



SYNTAX CONTROLLED OUTPUT PROCESS

JCTVC-H0567

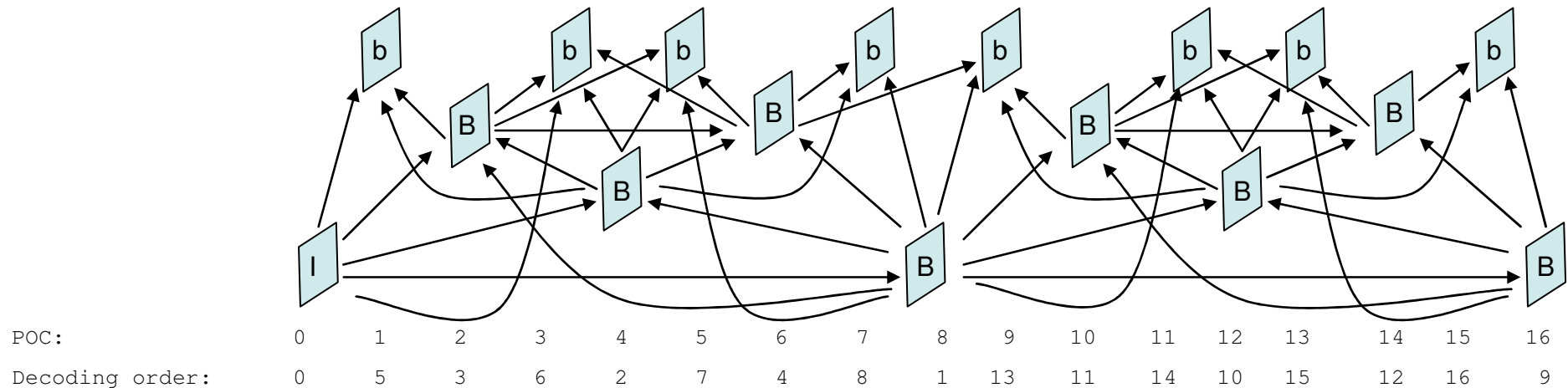
JONATAN SAMUELSSON

RICKARD SJÖBERG

THE FOLLOWING IS PROPOSED:

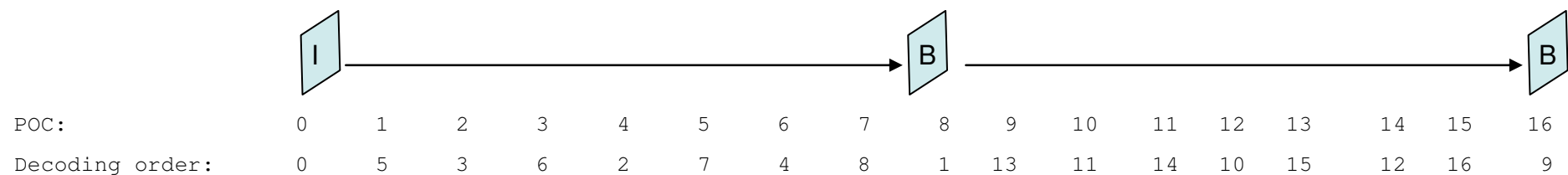
1. Add possibility to signal different values of the following SPS level parameters for each temporal layer:
 - a. max_dec_frame_buffering
 - b. max_num_ref_frames
2. Replace “bumping” process with a syntax controlled output process
 - a. Guarantees correct output order
 - b. An SPS flag to cover most common coding structures
 - c. When the SPS flag is not set, two slice header parameters are present

WHY DO WE WANT SMALLER DPB FOR A SUBSTREAM?



