



ABSOLUTE SIGNALING OF REFERENCE PICTURES

JCTVC-F493

RICKARD SJÖBERG, JONATAN SAMUELSSON
ERICSSON AB

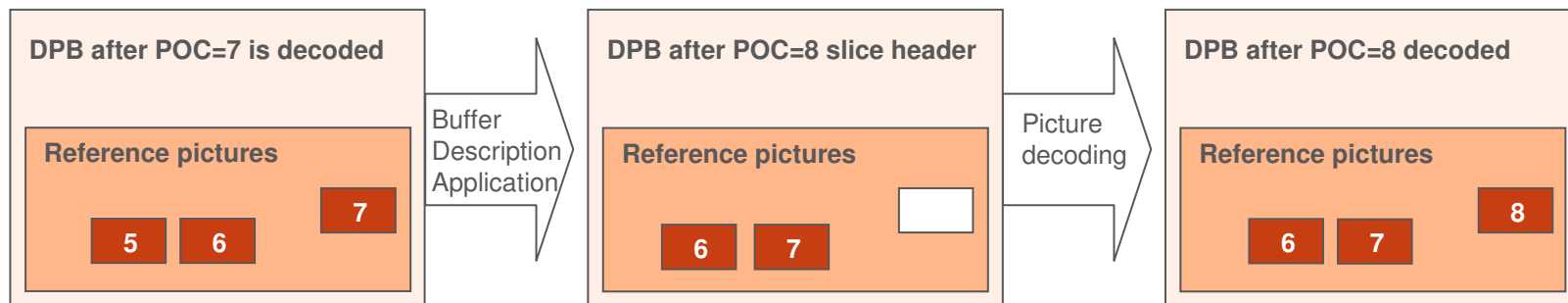
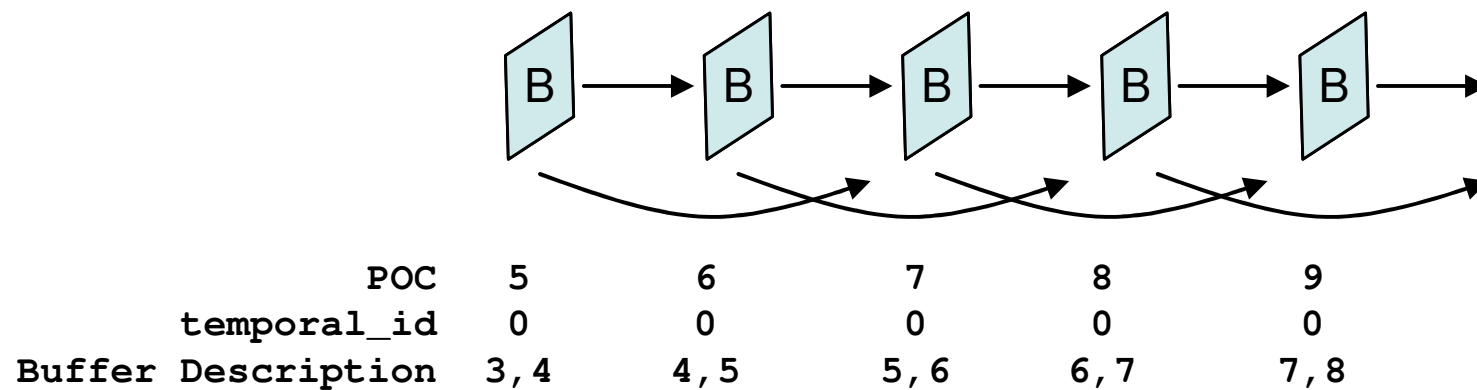
OUTLINE

- › Proposal introduction with examples
- › Proposed list of picture marking rules
- › Advantages of the proposal
- › Syntax and signaling
- › Results
- › Conclusion

PROPOSAL SUMMARY

- › We propose: Buffer Description
 - It is a list of reference pictures to be used for reference for the current picture or for any future picture
 - Each slice header contain a Buffer Description
 - The POC and the temporal_id of each picture in the Buffer Description is signaled
 - POC is signaled relative the current POC (e.g. -1, -2, -3 for LD)
- › Further, we propose to use POC as the only picture identifier
 - frame_num is no longer needed and is removed
 - We propose to use POC for reference picture list modification
- › This have the following effects
 - Improved detection of lost reference pictures
 - Improved concealment of lost reference pictures
 - MMCO and sliding window is removed
 - Every possible picture coding structure can be used with minimum number of reference frames and without using MMCO commands since the marking is explicit
 - Compact description, less rules
 - Simplifies implementation of both encoder and decoder
 - BDR=0.0% on our common test condition

EXAMPLE



DETAILS OF PROPOSAL

- › No Buffer Description is received for IDR pictures
 - Since IDR pictures refresh the decoder
- › For all non-IDR pictures a Buffer Description is decoded
 - This is the case also for non-reference pictures
- › A Buffer Description may contain at most `max_num_ref_frames` pictures
- › After decoding a picture it is marked ('used') according to `nal_ref_idc`
- › Picture marking ('unused') is done after the slice header is decoded, before reference picture list initialization

