DRAFT AMENDMENT
ISO/IEC 13818-1:2019 DAM 2

Information technology — Generic coding of moving pictures and associated audio information —

Part 1:
Systems

AMENDMENT 2: Carriage of VVC in MPEG-2 Systems

Partie 1: Systèmes

AMENDEMENT 2: Titre manque

ICS: 35.040.40

This document is circulated as received from the committee secretariat.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO/IEC 13818-1:2019/DAmd 2
https://standards.iteh.ai/catalog/standards/sist/2d958430-3878-48f1-8a07-
f96412311412/iso-iec-13818-1-2019-damd-2

COPYRIGHT PROTECTED DOCUMENT

Information technology — Generic coding of moving pictures and associated audio information —

Part 1:

Systems

AMENDMENT 2: Carriage of VVC in MPEG-2 Systems

In 1.2.2, add the following references:

– Recommendation ITU-T H.26X, Versatile video coding

– ISO/IEC 23090-3, Information technology – Coded Representation of Immersive Media – Part 3: Versatile video coding

After 2.1.136, add the following definitions

2.1.137 VVC video stream: A byte stream as specified in Rec. ITU-T H. 26X | ISO/IEC 23090-3 Annex B.

2.1.138 VVC access unit: An access unit as defined in Rec. ITU-T H. 26X | ISO/IEC 23090-3 with the constraints specified in section 2.23.1.

2.1.139 VVC 24-hour picture (system): An VVC access unit with a presentation time that is more than 24 hours in the future. For the purpose of this definition, VVC access unit n has a presentation time that is more than 24 hours in the future if the difference between the initial arrival time $t_{ai}(n)$ and the DPB output time $t_{o,dpb}(n)$ is more than 24 hours.

2.1.140 VVC slice: A slice as specified in Rec. ITU-T H. 26X | ISO/IEC 23090-3.

2.1.141 VVC subpicture: A subpicture as specified in Rec. ITU-T H. 26X | ISO/IEC 23090-3.

2.1.142 VVC tile of slices: One or more consecutive VVC slices which form the coded representation of a tile as defined in Rec. ITU-T H.26X | ISO/IEC 23090-3.

2.1.143 VVC still picture (system): A VVC still picture consists of a VVC access unit containing an IDR picture preceded by VPS, SPS and PPS NAL units, as defined in Rec. ITU-T H.26X | ISO/IEC 23090-3, that carry sufficient information to correctly decode this IDR picture. Preceding a VVC still picture, there shall be another VVC still picture or an End of Sequence NAL unit terminating a preceding coded video sequence as defined in Rec. ITU-T H.26X | ISO/IEC 23090-3.

2.1.144 VVC video sequence (system): coded video sequence as defined in Rec. ITU-T H.26X | ISO/IEC 23090-3.

2.1.145 VVC video sub-bitstream: A subset of the NAL units of a VVC video stream in their original order.

2.1.146 VVC temporal video sub-bitstream: An VVC video sub-bitstream that contains all VCL NAL units and associated non-VCL NAL units of the temporal sub-layer, as specified in Rec. ITU-T <mark>H.26X</mark> | ISO/IEC 23090-3, associated to TemporalId equal to 0 and which may additionally contain all VCL NAL units and associated non-VCL NAL units of all temporal sub-layers associated to a contiguous range of TemporalId from 1 to a value equal to or smaller than sps_max_sub_layers_minus1 included in the referred sequence parameter set, as specified in Rec. ITU-T <mark>H.26X</mark> | ISO/IEC 23090-3.

2.1.147 VVC temporal video subset: An VVC video sub-bitstream that contains all VCL NAL units and the associated non-VCL NAL units of one or more temporal sub-layers, as specified in Rec. ITU-T <mark>H.26X</mark> | ISO/IEC 23090-3, with each temporal sub-layer not being present in the corresponding VVC temporal video sub-bitstream and TemporalId associated with each temporal sub-layer forming a contiguous range of values.

NOTE – According to the constraints for the transport of VVC specified in 2.23.1, each temporal sub-layer of a VVC video stream is present either in the VVC temporal video sub-bitstream or in exactly one VVC temporal video subset which are carried in a set of elementary streams that are associated by hierarchy descriptors. This prevents multiple inclusion of the same temporal sub-layer and allows aggregation of the VVC temporal video sub-bitstream with associated VVC temporal video subsets according to the hierarchy descriptors as specified in 2.23.3.