- > max_dec_frame_buffering: 6
- > max_num_ref_frames: 6

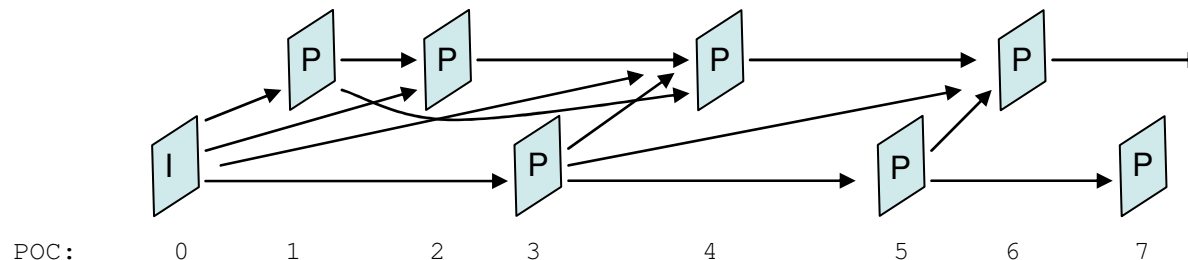
WHEN ONLY DECODING THE LOWEST TEMPORAL LAYER...



- > If max_dec_frame_buffering is 6, the first picture will be output when POC 56 is received.
- > max_dec_frame_buffering: 1
- > max_num_ref_frames: 1

PROBLEMS

- › With different DPB sizes for different temporal layers the complexity for an encoder that uses adaptive coding structure of reference pictures increases as it must ensure that DPB requirements are met for each temporal layer.
- › The output-status of decoded pictures might be different among the different temporal layers.



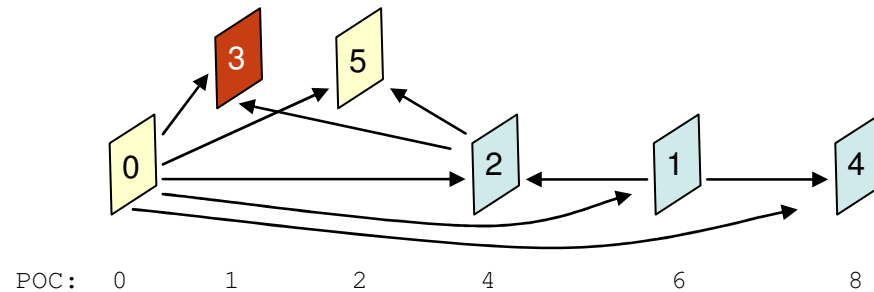
Two layers:

- › max_dec_frame_buffering: 4

One layer:

- › max_dec_frame_buffering: 1

MORE PROBLEMS



Two layers:

- › max_dec_frame_buffering: 3

One layer:

- › max_dec_frame_buffering: 2

- › What happens if temporal up-switching is performed between picture 4 and 5?
- › Picture 2 will already have been output when picture 5 is decoded. Output order will be violated.
- › To ensure that output order is not violated a network node may have to simulate the DPB status.
- › Unreasonably complex for a network node.

PROPOSED SOLUTION

- › Introduce a variable; OutputDistance for each picture
- › All pictures in the DPB with POC lower than $\text{POC}(\text{CurrPic}) - \text{OutputDistance}$ are output
- › Introduce an SPS flag; output_distance_always_zero_flag that can be used for most coding structures (e.g. AI, RA and LD in G1200)

A RESTRICTION

- › Add a restriction on the value of OutputDistance saying that if a picture has been output when decoding a picture in a higher layer it shall be output also when only the lower layer is decoded.
- › With this restriction pictures are output at the same time regardless of how many layers are decoded
- › A single DPB model is sufficient in the encoder
- › No complexity added at the network node

At any time instant, there shall be at most $\text{max_dec_frame_buffering}[i]$ pictures with temporal_id lower than or equal to i in the DPB for all i in the range of 0 to $\text{max_temporal_layers_minus1}$, inclusive.

- If the current picture is an IDR picture, the following applies.
 1. When the IDR picture is not the first IDR picture decoded and the value of `pic_width_in_luma_samples` or `pic_height_in_luma_samples` or `max_dec_frame_buffering` derived from the active sequence parameter set is different from the value of `pic_width_in_luma_samples` or `pic_height_in_luma_samples` or `max_dec_frame_buffering` derived from the sequence parameter set that was active for the preceding picture, respectively, `no_output_of_prior_pics_flag` is inferred to be equal to 1 by the HRD, regardless of the actual value of `no_output_of_prior_pics_flag`.