- › Proposed POC changes
 - POC wraps around
 - The POC types are removed
 - The number of bits to use for POC is signaled with `log2_max_pic_order_cnt_minus4` in the SPS
 - POC is explicitly coded in each slice header

PROPOSED SET OF RULES

Picture marking rule

- › Pictures in the DBP marked as “used for reference” that are not included in the Buffer Description are marked “unused for reference”

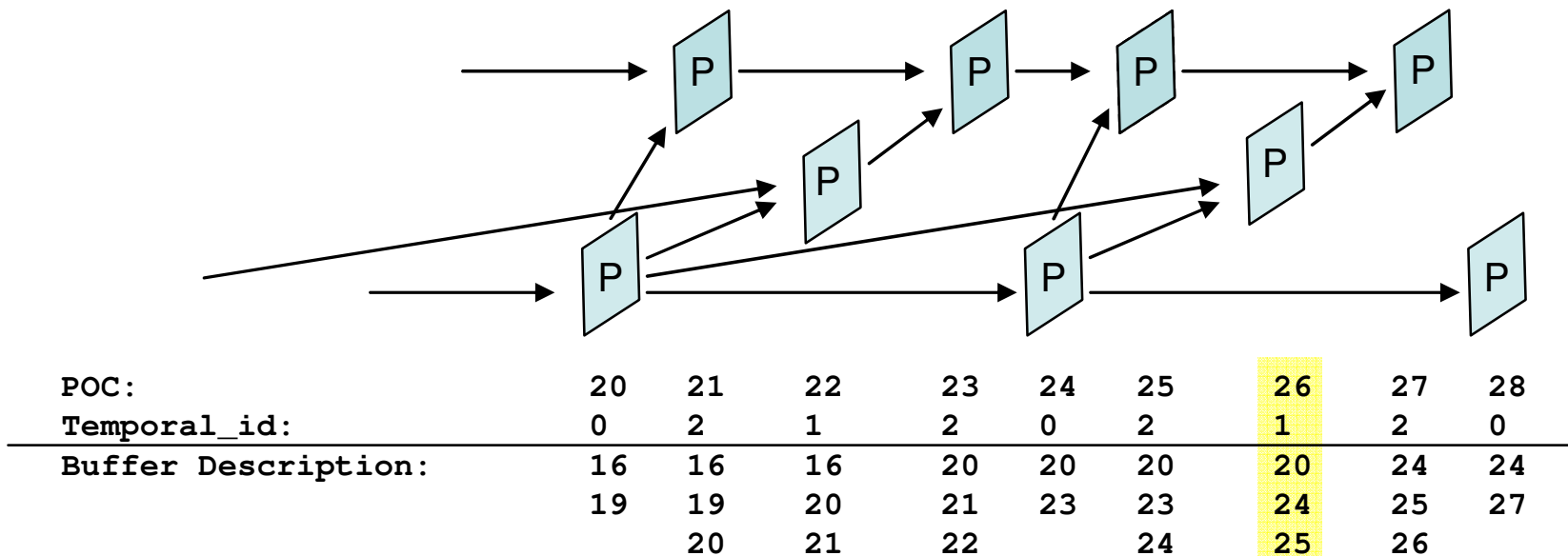
Temporal scalability rule

- › Pictures with higher temporal_id than current picture that are included in the Buffer Description but not present in the DPB marked as “used for reference” are ignored

Error resilience rule

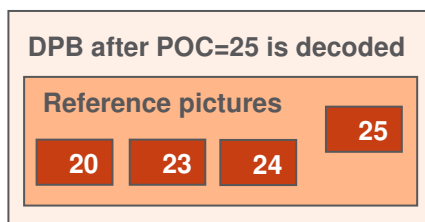
- › Pictures with same or lower temporal_id than current picture that are included in the Buffer Description but not present in the DPB marked as “used for reference” are detected as lost

PICTURE MARKING RULE

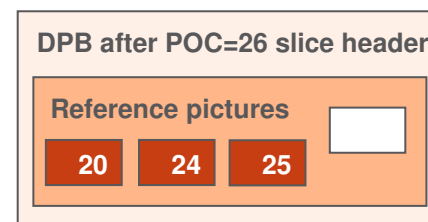


Picture marking rule

- › Pictures in the DBP marked as “used for reference” that are not included in the Buffer Description are marked “unused for reference”

