



Microsoft Lync

Unified Communication Specification for H.264 AVC and SVC UCConfig Modes V 1.1

Disclaimer: This document is provided "as-is". Information and views expressed in this document, including URL and other Internet Web site references, may change without notice. You bear the risk of using it.

This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.

Feedback: You have no obligation to give Microsoft any suggestions, comments or other feedback ("Feedback") relating to these Materials. However, any Feedback you voluntarily provide may be used in Microsoft products and related specifications or other documentation (collectively, "Microsoft Offerings") which in turn may be relied upon by other third parties to develop their own products. Accordingly, if you do give Microsoft Feedback on any version of these Materials or the Microsoft Offerings to which they apply, you agree: (a) Microsoft may freely use, reproduce, license, distribute, and otherwise commercialize your Feedback in any Microsoft Offering; (b) you also grant third parties, without charge, only those patent rights necessary to enable other products to use or interface with any specific parts of a Microsoft product that incorporate your Feedback; and (c) you will not give Microsoft any Feedback (i) that you have reason to believe is subject to any patent, copyright or other intellectual property claim or right of any third party; or (ii) subject to license terms which seek to require any Microsoft Offering incorporating or derived from such Feedback, or other Microsoft intellectual property, to be licensed to or otherwise shared with any third party.

© 2011 Microsoft Corporation. All rights reserved.

1. Table of Contents

1.	Table of Contents	2
2.	Introduction	3
3.	Normative references	3
4.	Terminology	3
4.1.	Abbreviations	3
4.2.	Definition	4
5.	Overview of the UCConfig Modes.....	4
6.	Elementary Stream Conformance.....	6
6.1.	Encoding Toolset	6
6.1.1.	UC Constrained High Toolset	7
6.1.2.	UC Scalable Constrained High Toolset	8
6.1.3.	UC Scalable Constrained Baseline Toolset	9
6.2.	Resolutions and Frame Rate	9
6.2.1.	General Constraints	9
6.2.2.	Conformance Requirement.....	10
7.	UCConfig Mode Definitions and Capabilities	11
7.1.	UCConfig Mode 0: H.264 AVC single layer mode with Simulcast	11
7.2.	UCConfig Mode 1: H.264 SVC temporal scalability with simulcast.....	14
7.3.	UCConfig Mode 2q: H.264 SVC temporal and SNR scalabilities with simulcast	19
7.4.	UCConfig Mode 2s: H.264 SVC temporal and spatial scalabilities with simulcast.....	25
7.5.	UCConfig Mode 3: H.264 Full SVC scalability with simulcast.....	30
8.	Conformance Test Configuration	35
8.1.	Resolutions.....	35
8.1.1.	Landscape Orientation	36
8.1.2.	Portrait Orientation	36
8.1.	Frame Rates	37

2. Introduction

This document contains a specification for H.264 UCConfig Modes for real-time point to point and conferencing products.

This specification assumes that the reader is familiar with the H.264/AVC standard specification and its SVC extensions in Annex G.

3. Normative references

The H.264/MPEG-4 AVC and SVC standards (referred to hereafter simply as H.264) are specified in the following document:

1. ITU-T Rec. H.264 | ISO/IEC 14496-10 Advanced video coding for generic audiovisual services. The standard is available at <http://www.itu.int/rec/T-REC-H.264>. Unless otherwise specified, this document refers to the edition approved by ITU-T in March 2010 (posted at the ITU-T web site link above).
2. The Scalable Video Coding (SVC) extensions to the H.264/MPEG-4 AVC standard (referred to hereafter simply as SVC) are specified in Annex G of the above document.

4. Terminology

4.1. Abbreviations

For the purposes of this specification, the following abbreviations apply:

ASO	Arbitrary Slice Ordering
CABAC	Context-based Adaptive Binary Arithmetic Coding
CAVLC	Context-based Adaptive Variable Length Coding
FMO	Flexible Macroblock Ordering
IDR	Instantaneous Decoding Refresh
MB	Macroblock
NAL	Network Abstraction Layer
POC	Picture Order Count
PPS	Picture Parameter Set
QP	Quantization Parameter
SEI	Supplemental Enhancement Information

SPS	Sequence Parameter Set
SSPS	Subset Sequence Parameter Set
SVC	Scalable Video Coding

4.2. Definition

For the purposes of this specification, the following definitions apply:

Bitstream	A sequence of bits that forms representation of a NAL unit stream.
DID	dependency_id defined in Annex G of the H.264 specification.
NAL unit	an H.264 syntax structure containing a one-byte header and the payload byte string.
PPSID	picture_parameter_set_id defined in the H.264 specification.
PRID	priority_id defined in Annex G of the H.264 specification.
QID	quality_id defined in Annex G the H.264 specification.
Reference frame	a frame that may be used for inter prediction in the decoding process of subsequent frame(s) in decoding order.
SPSID	seq_parameter_set_id defined in the H.264 specification.
SVC base layer	designate the layer that does not contain quality or spatial scalability (i.e. if exists, quality_id and dependency_id must be zero but temporal_id does not necessary have to be).
TID	temporal_id defined in Annex G of the H.264 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.

5. Overview of the UCConfig Modes

The purpose of this program is to support a wide range of hardware and software H.264, AVC and SVC, encoders with different capabilities targeting for various applications and scenarios (such as low-end mobile phone video chat up to high-end telepresence system). So a tiered approach, or UCConfig Modes, from low to high capabilities is designed and defined to allow these different encoders to use in the Unified Communication video systems.

Five UCConfig Modes are defined in this program, ranging from AVC single layer, SVC temporal scalability, SVC temporal and SNR scalability¹, SVC temporal and spatial scalability, to SVC full scalability, with the capability of generating multiple independent simulcast streams. Each simulcast stream may have different resolutions, different frame rates, or/and different number of scalable layers. The intention of including modes with incremental scalability is to allow the existing encoder chip and device (such as webcam) manufacturers who have produced, are working on, or are planning to produce, H.264/AVC encoders to be able to plug in the HW encoders into a conferencing system, and pave the way for transitioning to full SVC profiles which enables more scenarios for video conferencing.

Here is a summary of the UCConfig Modes.

- UCConfig Mode 0: Non-scalable single layer AVC bitstream with Simulcast (number of simulcast streams ≥ 1).
- UCConfig Mode 1: SVC temporal scalability with hierarchical P with Simulcast (number of simulcast streams ≥ 1).
- UCConfig Mode 2q: SVC temporal scalability + Quality/SNR scalability with Simulcast (number of simulcast streams ≥ 1).
- UCConfig Mode 2s: SVC temporal scalability + spatial scalability with Simulcast (number of simulcast streams ≥ 1).
- UCConfig Mode 3: Full SVC scalability (temporal scalability + SNR scalability + spatial scalability) with Simulcast (number of simulcast streams ≥ 1).

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 UCConfig Mode 2q must be able to generate a single layer AVC stream (i.e. UCConfig Mode 0). The following diagram is the superset/subset structure and relationship for all UCConfig Modes. Details of each UCConfig Mode are elaborated in Section 7.

¹ SNR scalability is also known as quality scalability

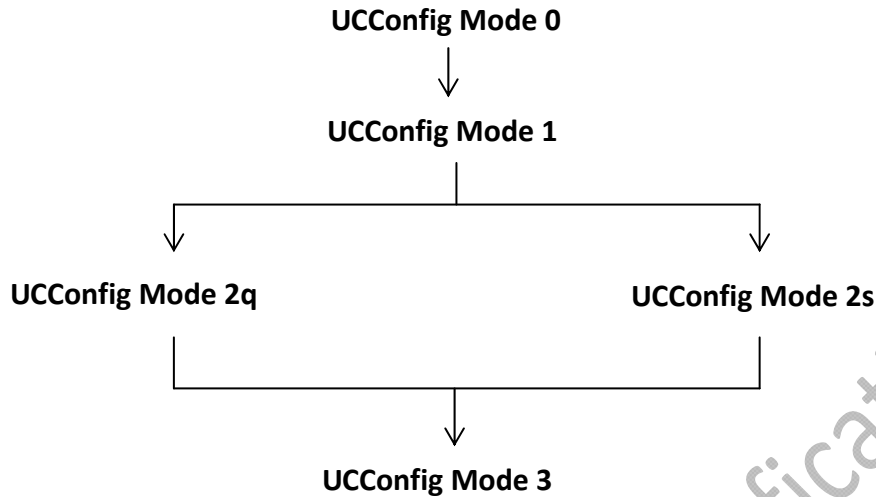


Figure 1. UCConfig Mode Superset/Subset Relationship

6. Elementary Stream Conformance

Encoders that conform to a particular UCConfig Mode must meet the constraints specified in this specification. Mode and capability negotiation between encoders and decoders is out of the scope of this specification.

6.1. Encoding Toolset

Here are the basic requirements that encoders shall meet in conformance to this specification:

- H.264 UCConfig Mode 0: Conformant to the Constrained Baseline profile or Unified Communication (UC) Constrained High toolset.
- H.264 UCConfig Mode 1: Conformant to the Unified Communication (UC) Constrained High toolset.
- H.264 UCConfig Modes 2q, 2s, and 3: Conformant to the UC Scalable Constrained Baseline toolset and the UC Scalable Constrained High toolset.

The UC Constrained High and UC Scalable Constrained High toolsets are a subset of High and Scalable High profiles respectively. The main reason for defining these two toolsets is to assure consistency between base and enhancement layers. Existing H.264 profiles use different sets of tools between Baseline and Scalable Baseline profiles, and between High and Scalable High profiles. For example, when using Scalable Baseline profile, CABAC is not allowed for the base layer, while it is allowed for the enhancement layers. The other reason for defining these two profiles is to have a set of tools needed for conferencing applications and exclude those that are less useful. For example, FMO and ASO are

excluded from the UC Constrained High and the UC Scalable Constrained High toolsets as they reduce coding efficiency. The UC Scalable Constrained Baseline toolset differs from the UC Scalable Constrained High toolset in that the base layer must obey the Constrained Baseline profile. This enables legacy AVC decoders that only support the Constrained Baseline profile with the opportunity of displaying base layer quality.

The details of the UC Constrained High, UC Scalable Constrained High and UC Scalable Constrained Baseline toolsets are elaborated in the following sections.

6.1.1. UC Constrained High Toolset

The UC Constrained High toolset contains the following encoding tools:

- All the tools defined in the Constrained Baseline profile:
 - I and P slices
 - 8 bits per sample
 - 4:2:0 format
 - 4x4 transform
 - CAVLC
- And these Main/High profile tools:
 - CABAC
 - 8x8 transform
 - Quantizer scaling matrices
 - QP for Cb/Cr

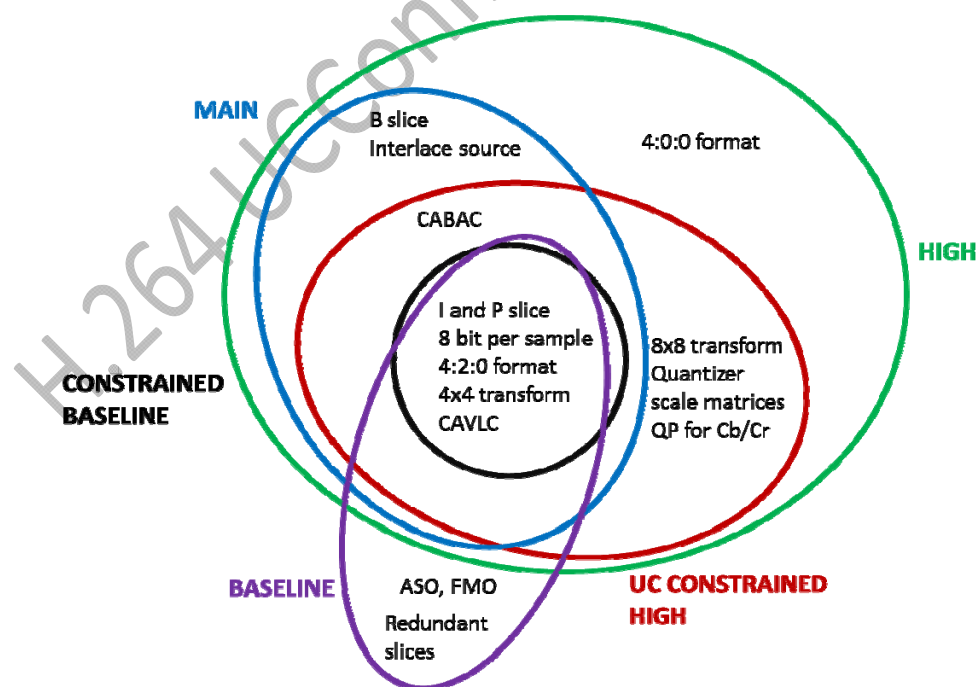


Figure 2. UC Constrained High Toolset

The use of UC Constrained High toolset for bitstreams is signaled via system level approaches that are outside the scope of this specification. For such bitstreams, profile_idc (and constraint set flags) should be equivalent to that corresponding to the High profile (profile_idc being equal to 100).

