Document Number:	**IMTC1016**	**Date:**	September 22, 2016
Working Group:	Scalable and Simulcast Video (SSV)	**Status:**	Approved
Title:	**Specification of Interoperable Scalability Modes for the HEVC Video Coding Standard for Unified Communication Applications**		
Purpose:	Technical Specification		

Version 1.1h – September 22, 2016

Copyright statement and disclaimer

Table of Contents

1. Introduction

This document contains a specification of configuration modes of H.265/HEVC High Efficiency Video Coding systems as used in real-time point-to-point and multipoint Unified Communication (UC) applications. A mode comprises a set of properties of the bitstream that is produced by an encoder or transmitted by a communication system.

The goal of this specification is to support the use of HEVC video in the entire gamut of UC applications that use video. Targeted application scenarios thus include low-end mobile phone video chat, all the way to high-end, multi-monitor telepresence systems.

This specification assumes that the reader is familiar with the H.265/HEVC standard specification and its scalable extension specified in Annexes F and H ("SHVC"). In the following, the term 'HEVC' refers to the H.265/HEVC specification excluding the scalability features of Annexes F and H, whereas 'SHVC' refers specifically to systems that use the scalability features.

2. Normative references

[1] ITU-T Rec. H.265 | ISO/IEC 23008-2 High efficiency coding and media delivery in heterogeneous environments – Part 2: High efficiency video coding, . The standard is available at http://www.itu.int/rec/T-REC-H.265. Unless otherwise specified, this document refers to the edition published by the ITU-T in April 2015 (posted at the ITU-T web site link above). Annexes F and H of this specification contain the SHVC extension.

[2] S. Bradner, "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119, March 1997.

3. Terminology

3.1. Abbreviations

For the purposes of this specification, the following abbreviations apply (those with an asterisk '*' are copied from the H.265/HEVC specification):

BLA*	Broken Link Access picture
CRA*	Clean Random Access picture
IDR*	Instantaneous Decoder Refresh
IRAP*	Intra Random Access Point
NAL*	Network Abstraction Layer
POC*	Picture Order Count
PPS*	Picture Parameter Set
RADL*	Random Access Decodable Leading picture or access unit
RASL*	Random Access Skipped Leading picture or access unit

SEI*	Supplemental Enhancement Information
SPS*	Sequence Parameter Set
UC	Unified Communications
VCL*	Video Coding Layer
VPS*	Video Parameter Set

3.2. Definition

For the purposes of this specification, the following definitions apply (those with an asterisk '*' are copied from the H.265/HEVC specification):

Bitstream*	A sequence of bits comprising NAL units that forms the representation of coded pictures and associated data forming one or more coded video sequences.
DID	DependencyId[ ] as defined in Annex F of the H.265/HEVC specification.
IDR_N_LP	IDR slice with no leading pictures as defined in the H.265/HEVC specification.
Leading picture*	A picture that precedes the associated IRAP picture in output order.
LID	nuhLayerId as defined in Annex F of the H.265/HEVC specification.
NAL unit	The basic encapsulation structure in H.265/HEVC.
Output operation point	A bitstream that is created from an input bitstream by operation of the sub-bitstream extraction process with the input bitstream, a target highest TemporalId, and a target layer identifier list as inputs, and that is associated with a set of output layers.
PPSID	pps_pic_parameter_set_id as defined in the H.265/HEVC specification.
Reference frame	A frame that may be used for inter prediction in the decoding process of subsequent frame(s) in decoding order.
SPSID	sps_seq_parameter_set_id as defined in the H.265/HEVC specification.
SHVC base layer	Designates the bitstream in which all VCL NAL units with nuh_layer_id greater than zero are removed.
TID	TemporalId as defined in Annex F of the H.265/HEVC specification.
Trailing picture*	A non-IRAP picture that follows the associated IRAP picture in output order.
TRAIL_N	Coded slice segment of a non-reference trailing picture as defined in the H.265/HEVC specification
TRAIL_R	Coded slice segment of a reference trailing picture as defined in the H.265/HEVC specification
TSA_R	Coded slice segment of a reference temporal switching access picture as defined in the H.265/HEVC specification
TSA_N	Coded slice segment of a non-reference temporal switching access picture as defined in the H.265/HEVC specification
VPSID	vps_video_parameter_set_id as defined in the H.265/HEVC specification.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [2], regardless if they appear in small or capital letters.

4. General Properties

Encoders and decoders that conform to a particular UC Mode must meet the constraints defined in this specification. Mode and capability negotiation between encoders and decoders is assumed to be available, but it is outside the scope of this specification.

4.1. Mode Structure

Three UC modes are defined in this specification. They include HEVC with temporal scalability, SHVC with two spatial scalable layers, and SHVC with three spatial scalable layers. The intention of including modes with incremental scalability capabilities is to allow encoder chip and device manufacturers to gradually incorporate the necessary support into their devices. This specification does not provide support for quality or any other form of scalability, such as hybrid or color gamut scalability.