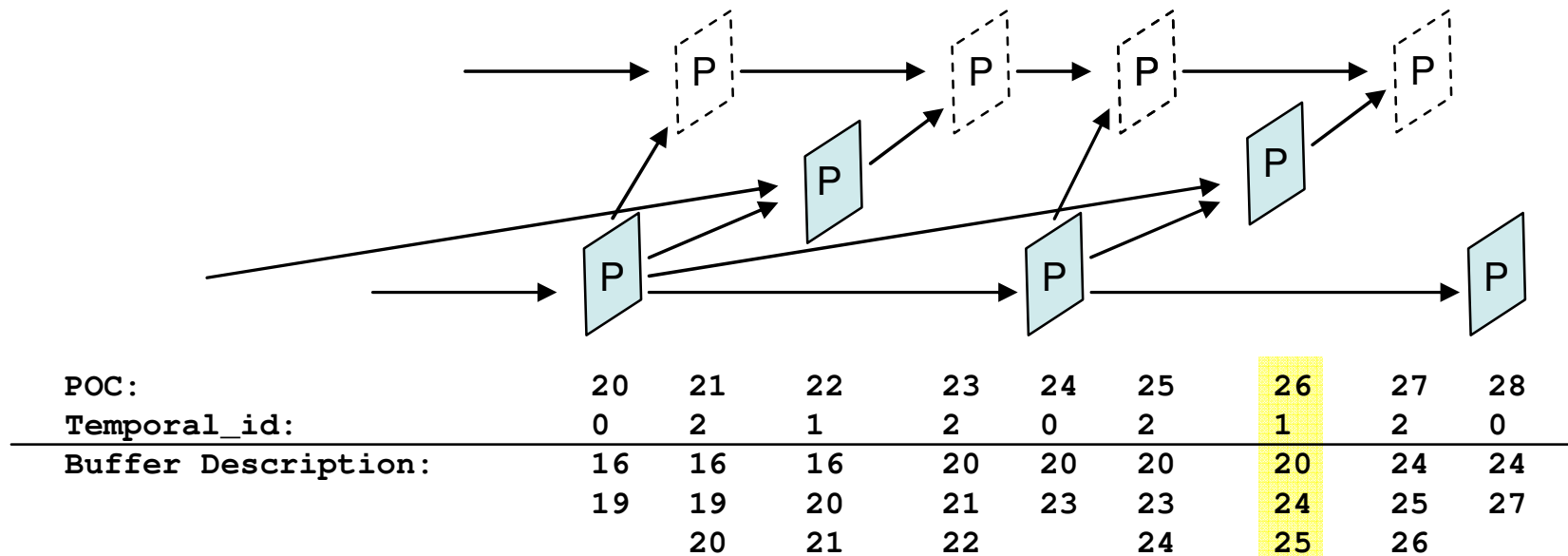


POC = 26 Buffer description 20, 24, 25



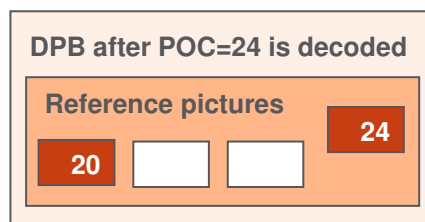
Note that 25 is not included in ref pic list for 26

TEMPORAL SCALABILITY RULE

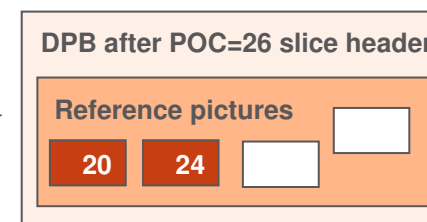


Temporal scalability rule

- › Pictures with higher temporal_id than current picture that are included in the Buffer Description but not present in the DPB marked as “used for reference” are ignored