6.1.2. UC Scalable Constrained High Toolset

The UC Scalable Constrained High toolset contains the following encoding tools:

- The base layer bitstream shall obey the UC Constrained High toolset:
- The enhancement layer bitstream shall contain all the tools defined in the UC Constrained High toolset plus the following tools:
 - Full temporal and SNR scalability,
 - Spatial scalability is restricted to resolution ratios of 1.5 and 2 between successive spatial layers in both horizontal and vertical direction.
 - No Extended Spatial Scalability Support

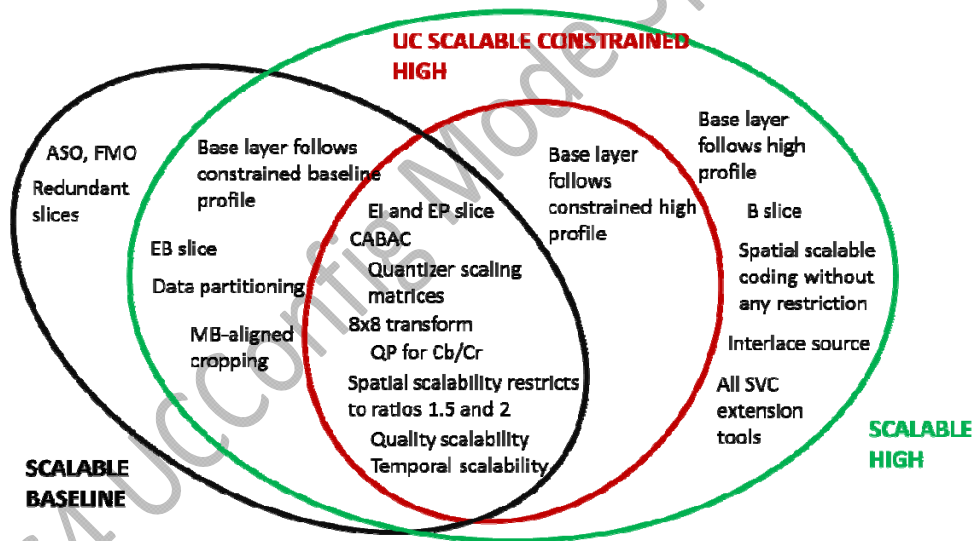


Figure 3. UC Scalable Constrained High Toolset

The use of UC Scalable Constrained High toolset is signaled via system level approaches that are outside the scope of this specification. For such bitstreams, profile_idc (and constraint set flags) in the base layer bitstream should be equivalent to that corresponding to the High profile (profile_idc being equal to 100) and profile_idc (and constraint set flags) in the enhancement layer bitstream should be equivalent to that corresponding to the Scalable High profile (profile_idc being equal to 86).

6.1.3. UC Scalable Constrained Baseline Toolset

The UC Scalable Constrained Baseline toolset contains the following encoding tools:

- The base layer bitstream contains all the tools defined in the Constrained Baseline profile.
- The enhancement layer bitstream contains all the tools defined in the UC Scalable Constrained High toolset.

The use of UC Scalable Constrained Baseline toolset is signaled via system level approaches that are outside the scope of this specification. For such bitstreams, `profile_idc` (and constraint set flags) in the base layer bitstream must be equivalent to that corresponding to the Constrained High profile (`profile_idc` is equal to 66 or `constraint_set0_flag` is equal to 1, and `constraint_set1_flag` is equal to 1) and `profile_idc` in the enhancement layer bitstream should be equivalent to that corresponding to the Scalable High profile (`profile_idc` being equal to 86).

Video encoded according to this specification shall use the H.264/AVC Annex B byte stream format. For purposes of this specification, any zero-valued bytes that appear at the end of an H.264 Annex B byte stream NAL unit (referred to as "trailing_zero_8bits" in the H.264 specification) are considered part of the NAL unit.

6.2. Resolutions and Frame Rate

Encoders that conform to this specification must be able to encode a bitstream specified in this section.

6.2.1. General Constraints

The maximum resolution allowed for this specification is 1080p for 16:9 aspect ratio, or VGA for 4:3 aspect ratio, or Pano288p for 20:3 aspect ratio in the landscape orientation, and 1080x1920 for 9:16 aspect ratio, or 480x640 for 3:4 aspect ratio in the portrait orientation. The minimum resolution allowed is QQVGA for 4:3 aspect ratio, or 180p for 16:9 aspect ratio, or Pano96p for 20:3 aspect ratio in the landscape orientation, and 120:160 for 3:4 aspect ratio, or 180x320 for 9:16 aspect ratio in the portrait orientation.

When other resolutions or other aspect ratios are supported, encoders that conform to this specification must generate bitstreams with the following constraints:

- When spatial scalability is supported, the vertical and horizontal resolution ratios between successive layers must be 1.5 or 2.
- When simulcast is supported, the vertical and horizontal resolution ratios between successive streams must be dividable by 1, 1.5 or 2.

The maximum and minimum frame rates are 60fps and 7.5fps respectively. When different frame rates are supported for temporal scalability, encoders that conform to this specification must generate bitstreams with dyadic frame rates.

Encoder developers are encouraged to generate bitstreams with constant frame rates. For encoders that do not support constant frame rates in certain situations (e.g. in low light conditions), frame rates lower than the nominal frame rate are allowed. In this case, the encoder must assure dyadic temporal scalability is preserved.

Only progressive video is allowed in this specification. Sample aspect ratio 1:1 (square) is required in this specification. Other sample aspect ratios may be supported later.

6.2.2. Conformance Requirement

Table 1 summarizes the set of resolutions and frame rates as well as the corresponding level that must be supported for each UCConfig Mode:

Table 1. Minimum Requirements for Resolutions and Frame rates

UCConfig Mode	Resolution and Frame Rate	Level
0	360p 30fps, 270p 30fps, 180p 15fps, 90p 15fps, VGA 30fps, QVGA 15fps, and QQVGA 15fps	3
1	720p 30fps, 540p 30fps, 360p 30fps, 270p 30fps, 180p 15fps, 90p 15fps, VGA 30fps, QVGA 15fps, QQVGA 15fps	3.1
2q	720p 30fps, 540p 30fps, 360p 30fps, 270p 30fps, 180p 15fps, 90p 15fps, VGA 30fps, QVGA 15fps, QQVGA 15fps	3.1
2s	720p 30fps, 540p 30fps, 360p 30fps, 270p 30fps, 180p 15fps, 90p 15fps, VGA 30fps, QVGA 15fps, and QQVGA 15fps	3.1
3	720p 30fps, 540p 30fps, 360p 30fps, 270p 30fps, 180p 15fps, 90p 15fps, VGA 30fps, QVGA 15fps, and QQVGA 15fps	3.1

Note that the resolutions, frame rates, and levels listed above are a minimum requirement. Encoders that conform to this specification may support other resolutions and/or frame rates, or a higher level value as long as the constraints specified in Section 6.2.1 are fulfilled. For a particular bitstream, encoders that conform to this specification shall use a level value that best describes the bitstream.

7. UCConfig Mode Definitions and Capabilities

7.1. UCConfig Mode 0: H.264 AVC single layer mode with Simulcast



Figure 4. AVC single layer mode

In this UCConfig Mode, one reference frame is used to constitute the single layer bitstream. Each P frame uses the immediate previously reconstructed frame as the reference frame.

This UCConfig Mode may support multiple independent simulcast streams generated from the encoder at the same time. Each independent stream has different bitrates and is conformant to UCConfig Mode 0. For each simulcast stream, the encoder shall insert a prefix NAL unit (type 14) in front of the first coded slice of each frame. The prefix NAL signals the relative importance of the stream among all in syntax element `priority_id`, with 0 representing the lowest bitrate stream, 1 the next lowest bitrate stream, and so forth (lower values mean higher priority).

Informative note: The values of `priority_id` in prefix NAL units are used by the Multipoint Control Unit (MCU) as a means to identify the priority of a simulcast stream.

For each prefix NAL, the values of `dependency_id`, `quality_id` and `temporal_id` must be equal to 0. The values of `no_inter_layer_pred_flag`, `discardable_flag` and `output_flag` must be 1. The value of `use_ref_base_pic_flag` must be 0. The prefix NAL units shall be discarded by legacy AVC decoders that are not SVC-compatible, so that the bitstream can still be decoded.

When interleaving of multiple streams into one single session is requested, the encoder must generate unique `seq_parameter_set_id` and `picture_parameter_set_id` across all simulcast streams so the extraction of one stream from the interleaved bitstream is feasible. When present, SEI messages are associated with the stream corresponding to the immediately following NAL unit. NAL units containing pictures captured at the same time should be ordered according to their `priority_id` in the bitstream. However, in cases that preserving such order increases encoding latency, encoders may order NALUs with larger value of `priority_id` before NALUs with smaller value of `priority_id`. Encoders shall not interleave NALUs corresponding to different timestamps.

Here is an example of multiple UCConfig Mode 0 streams. The encoder produces three independent streams with resolutions, bitrates, and maximum frame rates of: 720p 30fps at 1.5Mbps, 360p 30fps at 600Kbps, and 180p 15fps at 250Kbps. Each one is a non-scalable single-layer AVC stream.

Table 2 summarizes the three simulcast streams:

Table 2. UConfig Mode 0 Multiple Simulcast Stream Example

	Base layer
Simulcast streams	720p 30fps
	360p 30fps
	180p 15fps

Tables 3-5 illustrate the conformant bitstream structure when the three independent streams are transmitted via different sessions (i.e. Multi-Session Transmission, or MST). Even-numbered access units are shown in shaded cells:

Table 3. Conformant Bitstream Structure for Stream 180p 15fps

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 180p 15Hz
PPS (8)	PPSID = 0, SPSID = 0	PPS of stream 180p 15Hz
Prefix (14)	PRID = 0	Prefix NAL of stream 180p 15Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 180p 15Hz
Prefix (14)	PRID = 0	Prefix NAL of stream 180p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	P slice(s) of stream 180p 15Hz
Prefix (14)	PRID = 0	Prefix NAL of stream 180p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 4	P slice(s) of stream 180p 15Hz
...

