

**Joint Video Experts Team (JVET)
of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC
29/WG 11**

Document: JVET-R0414-v1

MPEG: m53769

18th Meeting: Teleconference, April 2020

Title: AHG9: A summary of proposals on mixed NAL unit types within a coded picture

Status: Input document to JVET

Purpose: Proposal

**Author(s) or
Contact(s):** [Ye-Kui Wang](#) | yekui.wang@bytedance.com | submitter (ByteDance)

Source: ByteDance

HEADER EXCERPT FROM SOURCE DOCUMENT

Abstract

This contribution provides a summary of the 10 proposals on mixed NAL unit types within a coded picture (the agenda item 4.1.9 in JVET-R0339).

It is suggested that this summary, in terms of a list of design questions, is used for the reviewing of these proposals, such that the discussions can be in a more structured and more efficient manner.

List of design questions on mixed NAL unit types within a coded picture

On the types of pictures with mixed NUTs and naming of NUTs:

Don't define types of pictures with mixed NUTs, define subpicture types (one corresponding to each picture type) instead. (R0042)

Built from JVET-R0414-v1.docx

CARD

<http://vexpertise.stbit.ru/documents/JVET-R0414>

JVET Search

Ye-Kui Wang

4

(m52769)

AH09: A summary of proposals on mixed NAL unit types within a coded picture

Authors:

ABSTRACT

MEETING NOTES



https://jvet-experts.org/doc_end_user/documents/18_Alpbach/wg11/JVET-R0414-v1.zip

A screenshot of a document viewer interface. At the top left, the word 'Archive' is written in a light blue font. Below it, a file named 'JVET-R0414-v1.docx' is displayed. To the left of the filename is a blue rounded square icon containing the white text 'DOC'. In the bottom right corner of the viewer, there is a square QR code.

https://jvet-experts.org/doc_end_user/current_document.php?id=10076

The screenshot shows a web application for previewing a document. At the top, a dark blue header contains the text 'JVET DOCUMENT MANAGEMENT SYSTEM' and 'Joint Video Experts Team'. Below the header, a light blue sidebar on the left contains a 'User' button, a 'Login' link, and links to 'Create new user', 'Claim your password', and 'Contact'. The main content area has a title 'Preview document JVET-R0414' in red. Below this is a section titled 'Document Information' with a red underline. It contains a 'Submitted by' field with the value 'Jens-Robert Klein', a 'Title' field with the value 'AFCD: A summary of proposals on mixed NAL unit types within a coded pic', and an 'Authors' field with the value 'V. Smiten'. To the right of this section is a link for 'Associated Resources'. Below the 'Document Information' section is a 'Document' section showing the file 'JVET-R0414' and a link to 'upload document'. At the bottom is a 'Comments' section with a red underline and a large empty text area for comments. On the right side of the page, there is a QR code.

https://www.itu.int/wftp3/av-arch/jvet-site/2020_04_R_Alpbach/JVET-R_Notes_dH.docx

</

MEETING NOTES

[JVET-R0414](#) AHG9: A summary of proposals on mixed NAL unit types within a coded picture [Y.-K. Wang (Bytedance)]

This contribution provides a summary of the 10 proposals on mixed NAL unit types within a coded picture.

It is suggested that this summary, in terms of a list of design questions, is used for the reviewing of these proposals, such that the discussions can be in a more structured and more efficient manner.

1. On the types of pictures with mixed NUTs and naming of NUTs:

1. Don't define types of pictures with mixed NUTs, define subpicture types (one corresponding to each picture type) instead. (JVET-R0042)

1. In addition, change the naming of the content of VCL NUTs as follows:

Name of nal_unit_type	Content of NAL unit and RBSP syntax structure	NAL unit type class
TRAIL_NUT	Coded slice of a trailing picture <i>or subpicture*</i> slice_layer_rbsp()	VCL
STSA_NUT	Coded slice of an STSA picture <i>or subpicture*</i> slice_layer_rbsp()	VCL
RADL_NUT	Coded slice of a RADL picture <i>or subpicture*</i> slice_layer_rbsp()	VCL
RASL_NUT	Coded slice of a RASL picture <i>or subpicture*</i> slice_layer_rbsp()	VCL
IDR_W_RADL IDR_N_LP	Coded slice of an IDR picture <i>or subpicture*</i> slice_layer_rbsp()	VCL
CRA_NUT	Coded slice of a CRA picture <i>or subpicture*</i> silce_layer_rbsp()	VCL
GDR_NUT	Coded slice of a GDR picture <i>or subpicture*</i> slice_layer_rbsp()	VCL
* indicates a property of a picture when mixed_nal_unit_types_in_pic is equal to 0 and a property of the subpicture when mixed_nal_unit_types_in_pic is equal to 1		