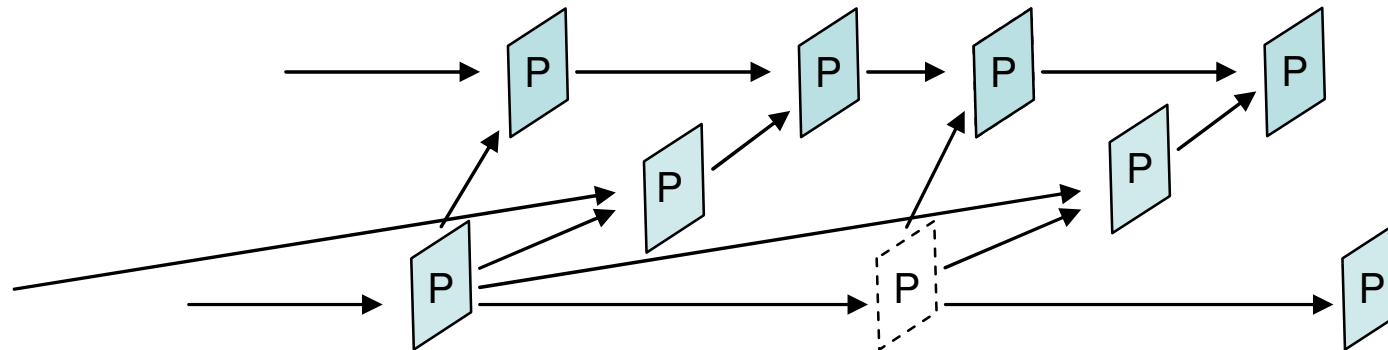


POC = 26 Buffer description 20, 24, 25



Note that 25 is not included in ref pic list for 26

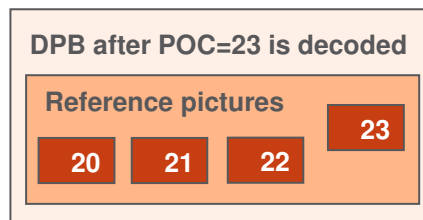
ERROR RESILIENCE RULE



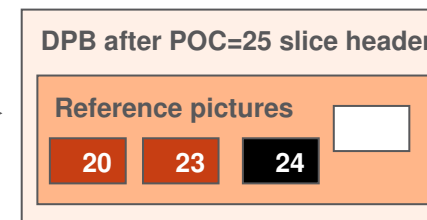
POC:	20	21	22	23	24	25	26	27	28
Temporal_id:	0	2	1	2	0	2	1	2	0
Buffer Description:	16	16	16	20	20	20	20	24	24
	19	19	20	21	23	23	24	25	27
		20	21	22		24	25	26	

Error resilience rule

- › Pictures with same or lower temporal_id than current picture that are included in the Buffer Description but not present in the DPB marked as “used for reference” are detected as lost



POC = 25 Buffer description 20, 23, 24



24 is detected as loss and is concealed

ERROR RESILIENCE ADVANTAGES

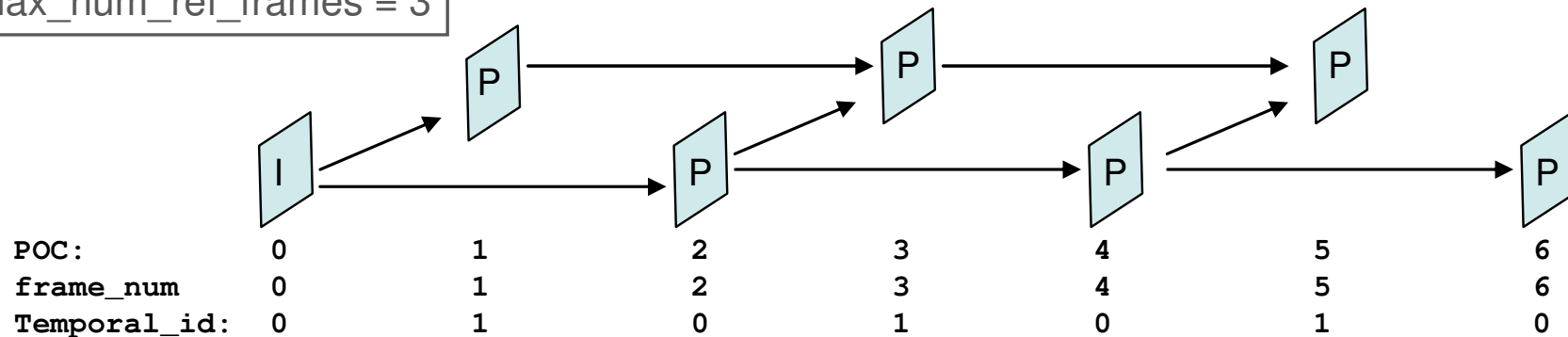
- › Losses of reference pictures are detected immediately
 - In HEVC, you may need to wait until a `ref_idx` identifies a “non-existing” picture before you know that a reference picture is lost. Call to concealment routines may be needed deep down in the decoder.
 - Also, in HEVC, you are not guaranteed to know that a picture is unintentionally lost at all
 - With the proposal, you will know after decoding the Buffer Description in the very next slice header if a reference picture is lost, for all bitstreams
- › Current HEVC MMCO picture marking is not robust to picture losses
 - If a picture with long-term MMCO commands is lost, there may be indefinite mismatches
 - With the proposal, lost long-term pictures are repeatedly detected as lost
- › The proposal help error concealment
 - POC is proposed to be signaled for all Buffer Description pictures
 - If a picture is lost, the decoder will know the POC of the lost picture
 - This is very good for “frame copy”, “motion copy” or other concealment methods

LOST PICTURE ADVANTAGE

- › Assume that a reference picture is lost. The loss is detected and a concealed picture is used in its place
- › HEVC reference picture list initialization uses temporal_id for picture list initialization, a picture list does not include pictures with higher temporal_id
 - Since the temporal_id of the lost picture is unknown, reference picture list initialization is not possible without explicit picture list initialization
 - With the proposal, temporal_id of all lost reference pictures are known
- › HEVC reference picture list initialization uses POC for picture list initialization, pictures are ordered in increasing or decreasing POC order
 - Since the POC of a lost picture is unknown, reference picture list initialization is not possible without explicit picture list initialization
 - With the proposal, POC of all lost reference pictures are known
- › The proposal handles these two issues without any additional rules in the WD or any additional bits in the bitstream