Table 4. Conformant Bitstream Structure for Stream 360p 30fps

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 360p 30Hz
PPS (8)	PPSID = 0, SPSID = 0	PPS of stream 360p 30Hz
Prefix (14)	PRID = 1,	Prefix NAL of stream 360p 30Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 360p 30Hz
Prefix (14)	PRID = 1	Prefix NAL of stream 360p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 1	P slice(s) of stream 360p 30Hz
Prefix (14)	PRID = 1	Prefix NAL of stream 360p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	P slice(s) of stream 360p 30Hz
Prefix (14)	PRID = 1	Prefix NAL of stream 360p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 3	P slice(s) of stream 360p 30Hz
...

Table 5. Conformant Bitstream Structure for Stream 720p 30fps

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 720p 30Hz

PPS (8)	PPSID = 0, SPSID = 0	PPS of stream 720p 30Hz
Prefix (14)	PRID = 2	Prefix NAL of stream 720p 30Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 720p 30Hz
Prefix (14)	PRID = 2	Prefix NAL of stream 720p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 1	P slice(s) of stream 720p 30Hz
Prefix (14)	PRID = 2	Prefix NAL of stream 720p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	P slice(s) of stream 720p 30Hz
Prefix (14)	PRID = 2	Prefix NAL of stream 720p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 3	P slice(s) of stream 720p 30Hz
...

Table 6 illustrates the conformant bitstream structure when the independent streams are interleaved in a single session (i.e. Single Session Transmission, or SST). Even-numbered access units are shown in shaded cells:

Table 6. Conformant Bitstream Structure when streams are interleaved in a single session

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 180p 15Hz
PPS (8)	PPSID = 0, SPSID = 0	PPS of stream 180p 15Hz
SPS (7)	SPSID = 1	SPS of stream 360p 30Hz
PPS (8)	PPSID = 1, SPSID = 1	PPS of stream 360p 30Hz
SPS (7)	SPSID = 2	SPS of stream 720p 30Hz
PPS (8)	PPSID = 2, SPSID = 2	PPS of stream 720p 30Hz
Prefix (14)	PRID = 0	Prefix NAL of stream 180p 15Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 180p 15Hz
Prefix (14)	PRID = 1,	Prefix NAL of stream 360p 30Hz
IDR slice (5)	PPSID = 1, POC = 0	Base layer IDR slice(s) in stream 360p 30Hz
Prefix (14)	PRID = 2	Prefix NAL of stream 720p 30Hz
IDR slice (5)	PPSID = 2, POC = 0	Base layer IDR slice(s) in stream 720p 30Hz
Prefix (14)	PRID = 1	Prefix NAL of stream 360p 30Hz
Non-IDR slice (1)	PPSID = 1, POC = 1	P slice(s) of stream 360p 30Hz
Prefix (14)	PRID = 2	Prefix NAL of stream 720p 30Hz
Non-IDR slice (1)	PPSID = 2, POC = 1	P slice(s) of stream 720p 30Hz
Prefix (14)	PRID = 0	Prefix NAL of stream 180p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	P slice(s) of stream 180p 15Hz
Prefix (14)	PRID = 1	Prefix NAL of stream 360p 30Hz
Non-IDR slice (1)	PPSID = 1, POC = 2	P slice(s) of stream 360p 30Hz
Prefix (14)	PRID = 2	Prefix NAL of stream 720p 30Hz
Non-IDR slice (1)	PPSID = 2, POC = 2	P slice(s) of stream 720p 30Hz
...

The possible combinations of number of simulcast streams, number of scalable layers, maximum frame rates, and resolutions are limited by the maximum macroblock processing rate (MB/sec) defined as a

part of the hardware capabilities. The sum of macroblocks per second across all simulcast streams must not exceed this limit. For example, Table 7 shows the macroblock processing rates of each stream in the above example.

Table 7. Macroblock Processing Rate Example

	MB/sec
Simulcast Streams	108000
	27600
	3600

In this example, the sum of macroblocks per second is 139200 MB/sec. This number must be smaller than or equal to the maximum macroblocks per second supported by the encoder.

7.2. UCConfig Mode 1: H.264 SVC temporal scalability with simulcast

In this UCConfig Mode, one reference frame is used for two temporal layers and two for three or more temporal layers. Hierarchical P prediction structure is used to achieve temporal scalability. The base layer is operated as in UCConfig Mode 0. The frames in the enhancement layer only use the immediate previously reconstructed frame in the lower layer as the reference frame. Each layer represents a temporal scale. In support with this mode, each coded slice NAL unit (type 1 or 5) shall follow a prefix NAL unit (type 14) for all the enhancement layer frames to form temporal scalability. The bitstream shall not use coded slice extension NAL units (type 20) in this UCConfig Mode. The prefix NAL signals the importance of a layer relative to all layers in syntax element `priority_id`, with 0 representing the base layer, 1 the next higher layer (or equivalently the first enhancement layer), and so forth.

For each prefix NAL, the value of `temporal_id` specifies the hierarchical dependency of a temporal layer relative to other layers, with 0 representing the base temporal layer, 1 the first temporal scalable layer, 2 the second temporal scalable layer, and so forth. The values of `dependency_id` and `quality_id` must be equal to 0. The values of `no_inter_layer_pred_flag`, `discardable_flag` and `output_flag` must be 1. The value of `use_ref_base_pic_flag` must be 0. The decoding order must be the same as display order. The prefix NAL units shall be discarded by legacy AVC decoders that are not SVC-compatible, so that the bitstream can still be decoded.

Informative note: The coded slice extension NAL units (type 20) are only recognized by SVC decoders. While temporal scalability can be realized by such NALs, this specification disallows the use of coded slice extension NAL units in UCConfig Mode 1 so that AVC decoders can decode bitstreams at highest frame rate.

Figure 5 shows an example for the two-layer mode:

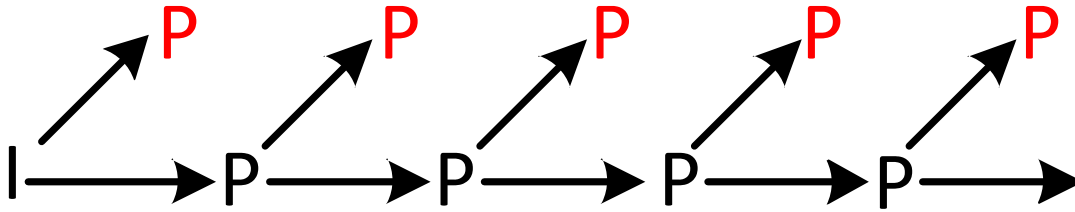


Figure 5. SVC two-layer temporal scalability mode

If the maximum frame rate of the source is 30fps, then Layer 0, or the base layer, has frame rate of 15fps. Layer 1, or the enhancement layer has frame rate of 15fps too. Decoding of Layer 0 and Layer 1 results in 30fps.

Figure 6 shows an example of the three-layer mode:

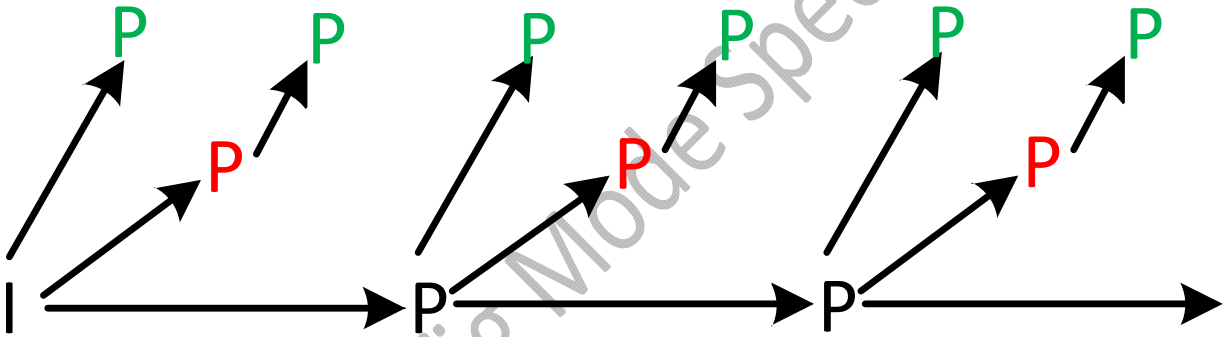


Figure 6. SVC three-layer temporal scalability mode

Figure 7 shows an example of the four-layer mode:

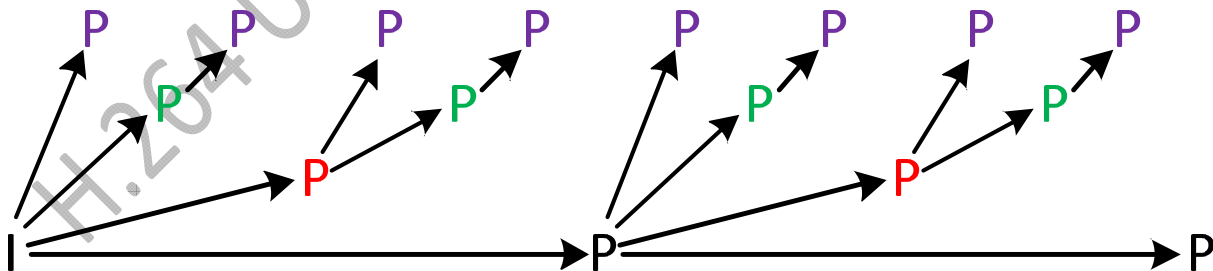


Figure 7. SVC four-layer temporal scalability mode

If the maximum frame rate of the source is 60fps, then Layer 0, or the base layer, has frame rate of 7.5fps. Layer 1, or the first enhancement layer, has frame rate of 7.5fps too. Decoding of Layer 0 and Layer 1 results in 15fps. Layer 2, or the second enhancement layer, has frame rate of 15fps. Decoding of

Layer 0, Layer1, and Layer 2 results in 30fps. Layer 3, or the third enhancement layer, has frame rate 30fps. Decoding of Layer 0, Layer 1, Layer 2 and Layer 3 results in 60fps.

Encoders that conform to this UCConfig Mode must be able to generate bitstreams with minimally two temporal layers.

Informative Note: As the maximum and minimum frame rates specified in this specification are 60fps and 7.5fps respectively, the maximum number of layers required is 4.

Informative Note: From encoder's perspective, the change of this mode from UCConfig Mode 0 is not significant. The major differences are: (1) the way the reference frame is used for P frames in support of hierarchical P structure and (2) one additional memory buffer for enhancement layers when three or more scalable layers are supported.

Similar to UCConfig Mode 0, this UCConfig Mode also supports multiple independent simulcast streams generated from the encoder at the same time. Each independent stream has different bitrates and is conformant to UCConfig Mode 1 with temporal scalability mode embedded.

When two or more independent streams are generated by the encoder, the encoder shall adjust the base value of `priority_id` (i.e. the value of `priority_id` associated with the base layer) of each stream in order to signal the importance of a temporal layer relative to all layers across all the streams. The base value of `priority_id` for the lowest bitrate is unchanged (i.e. shall be equal to 0). The base value of `priority_id` for the next higher bitrate stream is shifted up by one plus the value of `priority_id` for the highest scalable layer in the lowest bitrate stream. The base value of `priority_id` for the rest of the streams is adjusted in a similar manner. The adjustment assures a unique `priority_id` associated with each layer for each stream, with lower layers and lower bitrate streams being assigned a higher priority.

Encoders conforming to UCConfig Mode 1 must be able to generate bitstreams in UCConfig Mode 0 and UCConfig Mode 1. The run-time configuration is negotiated between decoders and encoders which is out of the scope of this specification.