2.1.148 VVC highest temporal sub-layer representation: The sub-layer representation of the temporal sub-layer with the highest value of TemporalId, as defined in Rec. ITU-T <mark>H.26X</mark> | ISO/IEC 23090-3, in the associated VVC temporal video sub-bitstream or VVC temporal video subset.

2.1.149 VVC complete temporal representation: A sub-layer representation as defined in Rec. ITU-T <mark>H.26X</mark> | ISO/IEC 23090-3 that contains all temporal sub-layers up to the temporal sub-layer with TemporalId equal to sps_max_sub_layers_minus1+1 as included in the referred sequence parameter set, as specified in Rec. ITU-T <mark>H.26X</mark> | ISO/IEC 23090-3.

In 2.4.2.7, replace the following 2 paragraphs.

Replace:

The delay of any data through the System Target Decoder buffers shall be less than or equal to one second except for still picture video data, ISO/IEC 14496 and ISO/IEC 23008-2 streams. Specifically: $td_n(j) – t(i) \le 1$ second for all j, and all bytes i in access unit $A_n(j)$.

with:

The delay of any data through the System Target Decoder buffers shall be less than or equal to one second except for still picture video data, ISO/IEC 14496, and ISO/IEC 23008-2 and ISO/IEC 23090-3 streams. Specifically: $td_n(j) – t(i) \le 1$ second for all j, and all bytes i in access unit $A_n(j)$.

Replace:

For ISO/IEC 14496 and ISO/IEC 23008-2 streams, the delay is constrained by $td_n(j) – t(i) \le 10$ seconds for all j, and all bytes i in access unit $A_n(j)$.

with:

For ISO/IEC 14496, ISO/IEC 23008-2 and ISO/IEC 23090-3 streams, the delay is constrained by $td_n(j) – t(i) \le 10$ seconds for all j, and all bytes i in access unit $A_n(j)$.

After 2.4.2.15, add 2.4.2.16:

2.4.2.16, T-STD extensions for carriage of VVC

T-STD extensions and T-STD parameters for decoding of VVC video streams are defined in 2.23.2 and 2.23.3. Program stream support including P-STD extensions and P-STD parameters are not specified for VVC video streams.

In 2.4.3.5, in the section specifying the discontinuity_indicator, at the end of the bulleted list introduced by "For the purpose of this clause, an elementary stream access point is defined as follows", add:

- VVC video streams or VVC temporal video sub-bitstreams – The first byte of a VVC access unit. The DPS, VPS, SPS and PPS parameter sets, as defined in Rec. ITU-T H.26X | ISO/IEC 23090-3, referenced in this and all subsequent VVC access units in the VVC video sequence shall be provided after this access point in the byte stream and prior to their activation.

In 2.4.3.5, in the section specifying the elementary_stream_priority_indicator, add:

In the case of VVC video streams or VVC temporal video sub-bitstreams or VVC temporal video subsets, this field may be set to '1' only if the payload contains one or more bytes from a slice with slice_type set to 2. A value of '0' indicates that the payload has the same priority as all other packets which do not have this bit set to '1'.

In 2.4.3.7, in Table 2-22, Stream_id assignments, replace the following line:

| 1110 xxxx | Rec. ITU-T H.262 \| ISO/IEC 13818-2, ISO/IEC 11172-2, ISO/IEC 14496-2, Rec. ITU-T H.264 \| ISO/IEC 14496-10 or Rec. ITU-T H.265 \| ISO/IEC 23008-2 video stream number xxxx |

With

| 1110 xxxx | Rec. ITU-T H.262 \| ISO/IEC 13818-2, ISO/IEC 11172-2, ISO/IEC 14496-2, Rec. ITU-T H.264 \| ISO/IEC 14496-10, Rec. ITU-T H.265 \| ISO/IEC 23008-2 or Rec. ITU-T H.26X \| ISO/IEC 23090-3 video stream number xxxx |

In 2.4.3.7, in the section specifying the PTS (presentation time stamp), add:

For VVC video streams, VVC temporal video sub-bitstreams and VVC temporal video subsets, if a PTS is present in the PES packet header, it shall refer to the first VVC access unit that commences in this PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of Rec. ITU-T H.26X | ISO/IEC 23090-3, for each VVC access unit the PTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal DPB output time in the HRD, as defined in Annex C of Rec. ITU-T H.26X | ISO/IEC 23090-3.

