

JVET-AD0161

AHG9: Alternative Output Timing Hint SEI

Han Boon Teo, Jingying Gao, Chong Soon Lim,  
Kiyofumi Abe

## Motivation

Typically, video content is captured at 60fps or 30fps, and the captured video content is set to output at the same frame rate.

Invented in the early 20<sup>th</sup> century, slow-motion, is an effect in film-making whereby time appears to be slowed down. This can be accomplished using high-speed cameras for encoding and playing the content at a lower rate. Slow-motion can also be achieved by playing a normally recorded content at a slower speed. Slow-motion is useful to make it easier to see detail movement in fast motion.

Camera systems from many manufacturers supports this capability. Mobile phones from various manufacturers also supports different slow-motion capturing rates.



Slow-motion

## Current Limitation

Based on latest VVC and VSEI specifications, only one set of decoding time and output time can be present in the bitstream.

In the case when video bitstream contains slow-motion scenes, the actual motion speed of the content at captured time is lost!



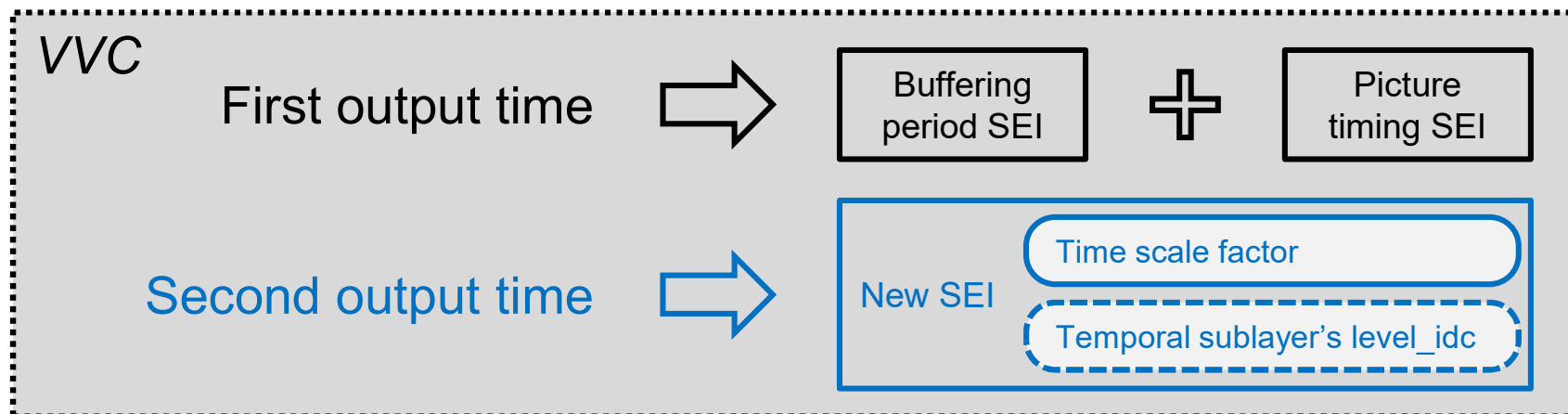
What is actual motion speed?!?