When interleaving of multiple streams into one single session is requested, the encoder must generate unique `seq_parameter_set_id` and `picture_parameter_set_id` across all simulcast streams so the extraction of one stream from the interleaved bitstream is feasible. When present, SEI messages are associated with the stream corresponding to the immediately following NAL unit. NAL units containing pictures captured at the same time should be ordered according to their `priority_id` in the bitstream. However, in cases that preserving such order increases encoding latency, encoders may order NALUs with larger value of `priority_id` before NALUs with smaller value of `priority_id`. Encoders shall not interleave NALUs corresponding to different timestamps.

Here is an example of multiple UCConfig Mode 1 streams. The encoder produces three independent streams with resolutions, bitrates, and maximum frame rates of: 720p 30fps at 1.5Mbps, 360p 30fps at 600Kbps, and 180p 15fps at 250Kbps. For both the 720p and 360p streams, there are two temporal

scalable layers. For the 180p, there is only one temporal scalable layer. There are a total of 8 layers, 3 in the 720p stream, 3 in the 360p stream, and 2 in the 180p stream.

Table 8 summarizes the three simulcast streams:

Table 8. UCConfig Mode 1 Multiple Simulcast Stream Examples

Simulcast streams	Scalable layers		
	720p 30fps	720p 15fps	720p 7.5fps
	360p 30fps	360p 15fps	360p 7.5fps
		180p 15fps	180p 7.5fps

Tables 9-11 illustrate the conformant bitstream when the independent streams are transmitted via different sessions (i.e. Multi-Session Transmission, or MST). Even-numbered access units are shown in shaded cells:

Table 9. Conformant Bitstream Structure for Stream 180p 15fps

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 180p 15Hz
PPS (8)	PPSID = 0, SPSID= 0	PPS of stream 180p 15Hz
Prefix (14)	PRID = 0, TID=0	Prefix NAL of stream 180p 7.5Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 180p 7.5Hz
Prefix (14)	PRID = 1, TID=1	Prefix NAL of stream 180p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	P slice(s) of stream 180p 15Hz
Prefix (14)	PRID = 0, TID=0	Prefix NAL of stream 180p 7.5Hz
Non-IDR slice (1)	PPSID = 0, POC = 4	P slice(s) of stream 180p 7.5Hz
Prefix (14)	PRID = 1, TID=1	Prefix NAL of stream 180p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 6	P slice(s) of stream 180p 15Hz
...

Table 10. Conformant Bitstream Structure for Stream 360p 30fps

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 360p 30Hz
PPS (8)	PPSID = 0, SPSID= 0	PPS of stream 360p 30Hz
Prefix (14)	PRID = 2, TID=0	Prefix NAL of stream 360p 7.5Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 360p 7.5Hz
Prefix (14)	PRID = 4, TID=2	Prefix NAL of stream 360p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 1	P slice(s) of stream 360p 30Hz
Prefix (14)	PRID = 3, TID=1	Prefix NAL of stream 360p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	P slice(s) of stream 360p 15Hz

Prefix (14)	PRID = 4, TID=2	Prefix NAL of stream 360p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 3	P slice(s) of stream 360p 30Hz
Prefix (14)	PRID = 2, TID=0	Prefix NAL of stream 360p 7.5Hz
Non-IDR slice (1)	PPSID = 0, POC = 4	P slice(s) of stream 360p 7.5Hz
...

Table 11. Conformant Bitstream Structure for Stream 720p 30fps

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 720p 30Hz
PPS (8)	PPSID = 0, SPSID= 0	PPS of stream 720p 30Hz
Prefix (14)	PRID = 5, TID=0	Prefix NAL of stream 720p 7.5Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 720p 7.5Hz
Prefix (14)	PRID = 7, TID=2	Prefix NAL of stream 720p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 1	P slice(s) of stream 720p 30Hz
Prefix (14)	PRID = 6, TID=1	Prefix NAL of stream 720p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	P slice(s) of stream 720p 15Hz
Prefix (14)	PRID = 7, TID=2	Prefix NAL of stream 720p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 3	P slice(s) of stream 720p 30Hz
Prefix (14)	PRID = 5, TID=0	Prefix NAL of stream 720p 7.5Hz
Non-IDR slice (1)	PPSID = 0, POC = 4	P slice(s) of stream 720p 7.5Hz
...

Table 12 illustrates the conformant bitstream structure when the streams are interleaved in a single session (i.e. Single Session Transmission, or SST). Even-numbered access units are shown in shaded cells:

Table 12. Conformant Bitstream Structure when streams are interleaved in a single session

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 180p 15Hz
PPS (8)	PPSID = 0, SPSID= 0	PPS of stream 180p 15Hz
SPS (7)	SPSID = 1	SPS of stream 360p 30Hz
PPS (8)	PPSID = 1, SPSID = 1	PPS of stream 360p 30Hz
SPS (7)	SPSID= 2	SPS of stream 720p 30Hz
PPS (8)	PPSID = 2, SPSID = 2	PPS of stream 720p 30Hz
Prefix (14)	PRID = 0, TID=0	Prefix NAL of stream 180p 7.5Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 180p 7.5Hz
Prefix (14)	PRID = 2, TID=0	Prefix NAL of stream 360p 7.5Hz
IDR slice (5)	PPSID = 1, POC = 0	Base layer IDR slice(s) in stream 360p 7.5Hz
Prefix (14)	PRID = 5, TID=0	Prefix NAL of stream 720p 7.5Hz
IDR slice (5)	PPSID = 2, POC = 0	Base layer IDR slice(s) in stream 720p 7.5Hz
Prefix (14)	PRID = 4, TID=2	Prefix NAL of stream 360p 30Hz
Non-IDR slice (1)	PPSID = 1, POC = 1	P slice(s) of stream 360p 30Hz
Prefix (14)	PRID = 7, TID=2	Prefix NAL of stream 720p 30Hz
Non-IDR slice (1)	PPSID = 2, POC = 1	P slice(s) of stream 720p 30Hz

Prefix (14)	PRID = 1, TID=1	Prefix NAL of stream 180p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	P slice(s) of stream 180p 15Hz
Prefix (14)	PRID = 3, TID=1	Prefix NAL of stream 360p 15Hz
Non-IDR slice (1)	PPSID = 1, POC = 2	P slice(s) of stream 360p 15Hz
Prefix (14)	PRID = 6, TID=1	Prefix NAL of stream 720p 15Hz
Non-IDR slice (1)	PPSID = 2, POC = 2	P slice(s) of stream 720p 15Hz
Prefix (14)	PRID = 4, TID=2	Prefix NAL of stream 360p 30Hz
Non-IDR slice (1)	PPSID = 1, POC = 3	P slice(s) of stream 360p 30Hz
Prefix (14)	PRID = 7, TID=2	Prefix NAL of stream 720p 30Hz
Non-IDR slice (1)	PPSID = 2, POC = 3	P slice(s) of stream 720p 30Hz
...

The possible combinations of number of simulcast streams, number of scalable layers, maximum frame rates, and resolutions are limited by the maximum macroblock processing rate (MB/sec) defined as a part of the hardware capabilities. The sum of macroblocks per second in the highest temporal scale across all simulcast streams must not exceed this limit. For example, Table 13 shows the macroblock processing rates for each scale in the above example.

Table 13. Macroblock Processing Rate Example

	MB/sec		
Simulcast streams	108000	54000	27000
	27600	13800	6900
		3600	1800

In this example, the sum of macroblocks per second is 139200 MB/sec. This number must be smaller than or equal to the maximum macroblocks per second supported by the encoder.

7.3. UCConfig Mode 2q: H.264 SVC temporal and SNR scalabilities with simulcast

In this UCConfig Mode, UCConfig Mode 1 is applied first. For the temporal base and enhancement layers, in addition to the temporal scalability mode representing different temporal layers, SNR scalability is applied to all the frames to further divide the streams into more quality layers. The bitstream in the base quality layer must conform to the constraints in Section 7.2. Here is the way to combine two-layer temporal scalability mode with one-layer SNR scalability mode.

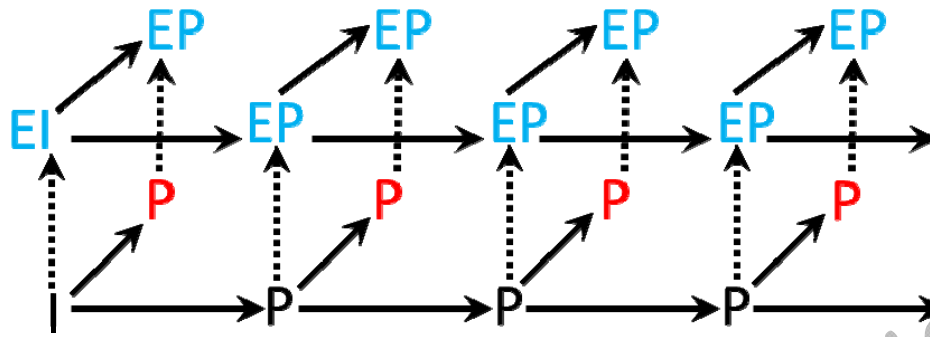


Figure 8. SVC UCConfig Mode 2q - temporal and SNR scalability mode with simulcast

Solid arrows represent temporal prediction and reference, and dashed arrows represent spatial quality prediction and reference.

Totally, there are 3 layers or scales of streams from this mode.

- I, and P's are frames for base layer,
- P's are frames for enhancement layer 1 from temporal scalability, and
- EI, and EP's are frames for enhancement layer 2 from SNR scalability.

Note that enhancement layer 2 includes frames in both base and enhancement temporal layers. To illustrate this, let's use the source materials of 720p 30fps and fixed quantization stepsize (Qp) as an example.

- 720p 15fps with Qp 38 as the base layer,
- 720p 15fps with Qp 38 as the enhancement layer 1 (or 30fps when combining with base layer), and
- 720p 30fps with Qp 34 as the enhancement layer 2 (combining frames in both base and enhancement temporal layers).

Three SNR scalability modes are supported:

- Coarse-grain scalability (CGS) non-rewrite mode only with two enhancement layers,
- CGS rewrite mode with two enhancement layers, and
- Medium-grain scalability (MGS) with one enhancement layer (multiple sub-layers).

For the MGS mode, the encoder can decide to use 4x4 or 8x8 transform mode and to decide the number of sub-layers (slices) it wishes to further divide into so that more scales can be created. The minimum number of sub-layers is 2 and the maximum is 4. The maximum number of temporal layers follows the guidelines defined in Section 7.2.

The rewrite mode is defined in the H.264 SVC specification. The encoded SVC SNR scalability bitstream utilizing the rewrite mode can be converted to an AVC bitstream without fully decoding the SVC bitstream.

Encoders that conform to this UCConfig Mode must support at least one of the three SNR capability modes. Decoders that conform to this UCConfig Mode must support all the three capability modes.

Similar to other UCConfig Modes, `priority_id` in SVC extension slice headers is used to signal the importance of a SNR layer relative to all layers across all streams. The value of `priority_id` for the first SNR scalable layer is the value of `priority_id` for the highest temporal layer plus one. The value of `priority_id` is incremented by one for the second SNR scalable layer and so forth. The decoding order must be the same as the display order in all scalable layers.

When the CGS mode is in use, the value of `quality_id` must be 0 in all SVC extension slice headers. The value of `dependency_id` is 0 for the base layer, 1 for the first SNR scalable layer, 2 for the second SNR scalable layer, and so forth. When the MGS mode is in use, the value of `dependency_id` is 0 in all SVC extension slice headers. The `quality_id` is 0 for the base layer, 1 for the first MGS sub-layer, 2 for the second MGS sub-layer, and so forth. The values of `temporal_id` must be the same as that in the corresponding SVC base layer frame in all SVC extension slice headers. The value of `no_inter_layer_pred_flag` must be 1, and the values of `output_flag`, `discardable_flag` and `use_ref_base_pic_flag` must be 0 for the SVC base layer (the values of `dependency_id` and `quality_id` are both equal to 0). The value of `use_ref_base_pic_flag` must be 0 unless the slice is part of an MGS key frame. The values of `discardable_flag` and `output_flag` are 1 for slices associated with the largest value of `dependency_id` in the coded picture and 0 otherwise.

