\_size\_minus3 + sps0.log2\_diff\_max\_min\_luma\_coding\_block\_size == sps1.log2\_min\_luma\_coding\_block\_size\_minus3 + sps1.log2\_diff\_max\_min\_luma\_coding\_block\_size$
    - These values are defined in [HEVC], 7.4.3.2.1 General sequence parameter set RBSP semantics
- The Video Parameter Set (VPS) for the whole sequence identifies `nuh_layer_id` values in this sequence
- The HEVC Decoder Configuration Record ('hvcC') shall contain the following parameter sets corresponding to the primary stereo eye:
  - One Video Parameter Set (VPS) with `nuh_layer_id` equal to 0, and with `vps_video_parameter_set_id` equal to 0, and containing a `vps_extension`.
    - The `vps_max_layers_minus1` shall be at least 1 to allow two layers, one for the left eye and one for the right eye. Only texture layers are supported.
    - The base layer should be present in the MV-HEVC bitstream - i.e, `vps_base_layer_internal_flag` and `vps_base_layer_available_flag` in VPS should have a value of 1.
    - The `vps_extension`, following the syntax in F.7.3.2.1.1, shall indicate values for `layer_id_in_nuh[]`, `scalability_mask_flag[]`, `dimension_id [][[]]`, `view_id_val[]`, and `direct_dependency_flag [][[]]` syntax elements to associate texture layers with the left eye and for the right eye and map to viewIDs used in the `three-dimensional_reference_displays_info` SEI message. See Annex A in this document for a description how the `left_view_id[0]` and `right_view_id[0]` are established from `nuh_layer_ids`.

- One Sequence Parameter Set (SPS) corresponding to the HEVC compatible base layer for the primary stereo eye
  - With `nuh_layer_id` equal to 0, and with `sps_seq_parameter_set_id` equal to 0.
- One Picture Parameter Set (PPS) corresponding to the primary stereo eye:
  - With `nuh_layer_id` equal to 0, with `pps_seq_parameter_set_id` equal to 0, and with `pps_pic_parameter_set_id` equal to 0.
- One 3D Reference Displays Information SEI message NAL unit (per G.14.2.3), with `nuh_layer_id` equal to 0, and with the following values:
  - To indicate the display width value is unspecified by using 0, use:
    - `prec_ref_display_width` = 31
    - `exponent_ref_display_width` = 0
    - `mantissa_ref_display_width` = 0
  - `ref_viewing_distance_flag` = 0
  - `prec_ref_viewing_dist` syntax element skipped
  - `num_ref_displays_minus1` = 0
  - `left_view_id[ 0 ]` is set to reference the texture layer used for the viewer's left eye using values in `vps_extension` (per F.7.3.2.1.1)
  - `right_view_id[ 0 ]` is set to reference the texture layer used for the viewer's right eye using values in `vps_extension` (per F.7.3.2.1.1)
  - `additional_shift_present_flag[ 0 ]` = 0
  - `exponent_ref_viewing_distance[ 0 ]`, `mantissa_ref_viewing_distance[ 0 ]`, `num_sample_shift_plus512[ 0 ]` syntax elements are skipped
- The L-HEVC Decoder Configuration Record ('lhvC') shall contain the following parameter sets corresponding to the secondary stereo eye:
  - One Sequence Parameter Set (SPS):
    - The `nuh_layer_id` shall be equal to the *secondary eye nuh\_layer\_id*, and with `sps_seq_parameter_set_id` equal to a non-zero number referred to here as the *secondary eye SPS ID*.
    - It is recommended to use 1 for the *secondary eye SPS ID*.
    - Readers should be prepared to accept a non-zero value for the *secondary eye layer SPS ID*, and recognize the value in the secondary eye PPS's `pps_seq_parameter_set_id` field.
  - One Picture Parameter Set (PPS):
    - The `nuh_layer_id` shall be equal to the *secondary eye nuh\_layer\_id*, and with `pps_seq_parameter_set_id` equal to the *secondary eye SPS ID*, and with `pp-`

`s_pic_parameter_set_id` equal to a non-zero number referred to here as the *secondary eye PPS ID*.