3

In 2.4.3.7, in the section specifying the DTS (decoding time stamp), add:

For VVC video streams, VVC temporal video sub-bitstreams and VVC temporal video subsets, if a DTS is present in the PES packet header, it shall refer to the first VVC access unit that commences in this PES packet. To achieve consistency between the STD model and the HRD model defined in Annex C of Rec. ITU-T H.26X | ISO/IEC 23090-3, for each VVC access unit the DTS value in the STD shall, within the accuracy of their respective clocks, indicate the same instant in time as the nominal CPB removal time t_r in the HRD, as defined in Annex C of Rec. ITU-T H.26X | ISO/IEC 23090-3.

In 2.4.4.9, in Table 2-34, Stream type assignments, replace the following line:

0x32-0x7E	ITU-T Rec. H.222.0	ISO/IEC 13818-1 Reserved

with:

0x32	VVC video stream or an VVC temporal video sub-bitstream conforming to one or more profiles defined in Annex A of Rec. ITU-T H.26X	ISO/IEC 23090-3
0x33	VVC temporal video subset of an VVC video stream conforming to one or more profiles defined in Annex A of Rec. ITU-T H.26X	ISO/IEC 23090-3
0x34-0x7E	ITU-T Rec. H.222.0	ISO/IEC 13818-1 Reserved

In 2.6.1, in Table 2-45, replace the following line:

57-62	n/a	n/a	Rec. ITU-T H.222.0	ISO/IEC 13818-1 Reserved

With:

57	X	n/a	VVC video descriptor
58-62	n/a	n/a	Rec. ITU-T H.222.0	ISO/IEC 13818-1 Reserved

In 2.6.6, in Table 2-50, replace the description for value 15:

Table 2-50 – Hierarchy_type field values

Value	Description
15	Base layer or MVC base view sub-bitstream or AVC video sub-bitstream of MVC or HEVC temporal video sub-bitstream or HEVC base sub-partition or Base layer of MVCD base view sub-bitstream or AVC video sub-bitstream of MVCD or VVC temporal video sub-bitstream.

In 2.6.11, add the following immediately after Table 2-56:

Table 2-xx describes the alignment type for VVC when the data_alignment_indicator in the PES packet header has a value of '1'.

Table 2-xx – VVC video stream alignment values

Alignment type	Description
00	Reserved
01	VVC access unit
02	VVC slice
03	VVC access unit or slice
04	VVC tile of slices
05	VVC access unit or tile of slices
06	VVC slice or tile of slices
07	VVC access unit or slice or tile of slices
08	VVC subpicture
09	VVC access unit or subpicture
10	VVC slice or subpicture
11	VVC subpicture or tile of slices
12	VVC access unit or slice or subpicture
13	VVC access unit or subpicture or tile of slices
14	VVC slice or subpicture or tile of slices
15	VVC access unit or slice or subpicture or tile of slices
16-255	Reserved

In 2.6.88, in Table 2-107, add the following extension_descriptor_tag for VVC timing_and_HRD_descriptor():

Table 2-107 – Extension descriptor

Syntax	No. of bits	Mnemonic
Extension_descriptor () {		
...		
else if (extension_descriptor_tag == 0x14) {		
VVC_timing_and_HRD_descriptor()		
}		
...		

In 2.6.89, before Table 2-108, add:

VVC_timing_and_HRD_descriptor() – This structure is defined in 2.6.127 and 2.6.128.

In 2.6.89, in Table 2-108, replace the following line:

0x14-0xFF	n/a	n/a	Rec. ITU-T H.222.0	ISO/IEC 13818-1 Reserved

with

0x14	X	n/a	VVC_timing_and_HRD_descriptor()
0x15-0xFF	n/a	n/a	Rec. ITU-T H.222.0	ISO/IEC 13818-1 Reserved

After 2.6.126 add 2.6.127 to 2.6.130:

2.6.127 VVC video descriptor

For a VVC video stream, the VVC video descriptor provides basic information for identifying coding parameters, such as profile and level parameters, of that VVC video stream. For an VVC temporal video sub-bitstream or a VVC temporal video subset, the VVC video descriptor provides information such as the associated VVC highest temporal sub-layer representation contained in the elementary stream to which it applies.