TEMPORAL SCALABILITY ADVANTAGE

HEVC/AVC example
max_num_ref_frames = 3



- › In order to have picture 0 and picture 2 available when decoding picture 4, picture 3 must have an MMCO marking picture 1 as unused
- › However, if layer 1 is removed, non-existing picture 3 will knock out picture 0 and there will be a mismatch
- › The proposal handles this without any problems, picture 4 will simply have pictures 3, 2 and 0 in its Buffer Description
- › AVC and HEVC do not support this picture structure with 3 picture buffers

SIMPLIFICATIONS

- › Working draft text
 - The proposed working draft text is more compact since two POC types, all six MMCO commands, the sliding window process, `frame_num`, `gaps_in_frame_num` and “non-existing” pictures are replaced by Buffer Description.
- › Robust encoder implementation
 - The proposal makes it a little easier to implement a robust encoder since the encoder need to worry less about picture list modifications, be careful with MMCO etc.

DECODER SIMPLIFICATION

Decoding steps

1. Parse the Buffer Description
 2. Mark pictures in DPB as unused if they are not part of Buffer Description
 3. Lost reference pictures are immediately identified
 4. Decode current picture
 5. Mark decoded picture according to `nal_ref_idc`
-
- › No “non-existing” pictures
 - › Error detection and concealment of reference pictures in one single place

SIGNALING

- › Explicit Buffer Description signaling
 - The slice header contains a list of pictures
 - Each list element consist of a delta_poc and a temporal_id

- › Buffer Description reference signaling
 - PPS contains a list of Buffer Descriptions
 - Each Buffer Description has an index assigned to it
 - The slice header contains an index to the Buffer Description to be used

SLICE HEADER SYNTAX

slice_header() {	Descriptor
...	
if(!IdrPicFlag){	
buffer_description_reference_flag	u(1)
if(buffer_description_reference_flag == 1) {	
buffer_description_idx	ue(v)
} else {	
number_of_explicitly_signalled_pictures	ue(v)
for(i = 0; i < number_of_explicitly_signalled_pictures; i++) {	
absolute_delta_poc_minus_one[i]	ue(v)
if(num_reorder_frames > 0)	
delta_poc_sign_flag[i]	u(1)
if(max_temporal_layers_minus1 > 0)	
temporal_id[i]	u(v)
}	
}	
}	

PPS SYNTAX

pic_parameter_set_rbsp() {	Descriptor
...	
number_of_buffer_descriptions	ue(v)
if(number_of_buffer_descriptions > 0){	
bits_for_temporal_id	u(2)
reorder_pictures_allowed_flag	u(1)
for(i = 0; i < number_of_buffer_descriptions; i++){	
number_of_pictures_in_buffer_description[i]	ue(v)
for(j = 0; j < number_of_pictures_in_buffer_description[i]; j++) {	
absolute_delta_poc_minus_one[i][j]	ue(v)
if(reorder_pictures_allowed_flag)	
delta_poc_sign_flag[i][j]	u(1)
if(bits_for_temporal_id > 0)	
temporal_id[i][j]	u(v)
}	
}	
}	

COMMON CONDITION RESULTS

- › Implementation on top of HM 3.1
- › Cross-verified results
- › Source code attached to proposal
- › HM 3.1 use 12 ref pics for RA
- › We use optimal 6 ref pics for RA
- › Note: HM 3.1 does not use frame_num, with 8 bit frame nums the proposal gives 0.0 or -0.1 for all classes.
- › Note: we have also implemented robust decoding that does not crash on picture losses (not uploaded)

	All Intra HE			All Intra LC		
	Y	U	V	Y	U	V
Class A	0,0	0,0	0,0	0,0	0,0	0,0
Class B	0,0	0,0	0,0	0,0	0,0	0,0
Class C	0,0	0,0	0,0	0,0	0,0	0,0
Class D	0,0	0,0	0,0	0,0	0,0	0,0
Class E	0,0	0,0	0,0	0,0	0,0	0,0
Overall	0,0	0,0	0,0	0,0	0,0	0,0
Enc Time[%]	101%			101%		
Dec Time[%]	103%			99%		
	Random Access HE			Random Access LC		
	Y	U	V	Y	U	V
Class A	0,0	0,0	0,0	0,0	0,0	0,0
Class B	0,0	0,0	0,0	0,0	0,0	0,0
Class C	0,0	0,0	0,0	0,0	0,0	0,0
Class D	0,1	0,1	0,1	0,1	0,1	0,1
Class E						
Overall	0,0	0,0	0,0	0,0	0,0	0,0
Enc Time[%]	100%			101%		
Dec Time[%]	106%			101%		
	Low delay B HE			Low delay B LC		
	Y	U	V	Y	U	V
Class A						
Class B	0,0	0,0	0,0	0,0	0,0	0,0
Class C	0,0	0,0	0,0	0,0	0,0	0,0
Class D	0,0	0,0	0,0	0,0	0,0	0,0
Class E	0,0	0,0	0,0	0,0	0,0	0,0
Overall	0,0	0,0	0,0	0,0	0,0	0,0
Enc Time[%]	100%			101%		
Dec Time[%]	103%			102%		

