|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  22nd Meeting: Geneva, CH, 15–21 October 2015 | Document: JCTVC-V1007 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **SHVC Test Model 11 (SHM 11) Introduction and Encoder Description** | | |
| *Status:* | Output Document of JCT-VC | | |
| *Purpose:* | SHVC Test Model Description | | |
| *Author(s) or Contact(s):* | Jianle Chen, Qualcomm  Jill Boyce, Vidyo  Yan Ye, InterDigital  Miska M. Hannuksela, Nokia  Guillaume Barroux, Fujitsu Laboratories | Email: | [cjianle@qti.qualcomm.com](mailto:cjianle@qti.qualcomm.com)  [jill@vidyo.com](mailto:jill@vidyo.com)  [Yan.Ye@interdigital.com](mailto:Yan.Ye@interdigital.com)  [miska.hannuksela@nokia.com](mailto:miska.hannuksela@nokia.com)  [guillaume.b@jp.fujitsu.com](mailto:guillaume.b@jp.fujitsu.com) |
| *Source:* | Editors | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

This document serves as a source of general tutorial information on SHVC Test Model 11 (SHM11).

**CONTENTS**

*Page*

[Abstract i](#_Toc411273502)

[1 Introduction 1](#_Toc411273503)

[2 Description of SHVC Test Model 1](#_Toc411273504)

[2.1 General overview of SHVC 1](#_Toc411273505)

[2.2 Coding of the enhancement layer in SHVC 2](#_Toc411273506)

[2.2.1 Inter layer texture prediction in SHVC 2](#_Toc411273507)

[2.2.2 Inter layer motion prediction in SHVC 3](#_Toc411273508)

[2.3 Resampling and colour mapping process of picture samples 4](#_Toc411273509)

[2.3.1 Downsampling process 4](#_Toc411273510)

[2.3.2 Upsampling process 5](#_Toc411273511)

[2.3.3 Cropping windows 7](#_Toc411273512)

[2.3.4 Colour mapping process for colour gamut scalability 8](#_Toc411273513)

[3 Profile, tier and level 10](#_Toc411273514)

[4 Software 11](#_Toc411273515)

[4.1 Software repository 11](#_Toc411273516)

[4.2 Build System 11](#_Toc411273517)

[4.3 Software Structure 11](#_Toc411273518)

[5 Reference 11](#_Toc411273519)

# Introduction

The HEVC scalable extension (SHVC) was developed by the Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11. The normative decoding process of SHVC is specified in the SHVC final draft text [1]. The SHM reference software is provided to demonstrate reference implementation of non-normative encoding techniques and normative decoding process. In addition, this document provides tutorial information for the SHVC test model, and describes key features of this standard.

# Description of SHVC Test Model

## General overview of SHVC

SHVC is the scalable extensions of HEVC. SHVC provides the following scalability features:

* Temporal scalability,
* Spatial scalability,
* Coarse grain SNR scalability,
* Bit depth scalability, Interlaced-to-progressive scalability,
* Colour gamut scalability,
* External base layer scalability,
* Combination of these scalabilities.

The design of HEVC enables temporal scalability when a hierarchical temporal prediction structure is used. Hence the JCTVC focused on developing tools to support bit depth, spatial, coarse grain SNR, interlaced-to-progressive, colour gamut and external base layer scalabilities in SHVC. In external base layer scalability, the codec for the base layer is not specified in the SHVC specification; rather, the base layer decoded picture is provided by external means. These scalabilities except temporal scalability are enabled using a layered approach in SHVC. A general block diagram of a three-layer SHVC encoder with spatial scalability is depicted in Figure 2‑1. SHVC adopts the multi-loop decoding framework. Pictures in different layers in the same access unit are coded into the bitstream in an ascending order of layer indices. The original input pictures are downsampled and coded into the base layer (BL) bitstream by using a conformant HEVC codec or a non-HEVC external (e.g., AVC) codec. To code the enhancement layer (EL) pictures, the already coded data in the lower reference layer(s) are used for inter-layer prediction to improve EL coding efficiency. In SHVC, both reconstructed picture samples and motion parameters from the reference layer(s) can be used for inter-layer prediction. The inter-layer prediction mechanisms currently employed by SHVC are described in the following sub-sections.