The UC Modes are as follows.

- UC Mode 1: HEVC with temporal scalability
- UC Mode 2: SHVC with two spatial layers and temporal scalability
- UC Mode 3: SHVC with three spatial layers and temporal scalability

Encoders that conform to higher level modes shall include the capabilities of encoding bitstreams associated with lower level modes. For example, encoders that conform to UC Mode 3 must be able to generate a two layer HEVC stream, i.e., UC Mode 2.

> NOTE: Change of UC Mode may be requested through signaling means that are outside the scope of this specification.

The UC Modes indicate properties of the bitstream that will be produced by an encoder or transmitted from a server without taking into account any adaptations that may happen dynamically at either the sender or during transport to the receiver. For example, a sender may elect to eliminate a layer to accommodate reduced available bit rate. Such adaptations are outside the scope of this specification.

4.2. Profiles

In order to facilitate interoperability among a large variety of UC systems, this specification includes both mandatory and optional profiles for video encoders. It is assumed that newer systems are capable of capability negotiation, and thus will enable the use of the optional and more sophisticated profiles.

Encoders and decoders conforming to this specification must support the Main profile. This requirement refers to UC Mode 1 and – due to the hierarchical construction of UC Modes – to all three UC Modes.

Encoders and decoders conforming to one of the scalable UC Modes (2 or 3) must support the Scalable Main profile.

> NOTE: Since decoders have to conform to one of the H.265/HEVC profiles indicated above they may have to support additional layering structures to those specified in this document (e.g., with additional temporal layers).

4.3. Resolutions and Frame Rates

Encoders and decoders that conform to this specification must be able to encode a bitstream with parameters as specified in this section.

> NOTE: Since decoders have to conform to one of the H.265/HEVC profiles indicated in Section 4.1, they may have to support additional parameters to those specified in this document (e.g., additional spatial resolutions or temporal frame rates).

4.3.1. Spatial Resolutions

Encoders and decoders conforming to this specification must be able to encode video sequences using all of the following spatial resolutions (in pixels):

1280x720, 960x540, 848x480, 640x360, 480x270, 424x240, 320x180

720x1280, 540x960, 480x848, 360x640, 270x480, 240x424, 180x320

Note that these resolutions have 16:9 and 9:16 aspect ratios.

Only progressive video is supported in this specification. Consequently, the SPS shall have general_progressive_source_flag equal to 1, general_interlaced_source_flag equal to 0, general_non_packed_constraint_flag equal to 1, general_frame_only_constraint_flag equal to 1.

A square sample aspect ratio (1:1) is required in this specification. Other sample aspect ratios may be supported in the future.

In case of a change from portrait to landscape orientation, or vice versa, the display orientation SEI message must be used to indicate portrait display of a landscape orientation coded frame, or the reverse. Section 6.1 specifies how the display orientation SEI message shall be used.

4.3.2. Temporal Resolutions

Encoders and decoders conforming to this specification must be able to encode and decode video at 30 frames per second.

When using temporal scalability, encoders that conform to this specification must generate bitstreams with dyadic frame rates (i.e., the ratio of the frame rates between any two temporal sub-layers is a power of 2).

Encoders should generate bitstreams with constant frame rates.

> NOTE: When there is adaptation, for example due to lighting level variations, packet losses, or rate control decisions, the frame rate may not end up being constant, or even dyadic (except when an entire layer is eliminated). The ratio of frames rates among temporal layers that is signaled, however, must be dyadic.

4.4. Levels

Encoders and decoders conforming to this specification must be able to operate at level 3.1, with the bit rate limited to 2 Mbps.

The resolutions, bit rates, frame rates, and levels listed in Sections 4.3 and 4.4 are a minimum set of requirements. Encoders that conform to this specification may support other resolutions, bit rates, and/or frame rates, or a higher level value than 3.1 as long as the constraints specified in Sections 4.3 and 4.4 are fulfilled. For a particular bitstream, encoders that conform to this specification shall use a level value that best describes the bitstream.

The same level must be used for all temporal sub-layers of a given layer.

5. UC Mode Definitions and Capabilities

5.1. UC Mode 1: HEVC with Temporal Scalability

This mode addresses the configuration of encoders that use HEVC with temporal scalability. Encoders conforming to this mode have to be able to generate two temporal sublayers. Decoders conforming to this mode have to be able to decode both two and three temporal sublayers. In the VPS, the value of vps_max_sub_layers_minus1 shall be equal to 1 or 2, to correspond to 2 or 3 temporal sub-layers, respectively.

In HEVC, temporal sub-layers are identified by the value of nuh_temporal_id_plus1 in the NAL unit header, and the TemporalId variable derived from it, TemporalId = nuh_temporal_id_plus1 - 1. In this specification, TID refers to the value of the TemporalId variable.