IMPROVED LOW DELAY RESULTS

	Low delay B HE			Low delay B LC		
	Y	U	V	Y	U	V
Class A						
Class B	-1,8	-2,8	-2,8	-2,1	-3,5	-4,4
Class C	-3,1	-3,3	-3,5	-3,2	-4,0	-4,1
Class D	-3,3	-3,0	-3,3	-3,7	-5,7	-5,7
Class E	-0,6	-0,5	-0,9	-1,1	-1,4	-1,9
Overall	-2,2	-2,6	-2,7	-2,6	-3,8	-4,2
Enc Time[%]	104%			104%		
Dec Time[%]	103%			103%		

- › Results for using two closest pictures for reference plus the two pictures with lowest QP in the range from POC(current)-3 to POC(current)-10
- › Only change is that the encoder uses a new set of Buffer Descriptions, no decoder change
- › No additional signaling (HEVC would need MMCO) or extra memory requirement
- › Anchor uses the four closest pictures and uses no frame_num

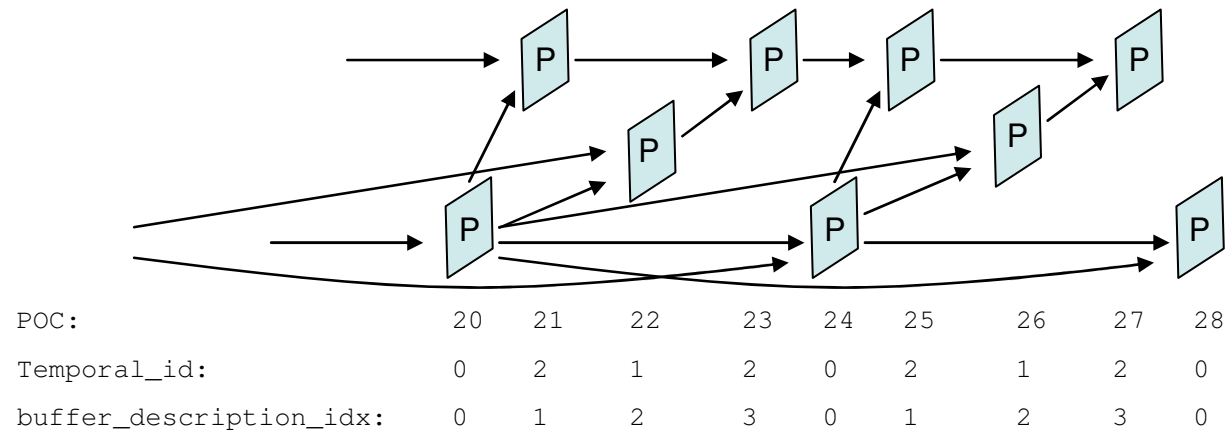
CONCLUSION

- › We suggest to adopt the proposal
 - The proposal brings significant advantages over current HEVC and at the same time simplifies picture marking
 - No additional bit cost for common condition
 - Source code is available in proposal for anyone to study and has been cross-checked
 - WD draft text is available in proposal and is ready for review



ERICSSON

WHY TEMPORAL ID IS NEEDED



Reason 1

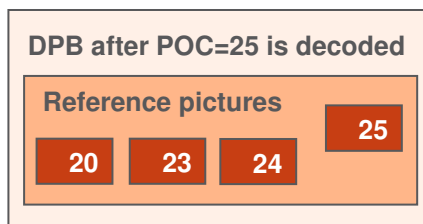
- Picture 22 will have 21 in its Buffer Description
- If layer 2 is kept, the decoder will know the temporal_id of 21
- If layer 2 is removed, the temporal_id of 21 is unknown and the reference list of 22 will be different

Reason 2

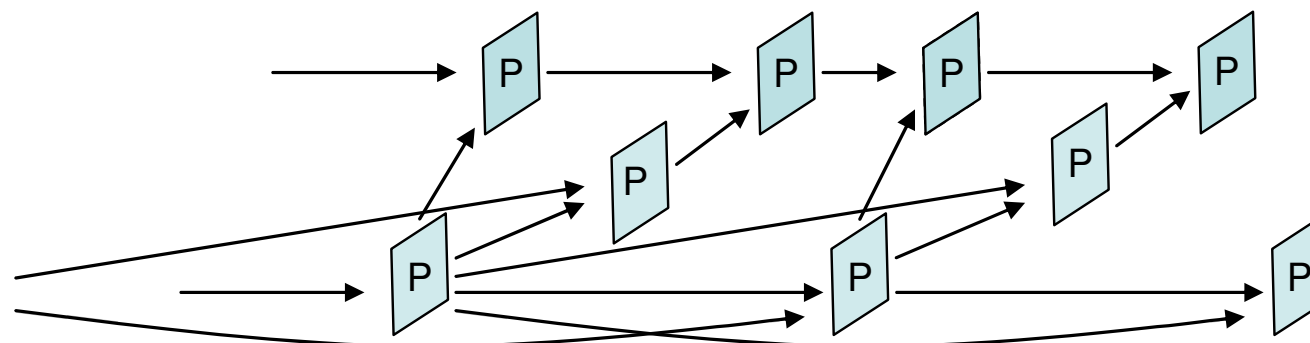
- If temporal_id is not signaled, the decoder will not know whether 21 was intentionally removed or was lost. Not good for error resilience.