Figure 2‑1 High-level block diagram of an SHVC encoder.

## Coding of the enhancement layer in SHVC

The design philosophy of the SHVC standard is to achieve high scalable coding efficiency using a system architecture that requires only high level syntax changes: that is, changes are restricted to slice header level and above. In this way the SHVC architecture design is maximally aligned with the Multi-View extensions of HEVC (MV-HEVC) [2]. Therefore, the EL codec in SHVC does not allow low-level (block-level) changes to the single layer HEVC design. Instead, SHVC high-level syntax design has been modified in such a way that the collocated reconstructed pictures (resampled if necessary) from the reference layers (that is, reference layer pictures with the same POC value as that of the current picture) can be used as inter-layer reference pictures when coding the current enhancement layer picture. This allows inter-layer prediction to be carried out without any low level coding process changes. The highlighted blocks shown in Figure 2‑1 are used to generate the inter layer reference pictures, which mainly includes 3 modules: resampling process, colour mapping process and motion field mapping process. These blocks may be the only major additions necessary to support the SHVC codec.

### Inter layer texture prediction in SHVC

In SHVC, inter-layer texture prediction is invoked by including the inter-layer reference pictures from the reference layers (with resampling and colour mapping process performed if necessary), together with the temporal reference pictures, in the reference picture lists of the enhancement layer picture. Further information on the resampling and colour mapping processes are provided in section 2.3. At the Prediction Unit (PU) level, the signalled one or two reference picture indices are used to indicate whether the current PU is predicted from temporal reference pictures, from inter-layer reference pictures, or from a combination of both. When a PU is predicted from at least one inter-layer reference picture, there is a bitstream conformance constraint that requires the motion vectors associated with the inter-layer reference picture(s) to be zero.

The initial reference picture lists in SHVC are constructed as follows. For reference picture list 0 (L0), the inter layer reference picture(s) is inserted between the set of short-term temporal reference pictures with positive delta POC values (that is, forward temporal reference pictures) and the set of short-term temporal reference pictures with negative delta POC values (that is, backward temporal reference pictures). For reference picture list 1 (L1), the temporal references are first added into the reference list in the same manner as the initial reference picture list construction in HEVC. After that, the inter-layer reference picture(s) is added at the end of L1 as long term reference picture(s). The inter-layer reference picture(s) is added to the reference picture list L0 when the current enhancement-layer picture is coded as P-Slice, and is added to both reference picture lists L0 and L1 when the current enhancement-layer picture is coded as B-Slice.

### Inter layer motion prediction in SHVC

In SHVC, inter-layer motion parameter prediction can be invoked by setting an inter-layer reference picture as the collocated reference picture for TMVP derivation. When spatial scalability is used between the current enhancement layer and its reference layer, motion field mapping process is performed to derive the motion field for the inter layer reference picture; no additional block level decoding process modification to TMVP derivation is needed at the enhancement layer.

In motion field mapping process, the motion field of the inter-layer reference picture is obtained based on the compressed motion field of the lower resolution reference layer picture. The motion parameters (including MVs and reference indices) and prediction mode for each 16×16 block of the inter-layer reference picture are derived from the corresponding motion parameters and prediction mode of the collocated block in the reference layer picture. The 16×16 block size is chosen to be compliant with the HEVC TMVP derivation process, where compressed motion field of 16×16 blocks of the temporal reference picture is used.