Table 2-129bis – VVC video descriptor

Syntax	No. Of bits	Mnemonic
VVC_descriptor() {		
descriptor_tag	8	uimsbf
descriptor_length	8	uimsbf
profile_idc	7	uimsbf
tier_flag	1	bslbf
num_sub_profiles	8	uimsbf
for (i=0; i<num_sub_profiles; i++) {		
sub_profile_idc[i]	32	uimsbf
}		
progressive_source_flag	1	bslbf
interlaced_source_flag	1	bslbf
non_packed_constraint_flag	1	bslbf
frame_only_constraint_flag	1	bslbf
no_mixed_nalu_types_in_pic_constraint_flag	1	bslbf
reserved_zero_3bits	3	bslbf
level_idc	8	uimsbf
temporal_layer_subset_flag	1	bslbf
VVC_still_present_flag	1	bslbf
VVC_24hr_picture_present_flag	1	bslbf
reserved	5	bslbf
HDR_WCG_idc	2	uimsbf
reserved	2	bslbf
video_properties_tag	4	uimsbf
if (temporal_layer_subset_flag == '1') {		
reserved	5	bslbf
temporal_id_min	3	uimsbf
reserved	5	bslbf
temporal_id_max	3	uimsbf
}		
}		

iTeh STANDARD PREVIEW
(standards.iteh.ai)
ISO/IEC 13818-1:2019/DAmd 2
https://standards.iteh.ai/catalog/standards/sist/2d958430-3878-48f1-8a07-
f96412311412/iso-iec-13818-1-2019-damd-2

2.6.128 Semantic definition of fields in VVC video descriptor

profile_idc, tier_flag, num_sub_profiles, sub_profile_idc[i], progressive_source_flag, interlaced_source_flag, non_packed_constraint_flag, frame_only_constraint_flag, reserved_zero_Xbits – These fields shall be coded according to the semantics defined in Rec. ITU-T H.26X | ISO/IEC 23090-3 for general_profile_idc, general_tier_flag, num_sub_profiles, general_sub_profile_idc[i], general_progressive_source_flag, general_interlaced_source_flag,

7

general_non_packed_constraint_flag, general_frame_only_constraint_flag, further bits in general_constraint_info(), respectively, for the corresponding VVC video stream or VVC complete temporal representation. When the corresponding VVC video stream or VVC complete temporal representation contains a DPS, the values of these fields shall be coded exactly the same as the corresponding fields in the DPS.

no_mixed_nalu_types_in_pic_constraint_flag – This 1-bit flag, when set to '1', indicates that it is a requirement of bitstream conformance that mixed_nalu_types_in_pic_flag shall be equal to 0. If no_mixed_nalu_types_in_pic_constraint_flag is set to '0', no such a constraint is imposed.

level_idc – When the VVC video descriptor applies to a VVC video stream or to a VVC complete temporal representation, this field shall be coded according to the semantics defined in Rec. ITU-T H.26X | ISO/IEC 23090-3 for general_level_idc, for the corresponding VVC video stream or VVC complete temporal representation. When the VVC video descriptor applies to an VVC temporal video sub-bitstream or VVC temporal video subset of which the corresponding VVC highest temporal sub-layer representation is not an VVC complete temporal representation, this field shall be coded according to semantics defined in Rec. ITU-T H.26X | ISO/IEC 23090-3 for sub_layer_level_idc, for the corresponding VVC highest temporal sub-layer representation. When the corresponding VVC video stream or VVC complete temporal representation contains a DPS, the values of this field shall be coded exactly the same as the corresponding fields in the DPS.

When the VVC video descriptor applies to a VVC video stream or to a VVC complete temporal representation, the entire VVC video stream or VVC complete temporal representation to which the VVC video descriptor is associated shall conform to the information signaled by these fields. When the VVC video descriptor applies to an VVC temporal video sub-bitstream or VVC temporal video subset of which the corresponding VVC highest temporal sub-layer representation is not an VVC complete temporal representation, the entire VVC highest temporal sub-layer representation to which the VVC video descriptor is associated shall conform to the information signalled by these fields.

NOTE – In one or more sequences in the VVC video stream the level may be lower than the level signalled in the VVC video descriptor, while also a profile may occur that is a subset of the profile signalled in the VVC video descriptor. However, in the entire VVC video stream, only subsets of the entire bitstream syntax shall be used that are included in the profile signalled in the VVC video descriptor, if present. When the VVC video stream contains a DPS NAL unit, the profile-tier-level indication signals the profile and level information to which all sequences in the VVC video stream conforms. When not, if the sequence parameter sets in an VVC video stream signal different profiles, and no additional constraints are signalled, then the stream may need examination to determine which profile, if any, the entire stream conforms to.

temporal_layer_subset_flag – This 1-bit flag, when set to '1', indicates that the syntax elements describing a subset of temporal layers are included in this descriptor. This field shall be set to 1 for VVC temporal video subsets and for VVC temporal video sub-bitstreams. When set to '0', the syntax elements temporal_id_min and temporal_id_max are not included in this descriptor.