NOTE – Decoder implementations should try to handle changes in the value of `pic_width_in_luma_samples` or `pic_height_in_luma_samples` or `max_dec_frame_buffering` more gracefully than the HRD.
 2. When `no_output_of_prior_pics_flag` is equal to 1 or is inferred to be equal to 1, all frame buffers in the DPB are emptied without output of the pictures they contain, and DPB fullness is set to 0.
- Otherwise (the current picture is not an IDR picture), frame buffers containing a picture which are marked as "not needed for output" and "unused for reference" are emptied (without output), and the DPB fullness is decremented by the number of frame buffers emptied. When ~~there is no empty frame buffer (i.e., DPB fullness is equal to DPB size), the "bumping"~~ **there is one or more pictures `picX` in the DPB marked as "needed for output" with $\text{PicOrderCnt}(\text{picX}) < \text{PicOrderCnt}(\text{CurrPic}) - \text{OutputDistance}$, the output process specified in subclause C.4.2.1 is invoked repeatedly until there is an empty frame buffer to store the current decoded picture all pictures `picX` with $\text{PicOrderCnt}(\text{picX}) < \text{PicOrderCnt}(\text{CurrPic}) - \text{OutputDistance}$ have been marked as "not needed for output".**

When the current picture is an IDR picture for which `no_output_of_prior_pics_flag` is not equal to 1 and is not inferred to be equal to 1, the following two steps are performed.

1. Frame buffers containing a picture that is marked as "not needed for output" and "unused for reference" are emptied (without output), and the DPB fullness is decremented by the number of frame buffers emptied.
2. All non-empty frame buffers in the DPB are emptied by repeatedly invoking the ~~"bumping"~~ **output** process specified in subclause C.4.2.1, and the DPB fullness is set to 0.

C.4.1.1 ~~"Bumping"~~ **Output** process

The ~~"bumping"~~ **output** process is invoked in the following cases.

- The current picture is an IDR picture and `no_output_of_prior_pics_flag` is not equal to 1 and is not inferred to be equal to 1, as specified in subclause C.4.2.
- ~~There is no empty frame buffer (i.e., DPB fullness is equal to DPB size) and an empty frame buffer is needed for storage of a decoded (non-IDR) picture~~ **There is one or more pictures `picX` in the DPB marked as "needed for output" with $\text{PicOrderCnt}(\text{picX}) < \text{PicOrderCnt}(\text{CurrPic}) - \text{OutputDistance}$, as specified in subclause C.4.2.**

The ~~"bumping"~~ **output** process consists of the following ordered steps:

1. The picture that is first for output is selected as the one having the smallest value of `PicOrderCntVal` of all pictures in the DPB marked as "needed for output".
2. The picture is cropped, using the cropping rectangle specified in the active sequence parameter set for the picture, the cropped picture is output, and the picture is marked as "not needed for output".
3. If the frame buffer that included the picture that was cropped and output contains a picture marked as "unused for reference", the frame buffer is emptied and the DPB fullness is decremented by 1.

7.3.2.1 Sequence parameter set RBSP syntax

seq_parameter_set_rbsp() {	Descriptor
...	
max_temporal_layers_minus1	u(3)
for(i = 0; i < max_temporal_layers_minus1; i++) {	
max_num_ref_frames[i]	ue(v)
max_dec_frame_buffering[i]	ue(v)
}	
output_distance_always_zero_flag	u(1)
...	
max_num_ref_frames	ue(v)
max_dec_frame_buffering	ue(v)
...	
}	

max_num_ref_frames[i] specifies the maximum number of short-term and long-term reference frames, complementary reference field pairs, and non-paired reference fields **pictures** that may be used by the decoding process for inter prediction of any picture in the sequence **with temporal_id equal to i**. ~~max_num_ref_frames also determines the size of the sliding window operation.~~ The value of max_num_ref_frames shall be in the range of 0 to MaxDpbFrames, inclusive.