The requirement for the temporal layers follows the ones defined in UCConfig Mode 1. Minimally, two temporal layers are required for each resolution/frame rate entry. The minimum SNR scalability required for each entry is two layers (one base layer and one SNR enhancement layer) in CGS rewrite and CGS non-rewrite modes and two sublayers in MGS mode.

This UCConfig Mode also supports multiple independent simulcast streams generated from the encoder at the same time. Each independent stream has different bitrates and is conformant to UCConfig Mode 3 with temporal and SNR scalability modes embedded.

When two or more independent streams are generated by the encoder, the encoder shall adjust the base value of `priority_id` (i.e. the value of `priority_id` associated with the base layer) of each stream in order to signal the importance of a layer relative to all layers across all the streams. The base value of `priority_id` for the lowest bitrate stream is unchanged (i.e. shall be equal to 0). The base value of `priority_id` for the next higher bitrate streams is shifted up by one plus the value of `priority_id` for the highest layer in the lowest bitrate stream. The base value of `priority_id` for the rest of the streams is adjusted in a similar manner. The adjustment assures a unique `priority_id` associated with each layer in each stream, with lower layers and lower bitrate streams being assigned a higher priority.

Encoders conforming to UCConfig Mode 2q must be able to generate bitstreams in UCConfig Mode 0, in UCConfig Mode 1, and in UCConfig Mode 2q. The run-time configuration is negotiated between decoders and encoders which is out of the scope of this specification.

When interleaving of multiple streams into one single session is requested, the encoder must generate unique seq_parameter_set_id and picture_parameter_set_id across all simulcast streams so the extraction of one stream from the interleaved bitstream is feasible. When present, SEI messages are associated with the stream corresponding to the immediately following NAL. NAL units containing pictures captured at the same time should be ordered according to their priority_id in the bitstream. However, in cases that preserving such order increases encoding latency, encoders may order NALUs with larger value of priority_id before NALUs with smaller value of priority_id. Encoders shall not interleave NALUs corresponding to different timestamps.

Here is an example of multiple UCConfig Mode 2q streams. The encoder produces three independent simulcast streams with resolutions, bitrates, and maximum frame rates of: 720p 30fps at 1.25Mbps, 360p 30fps at 750Kbps, and 180p 30fps at 300Kbps. For both the 720p and 360p streams, there are two temporal layers and two quality layers. For the 180p stream, there are two temporal scalable layers but not quality layer. There are a total of 8 scalable layers, 3 in the 720p stream, 3 in the 360p stream and 2 in the 180p stream. Table 14 summarizes the three simulcast streams:

Table 14. UCConfig Mode 2q Multiple Simulcast Stream Examples

Simulcast streams	Scalable layers		
	720p 30fps Quality 1	720p 30fps Quality 0	720p 15fps Quality 0
	360p 30fps Quality 1	360p 30fps Quality 0	360p 15fps Quality 0
		180p 30fps Quality 0	180p 15fps Quality 0

Tables 15-17 illustrate the conformant bitstream structure when the independent streams are transmitted via different sessions (i.e. Multi-Session Transmission, or MST). Even-numbered access units are shown in shaded cells. Assume CGS is used in this example:

Table 15. Conformant Bitstream Structure for Stream 180p 30fps

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 180p 30Hz
PPS (8)	PPSID = 0, SPSID = 0	PPS of stream 180p 30Hz
Prefix (14)	PRID = 0, TID=0, DID=0	Prefix NAL of stream 180p 15Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 180p 15Hz
Prefix (14)	PRID = 1, TID=1, DID=0	Prefix NAL of stream 180p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 1	P slice(s) of stream 180p 30Hz
Prefix (14)	PRID = 0, TID=0, DID=0	Prefix NAL of stream 180p 15Hz

Non-IDR slice (1)	PPSID = 0, POC = 2	Base layer P slice(s) in stream 180p 15Hz
Prefix (14)	PRID = 1, TID=1, DID=0	Prefix NAL of stream 180p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 3	P slice(s) of stream 180p 30Hz
...

Table 16. Conformant Bitstream Structure for Stream 360p 30fps

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 360p 30Hz
SSPS (15)	SPSID = 0	Subset SPS of stream 360p 30Hz
PPS (8)	PPSID = 0, SPSID = 0	PPS of stream 360p 30Hz SVC base layer
PPS (8)	PPSID = 1, SPSID = 0	PPS of stream 360p 30Hz SVC scalable layer
Prefix (14)	PRID = 2, TID=0, DID=0	Prefix NAL of stream 360p 15Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 360p 15Hz
SVC ext slice (20)	PPSID = 1, PRID = 4, TID=0, DID=1	EI slice(s) in stream 360p 15Hz
Prefix (14)	PRID = 3, TID=1, DID=0	Prefix NAL of stream 360p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 1	P slice(s) of stream 360p 30Hz
SVC ext slice (20)	PPSID = 1, PRID = 4, TID=1, DID=1	EP slice(s) in stream 360p 30Hz
Prefix (14)	PRID = 2, TID=0, DID=0	Prefix NAL of stream 360p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	Base layer P slice(s) in stream 360p 15Hz
SVC ext slice (20)	PPSID = 1, PRID = 4, TID=0, DID=1	EP slice(s) in stream 360p 15Hz
Prefix (14)	PRID = 3, TID=1, DID=0	Prefix NAL of stream 360p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 3	P slice(s) of stream 360p 30Hz
SVC ext slice (20)	PPSID = 1, PRID = 4, TID=1, DID=1	EP slice(s) in stream 360p 30Hz
...

Table 17. Conformant Bitstream Structure for Stream 720p 30fps

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 720p 30Hz
SSPS (15)	SPSID = 0	Subset SPS of stream 720p 30Hz
PPS (8)	PPSID = 0, SPSID = 0	PPS of stream 720p 30Hz SVC base layer
PPS (8)	PPSID = 1, SPSID = 0	PPS of stream 720p 30Hz SVC scalable layer
Prefix (14)	PRID = 5, TID=0, DID=0	Prefix NAL of stream 720p 15Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 720p 15Hz
SVC ext slice (20)	PPSID = 1, PRID = 7, TID=0, DID=1	EI slice(s) in stream 720p 15Hz
Prefix (14)	PRID = 6, TID=1, DID = 0	Prefix NAL of stream 720p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 1	P slice(s) of stream 720p 30Hz
SVC ext slice (20)	PPSID = 1, PRID = 7, TID=1, DID=1	EP slice(s) in stream 720p 30Hz
Prefix (14)	PRID = 5, TID=0, DID=0	Prefix NAL of stream 720p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	Base layer P slice(s) in stream 720p 15Hz
SVC ext slice (20)	PPSID = 1, PRID = 7, TID=0, DID=1	EP slice(s) in stream 720p 15Hz
Prefix (14)	PRID = 6, TID=1, DID = 0	Prefix NAL of stream 720p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 3	P slice(s) of stream 720p 30Hz
SVC ext slice (20)	PPSID = 1, PRID = 7, TID=1, DID=1	EP slice(s) in stream 720p 30Hz

...
-----	-----	-----

Table 18 illustrates the conformant bitstream structure when the streams are interleaved in a single session (i.e. Single Session Transmission, or SST). Even-numbered access units are shown in shaded cells. Assume CGS is used in this example:

Table 18. Conformant Bitstream Structure when streams are interleaved in a single session

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 180p 30Hz
PPS (8)	PPSID = 0, SPSID = 0	PPS of stream 180p 30Hz
SPS (7)	SPSID = 1	SPS of stream 360p 30Hz
SSPS (15)	SPSID = 0	Subset SPS of stream 360p 30Hz
PPS (8)	PPSID = 1, SPSID = 1	PPS of stream 360p 30Hz SVC base layer
PPS (8)	PPSID = 2, SPSID = 0	PPS of stream 360p 30Hz SVC scalable layer
SPS (7)	SPSID = 2	SPS of stream 720p 30Hz
SSPS (15)	SPSID = 1	Subset SPS of stream 720p 30Hz
PPS (8)	PPSID = 3, SPSID = 2	PPS of stream 720p 30Hz SVC base layer
PPS (8)	PPSID = 4, SPSID = 1	PPS of stream 720p 30Hz SVC scalable layer
Prefix (14)	PRID = 0, TID=0, DID=0	Prefix NAL of stream 180p 15Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 180p 15Hz
Prefix (14)	PRID = 2, TID=0, DID=0	Prefix NAL of stream 360p 15Hz
IDR slice (5)	PPSID = 1, POC = 0	Base layer IDR slice(s) in stream 360p 15Hz
SVC ext slice (20)	PPSID = 2, PRID = 4, TID=0, DID=1	EI slice(s) in stream 360p 15Hz
Prefix (14)	PRID = 5, TID=0, DID=0	Prefix NAL of stream 720p 15Hz
IDR slice (5)	PPSID = 3, POC = 0	Base layer IDR slice(s) in stream 720p 15Hz
SVC ext slice (20)	PPSID = 4, PRID = 7, TID=0, DID=1	EI slice(s) in stream 720p 15Hz
Prefix (14)	PRID = 1, TID=1, DID=0	Prefix NAL of stream 180p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 1	P slice(s) of stream 180p 30Hz
Prefix (14)	PRID = 3, TID=1, DID=0	Prefix NAL of stream 360p 30Hz
Non-IDR slice (1)	PPSID = 1, POC = 1	P slice(s) of stream 360p 30Hz
SVC ext slice (20)	PPSID = 2, PRID = 4, TID=1, DID=1	EP slice(s) in stream 360p 30Hz
Prefix (14)	PRID = 6, TID=1, DID = 0	Prefix NAL of stream 720p 30Hz
Non-IDR slice (1)	PPSID = 3, POC = 1	P slice(s) of stream 720p 30Hz
SVC ext slice (20)	PPSID = 4, PRID = 7, TID=1, DID=1	EP slice(s) in stream 720p 30Hz
Prefix (14)	PRID = 0, TID=0, DID=0	Prefix NAL of stream 180p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	Base layer P slice(s) in stream 180p 15Hz
Prefix (14)	PRID = 2, TID=0, DID=0	Prefix NAL of stream 360p 15H
Non-IDR slice (1)	PPSID = 1, POC = 2	Base layer P slice(s) in stream 360p 15Hz
SVC ext slice (20)	PPSID = 2, PRID = 4, TID=0, DID=1	EP slice(s) in stream 360p 15Hz
Prefix (14)	PRID = 5, TID=0, DID=0	Prefix NAL of stream 720p 15Hz
Non-IDR slice (1)	PPSID = 3, POC = 2	Base layer P slice(s) in stream 720p 15Hz
SVC ext slice (20)	PPSID = 4, PRID = 7, TID=0, DID=1	EP slice(s) in stream 720p 15Hz
...

The possible combinations of number of simulcast streams, number of scalable layers, maximum frame rates, and resolutions are limited by the maximum macroblock processing rate defined as a part of the hardware capabilities. The sum of macroblocks per second for all quality layers in the highest temporal scale across all simulcast streams must not exceed this limit. For example, Table 19 shows the macroblock processing rates for each scale in the above example.

Table 19. Macroblock Processing Rate Example

Simulcast streams	MB/sec		
	108000	108000	54000
	27600	27600	13800
		7200	3600

In this example, the sum of macroblocks per second is 278400 MB/sec. This number must be smaller than or equal to the maximum macroblocks per second supported by the encoder.

