

SONY

JCTVC-0272:  
SHVC: Upsampling with shorter-tap  
filters

[Kazushi.Sato@jp.sony.com](mailto:Kazushi.Sato@jp.sony.com)

# Agenda

- ◆ Introduction/Problem Statement
- ◆ Proposal
- ◆ Simulation
  - Simulation Condition
  - Simulation Result
- ◆ Proposed Change in SHVC WD
- ◆ Conclusion

# Introduction/Problem Statement

## ◆ Upsampling of the reconstructed reference layer picture is needed for spatial scalability.

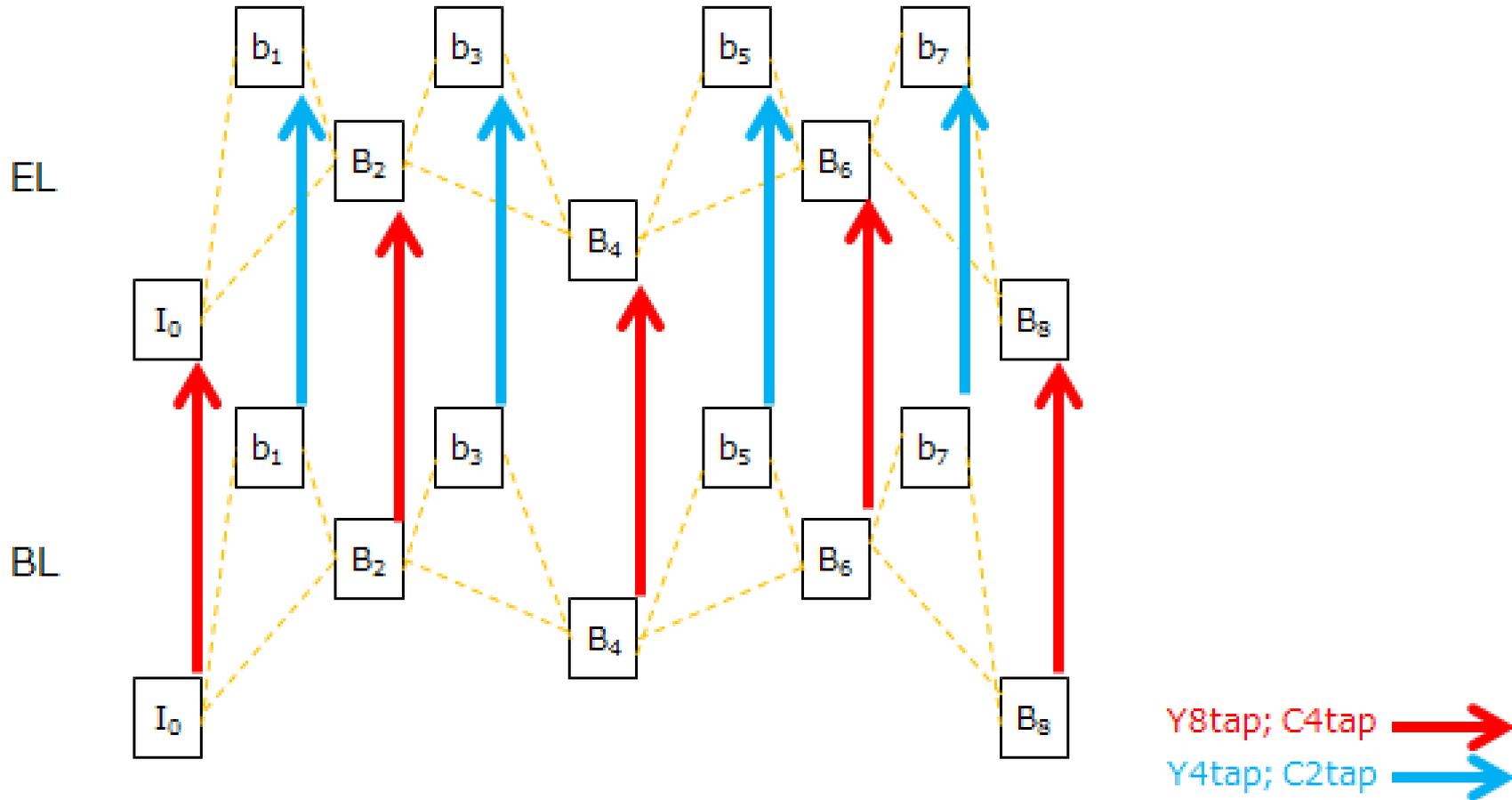
- Under the current SHVC specification same filter as the one for MC interpolation is applied also for upsampling.
- At the 14<sup>th</sup> JCTVC meeting in Vienna, it is proposed by JCTVC-N0265 to apply shorter-tap filters for complexity reduction.

## ◆ HEVC supports temporal scalability with hierarchical B-picture structure.

- With pictures in lower temporal hierarchy time distance between the current and the reference pictures are long.
  - With scalable coding more PUs are predicted from the reference layers, and in spatial scalability, image quality of the upsampled base-layer pictures do affect coding efficiency in the enhancement layers.
- With pictures in higher temporal hierarchy time distance between the current and the reference pictures are shorter.
  - With such pictures inter-layer prediction is not so helpful to improve coding efficiency as for pictures in lower temporal hierarchy.