VVC_still_present_flag – This 1-bit field, when set to '1', indicates that the VVC video stream or the VVC highest temporal sub-layer representation may include VVC still pictures. When set to '0', then the associated VVC video stream shall not contain VVC still pictures.

NOTE – According to Rec. ITU-T H.26X | ISO/IEC 23090-3, IDR pictures are always associated to a TemporalId value equal to 0, Consequently, if the VVC video descriptor applies to a VVC temporal video subset, VVC still pictures can only be present in the associated VVC temporal video sub-bitstream.

VVC_24_hour_picture_present_flag – This 1-bit flag, when set to '1', indicates that the associated VVC video stream or the VVC highest temporal sub-layer representation may contain VVC 24-hour pictures.

For the definition of an VVC 24-hour picture, see 2.1.X. If this flag is set to '0', the associated VVC video stream shall not contain any VVC 24-hour picture.

HDR_WCG_idc – This 2-bit field indicates the presence or absence of high dynamic range (HDR) and/or wide color gamut (WCG) video components in the associated PID according to Table 2-112. This field shall not be set to 2 unless bit_depth_luma_minus8 as defined in Rec. ITU-T H.265 | ISO/IEC 23008-2 in the associated video is greater than or equal to 2. This field also shall not be set to 1 or 2 unless bit_depth_chroma_minus8 as defined in [VVC Spec] in the associated video is greater than or equal to 2.

Table 2-129ter – Semantics of HDR_WGC_idc

HDR_WCG_idc	Description
0	SDR, i.e., video is based on the Rec. ITU-R BT.1886 reference EOTF with a color gamut that is contained within Rec. ITU-R BT.709 with a Rec. ITU-R BT.709 container (see Note 1)
1	WCG only, i.e., video color gamut in a Rec ITU-R BT.2020 container that exceeds Rec. ITU-R BT.709 (see Note 2)
2	Both HDR and WCG are to be indicated in the stream (see Note 3)
3	No indication is made regarding HDR/WCG or SDR characteristics of the stream

NOTE 1 – An example where it would be desirable to set HDR_WCG_idc to 0 would be when the colour_description_present_flag, as defined in Rec. ITU-T H.SEI | ISO/IEC 23002-7, is set to '0', with colour_primaries and transfer_characteristics not present in the video stream.

NOTE 2 – An example where it would be desirable to set HDR_WCG_idc to 1 would be when colour_primaries as defined in Rec. ITU-T H.SEI | ISO/IEC 23002-7 (which refers to the ColourPrimaries parameter in Rec. ITU-T H.273 | ISO/IEC 23091-2) is equal to 9 to indicate Rec. ITU-R BT.2020.

NOTE 3 – An example where it would be desirable to set HDR_WCG_idc to 2 would be when transfer_characteristics as defined in Rec. ITU-T H.SEI | ISO/IEC 23002-7 (which refers to the TransferCharacteristics parameter in Rec. ITU-T H.273 | ISO/IEC 23091-2) is equal to 16 to indicate BT.2100 PQ EOTF or to 18 to indicate BT.2100 HLG EOTF, and when colour_primaries as defined in ISO/IEC 23002-7 is equal to 9 to indicate Rec. ITU-R BT.2020.

video_properties_tag – This 4-bit field is used to indicate specific widely used video property CICP combinations as indicated by [23091-4/H-Series Supp 19] for SDR, WCG, or HDR/WCG streams depending on the value of HDR_WCG_idc. When HDR_WCG_idc is equal to 0, Table 2-129quater applies. When HDR_WCG_idc is equal to 1, Table 2-129quinquies applies. When HDR_WCG_idc is equal to 2, Table 2-129sexies applies. When HDR_WCG_idc is equal to 3, Table 2-129septies applies.

Table 2-129quater – SDR widely used video property combinations

video_properties_tag	CICP Values - System Identifier [ColourPrimaries, TransferCharacteristics, MatrixCoefficients, VideoFullRangeFlag]
0	Video property CICP combination not specified or unknown
1	[1,1,1,0]- BT709_YCC
2	[1,1,0,0]- BT709_RGB
3	[6,6,6,0]- BT601_525
4	[5,6,5,0]- BT601_625
5	[1,1,0,1]- FR709_RGB
6-15	Reserved