

# Draft: VP Codec ISO Media File Format Binding

Kilroy Hughes and David Ronca

2015.12.3

## Introduction

This document specifies a general ISO Base Media track and sample format for video encoded with Video Partition structured video codecs (“VP”), such as MPEG VCB (MPEG-4 Part 31), VP8, VP9, etc.

## Normative References

ISO/IEC 14496-12, Information technology — Coding of audio-visual objects — Part 12: ISO base media file format.

ISO/IEC 23001-7 second edition 2015-0401, Part 7: Information technology — MPEG systems technologies — Common encryption in ISO base media file format files

VP9 Bitstream and Decoding Process

## VP Codec Sample Entry Box

This section describes the sample entry and sample format for VP elementary streams.

### Definition

Box Type: ‘vpxx’ where ‘xx’ is one of ‘08’, ‘09’ or ‘10’

Container: Sample Description Box (‘stsd’)

Mandatory: Yes, for VP codec tracks

Quantity: Exactly One

The ‘vpxx’ Sample Entry Box specifies the coding of Video Partition Codec samples, and contains a ‘vpcC’ box that contains decoding and display configuration information. ‘vpxx’ indicates the generic class used to generate a box instance identified by the 4CC of the specific codec used. The 4CC codes currently defined by this spec are ‘vp08’, ‘vp09’, ‘vp10’.

### Syntax

```
class VP8SampleEntry extends VisualSampleEntry('vp08') {
    VPCodecConfigurationBox    config;
}

class VP9SampleEntry extends VisualSampleEntry('vp09') {
    VPCodecConfigurationBox    config;
}
```

```
class VP10SampleEntry extends VisualSampleEntry('vp10') {
    VPCodecConfigurationBox    config;
}
```

## Semantics

**compressorname** is a name, for informative purposes. It is formatted in a fixed 32-byte field, with the first byte set to the number of bytes to be displayed, followed by that number of bytes of displayable data, and then padding to complete 32 bytes total (including the size byte). The field may be set to 0. The value "\012VPC Coding" is recommended; the first byte is a count of the remaining bytes, here represented by \012, which (being octal 12) is 10 (decimal), the number of bytes in the rest of the string

**config** is defined in the following section.

## VP Codec Configuration Box

### Definition

Box Type: 'vpcC'  
 Container: VP Codec Sample Entry Box ('vpxx')  
 Mandatory: Yes  
 Quantity: Exactly One

### Description

The VP Codec Configuration Box is contained in every VP Codec Sample Entry Box. It exposes the general video parameters in standard fields, useful for track selection and display; and it contains decoder initialization information specific to the codec and sample format indicated by the 4CC code of the sample entry box that contains it. All parameters must be valid for every sample that references the sample entry, and equal the parameter value unless otherwise noted.

### Syntax

```
class VPCodecConfigurationBox extends FullBox('vpcC', version, 0){
    VPCodecConfigurationRecord() vpcConfig;
}
```

```
aligned (8) class VPCodecConfigurationRecord {
    unsigned int (8)    profile;
    unsigned int (8)    level;
    unsigned int (4)    bitDepth;
    unsigned int (4)    colorSpace;
    unsigned int (4)    chromaSubsampling;
    unsigned int (3)    transferFunction;
    unsigned int (1)    videoFullRangeFlag;
```

```

    unsigned int (16)    codecIntializationDataSize;
    unsigned int (8) []  codecIntializationData;
}

```

## Semantics

**profile** is an integer that specifies the VP codec profile. The value of profile must be valid for all samples that reference this sample entry, i.e. profile SHALL be equal or greater than the profile used to encode the sample.

**level** is an integer that specifies a VP codec level all samples conform to. The value is 0 if a codec level is not specified.