7.4. UCConfig Mode 2s: H.264 SVC temporal and spatial scalabilities with simulcast

In this UCConfig Mode, UCConfig Mode 1 is applied first. For the temporal base and enhancement layers, in addition to the temporal scalability mode representing different temporal layers, spatial scalability is applied to all the frames to further divide the streams into more spatial layers. The bitstream in the base spatial layer must conform to the constraints in Section 7.2. Here is the way to combine two-layer temporal scalability mode with one-layer spatial scalability mode.

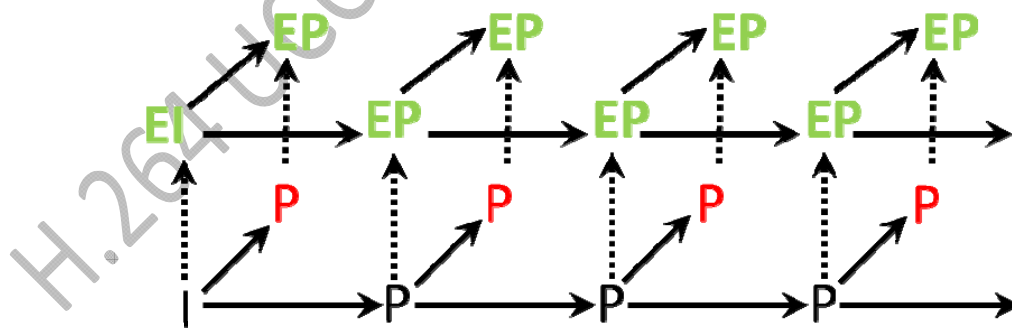


Figure 9. SVC UCConfig Mode 2s - temporal and spatial scalability mode with simulcast

Solid arrows represent temporal prediction and reference, and dashed arrows represent spatial quality prediction and reference.

Totally, there are 3 layers or scales of streams from this mode.

- I, and P's are frames for base layer,
- P's are frames for enhancement layer 1 from temporal scalability, and
- EI, and EP's are frames for enhancement layer 2 from spatial scalability.

Note that enhancement layer 2 includes frames in both base and enhancement temporal layers. To illustrate this, let's use the source materials of 720p 30fps and fixed quantization stepsize (Qp) as an example.

- 360p 15fps with Qp 34 as the base layer,
- 360p 15fps with Qp 34 as the enhancement layer 1 (or 30fps when combining with base layer), and
- 720p 30fps with Qp 34 as the enhancement layer 2 (combining frames in both base and enhancement temporal layers).

As mentioned in Section 6.1.2, the UC Scalable Constrained High toolset requires spatial scalability restricted to resolution ratios of 1.5 and 2 between successive spatial layers in both horizontal and vertical direction. In this UCConfig Mode, the minimum number of spatial layers is 2 (one base and one enhancement spatial layer). The requirement for the temporal layers follows the ones defined in UCConfig Mode 1.

Similar to other UCConfig Modes, priority_id in SVC extension slice headers is used to signal the importance of a spatial layer relative to all layers across all streams. The value of priority_id for the first spatial scalable layer is the value of priority_id for the highest temporal layer plus one. The value of priority_id is incremented by one for the second spatial scalable layer and so forth. The decoding order must be the same as the display order in all scalable layers.

This UCConfig Mode also supports multiple independent simulcast streams generated from the encoder at the same time. Each independent stream has different bitrates and is conformant to UCConfig Mode 2s with temporal and spatial scalability modes, embedded in each stream.

When two or more independent streams are generated by the encoder, the encoder shall adjust the base value of priority_id (i.e. the value of priority_id associated with the base layer) of each stream in order to signal the importance of a layer relative to all layers across all the streams. The base value of priority_id for the lowest bitrate stream is unchanged (i.e. shall be equal to 0). The base value of priority_id for the next higher bitrate streams is shifted up by one plus the value of priority_id for the highest layer in the lowest bitrate stream. The base value of priority_id for the rest of the streams is adjusted in a similar manner. The adjustment assures a unique priority_id associated with each layer in each stream, with lower layers and lower bitrate streams being assigned a higher priority.

Encoders conforming to UCConfig Mode 2s must be able to generate bitstreams in UCConfig Mode 0, in UCConfig Mode 1, and in UCConfig Mode 2s. The run-time configuration is negotiated between decoders and encoders which is out of the scope of this specification.

When interleaving of multiple streams into one single session is requested, the encoder must generate unique `seq_parameter_set_id` and `picture_parameter_set_id` across all simulcast streams so the extraction of one stream from the interleaved bitstream is feasible. When present, SEI messages are associated with the stream corresponding to the immediately following NAL. NAL units containing pictures captured at the same time should be ordered according to their `priority_id` in the bitstream. However, in cases that preserving such order increases encoding latency, encoders may order NALUs with larger value of `priority_id` before NALUs with smaller value of `priority_id`. Encoders shall not interleave NALUs corresponding to different timestamps.

Here is an example of multiple UCConfig Mode 2s streams. The encoder produces three independent simulcast streams with maximum resolutions, bitrates, and maximum frame rates of: 720p 30fps at 1.25Mbps and 360p 30fps at 750Kbps, and 180p 30fps at 300Kbps. For both the 720p and 360p streams, there are two temporal layers and two spatial layers. For the 180p stream, there are two temporal scalable layers but not spatial layer. There are a total of 8 scalable layers, 3 in the 720p stream, 3 in the 360p stream and 2 in the 180p stream. Table 20 summarizes the three simulcast streams:

Table 20. UCConfig Mode 2s Multiple Simulcast Stream Examples

Simulcast streams	Scalable layers		
	720p 30fps	480p 30fps	480p 15fps
	360p 30fps	240p 30fps	240p 15fps
		180p 30fps	180p 15fps

Table 21-Table 23 illustrate the conformant bitstream structure when the independent streams are transmitted via different sessions (i.e. Multi-Session Transmission, or MST). Even-numbered access units are shown in shaded cells.

Table 21. Conformant Bitstream Structure for Stream 180p 30fps

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 180p 30Hz
PPS (8)	PPSID = 0, SPSID = 0	PPS of stream 180p 30Hz
Prefix (14)	PRID = 0, TID=0, DID=0	Prefix NAL of stream 180p 15Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 180p 15Hz
Prefix (14)	PRID = 1, TID=1, DID=0	Prefix NAL of stream 180p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 1	P slice(s) of stream 180p 30Hz
Prefix (14)	PRID = 0, TID=0, DID=0	Prefix NAL of stream 180p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	Base layer P slice(s) in stream 180p 15Hz
Prefix (14)	PRID = 1, TID=1, DID=0	Prefix NAL of stream 180p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 3	P slice(s) of stream 180p 30Hz
...

Table 22. Conformant Bitstream Structure for Stream 360p 30fps

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 240p 30Hz
SSPS (15)	SPSID = 0	Subset SPS of stream 360p 30Hz
PPS (8)	PPSID = 0, SPSID = 0	PPS of stream 240p 30Hz SVC base layer
PPS (8)	PPSID = 1, SPSID = 0	PPS of stream 360p 30Hz SVC scalable layer
Prefix (14)	PRID = 2, TID=0, DID=0	Prefix NAL of stream 240p 15Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 240p 15Hz
SVC ext slice (20)	PPSID = 1, PRID = 4, TID=0, DID=1	EI slice(s) in stream 360p 15Hz
Prefix (14)	PRID = 3, TID=1, DID=0	Prefix NAL of stream 240p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 1	P slice(s) of stream 240p 30Hz
SVC ext slice (20)	PPSID = 1, PRID = 4, TID=1, DID=1	EP slice(s) in stream 360p 30Hz
Prefix (14)	PRID = 2, TID=0, DID=0	Prefix NAL of stream 240p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	Base layer P slice(s) in stream 240p 15Hz
SVC ext slice (20)	PPSID = 1, PRID = 4, TID=0, DID=1	EP slice(s) in stream 360p 15Hz
Prefix (14)	PRID = 3, TID=1, DID=0	Prefix NAL of stream 240p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 3	P slice(s) of stream 240p 30Hz
SVC ext slice (20)	PPSID = 1, PRID = 4, TID=1, DID=1	EP slice(s) in stream 360p 30Hz
...

Table 23. Conformant Bitstream Structure for Stream 720p 30fps

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 480p 30Hz
SSPS (15)	SPSID = 0	Subset SPS of stream 720p 30Hz
PPS (8)	PPSID = 0, SPSID = 0	PPS of stream 480p 30Hz SVC base layer
PPS (8)	PPSID = 1, SPSID = 0	PPS of stream 720p 30Hz SVC scalable layer
Prefix (14)	PRID = 5, TID=0, DID=0	Prefix NAL of stream 480p 15Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 480p 15Hz
SVC ext slice (20)	PPSID = 1, PRID = 7, TID=0, DID=1	EI slice(s) in stream 720p 15Hz
Prefix (14)	PRID = 6, TID=1, DID = 0	Prefix NAL of stream 480p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 1	P slice(s) of stream 480p 30Hz
SVC ext slice (20)	PPSID = 1, PRID = 7, TID=1, DID=1	EP slice(s) in stream 720p 30Hz
Prefix (14)	PRID = 5, TID=0, DID=0	Prefix NAL of stream 480p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	Base layer P slice(s) in stream 480p 15Hz
SVC ext slice (20)	PPSID = 1, PRID = 7, TID=0, DID=1	EP slice(s) in stream 720p 15Hz
Prefix (14)	PRID = 6, TID=1, DID = 0	Prefix NAL of stream 480p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 3	P slice(s) of stream 480p 30Hz
SVC ext slice (20)	PPSID = 1, PRID = 7, TID=1, DID=1	EP slice(s) in stream 720p 30Hz
...

Table 24 illustrates the conformant bitstream structure when the streams are interleaved in a single session (i.e. Single Session Transmission, or SST). Even-numbered access units are shown in shaded cells. Assume CGS is used in this example:

Table 24. Conformant Bitstream Structure when streams are interleaved in a single session