The hierarchical P prediction structure is used to achieve temporal scalability, with dyadic ratios of frame rates (i.e., the ratio of the frame rates between any two layers is a power of two).

> NOTE: Encoders may transition from one number of temporal sub-layers to another mid-stream, without any additional indication in the bitstream.

Figure 1 and Figure 2 depict the temporal picture coding structure for two and three sub-layer temporal scalability. These are the only temporal scalability structures allowed. The pictures are shown offset in

the vertical dimension to indicate their association with a different temporal sub-layer. Each picture is identified by a picture type letter ('I' or 'P') and a number indicating its temporal sub-layer (0 through 2, depending on the number of sub-layers available).

Figure 1 shows the picture coding structure for two temporal sub-layers. If the maximum frame rate of the source is 30 fps, then temporal sub-layer 0 consists of pictures I0 and P0 and has a frame rate of 15 fps. Temporal sub-layer 1 has a frame rate of 15 fps as well, and consists of pictures P1. Decoding of both sub-layers 0 and 1 results in 30 fps.

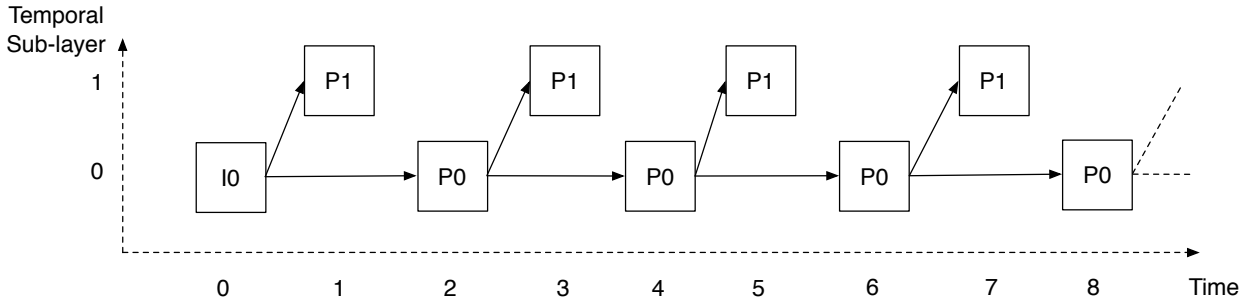

Figure 1. Two temporal sub-layer picture coding structure

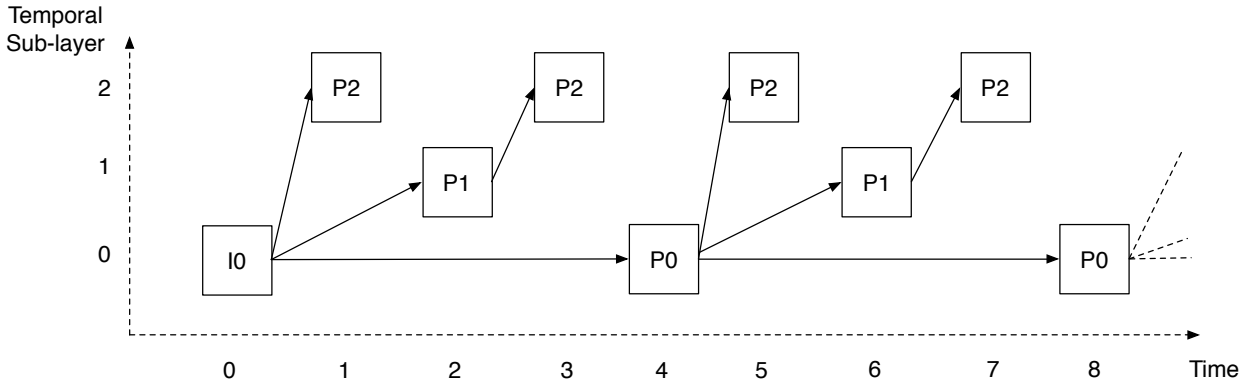

Figure 2. Three temporal sub-layer picture coding structure

Figure 2 shows the picture coding structure for the three temporal sub-layer mode. If the maximum frame rate of the source is 30 fps, then sub-layer 0 consisting of pictures I0 and P0 has a frame rate of 7.5 fps, sub-layer 1 consisting of pictures P1 has a frame rate of 7.5 fps, and sub-layer 2 consisting of pictures P2 has a frame rate of 15 fps.

For each NAL unit header, the value of TemporalId (TID) specifies the hierarchical dependency of a temporal sub-layer relative to other sub-layers, with 0 representing the lowest temporal sub-layer, 1 the next temporal sub-layer, and so forth.

The value of nuh_layer_id must be equal to 0.

The decoding order must be the same as the output order.

Multiple reference pictures as well as long-term reference pictures shall not be used.

Assuming source material of 360p 30 fps with three temporal sub-layers as an example, the coding structure will support the following output operation points, when formed by the sub-bitstream extraction process with a target temporal ID value.