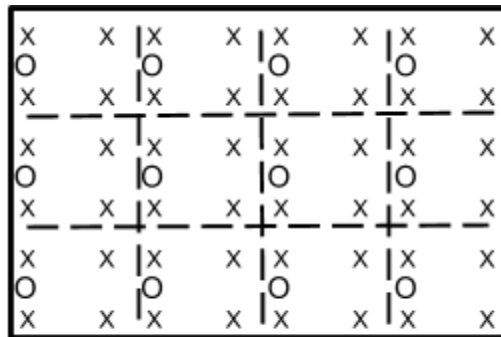
**bitDepth** is an integer that specifies the bit depth of the luma and color components. Valid values are 8, 10, 12.

**colorSpace** is an integer that specifies the color space of the video, enumerated in the following table:

Value	Color Space
0	Unspecified
1	Rec. ITU-R BT.601-7
2	Rec. ITU-R BT.709-6
3	SMPTE-170
4	SMPTE-240
5	Rec. ITU-R BT.2020 non-constant luminance
6	Rec. ITU-R BT. 2020 constant luminance
7	IEC 61966-2-1 (sRGB)
8..15	Reserved

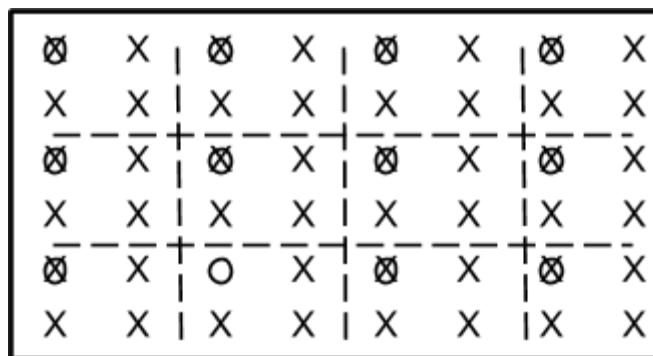
**chromaSubsampling** is an integer that specifies the chroma subsampling. Only the values in the following table are specified. If colorspace is 4 (RGB), then chroma subsampling must be 4 (4:4:4).

value	subsampling
0	4:2:0 vertical
1	4:2:0 collocated with luma (0,0)
2	4:2:2
3	4:4:4
4...15	Reserved



X represents luma samples  
O represents chroma samples

Figure: 4:2:0 Subsampling with vertical chroma samples



X represents luma samples  
O represents chroma samples

Figure: 4:2:0 chroma subsampling collocated with (0,0) luma

**transferFunction** is an integer that specifies the transfer function. Only the values in the following table are specified.

value	Transfer Function
0	Rec. ITU-R BT.709-6, Rec. ITU-R BT.601-7 525 or 625, Rec. ITU-R BT.2020.
1	SMPTE ST 2084:2014
2...7	Reserved

**videoFullRangeFlag** indicates the black level and range of the luma and chroma signals. 0 = legal range (e.g. 16-235 for 8 bit sample depth) 1 = full range (e.g. 0-255 for 8 bit sample depth).

**codecInitializationDataSize** For VP8 and VP9 this field must be 0.

**codecInitializationData** binary codec initialization data. Not used for VP8 and VP9.

## Video Samples

Video sample storage in the generic binding uses a simple mapping to frames defined in the codec specification.. The height and width in the Visual Sample Entry are specified in square pixels. If the video pixels are not square, then a 'pasp' box must be included. ALTREF frames must be part of a superframe structure.

Note: VP8 does not support superframes, and so it is not possible to carry VP8 using this specification if the VP8 stream includes ALTREF frames<sup>1</sup>.

## Common Encryption

### Scheme Info Box (sinf)

If the VP9 data is encrypted, the Protection Scheme Info box ('sinf') shall be present, and shall contain a Scheme Type ('schm') box. The scheme\_type field of the 'schm' box shall be 'cenc', indicating that AES-CTR encryption is used when samples are encrypted.