As shown in Figure 2‑2, where each grid in the enhancement layer (right) picture represents an 8×8 block and each grid in the reference (left) layer represent a 4×4 block, the collocated 16×16 block in the reference layer picture is derived as follows:

1. The collocated sample location of the center sample of the 16×16 block in the reference layer picture is denoted as ( *xRL*,  *yRL* ).
2. The location (  *xRL*,  *yRL* ) is then rounded to align with a 16×16 block by using an offset of 4, as follows,

*xRL* = ( ( *yRL* + 4 ) >> 4 ) << 4 (2‑1)

*xRL* = ( ( *yRL* + 4 ) >> 4 ) << 4 (2‑2)

As shown in Figure 2‑2, the values of ( xRL, yRL ) are rounded to align with a 16×16 block with top-left location indicated by 1, 2, 3 or 4. When the sample position ( xRL, yRL ) is located outside the reference layer picture, the motion information of the current 16x16 block is marked as unavailable by setting the block prediction mode to intra prediction mode.

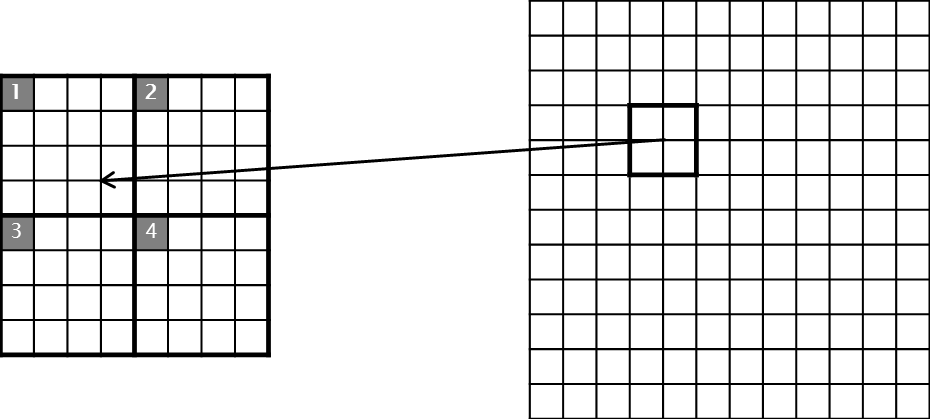


Figure 2‑2 Derivation of collocated 16×16 block in refrence layer

In SHVC, inter-layer motion parameter prediction is disabled between the reference layer and the current layer, when the reference layer is the base layer and is provided by external means.

## Resampling and colour mapping process of picture samples

SHVC uses the same inter layer prediction mechanism for both spatial scalability and SNR scalability. The only difference is that motion and texture information are resampled for inter-layer prediction in case of different resolutions between the layers. SHVC supports a generalized spatial relationship between two layers, including arbitrary spatial ratio and cropping mode. When cropping mode is enabled, picture of a lower layer may correspond to a cropped area of the higher layer picture, or vice versa. For colour gamut scalability, SHVC applies the colour mapping process to the reconstructed reference layer samples as part of the inter layer reference picture derivation process to achieve high coding efficiency.

### Downsampling process

This section introduces the informative downsampling process used in the SHVC reference software to generate the lower layer input video sequences from the enhancement layer input video sequences.

Figure 2‑3 (a) shows the default relative sampling grids in YCbCr4:2:0 chroma format of the current enhancement layer picture and the reference layer picture for ratio 2 spatial scalability, where white squares, white triangles, blue circles and blue triangles represent enhancement layer luma samples, enhancement layer chroma samples, reference layer luma samples, and reference layer chroma samples, respectively. In the default downsampling process, the locations of the luma sample grids of the two layers are aligned at the top-left sample position of the pictures. In addition to the default sample grid alignment, the SHVC resampling process also supports other sample position alignment schemes. Figure 2‑3 (b) shows such an example, where the locations of the luma sample grids of the two layers are aligned at the center sample position of the pictures. When a non-default sample grid alignment scheme is used in the downsampling process, the additional phase offsets are signalled in the bitstream to specify the sample grid alignment between the enhancement layer and its reference layer. These phase offset parameters will be used in upsampling to match the resampled reference layer sample locations with those of the enhancement layer sample locations.

