**ISO/IEC 14496-15:20xx(E)**

ISO/IEC jTC 1/SC 29/WG 03

Date: 2022-05

**Information technology — Coding of audio-visual objects — Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format — Amendment 2**

WD stage

**Warning for WDs and CDs**

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

© ISO 20XX

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO’s member body in the country of the requester.

ISO copyright office

CP 401 • Ch. de Blandonnet 8

CH-1214 Vernier, Geneva

Phone: +41 22 749 01 11

Email: copyright@iso.org

Website: www.iso.org

Published in Switzerland

Information technology — Coding of audio-visual objects — Part 15: Carriage of network abstraction layer (NAL) unit structured video in the ISO base media file format — Amendment 2

*Add subclause 4.16 as follows:*

* 1. **Storage of SEI manifest and SEI prefix indication SEI messages**

When one or more SEI NAL units containing an SEI manifest SEI message and/or an SEI prefix indication SEI message are available, they should be stored in the decoder configuration information as follows:

* For AVC, SVC, and MVC, these SEI NAL units should be stored as instances of pictureParameterSetNALUnit in AVCDecoderConfigurationRecord, SVCDecoderConfigurationRecord, and MVCDecoderConfigurationRecord, respectively. [Ed. (YK): Regarding the SPS vs PPS array, no big difference, but since the SEI manifest and SEI prefix indication SEI messages are more of sequence-level than picture-level, it should be slightly better to put them in the SPS array. Currently, the above says the PPS array.]
* For HEVC and L-HEVC, these SEI NAL units should be stored as instances of nalUnit with NAL\_unit\_type indicating a prefix SEI NAL unit in HEVCDecoderConfigurationRecord and LHEVCDecoderConfigurationRecord, respectively.
* For VVC, these SEI NAL units should be stored as instances of nal\_unit with NAL\_unit\_type indicating a prefix SEI NAL unit in VvcDecoderConfigurationRecord.

When a parameter set elementary stream is in use for AVC, SVC, MVC, or MVD, SEI NAL units containing an SEI manifest SEI message and/or an SEI prefix indication SEI message should also be placed in the parameter set elementary stream.

*Add subclause 4.17 as follows:*

* 1. **Picture-in-picture region replacement sample group**

### Definition

Picture-in-picture (PiP) services offer the ability to include a video with a smaller spatial resolution within a video with a bigger spatial resolution, referred to as the PiP video and the main video, respectively. A video track containing a 'subt' track reference indicates that the track contains PiP video and the main video is contained in the referenced track or any track in the alternate group to which the referenced track belongs, if any. [Ed. (CC): The registration authority indicates that 'subt' is used for subtitles. 14496-12 says "*subtitle, timed text or overlay graphical information.* Is this the right use? (YK): Good question. At a previous MPEG meeting, I asked whether PiP can be considered as a type of overlay graphical information such that the 'subt' track reference can be used, and David (Singer) said yes. Maybe it's better to use a different 4CC for this, to avoid confusion. And to remove " *or overlay graphical information*" from 14496-12, now that does not appear in the registration authority.]

For each pair of PiP video and main video, using the tools defined in ISO/IEC 14496-12, the window in the main video for embedding/overlaying the PiP video, which is smaller in size than the main video, is indicated by the values of the matrix fields of the TrackHeaderBoxes of the PiP video track and the main video track, and the value of the layer field of the TrackHeaderBox of the PiP video track shall be less than that of the main video track, to layer the PiP video in front of the main video.

When PicInPicRegionReplacementEntry is present in a PiP video track, it indicates that the NAL units representing the target PiP region in the main video can be replaced with the corresponding NAL units of the PiP video track. In this case, it is required that the same video codec is used for coding of the PiP video and the main video. The absence of this sample group indicates that it is unknown whether such replacement is possible. [Ed. (CC): In that case do you need the track reference? (YK): I think so, as the track reference associates the two tracks and indicates that they together can be used to provide a PiP experience.]

When this sample group is present, the player may choose to replace the NAL units representing the target PiP region in the main video with the corresponding NAL units of the PiP video before sending to the video decoder for decoding. In this case, for a particular picture in the main video, the corresponding NAL units of the PiP video are all the NAL units in the decoding-time-synchronized sample in the PiP video track.

### Syntax

class PicInPicRegionReplacementEntry() extends VisualSampleGroupEntry ('pprr') {  
 unsigned int(8) region\_id\_type;  
 unsigned int(8) num\_region\_ids;  
 for(i=0; i<num\_region\_ids; i++)  
 unsigned int(16) region\_id[i];  
}

### Semantics

region\_id\_type indicates the type for the value taken by the region\_id. When region\_id\_type is equal to 0, the region IDs are VVC subpicture IDs. When region\_id\_type is equal to 1, the region IDs are the groupID values in the NAL unit map sample group for the NAL units that may be replaced by the NAL units of the PiP track. region\_id\_type values greater than 1 are reserved.

num\_region\_ids specifies the number of the following region\_id[i] fields.

region\_id[i] specifies the i-th ID for the NAL units representing the target picture-in-picture region.