### Sample Encryption

VP8/9 samples packaged using this specification use sub-sample encryption as specified in section 10.6 of "ISO/IEC 23001-7 Part 7: Common encryption in ISO base media file format files". The subsample encryption table may be implemented using the 'senc' box described in section 8.1 of "ISO/IEC 23001-7 Part 7" or the 'saio' and 'saiz' boxes described in section 8.7 of "14496-12".

---

<sup>1</sup> A model for carriage of VP8 ALTREF frames may be considered for a future version of this specification.

When encrypting VP9 video frames, the uncompressed header must be unencrypted. A subsample encryption (SENC) map must be used to identify the clear and encrypted bytes of each video sample. This is illustrated in figure 1.

When encrypting superframes, the uncompressed headers of the displayed frame, the uncompressed headers for all ALTREF frames, and the superframe header must be clear. The encrypted bytes of each frame within the superframe must be block aligned so that the counter state can be computed for each frame within the superframe. Block alignment is achieved by adjusting the size of the unencrypted bytes that precede the encrypted bytes for that frame.

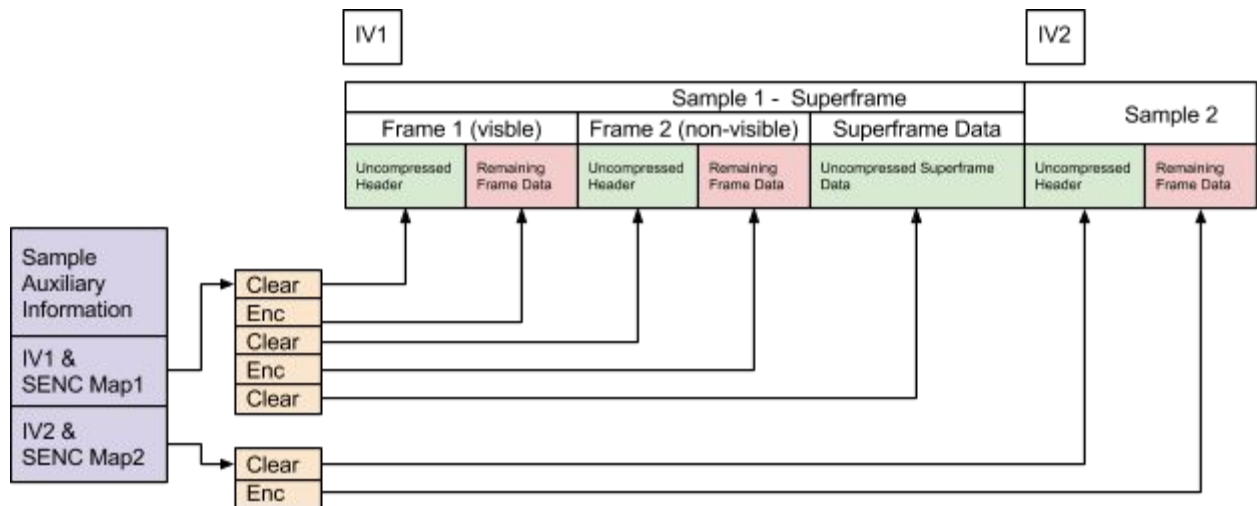


Figure 1: Sample-based VP9 encryption with clear uncompressed header

## DASH Application

DASH and other applications require defined values for the “codecs” parameter specified in RFC-6381 for ISO Media tracks. A Suggested codecs parameter of VP codecs is:

<sample entry

4CC>.<profile>.<level>.<bitDepth>.<colorSpace>.<chromaSubsampling>.<transferFunction>.<videoFullRangeFlag>

Numbers are expressed in decimal. The string may be truncated on any parameter in sequence following the sample entry, and missing values are indicated by a sequence of two periods with no parameter value between them.

For example:

```
codecs="vp09.01.01.02.01.01.00"
```

to indicate 10 bit 4:2:0 Rec. ITU-R BT.2020 video encoded using VP9 profile 1 and level 1, 4:2:0 collocated subsampling, st02-84 EOTF, and

```
codecs="vp09" to indicate only the codec and sample format.
```