(a) Zero position alignment (b) Center position alignment  
Figure 2‑3 Sampling grids location between the current layer picture and the reference layer picture for 2x spatial ratio

To generate the lower resolution pictures, 12-tap, 2D separable downsampling filters are applied to the higher resolution pictures. The downsampling filter coefficients are shown in Table 2‑1 and Table 2‑2, where filter coefficients in Table 2‑1 are used for downsampling ratios in the range of (5/4, 5/3], and filter coefficients in Table 2‑2 are used for downsampling ratios in the range of (5/3, 2]. Filter coefficients for downsampling ratios beyond these ranges are not available currently, and may be developed in the future. The same downsampling filters are applied to luma and to chroma components. For each sample location (*x, y*) in the lower resolution picture, the collocated fractional sample location in the higher resolution picture is derived and quantized to 1/16 pixel accuracy. Then the filters corresponding to the quantized fractional position are selected from Table 2‑1 or Table 2‑2 and applied in cascade, first in the horizontal dimension and then in the vertical dimension.

This informative downsampling process with zero position alignment is used to generate the two layer inputs in SHVC common test conditions for ratio-1.5 and ratio-2 spatial scalability. For ratio-1.5, 4 filters from Table 2‑1 are used. For ratio-2, only 2 filters from Table 2‑2 are used.

Additional detailed information on the derivation of filter coefficients can be found in [3] and information on the downsampling process, including the source code, can be found in [4].

Table 2‑1 –Downsampling filter coefficients for ratios in the range of (5/4, 5/3]

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| phase i | Downsampling filter coefficients | | | | | | | | | | | |
| f15[i,0] | f15[i,1] | f15[i,2] | f15[i,3] | f15[i,4] | f15[i,5] | f15[i,6] | f15[i,7] | f15[i,8] | f15[i,9] | f15[i,10] | f15[i,11] |
| 0 | 0 | 5 | –6 | –10 | 37 | 76 | 37 | –10 | –6 | 5 | 0 | 0 |
| 1 | 0 | 5 | –4 | –11 | 33 | 76 | 40 | –9 | –7 | 5 | 0 | 0 |
| 2 | –1 | 5 | –3 | –12 | 29 | 75 | 45 | –7 | –8 | 5 | 0 | 0 |
| 3 | –1 | 4 | –2 | –13 | 25 | 75 | 48 | –5 | –9 | 5 | 1 | 0 |
| 4 | –1 | 4 | –1 | –13 | 22 | 73 | 52 | –3 | –10 | 4 | 1 | 0 |
| 5 | –1 | 4 | 0 | –13 | 18 | 72 | 55 | –1 | –11 | 4 | 2 | –1 |
| 6 | –1 | 4 | 1 | –13 | 14 | 70 | 59 | 2 | –12 | 3 | 2 | –1 |
| 7 | –1 | 3 | 1 | –13 | 11 | 68 | 62 | 5 | –12 | 3 | 2 | –1 |
| 8 | –1 | 3 | 2 | –13 | 8 | 65 | 65 | 8 | –13 | 2 | 3 | –1 |
| 9 | –1 | 2 | 3 | –12 | 5 | 62 | 68 | 11 | –13 | 1 | 3 | –1 |
| 10 | –1 | 2 | 3 | –12 | 2 | 59 | 70 | 14 | –13 | 1 | 4 | –1 |
| 11 | –1 | 2 | 4 | –11 | –1 | 55 | 72 | 18 | –13 | 0 | 4 | –1 |
| 12 | 0 | 1 | 4 | –10 | –3 | 52 | 73 | 22 | –13 | –1 | 4 | –1 |
| 13 | 0 | 1 | 5 | –9 | –5 | 48 | 75 | 25 | –13 | –2 | 4 | –1 |
| 14 | 0 | 0 | 5 | –8 | –7 | 45 | 75 | 29 | –12 | –3 | 5 | –1 |
| 15 | 0 | 0 | 5 | –7 | –9 | 40 | 76 | 33 | –11 | –4 | 5 | 0 |