### Possible problems

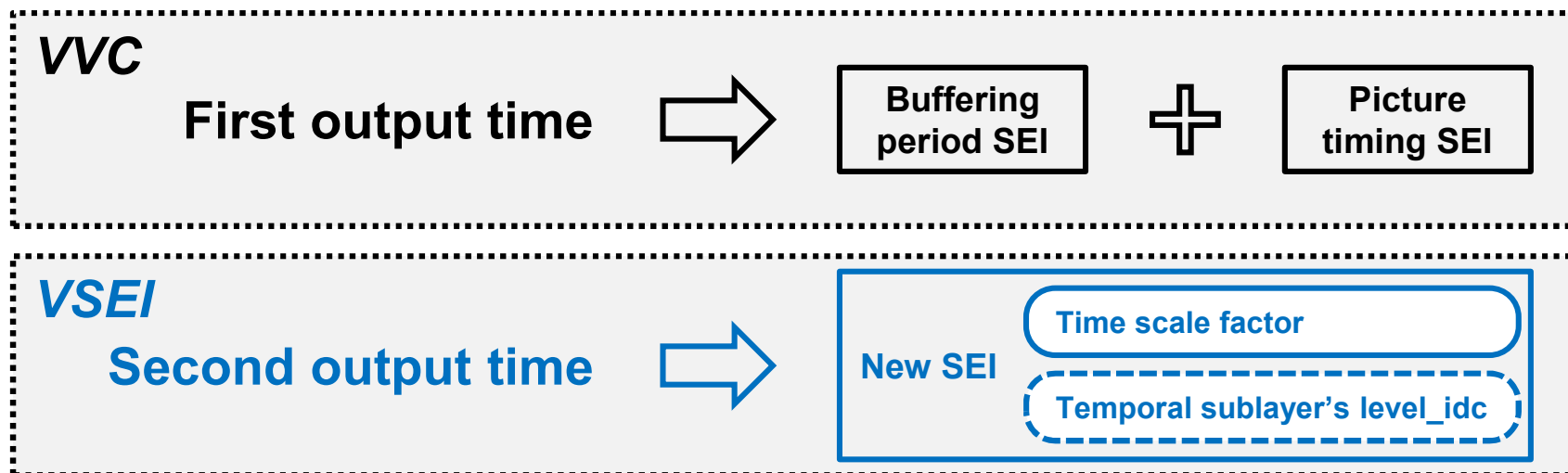
- For AI system, the training and inference accuracy of machine tasks would drop
- For human viewing, cannot playback at actual speed