- 1. Consolidate picture types into 5 types: IRAP, GDR, leading, trailing and STSA picture, including for pictures with mixed NUTs, through changing the naming of the content of VCL NUTs as follows (JVET-R0267):

Name of nal_unit_type	Content of NAL unit and RBSP syntax structure	NAL unit type class
TRAIL_NUT	Coded slice of a trailing picture slice_layer_rbsp()	VCL
STSA_NUT	Coded slice of an STSA picture <i>or a trailing picture</i> slice_layer_rbsp()	VCL
RADL_NUT	Coded slice of a <i>leading or a trailing picture</i> slice_layer_rbsp()	VCL
RASL_NUT	Coded slice of a <i>leading or a trailing picture</i> slice_layer_rbsp()	VCL
IDR_W_RADL IDR_N_LP	Coded slice of an IDR picture, <i>a leading or a trailing picture</i> slice_layer_rbsp()	VCL
CRA_NUT	Coded slice of a CRA picture, <i>a leading or a trailing picture</i> silce_layer_rbsp()	VCL
GDR_NUT	Coded slice of a GDR picture slice_layer_rbsp()	VCL
...	...	

Both proposals are essentially only editorial, as they are just expressions of the existing technical intent, and they would not change the intended output behaviour of a decoder or the conformance of a bitstream. They are just a matter of terminology.

One participant commented that the NAL unit types seem less useful to a decoder than seems desirable in the mixed case.

From a technical perspective, the table is more necessary for the other NAL unit types that are not shown above, as it specifies the RBSP syntax structure that corresponds to each NAL unit type.

It was commented that the “a” expression seems more clear and useful to the reader. Adding an explanatory footnote to the table was suggested, and suggested text for this was integrated above.

Decision (expression of existing intent): Adopt approach “a”, as amended.

1. On requiring `subpic_treated_as_pic_flag[ ]` to be equal to 1 for subpictures with different NUTs in a picture, replace the existing constraint on this specified only for the mixing of an IRAP NUT and another NUT with the following constraint to cover all allowed NUT mixes (JVET-R0042):

Any two neighbouring subpictures with different NAL unit types within a picture shall both have the `subpic_treated_as_pic_flag[ ]` equal to 1.

This was agreed to be the minimum we should do.

- 1. Add the following constraint on this for the mixing of RASL_NUT and RADL_NUT (JVET-R0270):

A RADL picture (i.e., a picture that contains one or more slices all with `nal_unit_type` equal to RADL_NUT) may reference a RASL picture with `mixed_nalu_types_in_pic_flag` equal to 1. When a slice of such a RADL picture refer to a RASL picture as an active reference picture, the following applies:

- ○ ■ The current subpicture (i.e., the subpicture that contains the slice) and the referenced subpicture of the RASL picture shall have `subpic_treated_as_pic_flag[ i ]` equal to 1.

Alternatively it was suggested that *always* when any two subpictures have different NAL unit types, `subpic_treated_as_pic_flag[ ]` shall be equal to 1 for all subpictures in the picture that contain at least one P or B slice. Decision (cleanup): Express this modified constraint (exact text to be provided by YKW).

1. On constraints on prediction relationship (through constraints on RPLs): Similar constraints for all different types of pictures and the same-layer pictures in the preceding and succeeding AUs are specified in the subpicture domain for all the different types of subpictures and the same-layer subpictures with the same subpicture index in preceding and succeeding AUs. (JVET-R0042)

1. Add the following constraint for the mixing of RASL_NUT and RADL_NUT (JVET-R0270):

A RADL picture (i.e., a picture that contains one or more slices all with `nal_unit_type` equal to RADL_NUT) may reference a RASL picture with `mixed_nalu_types_in_pic_flag` equal to 1. When a slice of such a RADL picture refer to a RASL picture as an active reference picture, the following applies:

- ○ ■ The referenced subpicture of the RASL picture shall only contain slices with `nal_unit_type` equal to RADL_NUT.

Item “a” was said to be resolved by the proposal in JVET-R0042.

It was commented that it should not be necessary to express the constraints both for pictures and for subpictures. If they are specified for subpictures, this may be sufficient.

Decision (expression of existing intent): Adopt as proposed in JVET-R0042.