WHY THERE IS NO ADDITIONAL DELAY

- › Example: Assume $\text{max_num_ref_frames} = 3$ and $\text{num_reorder_frames} = 3$
- › Assume that when picture with POC = 25 has just been decoded the DPB looks as indicated below.
- › There are now two possible scenarios:
 1. The next Buffer Description does not contain the picture with POC = 20.
 2. The next Buffer Description does contain the picture with POC = 20 but does not contain one of the pictures with POC 23, 24 or 25 (since it can not contain more than 3 pictures).
- › Either case, the picture with POC = 20 will be displayed first and that is already known as soon as the POC and buffer description of picture 25 has been parsed.



BITCOST EXAMPLE



POC: 20 21 22 23 24 25 26 27 28
 Temporal_id: 0 2 1 2 0 2 1 2 0
 buffer_description_idx: 0 1 2 3 0 1 2 3 0

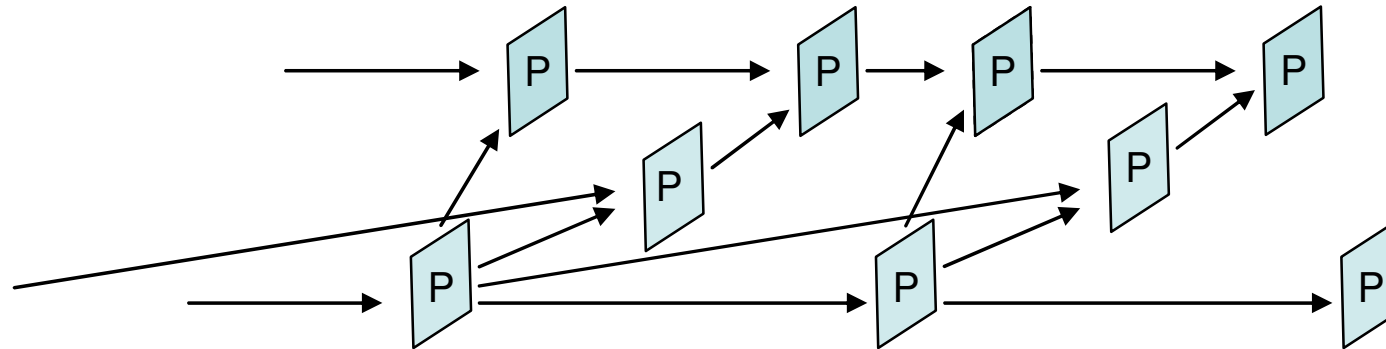
Buffer Description Index	Picture buffer 1	Picture buffer 2	Picture buffer 3	Picture buffer 4
0	dP: -1 tId: 2	dP: -2 tId: 1	dP: -4 tId: 0	dP: -8 tId: 0
1	dP: -1 tId: 0	dP: -2 tId: 2	dP: -5 tId: 0	dP: -9 tId: 0
2	dP: -1 tId: 2	dP: -2 tId: 0	dP: -6 tId: 0	dP: -10 tId: 0
3	dP: -1 tId: 1	dP: -2 tId: 2	dP: -3 tId: 0	dP: -7 tId: 0