# Proposal

JVET-AC0141  
*Proposed to*  
VVC

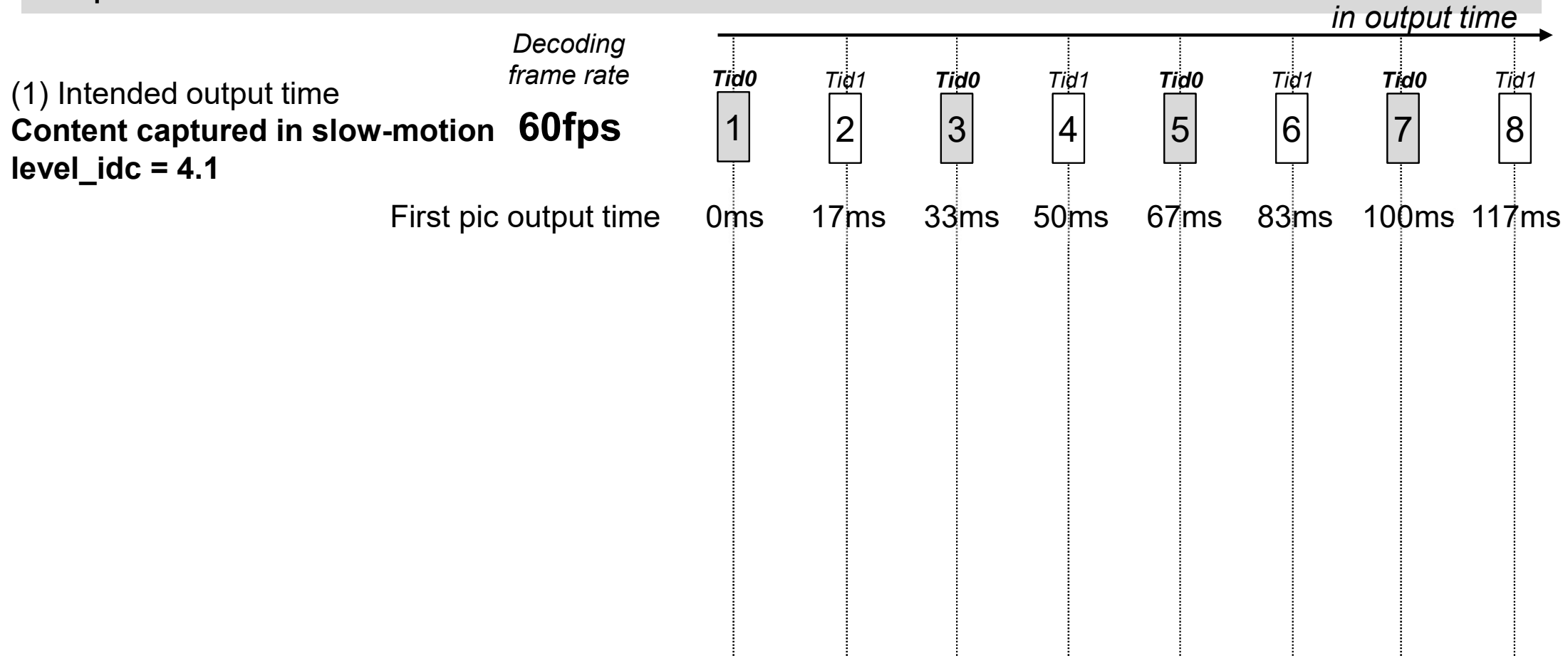


**This proposal**  
**for VSEI**



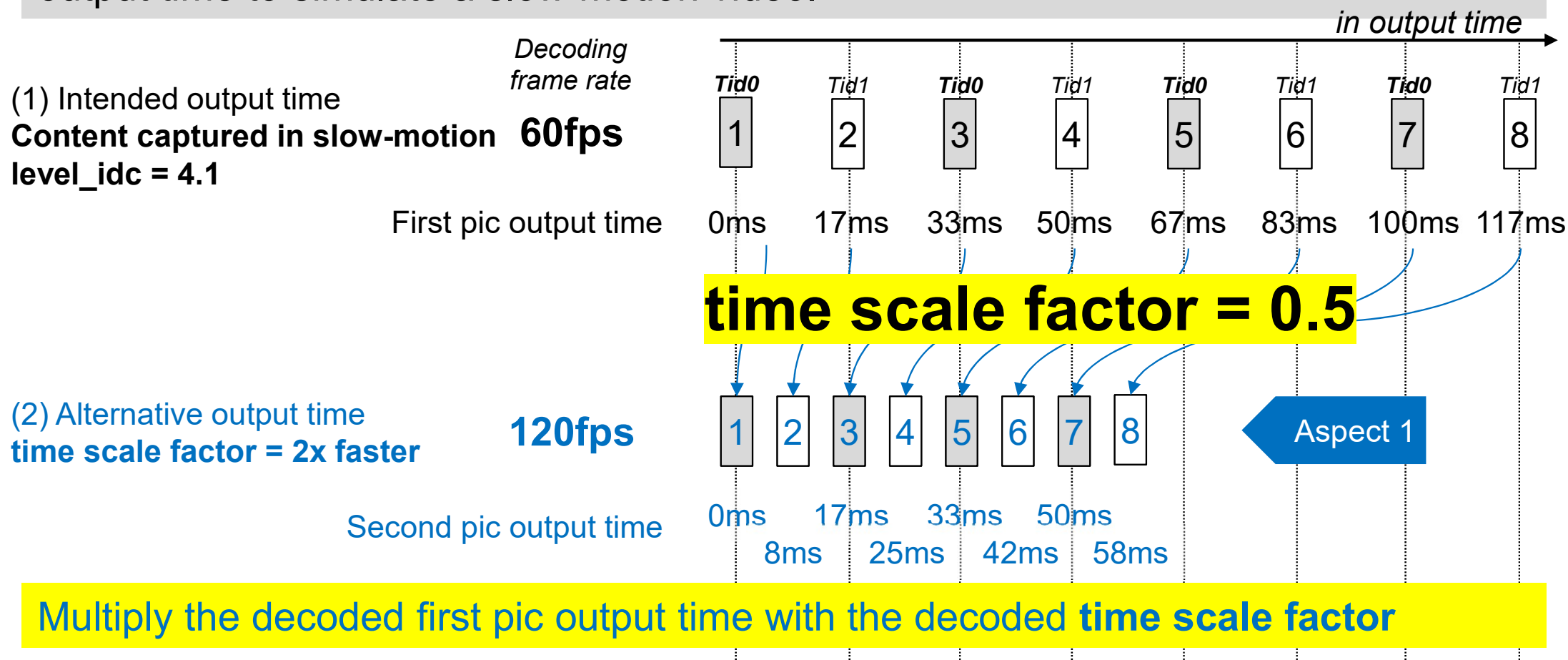
# Proposal

Example: Video content is captured in high frame rate (120fps) and stored as 60fps output time to simulate a slow-motion video.