max_dec_frame_buffering[i] specifies the required size of the HRD decoded picture buffer (DPB) in units of frame buffers **when decoding all temporal layers lower than or equal to i**. The coded video sequence shall not require a decoded picture buffer with size of more than Max(1, max_dec_frame_buffering[i]) frame buffers **for pictures with temporal_id lower than or equal to i** to enable the output of decoded pictures at the output times specified by dpb_output_delay of the picture timing SEI messages. The value of max_dec_frame_buffering shall be greater than or equal to max_num_ref_frames. ~~An upper limit for the value of max_dec_frame_buffering is specified by the level limits in subclauses A.3.1, A.3.2, G.10.2.1, and H.10.2.~~

output_distance_always_zero_flag equal to 1 specifies that the variable OutputDistance is always 0 and that output_all_preceding_pics_flag is not present in the slice headers of any picture in the coded video sequence. output_distance_always_zero_flag equal to 0 specifies that output_all_preceding_pics_flag is present in the slice headers of all pictures in the coded video sequence and is used to calculate the value of the variable OutputDistance for each picture.

SLICE HEADER SYNTAX

7.3.3. Slice header syntax

slice_header() {	Descriptor
entropy_slice_flag	u(1)
if(!entropy_slice_flag) {	
...	
if(IdrPicFlag) {	
idr_pic_id	ue(v)
no_output_of_prior_pics_flag	u(1)
}	
else {	
pic_order_cnt_lsb	u(v)
if(output_distance_always_zero == 0) {	
output_all_preceding_pics_flag	u(1)
if(output_all_preceding_pics_flag == 0) {	
output_distance_idc	ue(v)
}	
}	
...	
}	

SEMANTICS

output_all_preceding_pics_flag equal to 1 specifies that all decoded pictures that are marked as "needed for output" shall be outputted and marked as "not needed for output" when their picture order count value is lower than the current picture before the decoding of the current picture as specified in Annex C. **output_all_preceding_pics_flag** equal to 0 specifies that **output_distance_idc** is present and will be used to specify which pictures shall be outputted before decoding of the current picture.

output_distance_idc is used to specify which pictures shall be outputted before decoding of the current picture. **output_distance_idc** shall be in the range of 0 to $\text{MaxPicOrderCntLsb}/2-1$.

The variable **OutputDistance** is calculated as follows:

If **output_distance_always_zero** equals 1, $\text{OutputDistance} = 0$

Otherwise, if **output_all_preceding_pics_flag** equals 0, $\text{OutputDistance} = 0$

Otherwise, if **output_distance_idc** equals 0, $\text{OutputDistance} = \text{MaxPicOrderCntLsb}/2$

Otherwise, (**output_distance_always_zero** equals 0, **output_all_preceding_pics_flag** equals 1 and **output_distance_idc** is not equal to 0) $\text{OutputDistance} = \text{output_distance_idc}$.

When at least one picture with **temporal_id** lower than or equal to the **temporal_id** of the current picture has been outputted the following applies:

Let **X** be the highest **PicOrderCntVal** of the **PicOrderCntVal** of all decoded pictures with **temporal_id** lower than or equal to the **temporal_id** of the current picture that have been outputted before invoking subclause C.4.2 for the current picture. Let **X'** be the highest **PicOrderCntVal** of the **PicOrderCntVal** of all decoded pictures with **temporal_id** lower than or equal to the **temporal_id** of the current picture that was outputted after invoking subclause C.4.2 for **prevPic**, where **prevPic** is the previous picture in decoding order with **temporal_id** lower than or equal to the **temporal_id** of the current picture. When $\text{X} \neq \text{X}'$ then **OutputDistance** shall be less than $\text{PicOrderCntVal} - \text{X}$.

NOTE – The above restriction ensures that if a picture has been outputted before the decoding of a picture in a higher temporal layer then it will be outputted by the current picture if the higher temporal layer is removed.

-
- › The WD text contains five instances of `max_num_ref_frames` that will be replaced with `max_num_ref_frames[temporal_id]`. The WD text contains seven instances of `max_dec_frame_buffering` that will be replaced by `max_dec_frame_buffering[temporal_id]`.

SUMMARY

1. Add possibility to signal different values of the following SPS level parameters for each temporal layer:
 - a. max_dec_frame_buffering
 - b. max_num_ref_frames
2. Replace “bumping” process with a syntax controlled output process
 - a. Guarantees correct output order
 - b. An SPS flag to cover most common coding structures
 - c. When the SPS flag is not set, two slice header parameters are present
 - d. Gives encoder control over output delay on frame by frame level



ERICSSON