1. 360p 7.5 fps as the sub-bitstream with target temporal ID of 0, includes the NAL units with TID=0.
2. 360p 15 fps as the sub-bitstream with target temporal ID of 1, includes the NAL units with TID=0 and TID=1.
3. 360p 30 fps as the full bitstream, equivalent to a sub-bitstream with target temporal ID of 2, includes the NAL units with TID=0, TID=1, and TID=2.

Table 1 provides an example bitstream structure for a UC Mode 1 stream, assuming a 360p 30fps resolution, with 3 temporal sub-layers. Even-numbered access units are shown in shaded cells.

Table 1. Mode 1 Bitstream Structure Example (3 temporal layers)

NAL unit (type)	Relevant fields in the NAL	Description
VPS (32)	VPSID = 0	VPS
SPS (33)	SPSID = 0, VPSID = 0	SPS
PPS (34)	PPSID = 0, SPSID = 0	PPS
IDR_N_LP slice (20)	PPSID = 0, POC = 0, TID = 0	IDR slice(s) at 7.5Hz
TSA_N slice (2)	PPSID = 0, POC = 1, TID = 2	P slice(s) at 30Hz
TSA_R slice (3)	PPSID = 0, POC = 2, TID = 1	P slice(s) at 15Hz
TSA_N slice (2)	PPSID = 0, POC = 3, TID = 2	P slice(s) at 30Hz
TRAIL_R slice (3)	PPSID = 0, POC = 4, TID = 0	P slice(s) at 7.5Hz
…	…	…

5.2. UC Mode 2: SHVC with 2 Spatial Layers

This mode addresses the configuration of encoders that use SHVC with two spatial layers with temporal scalability. Encoders conforming to this mode may generate any combination of one to three temporal sub-layers, and with a base layer and a spatial scalable enhancement layer as specified below.

Encoders conforming to this mode must be able to generate bitstreams with at least two temporal sub-layers, a base layer, and a spatial scalable enhancement layer.

The number of temporal sub-layers follows the constraints specified in Section 5.1.

Each access unit in the bitstream shall contain a picture from each of the two spatial layers. As a result the frame rate for both spatial layers will be the same.

The vertical and horizontal resolution ratios between successive spatial scalability layers must be 1.5 or 2 and identical in both dimensions.

Encoders conforming to Mode 2 must be able to generate bitstreams in Mode 1 or Mode 2. The run-time configuration is negotiated between decoders and encoders through a process which is outside the scope of this specification.

Let DId be the DependencyId[i] value of a layer, i, of a coded video sequence. DId shall be equal to 1 for the spatial enhancement layer.

An output layer set must be present in the VPS extension which includes the layer with the nuh_layer_id value of the spatial enhancement layer as a target output layer.

Figure 3 shows the coding structure where temporal scalability with two temporal sub-layers is combined with spatial scalability with one enhancement layer. Solid arrows represent temporal prediction and reference, and dashed arrows represent inter-layer prediction and reference. Each picture is identified by a one- or two-character type indication, followed by its temporal layer. The type indications are I for intra, P for predicted, EI for intra enhancement, and EP for P picture enhancement. The temporal sub-layers here are 0 and 1.

Figure 4 shows the coding structure where temporal scalability with three temporal sub-layers is combined with spatial scalability with one enhancement layer. These are the only two spatial and temporal prediction structures allowed in Mode 2.

Assuming source material of 720p 30 fps with two temporal sub-layers as an example and using a 2:1 resolution ratio, the coding structure of Figure 3 will support the following output operation points:

1. 360p 15 fps as the sub-bitstream with the base layer as the target output layer and target temporal ID of 0, includes the NAL units with (TID=0, DID=0).
2. 360p 30 fps as the sub-bitstream with the base layer as the target output layer and target temporal ID of 1, includes the NAL units with (TID=0, DID=0) and (TID=1, DID=0).