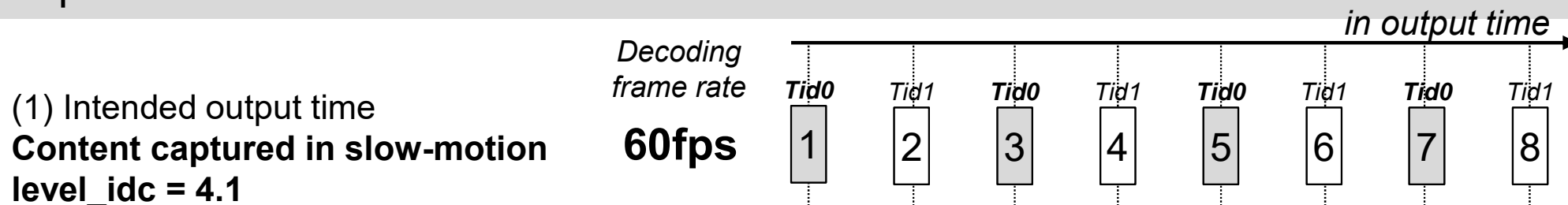
## Proposal (Aspect 1)

Example: Video content is captured in high frame rate (120fps) and stored as 60fps output time to simulate a slow-motion video.



## Proposal (Aspect 2)

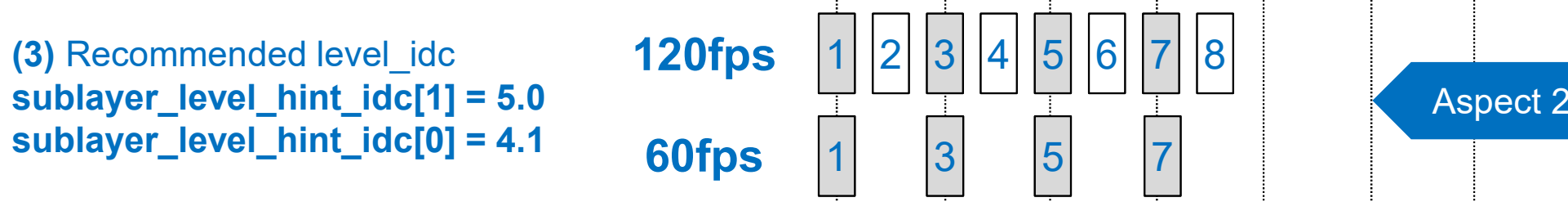
Example: Video content is captured in high frame rate (120fps) and stored as 60fps output time to simulate a slow-motion video. Assume initial CPB removal time = 0.



In order to view video content at actual speed:



To recommend suitable sublayer's level\_idc value for decoding the specific layer



## Proposed addition to VSEI: syntax

### 8.x.1 Alternative output timing hint SEI message syntax

		Descriptor
Aspect 1	alternative output timing hint( payloadSize ) {	
	aot_time_scale_factor_nominator_minus1	u(10)
	aot_time_scale_factor_denominator_minus1	u(10)
Aspect 2	aot_live_capture_flag	u(1)
	aot_sublayer_info_hint_present_flag	u(1)
	if( aot_sublayer_info_hint_present_flag ) {	
	aot_max_sublayer_info_hint_minus1	u(3)
	for( i = aot_max_sublayer_info_hint_minus1; i >= 0; i-- ) {	
	aot_sublayer_level_hint_present_flag[ i ]	u(1)
	if( aot_sublayer_level_hint_present_flag[ i ] )	
	aot_sublayer_level_hint_idc[ i ]	u(8)
	}	
	}	
	while( !byte_aligned( ) )	
	aot_reserved_zero_bit	u(1)
	}	

## Proposed addition to VSEI: persistence scope

Table 4 — Persistence scope of SEI messages

SEI message	Persistence scope
...	...
Alternative output timing hint	The CLVS containing the SEI message

#### 8.x.2 Alternative output timing hint SEI message semantics

The alternative output timing hint SEI message provides information to obtain an alternative output time of the content. It defines a scaling factor for use in calculating the new output time to indicate a time of origin capture, or alternative display time.