The cost of signaling this table is 117 bits in the PPS

If you signal this explicitly instead it cost 26.5 bits per slice on average

- Less than 117/4 since you do not send the table size

BITSTREAM RESTRICTION RULE



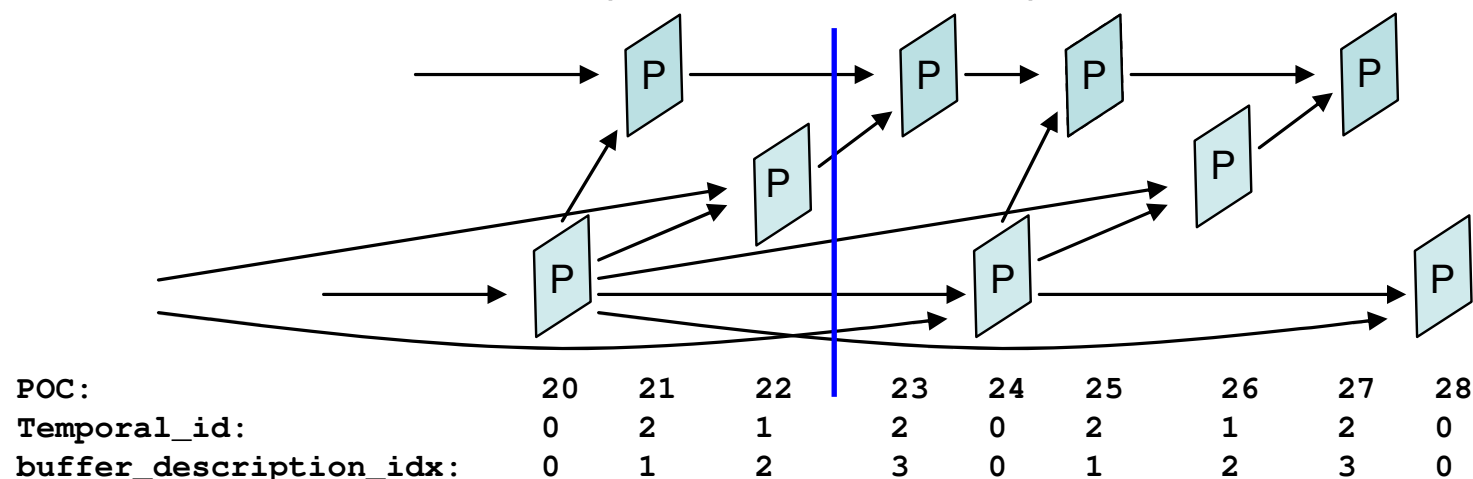
POC:	20	21	22	23	24	25	26	27	28
Temporal_id:	0	2	1	2	0	2	1	2	0
Buffer Description:	16	16	16	20	20	20	20	24	24
	19	19	20	21	23	23	24	25	27
		20	21	22		24	25	26	

Bitstream restriction rule

- > A Buffer Description must not contain a picture P if the temporal_id of P is less than or equal to the temporal_id of the current picture unless that picture P is available in the decoded picture buffer and marked as “used for reference”
- Assume that the Buffer Description of 26 is {20, 24, 25}
- The rule says that 25 does not have to be marked as “used for reference” in DPB. This is ok and can happen if layer 2 is removed
- However, picture 24 has to be marked as “used for reference” in DPB since otherwise reference picture list construction will be incorrect.

ENCODER SIMPLIFICATION

- › In the case of more complex picture structures it is much easier to signal the Buffer Description than using MMCO
- › Consider the example below with max_num_ref_frames=4
- › We believe the Buffer Descriptions are easier to implement than the MMCO syntax



Buffer Description Index	Picture buffer 1	Picture buffer 2	Picture buffer 3	Picture buffer 4
0	dP: -1 tId: 2	dP: -2 tId: 1	dP: -4 tId: 0	dP: -8 tId: 0
1	dP: -1 tId: 0	dP: -2 tId: 2	dP: -5 tId: 0	dP: -9 tId: 0
2	dP: -1 tId: 2	dP: -2 tId: 0	dP: -6 tId: 0	dP: -10 tId: 0
3	dP: -1 tId: 1	dP: -2 tId: 2	dP: -3 tId: 0	dP: -7 tId: 0

ON LONG-TERM PICTURES

- › Explicit marking is preferably used for “long-term” pictures
- › An addition to the proposed scheme is to allow for both referenced coding of a Buffer Description and explicit coding
 - First there is a flag whether reference or explicit Buffer Description is used
 - If explicit signaling is signaled, there is a second flag whether combined coding is used or not.
 - › If combined coding is off, there is only explicit coding of the Buffer Description
 - › If combined coding is on, there is both reference and explicit coding. First the referenced Buffer Description is decoded and then explicitly coded pictures replace entries in the Buffer Description starting from the end. The POC for the explicitly coded Buffer Description entries are coded with $\log_2_max_pic_order_cnt_minus4+4$ bits instead of UVLC

ALTERNATIVE SLICE HEADER SYNTAX

slice_header() {	Descriptor
...	
if(!IdrPicFlag){	
buffer_description_reference_flag	u(1)
if(buffer_description_reference_flag == 1) {	
buffer_description_idx	ue(v)
} else {	
buffer_description_combination_flag	u(1)
if(buffer_description_combination_flag){	
{	
buffer_description_combination_idx	u(1)
num_explicit_pictures_for_combination_minus1	ue(v)
for(i = 0; i < num_explicit_pictures_for_combination_minus1; i++){	
combination_poc[i]	u(v)
if(max_temporal_layers_minus1 > 0)	
combination_temporal_id[i]	u(v)
}	

Note
 1) POC and not deltaPOC
 2) u(v) and not ue(v)

ALT. SLICE HEADER SYNTAX CONT.

} else {	
number_of_explicitly_signalled_pictures	ue(v)
for(i = 0; i < number_of_explicitly_signalled_pictures; i++) {	
absolute_delta_poc_minus_one[i]	ue(v)
if(num_reorder_frames > 0)	
delta_poc_sign_flag[i]	u(1)
if(max_temporal_layers_minus1 > 0)	
temporal_id[i]	u(v)
}	
}	
}	
}	
...	
}	