1. On constraints on relative decoding order and output order:
 1. Similar constraints for all the different types of pictures and the same-layer pictures in the preceding and succeeding AUs are specified in the subpicture domain for all the different types of subpictures

and the same-layer subpictures with the same subpicture index in preceding and succeeding AUs. (JVET-R0042)

2. Similar intent as item 4.a, through the following text changes (JVET-R0136):

For a single-layer bitstream which include all picture with `mixed_nalu_types_in_pic_flag` equal to 0 the following constraints apply:

Or for a single-layer bitstream which include one or more pictures with `mixed_nalu_types_in_pic_flag` equal to 1, the following constraints apply to a bitstream of a subpicture sequence which could be extracted:

- ○ ■ Each picture, other than the first picture in the bitstream in decoding order, is considered to be associated with the previous IRAP picture in decoding order.
- When a picture is a leading picture of an IRAP picture, it shall be a RADL or RASL picture.
- ...

Decision (expression of existing intent): Adopt as proposed in JVET-R0042.

1. Replace the following constraint requiring that a subpicture with a different subpicture ID compared to the collocated subpicture in the previous picture in the CLVS needing to have IRAP NUTs (JVET-R0276):

When the current picture is not the first picture of the CLVS, for each value of `i` in the range of 0 to `sps_num_subpics_minus1`, inclusive, if the value of `SubpicIdVal[ i ]` is not equal to the value of `SubpicIdVal[ i ]` of the previous picture in decoding order in the same layer, the `nal_unit_type` for all coded slice NAL units of the subpicture in the current picture with subpicture index `i` shall be equal to a particular value in the range of `IDR_W_RADL` to `CRA_NUT`, inclusive.

to be as follows (basically to only allow reordering/change of the subpicture IDs at pictures that are not used for inter-prediction reference):

For each value of `i` in the range of 0 to `sps_num_subpics_minus1`, inclusive, if the value of `SubpicIdVal[ i ]` is not equal to the value of `SubpicIdVal[ i ]` of a reference picture, such reference picture shall not be used for predicting the slice NAL units of the subpicture in the current picture with subpicture index `i`.

This is a relaxation to allow change of subpicture ID for non-reference pictures.

After offline study, this was discussed further on Friday 24 April at 0525 (chaired by GJS).

Editorially, it was commented that this should be rephrased, in the spirit of:

For each value of `i` in the range of 0 to `sps_num_subpics_minus1`, inclusive, if the value of `SubpicIdVal[ i ]` is not equal to the value of `SubpicIdVal[ i ]` of a reference picture, the active entries of the RPL of the coded slices in the `i`-th subpicture of the current picture shall not include that reference picture.

An example use case that was discussed is 360° real-time streaming with a background LTRP.

It was commented that the use of this may involve a higher degree of coordination of the encoding than in most BEAM uses. Another participant said the potential use case seems hypothetical and not likely to be commonly used. The proponent said the use of this feature is voluntary for the encoder, so it would not pose a problem if an encoder chose not to take advantage of the slightly greater flexibility.

Decision (cleanup): Adopt with editorial clarification as described.

1. On signalling and semantics of the mixed NUT flag:

1. Move the `mixed_nalu_types_in_pic_flag` from the PPS to the PH. (JVET-R0085, JVET-R0315)
 1. Condition it under "`!gdr_or_irap_pic_flag`)" in the PH. (JVET-R0085, JVET-R0315)
 2. Add a constraint such that when `subpic_info_present_flag` is equal to 0 or `sps_num_subpics_minus1` is equal to 0, the value of `mixed_nalu_types_in_pic_flag` shall be 0. (JVET-R0085)

It was asked why the flag should be moved. The proponent said the desire was to minimize the number of PPSs. It was asked whether this would require changing the PH for BEAMing, and proponent suggested to define this flag as a “one way” constraint. However, a “one way” definition would reduce the ability to clearly identify non-mixed pictures – an encoder could just label everything as mixed. This had been previously considered in the context of JVET-Q0284.

No action was thus taken on this.

- 1. Add an SPS flag to indicate presence of pictures with mixed NUTs in the CLVS. (JVET-R0085, JVET-R0267, JVET-R0315)
 1. When item 5.a is done, use this new SPS flag to gate the flag moved to the PH. (JVET-R0085, JVET-R0315)

It was noted that there is a general constraint flag for this already, and the move of a flag to the PH was not done per item “a” above, so no action was taken on this.

- 1. Specify the semantics of `mixed_nalu_types_in_pic_flag` to be one-way, as follows (JVET-R0315):

`mixed_nalu_types_in_pic_flag` equal to 1 specifies that each picture referring to the PPS has more than one VCL NAL unit and the VCL NAL units *may* not have the same value of `nal_unit_type`.