- It is recommended that writers use 1 for the *secondary eye PPS ID*.
- Readers should be prepared to accept a non-zero value for the *secondary eye PPS ID*, and recognize the value in the secondary eye portion of video frames, in video slice `slice_pic_parameter_set_id` fields.

## Constraints on Video Frame Samples

- Every video frame in the video track shall contain a HEVC compatible base layer NAL unit sequence and a secondary stereo eye layer NAL unit sequence. While the frame type and reference scheme differ between the base and secondary layer, dependency information shall stay the same.
- Both layers of every video frame shall have the same frame type and dependency structure. (This is necessary because there is only one sample table for the track, and readers need to use that single sample table's set of dependency information to determine which frames need to be decoded during random access or any kind of trick play.)
- All base layer NAL units shall have `nuh_layer_id` 0.
- All secondary stereo layer NAL units shall have `nuh_layer_id` equal to the *secondary eye nuh\_layer\_id* specified in the VPS and `vps_extension`.
- All video slices in the base layer NAL units shall have `slice_pic_parameter_set_id` set to the the base layer PPS ID (0).
- All video slices in the secondary stereo view layer NAL units shall have `slice_pic_parameter_set_id` set to the the *secondary eye layer PPS ID*.



## Diagram of decoder configurations and a sample

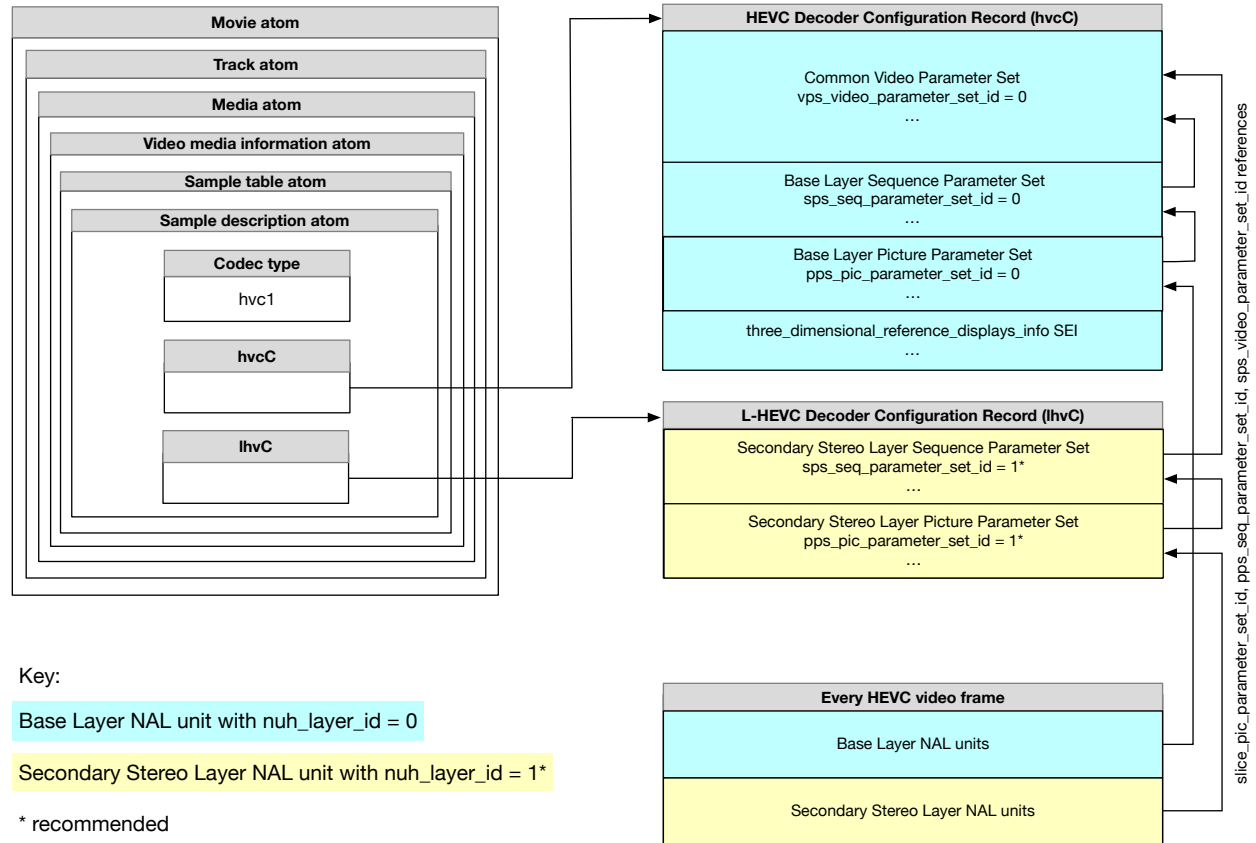


Figure 1. HEVC and L-HEVC Decoder Configuration Records, Parameter Sets and Video Frames

## File Type Brand

Movie files and MPEG-4 files containing tracks conforming to this profile are not required to signal this by including a specific minor brand in the File Type Brand atom but are not restricted in doing so.

# Annex A—Establishing left and right view IDs

## Overview

The two images corresponding to the stereo left eye and stereo right eye are signaled in the `three_dimensional_reference_displays_info` SEI using the `left_view_id[0]` and `right_view_id[0]` syntax elements. While the HEVC specification [HEVC] provides detail sufficient to specify relevant elements of the Video Parameter Set (VPS) extension, it might not be obvious. This informative annex attempts to describe the mapping from `nuh_layer_id` values for the left and right eye layers to the relevant values for `left_view_id[0]` and `right_view_id[0]`. A layer is used for each of the two stereo images and specified with one of two assigned `nuh_layer_id` values that occur in relevant syntax elements such as NAL units. One layer is termed the *base layer* corresponding to the base stereo layer. The other is the *secondary stereo layer* and corresponds to the secondary stereo eye. The base layer and view may be the left or the right stereo eye with the secondary stereo layer being for the other stereo eye. Requirements and recommendations are:

- All base layer NAL units shall have `nuh_layer_id` 0. This is the *base layer nuh\_layer\_id*.
- All secondary stereo layer NAL units shall have `nuh_layer_id` equal to the secondary eye `nuh_layer_id`. This is the *secondary layer nuh\_layer\_id*. It is recommended that this be 1 for writers but other values can be chosen. Readers must be prepared for any value of secondary stereo layer `nuh_layer_id`.

In the HEVC specification [HEVC], Section G.14.2.3 *3D reference displays information* SEI message syntax carries `left_view_id[]` and `right_view_id[]` fields that must be established to identify the relevant portions of video frame samples for each stereo eye to be processed to produce the corresponding stereo eye image. These are established by syntax elements of the Video Parameter Set extension.

Section F.7.3.2.1.1 *Video parameter set extension syntax* [HEVC] carries the relevant syntax elements to establish the video is multiview and to indicate the left and right view IDs. These VPS extension syntax elements are used in the algorithm description:

- `layer_id_in_nuh[]`
- `scalability_mask_flag[]`
- `dimension_id [][]`
- `view_id_val[]`
- `direct_dependency_flag[][]`

Other intermediate values are calculated for the purposes of the algorithm described next.

## Algorithm description

1. Associate each layer's `nuh_layer_id` to a layer number:

- `layer_id_in_nuh[0] = 0` where the assignment of 0 is for the base layer `nuh_layer_id` and the indexed 0 is the 0th layer index

- $layer\_id\_in\_nuh[1] = secondary\ nuh\_layer\_id$
2. Establish a reverse mapping by computing an algorithm  $LayerIdxInVps[nuh\_layer\_id]$  to map a  $nuh\_layer\_id$  to the layer index from  $layer\_id\_in\_nuh[ ]$ :
    - $LayerIdxInVps[0] = 0$
    - $LayerIdxInVps[secondary\ nuh\_layer\_id] = 1$
  3. Specify  $scalability\_mask\_flag[ ]$  values to indicate that it is multiview:
    - $scalability\_mask\_flag[1] = 1$  where the index is for Multiview
  4. Associate each scalability value to each Layer number using  $dimension\_id[Layer\ \#][Scalability\ \#]$ :
    - $dimension\_id[0][1] = 0$  to indicate this is a new view order index (i.e., 0)
    - $dimension\_id[1][1] = 1$  to indicate this is a new view order index (i.e., 1)
  5. Associate each view order index  $ViewOrderIdx$  to arbitrary  $view\_id\_val$  using  $view\_id\_val[ViewOrderIdx]$ :
    - $view\_id\_val[0] = base\ view\ ID$
    - $view\_id\_val[1] = secondary\ view\ ID$

Assigning 0 to *base view ID* and 1 to *secondary view ID* is recommended but not required.