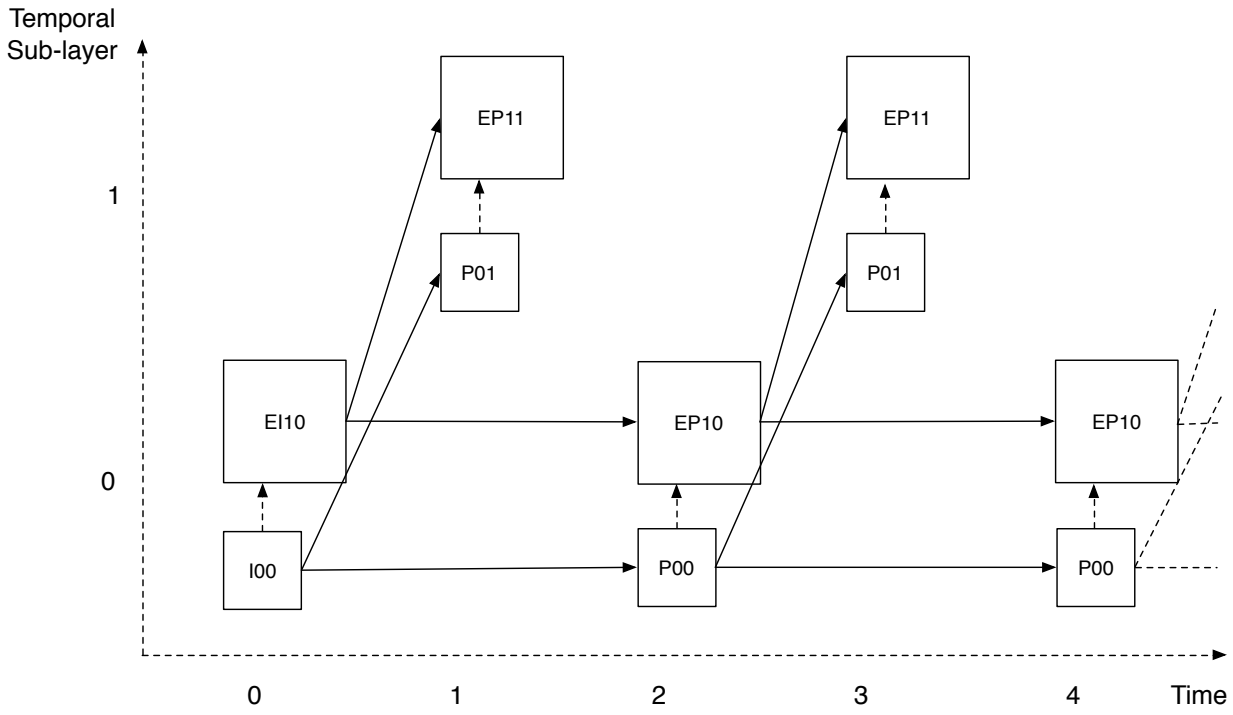

Figure 3. Example of 2-layer temporal combined with 2-layer spatial scalability (UC Mode 2)

3. 720p 15 fps as the sub-bitstream with the enhancement layer as the target output layer and target temporal ID of 0, includes the NAL units with (TID=0, DID=0) and (TID=0, DID=1).
4. 720p 30 fps as the full bitstream, equivalent to a sub-bitstream with the enhancement layer as the target output layer and target temporal ID of 1, includes the NAL units with (TID=0, DID=0), (TID=0, DID=1), (TID=1, DID=0), and (TID=1 DID=1).