`mixed_nalu_types_in_pic_flag` equal to 0 specifies that each picture referring to the PPS has one or more VCL NAL units and the VCL NAL units of each picture referring to the PPS have the same value of `nal_unit_type`.

It was commented that there is a constraint that IRAP pictures shall have the mixed flag equal to 0. The proponent said that there should still be a constraint that if all NAL units of a picture have an IRAP (maybe or GDR) NUT, then the mix flag shall be 0.

Another participant said the IRAP case is not the only one that is of concern; leading pictures are also assisted by having two-way semantics. A mixture of RADL and RASL subpictures would be expected to result in RASL behaviour, and currently the output flag is set without checking all subpictures.

It was discussed whether rewriting of the PH would be needed for BEAMing.

The proponent said that BEAMing with merging an IRAP with a non-IRAP would involve having two picture headers in source content and selecting the one with the mixing flag equal to 1.

No action was taken, as most participants desired “two way” indication assurance for the mixture flag behaviour.

1. Disallow the mix of an IRAP NUT and a leading picture NUT. (JVET-R0203)

It was said that BEAMing is generally intended only with alignment of GOP structures such that the decoding order and output order of pictures in the source bitstreams are aligned.

Decision (an overlooked sensibility constraint): Adopt.

Discussion stopped here for JVET Track A on 18 April at 0915 UTC.

Discussion began here for JVET Track A Monday 20 April at approximately 1300 UTC (chaired by GJS).

1. Allow mixing of more than two NUTs within a coded picture. (JVET-R0203). The constraint of only two had been imposed out of conservatism, and after further study, appeared not to be necessary. Multisource bitstream merging was discussed. There was no reason identified to have this constraint.

Decision (cleanup): Adopt.

2. On mixing of RASL_NUT with another NUT, do either of the following two options (JVET-R0267)

1. Prohibit mixing of RASL_NUT with another NUT.
2. Add a new PPS or PH flag to indicate the presence of mixing of RASL_NUT with another NUT.

The concept is mix bitstreams that have aligned CRAs for which some bitstream has RADL and another has RASL. It was originally intended for such a mixed picture to be treated as a RASL picture. There is a PPS flag to indicate that the picture is mixed, and the POC indicates that the picture is a leading picture.

There was discussion of a mixing of RASL and trailing picture (because of differing IRAP periods, esp. mixing open GOP and closed GOP) – see the prior contribution JVET-Q0396. Such a picture is treated as an ordinary trailing picture (associated with a different IRAP than in the source bitstream of the RASL picture).

Except for identifying an IRAP picture based on a PPS flag and the NUT of the first VCL NAL unit, the decoder does not need to care about the different NUTs of a picture in terms of the decoding process – the mixture only affects whether to output the decoded result or not. In the IRAP case, there is a PPS flag that indicates whether there is a mixture, and if there is a mixture, the picture is not an IRAP picture.

There seemed to be no need for the extra prohibition or additional flag as proposed, so no action was taken on this.

1. On treating a picture with mixed RASL_NUT and RADL_NUT in the output of the decoding process and bitstream conformance tests (JVET-R0270)

1. A picture with mixed RASL_NUT and RADL_NUT is treated as a RASL picture during the decoding process, i.e., regardless of whether a RASL picture has slices with nal_unit_type equal to RADL_NUT, the RASL picture is not output when the associated IRAP picture has NoOutputBeforeRecoveryFlag equal to 1.

2. A picture with mixed RASL_NUT and RADL_NUT is treated as a RADL picture in bitstream conformance tests, i.e., only those RASL pictures for which all slices have nal_unit_type equal to RASL_NUT associated with the first IRAP picture are removed from the bitstream to be decoded when the alternative HRD timing is used in the particular bitstream conformance test. These pictures can still serve as references for other pictures that are to be output, so they need to be kept in the bitstream.

Decision (expression of existing intent): Adopt (both aspects).

As an off-topic matter, there was discussion of whether the decoding process for GDR should output “partially dirty” pictures that precede the recovery point. It was commented that this is commonly done, but our standard says these are not output (and it may say that they shall not be output, just as the area outside a cropping window is also not output). This issue is for further study.

It was commented that there are two potential uses of GDR – one is trying to tune in rapidly, which is not always intended. The other use is to just keep “glass to glass” latency at a minimum while enabling (potentially slow) random access.

1. Add the following paragraph in the general decoding processs in clause 8.1.1 (JVET-R0120):

When mixed_nalu_types_in_pic_flag is equal to 1 and at least one VCL NAL unit of a picture has nal_unit_type equal to CRA_NUT, HandleCraAsCvsStartFlag and NoOutputBeforeRecoveryFlag for the picture are both set equal to 0.