When an alternative output timing hint SEI message is present for any picture of a CLVS of a particular layer, an alternative output timing hint SEI message shall be present for the first picture of the CLVS. The alternative output timing hint SEI message persists for the current layer in decoding order from the current picture until the end of the CLVS. All alternative output timing hint SEI messages that apply to the same CLVS shall have the same content.

NOTE – It is not necessary to conduct HRD conformance checks with the alternative output timing hint SEI message.

## Proposed addition to VSEI: semantics (2 of 3)

### Aspect 1

**aot\_time\_scale\_factor\_nominator\_minus1** plus 1 and **aot\_time\_scale\_factor\_denominator\_minus1** plus 1 specifies the value for nominator and denominator respectively, for the calculation of timeScaleFactor. When **aot\_time\_scale\_factor\_nominator\_minus1** and **aot\_time\_scale\_factor\_denominator\_minus1** are not present, they are inferred to be equal to 1. The variable timeScaleFactor is derived as follows:

$$\text{timeScaleFactor} = (\text{aot\_time\_scale\_factor\_nominator\_minus1} + 1) \div (\text{aot\_time\_scale\_factor\_denominator\_minus1} + 1)$$

The alternative output time is obtained by multiplying the DPB output time with the timeScaleFactor.

**aot\_live\_capture\_flag** equals to 1 indicates the alternative output time is live capture time. **aot\_live\_capture\_flag** equals to 0 indicates the alternative output time is unspecified. When **aot\_live\_capture\_flag** is not present, it is inferred to be equal to 0.

## Proposed addition to VSEI: semantics (3 of 3)

### Aspect 2

**aot\_sublayer\_info\_hint\_present\_flag** equal to 1 specifies that information about recommended temporal sublayers is present in the `alternative_output_timing_hint( )` syntax structure. **aot\_sublayer\_info\_hint\_present\_flag** equal to 0 specifies that information about recommended temporal sublayers is not present in the `alternative_output_timing_hint( )` syntax structure.

**aot\_max\_sublayer\_info\_hint\_minus1** plus 1 specifies the maximum number of temporal sublayers that may be present in each CLVS referring to the SPS. The value of **aot\_max\_sublayer\_info\_hint\_minus1** shall be equal to the value of **sps\_max\_sublayers\_minus1**.

**aot\_sublayer\_level\_hint\_present\_flag[ i ]** equal to 1 specifies that recommended level information is present in the `alternative_output_timing_hint( )` syntax structure for the sublayer representation with **TemporalId** equal to **i**. **aot\_sublayer\_level\_hint\_present\_flag[ i ]** equal to 0 specifies that recommended level information is not present in the `alternative_output_timing_hint( )` syntax structure for the sublayer representation with **TemporalId** equal to **i**.

**aot\_sublayer\_level\_hint\_idc[ i ]** indicates the recommended **level\_idc** for sublayer with **TemporalId** equal to **i**. It is recommended that decoders with the same or higher **level\_idc** to decode this sublayer when the alternative output time is used. It is recommended that decoders with **level\_idc** lower than **aot\_sublayer\_level\_hint\_idc[ i ]** to avoid decoding this sublayer when the alternative output time is used.

**aot\_reserved\_zero\_bit** shall be equal to 0. The value 1 for **aot\_reserved\_zero\_bit** is reserved for future use by ITU-T | ISO/IEC. Decoders conforming to this version of this Specification shall ignore the value of **aot\_reserved\_zero\_bit**.

## Conclusion

The increased in capabilities of cameras to capture in high frame rate, and the popularity in slow-motion video leads to new video use cases which require more than one set of picture timings.

This SEI could also benefit future uses of the video involving AI or machine analysis.

### Benefits

- ✓ No transcoding needed to support 2 output times
- ✓ Actual capture time is closely coupled within the video bitstream
- ✓ Future proof

We propose to adopt the solution with either

- 1) Aspect 1, or
- 2) Aspect 1 and Aspect 2  
into VSEI specification