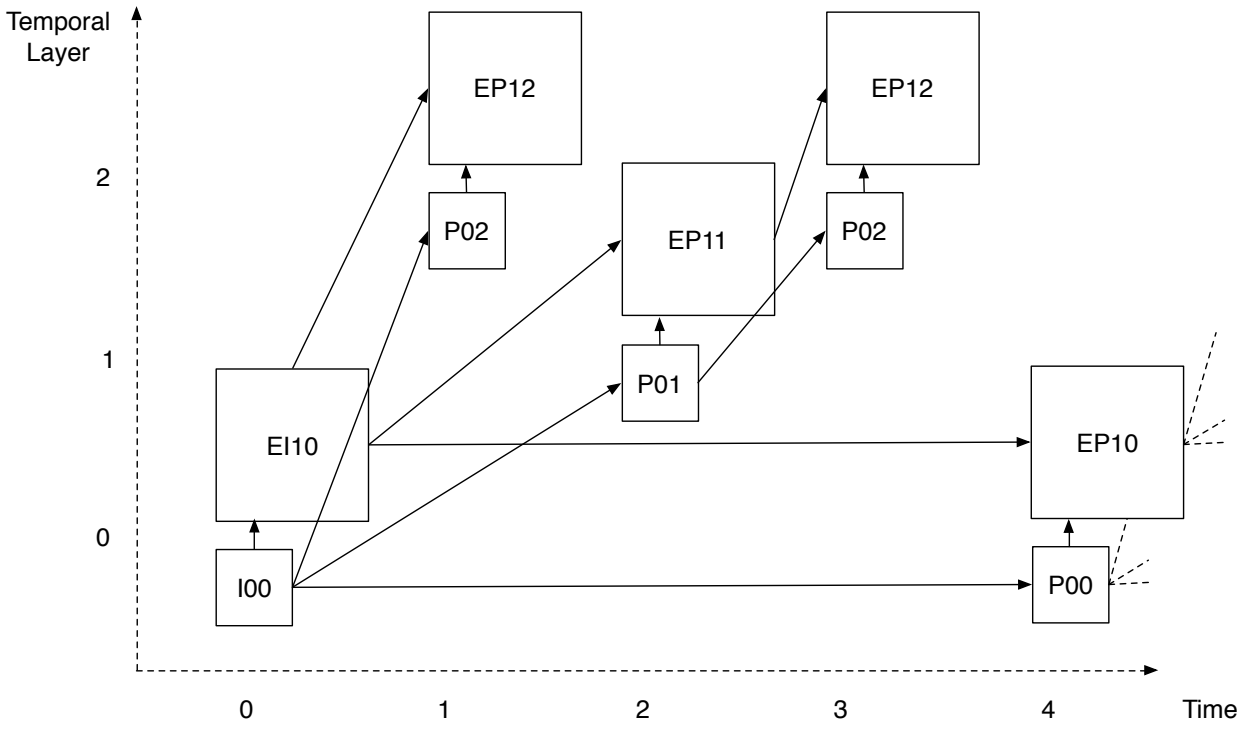

Figure 4. Example of 3-layer temporal combined with 2-layer spatial scalability (UC Mode 2)

Table 2 shows the output operation points supported by a bitstream with two temporal sub-layers and two spatial layers with 2:1 spatial scalability ratio.

Table 2. Mode 2 Output Operation Points Example

		Target output layer	
		Layer with DID = 0	Layer with DID = 1
target TID	0	360p 15fps	720p 15 fps
	1	360p 30fps	720p 30 fps

Table 3 illustrates the bitstream structure for the stream of Table 2. Even-numbered access units are shown in shaded cells.

Table 3. Mode 2 Bitstream Structure Example (3 temporal layers)

NAL unit (type)	Relevant fields in the NAL	Description
VPS (32)	VPSID = 0	VPS
SPS (33)	SPSID = 0, DID = 0	SPS of 360p layer
PPS (34)	PPSID = 0, SPSID = 0, DID = 0	PPS of 360p layer
SPS (33)	SPSID = 1, DID = 1	SPS of 720p layer

PPS (34)	PPSID = 1, SPSID= 1, DID = 1	PPS of 720p layer
IDR_N_LP slice (20)	PPSID = 0, POC = 0, TID = 0, DID = 0	IDR slice(s) in 360p layer at 7.5 fps
IDR_N_LP slice (20)	PPSID = 1, POC = 0, TID = 0, DID = 1	IDR slice(s) in 720p layer at 7.5 fps
TSA_N slice (2)	PPSID = 0, POC = 1, TID = 2, DID = 0	P slice(s) of 360p layer at 30 fps
TSA_N slice (2)	PPSID = 1, POC = 1, TID = 2, DID = 1	P slice(s) of 720p layer at 30 fps
TSA_R slice (3)	PPSID = 0, POC = 2, TID = 1, DID = 0	P slice(s) of 360p layer at 15 fps
TSA_R slice (3)	PPSID = 1, POC = 2, TID = 1, DID = 1	P slice(s) of 720p layer at 15 fps
TSA_N slice (2)	PPSID = 0, POC = 3, TID = 2, DID = 0	P slice(s) of 360p layer at 30 fps
TSA_N slice (2)	PPSID = 1, POC = 3, TID = 2, DID = 1	P slice(s) of 720p layer at 30 fps
TRAIL_R slice (1)	PPSID = 0, POC = 4, TID = 0, DID = 0	P slice(s) of 360p layer at 7.5 fps
TRAIL_R slice (1)	PPSID = 1, POC = 4, TID = 0, DID = 1	P slice(s) of 720p layer at 7.5 fps
…	…	…

5.3. UC Mode 3: SHVC with 3 layer Spatial Scalability

This mode addresses the configuration of encoders that use SHVC with three spatial layers with temporal scalability. Encoders conforming to this mode may generate any combination of one to three temporal sub-layers, and with a base layer and two spatial scalable enhancement layers as specified below.

Encoders conforming to this mode must be able to generate bitstreams with at least two temporal sub-layers, and with a base layer and two spatial scalable enhancement layers.

The number of temporal sub-layers follows the constraints specified in Section 5.1.

Each access unit in the bitstream shall contain a picture from each of the three spatial layers. As a result, the frame rate for all spatial layers will be the same.

The vertical and horizontal resolution ratios between successive spatial scalability layers must be 1.5 or 2 and identical in both dimensions.

Encoders conforming to Mode 3 must be able to generate bitstreams in Mode 1, Mode 2, and Mode 3. The run-time configuration is negotiated between decoders and encoders through a process which is outside the scope of this specification.

Let DId be the DependencyId[i] value of a layer, i, of a coded video sequence. DId shall be equal to 1 for the first spatial enhancement layer, and equal to 2 for the second spatial enhancement layer.

Figure 5 shows the coding structure where temporal scalability with two temporal sub-layers is combined with spatial scalability with two enhancement layers. Solid arrows represent temporal prediction and reference, and dashed arrows represent inter-layer prediction and reference. Each picture is identified by a one- or two-character type indication, followed by its temporal layer. The type indications are I for intra, P for predicted, EI for intra enhancement, and EP for P picture enhancement. The temporal sub-layers here are 0 and 1.