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 180p 30Hz
PPS (8)	PPSID = 0, SPSID = 0	PPS of stream 180p 30Hz
SPS (7)	SPSID = 1	SPS of stream 240p 30Hz
SSPS (15)	SPSID = 0	Subset SPS of stream 360p 30Hz
PPS (8)	PPSID = 1, SPSID = 1	PPS of stream 240p 30Hz SVC base layer
PPS (8)	PPSID = 2, SPSID = 0	PPS of stream 360p 30Hz SVC scalable layer
SPS (7)	SPSID = 2	SPS of stream 480p 30Hz
SSPS (15)	SPSID = 1	Subset SPS of stream 720p 30Hz
PPS (8)	PPSID = 3, SPSID = 2	PPS of stream 480p 30Hz SVC base layer
PPS (8)	PPSID = 4, SPSID = 1	PPS of stream 720p 30Hz SVC scalable layer
Prefix (14)	PRID = 0, TID=0, DID=0	Prefix NAL of stream 180p 15Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 180p 15Hz
Prefix (14)	PRID = 2, TID=0, DID=0	Prefix NAL of stream 240p 15Hz
IDR slice (5)	PPSID = 1, POC = 0	Base layer IDR slice(s) in stream 240p 15Hz
SVC ext slice (20)	PPSID = 2, PRID = 4, TID=0, DID=1	EI slice(s) in stream 360p 15Hz
Prefix (14)	PRID = 5, TID=0, DID=0	Prefix NAL of stream 480p 15Hz
IDR slice (5)	PPSID = 3, POC = 0	Base layer IDR slice(s) in stream 480p 15Hz
SVC ext slice (20)	PPSID = 4, PRID = 7, TID=0, DID=1	EI slice(s) in stream 720p 15Hz
Prefix (14)	PRID = 1, TID=1, DID=0	Prefix NAL of stream 180p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 1	P slice(s) of stream 180p 30Hz
Prefix (14)	PRID = 3, TID=1, DID=0	Prefix NAL of stream 240p 30Hz
Non-IDR slice (1)	PPSID = 1, POC = 1	P slice(s) of stream 240p 30Hz
SVC ext slice (20)	PPSID = 2, PRID = 4, TID=1, DID=1	EP slice(s) in stream 360p 30Hz
Prefix (14)	PRID = 6, TID=1, DID=0	Prefix NAL of stream 480p 30Hz
Non-IDR slice (1)	PPSID = 3, POC = 1	P slice(s) of stream 480p 30Hz
SVC ext slice (20)	PPSID = 4, PRID = 7, TID=1, DID=1	EP slice(s) in stream 720p 30Hz
Prefix (14)	PRID = 0, TID=0, DID=0	Prefix NAL of stream 180p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	Base layer P slice(s) in stream 180p 15Hz
Prefix (14)	PRID = 2, TID=0, DID=0	Prefix NAL of stream 240p 15H
Non-IDR slice (1)	PPSID = 1, POC = 2	Base layer P slice(s) in stream 240p 15Hz
SVC ext slice (20)	PPSID = 2, PRID = 4, TID=0, DID=1	EP slice(s) in stream 360p 15Hz
Prefix (14)	PRID = 5, TID=0, DID=0	Prefix NAL of stream 480p 15Hz
Non-IDR slice (1)	PPSID = 3, POC = 2	Base layer P slice(s) in stream 480p 15Hz
SVC ext slice (20)	PPSID = 4, PRID = 7, TID=0, DID=1	EP slice(s) in stream 720p 15Hz
...

The possible combinations of number of simulcast streams, number of scalable layers, maximum frame rates, and resolutions are limited by the maximum macroblock processing rate defined as a part of the hardware capabilities. The sum of macroblocks per second for all quality layers in the highest temporal scale across all simulcast streams must not exceed this limit. For example, Table 25 shows the macroblock processing rates for each scale in the above example.

Table 25. Macroblock Processing Rate Example

Simulcast streams	MB/sec		
	108000	108000	54000
	27600	27600	13800
		7200	3600

In this example, the sum of macroblocks per second is 278400 MB/sec. This number must be smaller than or equal to the maximum macroblocks per second supported by the encoder.

7.5. UCConfig Mode 3: H.264 Full SVC scalability with simulcast

In this UCConfig Mode, UCConfig Mode 2q is applied first. In addition to the temporal and SNR scalability modes, spatial scalability is used to all the frames to further divide the streams into more spatial layers. The bitstream in the base spatial layer must conform to the constraints in Section 7.3. In each spatial enhancement layer, additional SNR scalability (CGS or MGS) can be applied when more layers are needed. The decoding order must be the same as the display order in all scalable layers. Here is the way to combine two-layer temporal scalability mode, two-layer spatial scalability mode, and two-layer SNR scalability mode within each spatial layer.

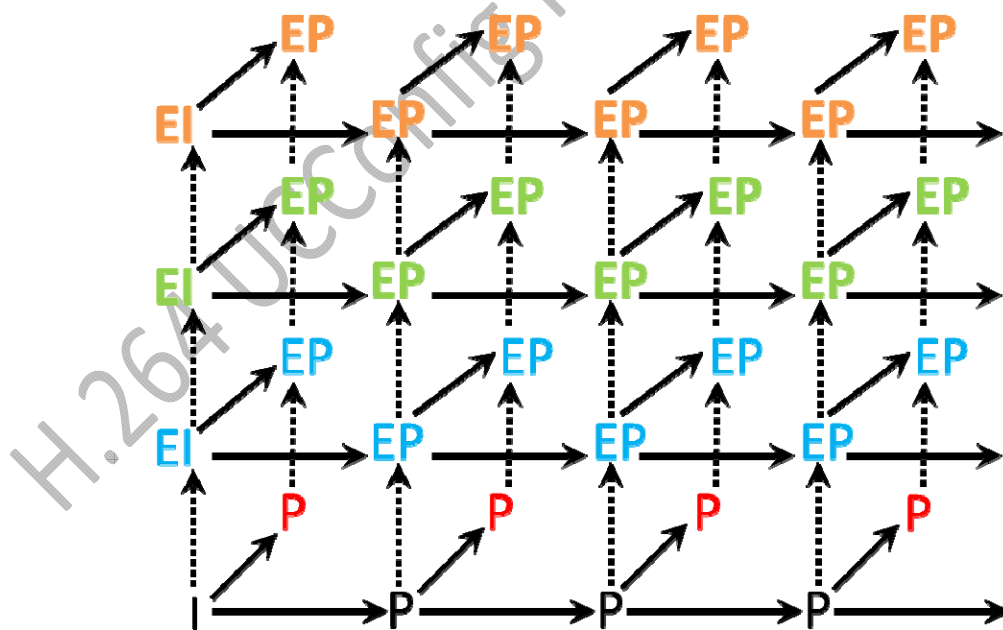


Figure 10. SVC UCConfig Mode 3 – full SVC scalability mode with Simulcast

Solid arrows represent temporal prediction and reference, and dashed arrows represent spatial quality prediction and reference. The bold type represents frames with high resolution.

Totally, there are 5 layers or scales of streams from this mode.

- I, and P's are frames for base layer,
- **P**'s are frames for enhancement layer 1 from temporal scalability, and
- **I**, and **P**'s are frames for enhancement 2 from SNR scalability, and
- **EL**, and **EP**'s are frames for enhancement layer 3 from spatial scalability, and
- **EL**, and **EP**'s are frames for enhancement layer 4 from additional SNR scalability in the spatial enhancement layer.

To illustrate this, let's use the source materials of 720p 30fps and fixed quantization step size (Qp) as an example.

- 360p 15fps with Qp 38 as the base layer, and
- 360p 15fps with Qp 38 as the enhancement layer 1 (or 30fps when combining with the base layer), and
- 360p 30fps with Qp 34 as the enhancement layer 2, and
- 720p 30fps with Qp 34 as the enhancement layer 3, and
- 720p 30fps with Qp 30 as the enhancement layer 4.

Similar to other UCConfig Modes, priority_id in SVC extension slice headers is used to signal the importance of a spatial layer relative to all layers across all streams. The value of priority_id for the first spatial scalable layer must be the highest value of priority_id in the base layer resolution plus one. If additional SNR scalability is introduced inside a spatial scalable layer, the value of priority_id increments by one for each additional SNR layer, from the lowest to the highest. The value of priority_id of the second spatial scalable layer is one plus the highest value of priority_id in the first spatial scalable layer. If additional SNR scalability is introduced, the same rule for assigning the values of priority_id applies. The same procedure applied to higher spatial layers.

The values of temporal_id must be the same as that in the corresponding SVC base layer frame in all SVC extension slice headers. The value of dependency_id is incremented by one in successive spatial or CGS layers. When additional MGS sub-layers are introduced inside a spatial scalable layer, the value of quality_id is assigned in the same way as specified in Section 7.3. When present, the minimum number of MGS sub-layers in each spatial scalable layer is 2 and the maximum is 4. When additional CGS scalability is introduced inside a spatial scalable layer, the value of dependency_id for the first CGS scalable layer is one plus the value of dependency_id of the base quality layer with the same resolution. The value of dependency_id is incremented by one for additional CGS layers with the same resolution. The value of use_ref_base_pic_flag must be 1 for slices that are part an MGS key picture, and 0 otherwise. The values of output_flag and discardable_flag are 1 for slices associated with the largest value of dependency_id in the coded picture and 0 otherwise.

As mentioned in Section 6.1.2, the UC Scalable Constrained High toolset requires spatial scalability restricted to resolution ratios of 1.5 and 2 between successive spatial layers in both horizontal and vertical direction. In this UCConfig Mode, the minimum number of spatial layers is 2 (one base and one enhancement spatial layer). The requirement for the temporal layers follows the ones defined in UCConfig Mode 1. The requirement for the quality layers follows the ones defined in UCConfig Mode 2q.

This UCConfig Mode also supports multiple independent simulcast streams generated from the encoder at the same time. Each independent stream has different bitrates and is conformant to UCConfig Mode 3 with temporal, spatial and SNR scalability modes, embedded in each stream.

When two or more independent streams are generated by the encoder, the encoder shall adjust the base value of `priority_id` (i.e. the value of `priority_id` associated with the base layer) of each stream to signal the importance of a layer relative to all layers across all streams. The base value of `priority_id` for the lowest bitrate stream is unchanged (i.e. shall be equal to 0). The base value of `priority_id` for the next higher bitrate stream is shifted up by one plus the value of `priority_id` for the highest scalable layer in the lowest bitrate stream. The base value of `priority_id` for the rest of the streams is adjusted in a similar manner. The adjustment assures a unique `priority_id` associated with each layer for each stream, with lower layers and lower bitrate streams being assigned a higher priority.

Encoders conforming to UCConfig Mode 3 must be able to generate bitstreams in UCConfig Mode 0, in UCConfig Mode 1, in UCConfig Mode 2q, in UCConfig Mode 3, bitstreams that support SNR scalability only, bitstreams that support spatial scalability only, and bitstreams that support SNR and spatial scalability only. The run-time configuration is negotiated between decoders and encoders which is out of the scope of this specification.

When interleaving of multiple streams into one single session is requested, the encoder must generate unique `seq_parameter_set_id` and `picture_parameter_set_id` across all simulcast streams so the extraction of one stream from the interleaved bitstream is feasible. When present, SEI messages are associated with the stream corresponding to the immediately following NAL unit. NAL units containing pictures captured at the same time should be ordered according to their `priority_id` in the bitstream. However, in cases that preserving such order increases encoding latency, encoders may order NALUs with larger value of `priority_id` before NALUs with smaller value of `priority_id`. Encoders shall not interleave NALUs corresponding to different timestamps.

Here is an example of multiple UCConfig Mode 3 streams. The encoder produces two independent simulcast streams with highest resolutions, bitrates, and maximum frame rates of: 720p 30fps at 1.5Mbps and 360p 30fps at 600Kbps. For the 720p stream, there are two temporal layers and two spatial layers. Each spatial layer contains one quality enhancement layer. For the 360p stream, there are two temporal scalable layers, two spatial layers, but no quality enhancement layer (same Qp across all layers). There are a total of 8 scalable layers, 5 in the 720p stream and 3 in the 360p stream. Table 26 summarizes these two simulcast streams:

Table 26. UCConfig Mode 3 Multiple Simulcast Stream Examples

Simulcast streams	Scalable layers				
	720p 30fps Quality 2	720p 30fps Quality 1	480p 30fps Quality 1	480p 30fps Quality 0	480p 15fps Quality 0
			360p 30fps Quality 0	180p 30fps Quality 0	180p 15fps Quality 0

Table 27-Table 29 illustrate the conformant bitstream structure when the independent streams are transmitted via different sessions (i.e. Multi-Session Transmission, or MST). Even-numbered access units are shown in shaded cells. Assume CGS is used in this example:

Table 27. Conformant Bitstream Structure for Stream 360p 30fps

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 180p 30Hz
SSPS (15)	SSPID = 0	Subset SPS of stream 360p 30Hz
PPS (8)	PPSID = 0, SPSID = 0	PPS of stream 180p 30Hz SVC base layer
PPS (8)	PPSID = 1, SPSID = 0	PPS of stream 360p 30Hz SVC scalable layer
Prefix (14)	PRID = 0, TID=0, DID=0	Prefix NAL of stream 180p 15Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 180p 15Hz
SVC ext slice (20)	PPSID = 1, PRID = 2, TID=0, DID=1	EI slice(s) in stream 360p 15Hz
Prefix (14)	PRID = 1, TID=1, DID=0	Prefix NAL of stream 180p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 1	P slice(s) of stream 180p 30Hz
SVC ext slice (20)	PPSID = 1, PRID = 2, TID=1, DID=1	EP slice(s) in stream 360p 30Hz
Prefix (14)	PRID = 0, TID=0, DID=0	Prefix NAL of stream 180p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	P slice(s) in stream 180p 15Hz
SVC ext slice (20)	PPSID = 1, PRID = 2, TID=0, DID=1	EP slice(s) in stream 360p 15Hz
Prefix (14)	PRID = 1, TID=1, DID=0	Prefix NAL of stream 180p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 3	P slice(s) of stream 180p 30Hz
SVC ext slice (20)	PPSID = 1, PRID = 2, TID=1, DID=1	EP slice(s) in stream 360p 30Hz
...