# Proposal

- ◆ **It is proposed that shorter-tap filters for upsampling be applied for higher temporal layers.**
  - In the example shown in the next slide, shorter-tap filters are applied for the highest temporal layer.
- ◆ **It is also proposed to add a syntax element in `vps_extension()` that specifies from which temporal layer shorter-tap filter is applied for upsampling.**



# Simulation Condition

- ◆ **SHM-3.0.1 is used as anchor.**
- ◆ **On top of the anchor software the proposed method has been implemented by the author.**
- ◆ **As for the shorter-tap filters 4tap FIR used for chroma MC interpolation is applied for luma, and 2tap FIR (linear interpolation) is applied for chroma.**
- ◆ **2 tests have been conducted as:**
  - Experiment -1: Shorter-tap upsampling filter is applied just for the 1<sup>st</sup> highest temporal layer
  - Experiment -2: Shorter-tap upsampling filter is applied for the 1<sup>st</sup> and the 2<sup>nd</sup> highest temporal layers
- ◆ **Class A&B sequences are tested with RA\_2x and RA\_1.5x conditions.**

# Simulation Result

## Experiment -1:

	RA HEVC 2x			RA HEVC 1.5x		
	Y	U	V	Y	U	V
Class A	0.1%	0.0%	0.0%			
Class B	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
<b>Overall (Test vs Ref)</b>	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
<b>Overall (Test vs single layer)</b>	19.0%	33.1%	31.9%	16.2%	28.9%	29.3%
<b>Overall (Ref vs single layer)</b>	19.0%	33.1%	31.8%	16.2%	28.9%	29.2%
EL only (Test vs Ref)	0.0%	0.0%	0.1%	0.2%	0.2%	0.2%
Enc Time[%]	100.3%			100.1%		
Dec Time[%]	99.0%			100.3%		
Enc Mem[%]	#NUM!			#NUM!		
BL Match	Matched			Matched		

## Experiment -2:

	RA HEVC 2x			RA HEVC 1.5x		
	Y	U	V	Y	U	V
Class A	0.4%	0.2%	0.0%			
Class B	0.0%	0.0%	0.0%	0.2%	0.1%	0.2%
<b>Overall (Test vs Ref)</b>	0.1%	0.0%	0.0%	0.2%	0.1%	0.2%
<b>Overall (Test vs single layer)</b>	19.2%	33.2%	31.9%	16.4%	29.0%	29.4%
<b>Overall (Ref vs single layer)</b>	19.0%	33.1%	31.8%	16.2%	28.9%	29.2%
EL only (Test vs Ref)	0.3%	0.2%	0.2%	0.6%	0.5%	0.7%
Enc Time[%]	100.3%			100.1%		
Dec Time[%]	99.5%			100.3%		
Enc Mem[%]	#NUM!			#NUM!		
BL Match	Matched			Matched		

The author would like to thank Sharp for crosschecking! (JCTVC-00287).

# Proposed Change in SHVC WD

▪ <code>vps_extension()</code> { <sup>↵</sup>	Descriptor <sup>↵</sup>
▪ ... <sup>↵</sup>	u(1) <sup>↵</sup>
▪ <code>max_tid_ref_present_flag</code> <sup>↵</sup>	u(1) <sup>↵</sup>
▪ <code>if(max_tid_ref_present_flag)</code> <sup>↵</sup>	<sup>↵</sup>
▪ <code>for(i=0; i &lt; vps_max_layers_minus1; i++)</code> <sup>↵</sup>	<sup>↵</sup>
▪ <code>max_tid_il_ref_pics_plus1[i]</code> <sup>↵</sup>	u(3) <sup>↵</sup>
▪ <code>all_ref_layers_active_flag</code> <sup>↵</sup>	u(1) <sup>↵</sup>
▪ <code>for(i=0; i &lt; vps_max_layers_minus1; i++)</code> <sup>↵</sup>	<sup>↵</sup>
▪ <code>max_tid_il_upsampling_with_longer_filter_plus1[i]</code> <sup>↵</sup>	u(3) <sup>↵</sup>
▪ ... <sup>↵</sup>	u(10) <sup>↵</sup>
▪ } <sup>↵</sup>	<sup>↵</sup>

**max\_tid\_il\_upsampling\_with\_longer\_filter\_plus1** minus1 specifies the max tid where upsampling with longer filters are applied. Longer filters mean that 8/7-tap FIR for luma and 4-tap FIR for chroma specified for fractional pixel extraction process for inter prediction. Above the tid specified by this syntax element, 4-tap FIR originally for chroma fractional pixel extraction for inter prediction is applied for upsampling of luma, and linear interpolation is applied for upsampling of chroma. The value shall be smaller or equal to `vps_max_sub_layers_minus1 plus1`. When the value of `max_tid_ref_present_flag` is equal to 1, the value shall be smaller or equal to `max_tid_il_ref_pics_plus1[ i ]`.

# Conclusion

- ◆ **This contribution proposes to add option that enables shorter-tap filters for up-sampling be applied for higher temporal layers.**
  - Simulation results show that
  - by applying shorter-tap filter just for the 1<sup>st</sup> highest temporal layer loss in coding efficiency is observed by 0.0%, 0.0%, and 0.0% for Y,U and V component with the RA\_2x case, and 0.0%, 0.0%, and 0.1% for Y,U and V component with the RA\_1.5x case.
  - by applying shorter-tap filter for the 1<sup>st</sup> and 2<sup>nd</sup> highest temporal layers loss in coding efficiency is observed by 0.1%, 0.0%, and 0.0% for Y,U and V component with the RA\_2x case, and 0.2%, 0.1%, and 0.2% for Y,U and V component with the RA\_1.5x case.
- ◆ **The proposed method provides a trade-off between coding efficiency and complexity to the encoder manufactures.**
- ◆ **It is recommended that the proposed method be adopted into SHVC WD.**

# SONY

“Sony”はソニー株式会社の商標です。

各ソニー製品の商品名・サービス名はソニー株式会社またはグループ各社の登録商標です。その他の製品および会社名は、各社の商号、登録商標または商標です。