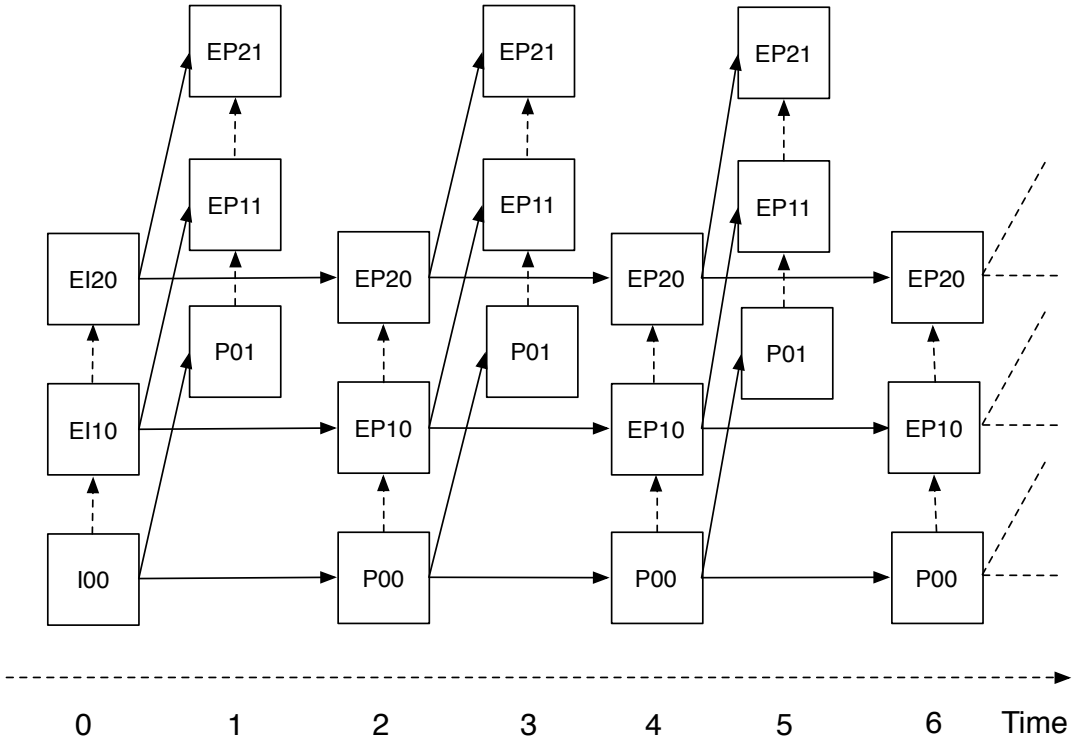

Figure 5. Example of 2-layer temporal combined with 3-layer spatial scalability (UC Mode 3)

Assuming source material of 720p 30 fps with two temporal sub-layers as an example and using a 2:1 resolution ratio, the coding structure of Figure 5 will support the following output operation points:

1. 180p 15 fps as the sub-bitstream with the base layer as the target output layer and target temporal ID of 0, includes the NAL units with (TID=0, DID=0).
2. 180p 30 fps as the sub-bitstream with the base layer as the target output layer and target temporal ID of 1, includes the NAL units with (TID=0, DID=0) and (TID=1, DID=0).
3. 360p 15 fps as the sub-bitstream with the first enhancement layer as the target output layer and target temporal ID of 0, includes the NAL units with (TID=0, DID=0) and (TID=0, DID=1).
4. 360p 30 fps as the sub-bitstream with the first enhancement layer as the target output layer and target temporal ID of 1, includes the NAL units with (TID=0, DID=0), (TID=0, DID=1), (TID=1, DID=0), and (TID=1 DID=1).
5. 720p 15 fps as the sub-bitstream with the second enhancement layer as the target output layer and target temporal ID of 0, includes the NAL units with (TID=0, DID=0), (TID=0, DID=1), and (TID=0 DID=2).
6. 720p 30 fps as the full bitstream, equivalent to a sub-bitstream with the second enhancement layer as the target output layer and target temporal ID of 1, includes the NAL units with (TID=0, DID=0), (TID=0, DID=1), (TID=0 DID=2), (TID=1, DID=0), (TID=1 DID=1), and (TID=1 DID=2).

Table 4 shows the partitioning into two temporal layers and three spatial resolutions (180p, 360p, and 720p).

Table 4. Mode 3 Output Operation Points Example

		Target output layer		
		Layer with DID = 0	Layer with DID = 1	Layer with DID = 2
target TID	0	180p 15 fps	360p 15 fps	720p 15 fps
	1	180p 30 fps	360p 30 fps	720p 30 fps

Table 5 illustrates the bitstream structure for the stream of Table 4. Even-numbered access units are shown in shaded cells.