Table 28. Conformant Bitstream Structure for Stream 720p 30fps

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 480p 30Hz
SSPS (15)	SSPID = 0	Subset SPS of stream 480p 30Hz
SSPS (15)	SSPID = 1	Subset SPS of stream 720p 30Hz
PPS (8)	PPSID = 0, SPSID = 0	PPS of stream 480p 30Hz SVC base layer
PPS (8)	PPSID = 1, SPSID = 0	PPS of stream 480p 30Hz SVC scalable layer
PPS (8)	PPSID = 2, SPSID = 1	PPS of stream 720p 30Hz SVC scalable layer
Prefix (14)	PRID = 3, TID=0, DID=0	Prefix NAL of stream 480p 15Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 480p 15Hz
SVC ext slice (20)	PPSID = 1, PRID = 5, TID=0, DID=1	EI slice(s) in stream 480p 15Hz
SVC ext slice (20)	PPSID = 2, PRID = 6, TID=0, DID=2	EI slice(s) in stream 720p 15Hz
SVC ext slice (20)	PPSID = 2, PRID = 7, TID=0, DID=3	EI slice(s) in stream 720p 15Hz

Prefix (14)	PRID = 4, TID=1, DID=0	Prefix NAL of stream 480p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 1	P slice(s) of stream 480p 30Hz
SVC ext slice (20)	PPSID = 1, PRID = 5, TID=1, DID=1	EP slice(s) in stream 480p 30Hz
SVC ext slice (20)	PPSID = 2, PRID = 6, TID=1, DID=2	EP slice(s) in stream 720p 30Hz
SVC ext slice (20)	PPSID = 2, PRID = 7, TID=1, DID=3	EP slice(s) in stream 720p 30Hz
Prefix (14)	PRID = 3, TID=0, DID=0	Prefix NAL of stream 480p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	P slice(s) in stream 480p 15Hz
SVC ext slice (20)	PPSID = 1, PRID = 5, TID=0, DID=1	EP slice(s) in stream 480p 15Hz
SVC ext slice (20)	PPSID = 2, PRID = 6, TID=0, DID=2	EP slice(s) in stream 720p 15Hz
SVC ext slice (20)	PPSID = 2, PRID = 7, TID=0, DID=3	EP slice(s) in stream 720p 15Hz
Prefix (14)	PRID = 4, TID=1, DID=0	Prefix NAL of stream 480p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 3	P slice(s) of stream 480p 30Hz
SVC ext slice (20)	PPSID = 1, PRID = 5, TID=1, DID=1	EP slice(s) in stream 480p 30Hz
SVC ext slice (20)	PPSID = 2, PRID = 6, TID=1, DID=2	EP slice(s) in stream 720p 30Hz
SVC ext slice (20)	PPSID = 2, PRID = 7, TID=1, DID=3	EP slice(s) in stream 720p 30Hz
...

Table 29 illustrates the conformant bitstream structure when the streams are interleaved in a single session (i.e. Single Session Transmission, or SST). Even-numbered access units are shown in shaded cells. Assume CGS is used in this example:

Table 29. Conformant Bitstream Structure when streams are interleaved in a single session

NAL unit (type)	Relevant fields in the NAL	Description
SPS (7)	SPSID = 0	SPS of stream 180p 30Hz
SSPS (15)	SSPID = 0	Subset SPS of stream 360p 30Hz
PPS (8)	PPSID = 0, SPSID = 0	PPS of stream 180p 30Hz SVC base layer
PPS (8)	PPSID = 1, SPSID = 0	PPS of stream 360p 30Hz SVC scalable layer
SPS (7)	SPSID = 1	SPS of stream 480p 30Hz
SSPS (15)	SSPID = 1	Subset SPS of stream 480p 30Hz
SSPS (15)	SSPID = 2	Subset SPS of stream 720p 30Hz
PPS (8)	PPSID = 2, SPSID = 1	PPS of stream 480p 30Hz SVC base layer
PPS (8)	PPSID = 3, SPSID = 1	PPS of stream 480p 30Hz SVC scalable layer
PPS (8)	PPSID = 4, SPSID = 2	PPS of stream 720p 30Hz SVC scalable layer
Prefix (14)	PRID = 0, TID=0, DID=0	Prefix NAL of stream 180p 15Hz
IDR slice (5)	PPSID = 0, POC = 0	Base layer IDR slice(s) in stream 180p 15Hz
SVC ext slice (20)	PPSID = 1, PRID = 2, TID=0, DID=1	EI slice(s) in stream 360p 15Hz
Prefix (14)	PRID = 3, TID=0, DID=0	Prefix NAL of stream 480p 15Hz
IDR slice (5)	PPSID = 2, POC = 0	Base layer IDR slice(s) in stream 480p 15Hz
SVC ext slice (20)	PPSID = 3, PRID = 5, TID=0, DID=1	EI slice(s) in stream 480p 15Hz
SVC ext slice (20)	PPSID = 4, PRID = 6, TID=0, DID=2	EI slice(s) in stream 720p 15Hz
SVC ext slice (20)	PPSID = 4, PRID = 7, TID=0, DID=3	EI slice(s) in stream 720p 15Hz
Prefix (14)	PRID = 1, TID=1, DID=0	Prefix NAL of stream 180p 30Hz
Non-IDR slice (1)	PPSID = 0, POC = 1	P slice(s) of stream 180p 30Hz
SVC ext slice (20)	PPSID = 1, PRID = 2, TID=1, DID=1	EP slice(s) in stream 360p 30Hz

Prefix (14)	PRID = 4, TID=1, DID=0	Prefix NAL of stream 480p 30Hz
Non-IDR slice (1)	PPSID = 2, POC = 1	P slice(s) of stream 480p 30Hz
SVC ext slice (20)	PPSID = 3, PRID = 5, TID=1, DID=1	EP slice(s) in stream 480p 30Hz
SVC ext slice (20)	PPSID = 4, PRID = 6, TID=1, DID=2	EP slice(s) in stream 720p 30Hz
SVC ext slice (20)	PPSID = 4, PRID = 7, TID=1, DID=3	EP slice(s) in stream 720p 30Hz
Prefix (14)	PRID = 0, TID=0, DID=0	Prefix NAL of stream 180p 15Hz
Non-IDR slice (1)	PPSID = 0, POC = 2	P slice(s) in stream 180p 15Hz
SVC ext slice (20)	PPSID = 1, PRID = 2, TID=0, DID=1	EP slice(s) in stream 360p 15Hz
Prefix (14)	PRID = 3, TID=0, DID=0	Prefix NAL of stream 480p 15Hz
Non-IDR slice (1)	PPSID = 2, POC = 2	P slice(s) in stream 480p 15Hz
SVC ext slice (20)	PPSID = 3, PRID = 5, TID=0, DID=1	EP slice(s) in stream 480p 15Hz
SVC ext slice (20)	PPSID = 4, PRID = 6, TID=0, DID=2	EP slice(s) in stream 720p 15Hz
SVC ext slice (20)	PPSID = 4, PRID = 7, TID=0, DID=3	EP slice(s) in stream 720p 15Hz
...

The possible combinations of number of simulcast streams, number of scalable layers, maximum frame rates, and resolutions are limited by the maximum macroblock processing rate (MB/sec) defined as a part of the hardware capabilities. The sum of macroblocks per second for all quality/spatial layers in the highest temporal scale across all simulcast streams must not exceed this limit. For example, Table 30 shows the macroblock processing rates for each scale in the above example.

Table 30. Macroblock Processing Rate Example

	MB/sec				
Simulcast streams	108000	108000	48600	48600	24300
			27600	7200	3600

In this example, the sum of macroblocks per second is 348000 MB/sec. This number must be smaller than or equal to the maximum macroblocks per second supported by the encoder.

8. Conformance Test Configuration

8.1. Resolutions

This section describes the resolutions that are considered for conformance test. Encoders that conform to this specification may support resolution not listed here as long as the constraints specified in Section 6.2 are fulfilled.

8.1.1. Landscape Orientation

- (Approximate) 16:9 aspect ratio:
 - 1080p (1920x1080 pixels) (coded as 1920x1088, or 120x68 MB's),
 - 720p (1280x720 pixels) (80x45 MB's),
 - 540p (960x540 pixels) (coded as 960x544, or 60x34 MB's),
 - 480p (848x480 pixels) (53x30 MB's),
 - 360p (640x360 pixels) (coded as 640x368, or 40x23 MB's),
 - 270p (480x270 pixels) (coded as 480x272, or 30x17 MB's),
 - 240p (424x240 pixels) (coded as 432x240, or 27x15 MB's),
 - 180p (320x180 pixels) (coded as 320x192, or 20x12 MB's),
 - 90p (160x90 pixels) (coded as 160x96, or 10x6 MB's)
- (Approximate) 4:3 aspect ratio:
 - VGA (640x480 pixels) (40x30 MB's),
 - 424x320 pixels (coded as 432x320, or 27x20 MB's),
 - QVGA (320x240 pixels) (20x15 MB's),
 - QQVGA (160x120 pixels) (coded as 160x128, or 10x8 MB's)
- 20:3 aspect ratio (panorama):
 - Pano288p (1920x288 pixels) (120x18 MB's),
 - Pano192p (1280x192 pixels) (80x12 MB's),
 - Pano144p (960x144 pixels) (60x9 MB's),
 - Pano96p (640x96 pixels) (40x6 MB's).

8.1.2. Portrait Orientation

- (Approximate) 9:16 aspect ratio:
 - 1080x1920 pixels (coded as 1088x1920, or 68x120 MB's),
 - 720x128 pixels (45x80 MB's),
 - 540x960 pixels (coded as 544x960, or 34x60 MB's),
 - 480x848 pixels (30x53 MB's),
 - 360x640 pixels (coded as 368x640, or 23x40 MB's),
 - 270x480 pixels (coded as 272x480, or 17x30 MB's),
 - 240x424 pixels (coded as 240x432, or 15x27 MB's),
 - 180x320 pixels (coded as 192x320, or 12x20 MB's),
 - 90x160 pixels (coded as 96x160, or 6x10 MB's)
- (Approximate) 3:4 aspect ratio:
 - 480x640 pixels (30x40 MB's),
 - 320x424 pixels (coded as 320x432, or 20x27 MB's),
 - 240x320 pixels (15x20 MB's),
 - 120x160 pixels (coded as 128x60, or 8x10 MB's)

8.1. Frame Rates

Here are the frame rates that are considered for conformance test. Encoders that conform to this specification may support frame rates not listed here as long as the constraints specified in Section 6.2 are fulfilled.

- 30 Hz
- 15 Hz
- 7.5 Hz

H.264 UCConfig Mode Specification