Also, setting the  $left\_view\_id[0]$  as the base view ID and  $right\_view\_id[0]$  as the secondary view ID of the `three_dimensional_reference_displays_info` SEI is recommended but again not required.
  6. Compute an algorithm  $ViewId[nuh\_layer\_id]$  corresponding to  $ViewId[ nuhLayerId ] = view\_id\_val[ViewOrderIdx[nuhLayerId]]$ :
    - $ViewId[0] = base\ view\ ID$
    - $ViewId[secondary\ nuh\_layer\_id] = secondary\ view\ ID$
  7. Specify the base view ID and secondary view ID in the `three_dimensional_reference_displays_info` SEI  $left\_view\_id[0]$  and  $right\_view\_id[0]$  syntax elements:
    - $left\_view\_id[0] = base\ view\ ID$

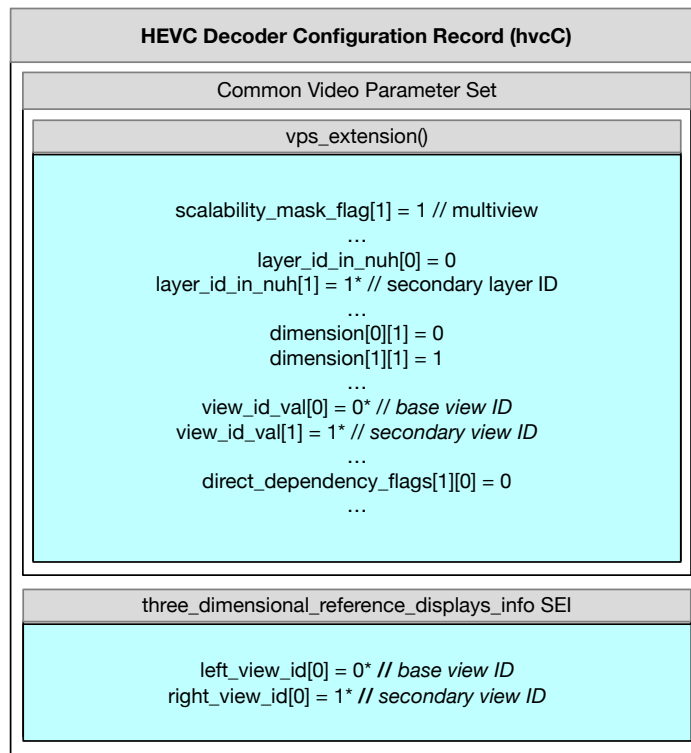
June 21, 2023

- `right_view_id[0]` = *secondary view ID*

As mentioned in step 5, assigning 0 to *base view ID* and 1 to *secondary view ID* is recommended but not required. Also, setting the `left_view_id[0]` as the base view ID or the recommended 0 and `right_view_id[0]` as the secondary view ID or the recommended 1 is recommended but not required.

Figure A-1 shows the `vps_extension()` and `three_dimensional_reference_displays_info` SEI fields set with `nuh_layer_id = 0` and `nuh_layer_id = 1` for the base and secondary layers and assignments of a *base video ID* and *secondary view ID*.

## Diagram of view ID connection



**Key:**

Base Layer NAL unit with nuh\_layer\_id = 0

Secondary Stereo Layer NAL unit with nuh\_layer\_id = 1\*

Base view ID = 0\*

Secondary view ID = 1\*

\* recommended

Figure A-1. Setting up left\_view\_id[0] and right\_view\_id[0] in HEVC Decoder Configuration Record

# Document Revision History

This table describes the changes to Apple *HEVC Stereo Video Interoperability Profile*

Date	Revision	Notes
2023-06-21	0.9	First publication