Table 5. Mode 3 Bitstream Structure Example

NAL unit (type)	Relevant fields in the NAL	Description
VPS (32)	VPSID = 0	VPS
SPS (33)	SPSID = 0, DID = 0	SPS of 180p layer
PPS (34)	PPSID = 0, SPSID = 0, DID = 0	PPS of 180p layer
SPS (33)	SPSID = 1, DID = 1	SPS of 360p layer
PPS (34)	PPSID = 1, SPSID= 1, DID = 1	PPS of 360p layer
SPS (33)	SPSID = 2, DID = 2	SPS of 720p layer
PPS (34)	PPSID = 2, SPSID= 2, DID = 2	PPS of 720p layer
IDR_N_LP slice (20)	PPSID = 0, POC = 0, TID = 0, DID = 0	IDR slice(s) in 180p layer at 7.5 fps
IDR_N_LP slice (20)	PPSID = 1, POC = 0, TID = 0, DID = 1	IDR slice(s) in 360p layer at 7.5 fps
IDR_N_LP slice (20)	PPSID = 2, POC = 0, TID = 0, DID = 2	IDR slice(s) in 720p layer at 7.5 fps
TSA_N slice (2)	PPSID = 0, POC = 1, TID = 2, DID = 0	P slice(s) of 180p layer at 30 fps
TSA_N slice (2)	PPSID = 1, POC = 1, TID = 2, DID = 1	P slice(s) of 360p layer at 30 fps
TSA_N slice (2)	PPSID = 2, POC = 1, TID = 2, DID = 2	P slice(s) of 720p layer at 30 fps
TSA_R slice (3)	PPSID = 0, POC = 2, TID = 1, DID = 0	P slice(s) of 180p layer at 15 fps
TSA_R slice (3)	PPSID = 1, POC = 2, TID = 1, DID = 1	P slice(s) of 360p layer at 15 fps
TSA_R slice (3)	PPSID = 2, POC = 2, TID = 1, DID = 2	P slice(s) of 720p layer at 15 fps
TSA_N slice (2)	PPSID = 0, POC = 3, TID = 2, DID = 0	P slice(s) of 180p layer at 30 fps
TSA_N slice (2)	PPSID = 1, POC = 3, TID = 2, DID = 1	P slice(s) of 360p layer at 30 fps
TSA_N slice (2)	PPSID = 2, POC = 3, TID = 2, DID = 2	P slice(s) of 720p layer at 30 fps
TRAIL_R slice (1)	PPSID = 0, POC = 4, TID = 0, DID = 0	P slice(s) of 180p layer at 7.5 fps
TRAIL_R slice (1)	PPSID = 1, POC = 4, TID = 0, DID = 1	P slice(s) of 360p layer at 7.5 fps
TRAIL_R slice (1)	PPSID = 2, POC = 4, TID = 0, DID = 2	P slice(s) of 720p layer at 7.5 fps
…	…	…

6. General Constraints

6.1. Constraints for All Modes

The following constraints apply to all modes.

In the VPS, the value of vps_max_layers_minus1 shall be equal to 0, 1, or 2, depending on if the mode is 1, 2, or 3. The value of vps_base_layer_internal_flag shall be equal to 1, and the value of vps_base_layer_available_flag must be equal to 1. The value of vps_max_sublayers_minus1 must be 0, 1, or 2, depending on if the number of temporal layers is 1, 2, or 3.

B frames may be used, but picture reordering is not allowed, e.g. the picture output order and decoding order must be the same. In the SPS sps_max_num_reorder_pics[i] equal to 0 for all i in the range of 0 to sps_max_sub_layers_minus1, inclusive.

The bitstream may not contain BLA (16, 17, 18), CRA (21), RASL (8, 9) or RADL (6, 7) pictures or slices, as defined in the H.265/HEVC specification.vps_temporal_id_nesting_flag in the VPS and sps_temporal_id_nesting_flag in the SPS shall be equal to 1. This requirement enables decoders to perform temporal sub-layer up-switching at any picture.

If the display orientation SEI message is present ([1], Section D.2.17), the value of the anticlockwise_rotation parameter shall be a multiple of 16,384 (corresponding to degrees that are a multiple of 90), and the value of display_orientation_persistence_flag and display_orientation_cancel_flag shall be zero.

6.2. Constraints for Modes 2 and 3

The following bitstream constraints apply to UC Mode 2 and 3 bitstreams.

Encoders shall create a bitstream with a VPS extension containing an output layer set that includes the highest spatial layer in the bitstream as a target output layer. Specifically, in the VPS extension the default_output_layer_idc must be equal to 1. Furthermore, for different layer sets alt_output_layer_flag [i] should be equal to 1, where i is the i-th output layer set, to indicate that if the target output layer is not available, the next highest layer available will be output.

When a layer is removed from a bitstream, so that the number of layers present in the bitstream is not equal to vps_max_layer_minus1 + 1, the layers not present SEI message shall be present in the bitstream and indicate the missing layers.

# # #