The intent is that a mixed picture is not treated as an IRAP picture. Instead, it is treated as a trailing picture. This is agreed as the existing design intent. However, the proponent said this may not be sufficiently clear to the reader, and this is a valid concern.

It was commented that the identified flag variables are not used under these circumstances, so specifying values for them is not necessary or helpful, as they have no effect.

The editor may consider adding clarifying text about this (e.g., in one or more NOTES), to make sure it is clear to the reader, but no normative action seemed necessary.

1. Change the semantics of gdr_or_irap_pic_flag as follows such that when mixed_nalu_types_flag is equal to 1 the no_output_of_prior_pics_flag can be signalled for a non-IRAP and non-GDR picture that contains at least one IRAP or GDR subpicture ([JVET-R0124](#)):

gdr_or_irap_pic_flag equal to 1 specifies that the current picture is a GDR or IRAP picture *or a picture with mixed_nalu_types_in_pic_flag equal to 1 and VCL NAL units with NAL unit type equal to IDR_W_RADL, IDR_N_LP or CRA_NUT*. gdr_or_irap_pic_flag equal to 0 specifies that the current picture may or may not be a GDR or IRAP picture.

(italics for added phrase)

It was commented that this does not seem quite correct because the gdr_or_irap_pic_flag flag has “one way” semantics.

It was commented that we don't allow a mixture of IDR and CRA NAL unit types within a picture.

It was suggested to add a constraint that when `mixed_nalu_types_in_pic_flag` is equal to 1 (i.e., a mixed picture) `gdr_or_irap_pic_flag` shall be equal to 0. Such a PH should be available from the one of the pre-merged picture, so rewriting the PH content should not be necessary.

This was further discussed on Friday 24 April at 0605 UTC (chaired by GJS) after offline study. No action was recommended since the use case seemed too hypothetical.

Discussion stopped here for JVET Track A Monday 20 April at 1605 UTC (chaired by GJS).

CITATION

Y.-K. Wang, "AHG9: A summary of proposals on mixed NAL unit types within a coded picture", JVET-R0414, 18th Meeting: Teleconference, April 2020.

REFERENCES

1. [Y.-K. Wang, "AHG8/AHG9/AHG12: On mixed subpicture types within a picture", JVET-R0042, 18th Meeting: Teleconference, April 2020.](#)
2. [L. Chen, S.-T. Hsiang, O. Chubach, Y.-W. Huang, S.-M. Lei, "AHG9: On signalling the mixed NAL unit type flag", JVET-R0085, 18th Meeting: Teleconference, April 2020.](#)
3. [B. Choi, S. Wenger, S. Liu, "AHG9/AHG12: On mixed NAL unit types", JVET-R0120, 18th Meeting: Teleconference, April 2020.](#)
4. [B. Choi, S. Wenger, S. Liu, "AHG9: Clean-ups on picture header", JVET-R0124, 18th Meeting: Teleconference, April 2020.](#)
5. [M. Katsumata, M. Hirabayashi, T. Suzuki, "AHG9/AHG12: Improvements on sps independent subpics flag and nal unit type constraint", JVET-R0136, 18th Meeting: Teleconference, April 2020.](#)
6. [Hendry, S. Kim, "AHG9/AHG12: On combination of NAL unit types in a picture", JVET-R0203, 18th Meeting: Teleconference, April 2020.](#)
7. [Y. He, M. Coban, V. Seregin, A.K. Ramasubramonian, M. Karczewicz, "AHG9/AHG12: On mixed NAL unit types", JVET-R0267, 18th Meeting: Teleconference, April 2020.](#)
8. [Hendry, S. Kim, R. Skupin, Y. Sanchez, K. Suehring, "AHG9: On mixing of RASL and RADL NAL unit types", JVET-R0270, 18th Meeting: Teleconference, April 2020.](#)
9. [V. Seregin, Y. He, M. Coban, M. Karczewicz, "AHG9: On IRAP NAL constraint for reordered sub-pictures", JVET-R0276, 18th Meeting: Teleconference, April 2020.](#)
10. [M. Coban, V. Seregin, Y. He, Y.-J. Chang, M. Karczewicz, "AHG9: On mixed nal unit type signalling and PPS cleanup", JVET-R0315, 18th Meeting: Teleconference, April 2020.](#)
11. [G. J. Sullivan, Y.-K. Wang, "Agenda and report of the Category 1 AHG pre-meeting for the 18th JVET meeting", JVET-R0339, 18th Meeting: Teleconference, April 2020.](#)