When region\_id\_type is equal to 1, the main video track shall have a 'nalm' sample group with grouping\_type\_parameter equal to 'pprr' indicating the NAL units in the main track that may be replaced by the NAL units in the PiP track with the same groupID values.

When region\_id\_type is equal to 1 and num\_region\_ids is equal to 1, 'nalm' sample group shall not be present in the PiP track and all the NAL units of the PiP track are inferred to have groupID equal to region\_id[0].

When region\_id\_type is equal to 1 and num\_region\_ids is greater than 1, a 'nalm' sample group with grouping\_type\_parameter equal to 'pprr' shall be present in the PiP track and provide a mapping of groupID values to NAL units.

*In subclause 5.4.3.2.3, replace*

NALUnit contains a single NAL unit. The syntax of a NAL unit is defined in ISO/IEC 14496-10 and includes both the one byte NAL header and the variable length encapsulated byte stream payload.

*with*

NALUnit contains a single NAL unit. The syntax of a NAL unit is defined in ISO/IEC 14496-10 and includes both the one-byte NAL unit header and the variable length NAL unit payload.

*In subclause 8.3.2.1.1, replace*

* The level indication general\_level\_idc shall indicate a level of capability equal to or greater than the highest level indicated for the highest tier in all the parameter sets.

*with*

* The level indication general\_level\_idc shall indicate a level of capability greater than or equal to the highest level in all the parameter sets.

*In subclause 8.3.2.1.3, replace*

array\_completeness when equal to 1 indicates that all NAL units of the given type are in the following array and none are in the stream; when equal to 0 indicates that additional NAL units of the indicated type may be in the stream; the default and permitted values are constrained by the sample entry name.

*with*

array\_completeness when equal to 1 indicates that all NAL units of the given type are in the following array and none are in the stream; when equal to 0 indicates that additional NAL units of the indicated type may be in the stream; the permitted values are constrained by the sample entry name.

*In subclause 8.3.2.1.3, remove the following from the semantics of nalUnit:*

When one or more SEI NAL units containing an SEI manifest SEI message and/or an SEI prefix indication SEI message are available, they should be stored as instances of nalUnit.

*In subclause 8.4.1.1.1, replace*

When the sample entry name is 'hvc1', the default and mandatory value of array\_completeness is 1 for arrays of all types of parameter sets, and 0 for all other arrays. When the sample entry name is 'hev1', the default value of array\_completeness is 0 for all arrays.

*with*

When the sample entry name is 'hvc1', the value of array\_completeness shall be equal to 1 for the arrays of all types of parameter sets. When the sample entry name is 'hev1', the value of array\_completeness shall be equal to 0 for the arrays of all types of parameter sets.

[Ed. (YK): I think it might be better that the sentence on 'hev1' be completely removed. Because the main purpose of the entire change, for both sentences, was to remove the unclear phrasing of "the default value" while keeping only the mandatory part. The original sentence on 'hvc1' does have the mandatory part, while the original sentence on 'hev1' does not. Technically, indeed the value of array\_completeness for 'hev1' should be allowed to be equal to either 0 or 1 for any type of the parameter set arrays, because either and/or both in-band and out-of-band parameter sets storage are allowed herein.]

[Ed. (MH): For VVC, the following phrasing is used: "When the sample entry name is 'vvi1', the value of array\_completeness of at least one of the following arrays shall be equal to 0: [bullet list of different types of parameter set NAL units]". Consider using similar phrasing also for HEVC. (YK): Agreed, although we should probably compare between completely removing the sentence on 'hev1' and rephrasing it similarly as in VVC, with the determining factor to consider being whether the backward compatibility issue introduced is ignorable (actually this would also apply to the option of just replacing "shall be equal to 1" with "shall be equal to 0" for that sentence.]

*In subclause 9.5.3.1.1, replace*

When the sample entry name is 'lhv1', the default and mandatory value of array\_completeness is 1 for arrays of all types of parameter sets, and 0 for all other arrays. When the sample entry name is 'lhe1', the default value of array\_completeness is 0 for all arrays.

*with*

When the sample entry name is 'lhv1', the value of array\_completeness shall be equal to 1 for the arrays of all types of parameter sets. When the sample entry name is 'lhe1', the value of array\_completeness shall be equal to 0 for the arrays of all types of parameter sets.

[Ed. (YK): Same comment as above for 'hev1'.]

[Ed. (MH): For VVC, the following phrasing is used: "When the sample entry name is 'vvi1', the value of array\_completeness of at least one of the following arrays shall be equal to 0: [bullet list of different types of parameter set NAL units]". Consider using similar phrasing also for L-HEVC. (YK): Same comment as above for 'hev1'.]