Table 2‑2 –Downsampling filter coefficients for ratios in the range of (5/3, 2]

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| phase i | Downsampling filter coefficients | | | | | | | | | | | |
| f20[i,0] | f20[i,1] | f20[i,2] | f20[i,3] | f20[i,4] | f20[i,5] | f20[i,6] | f20[i,7] | f20[i,8] | f20[i,9] | f20[i,10] | f20[i,11] |
| 0 | 2 | –3 | –9 | 6 | 39 | 58 | 39 | 6 | –9 | –3 | 2 | 0 |
| 1 | 2 | –3 | –9 | 4 | 38 | 58 | 43 | 7 | –9 | –4 | 1 | 0 |
| 2 | 2 | –2 | –9 | 2 | 35 | 58 | 44 | 9 | –8 | –4 | 1 | 0 |
| 3 | 1 | –2 | –9 | 1 | 34 | 58 | 46 | 11 | –8 | –5 | 1 | 0 |
| 4 | 1 | –1 | –8 | –1 | 31 | 57 | 47 | 13 | –7 | –5 | 1 | 0 |
| 5 | 1 | –1 | –8 | –2 | 29 | 56 | 49 | 15 | –7 | –6 | 1 | 1 |
| 6 | 1 | 0 | –8 | –3 | 26 | 55 | 51 | 17 | –7 | –6 | 1 | 1 |
| 7 | 1 | 0 | –7 | –4 | 24 | 54 | 52 | 19 | –6 | –7 | 1 | 1 |
| 8 | 1 | 0 | –7 | –5 | 22 | 53 | 53 | 22 | –5 | –7 | 0 | 1 |
| 9 | 1 | 1 | –7 | –6 | 19 | 52 | 54 | 24 | –4 | –7 | 0 | 1 |
| 10 | 1 | 1 | –6 | –7 | 17 | 51 | 55 | 26 | –3 | –8 | 0 | 1 |
| 11 | 1 | 1 | –6 | –7 | 15 | 49 | 56 | 29 | –2 | –8 | –1 | 1 |
| 12 | 0 | 1 | –5 | –7 | 13 | 47 | 57 | 31 | –1 | –8 | –1 | 1 |
| 13 | 0 | 1 | –5 | –8 | 11 | 46 | 58 | 34 | 1 | –9 | –2 | 1 |
| 14 | 0 | 1 | –4 | –8 | 9 | 44 | 58 | 35 | 2 | –9 | –2 | 2 |
| 15 | 0 | 1 | –4 | –9 | 7 | 43 | 58 | 38 | 4 | –9 | –3 | 2 |

### Upsampling process

This section reviews the normative upsampling process applied to the reference layer reconstructed picture, in order to provide inter layer texture prediction for the enhancement layer, when the reference layer and the enhancement layer pictures have different spatial resolutions.

Given an upsampling ratio of N in both directions, the upsampling process is conceptually approximated by interpolating the reference layer reconstructed picture to 16 times its size in both directions and decimating the 16x picture with the ratio M, where M=16/N. To perform 16x upsampling, 16-phase interpolation filters are used. Detailed filter coefficients are provided in Table H‑1 and Table H‑2 in [1], for luma and for chroma, respectively. The upsampling filters are designed to be backward compatible with the filters used for motion compensation interpolation process in HEVC, where 8-tap filters are applied for the luma component and 4-tap filters are applied for the chroma components. The filters used in HEVC motion compensation interpolation process are kept unchanged in the upsampling process at the corresponding ½- and ¼-pixel fractional positions, and new filter coefficients are added for the new phases that do not exist in the HEVC motion compensation interpolation process. By defining a full set of 16-phase interpolation filters, SHVC is able to support arbitrary spatial ratios between two layers.

In actual implementation, the 16x interpolation is not performed for every sample; instead, only those samples in the 16x picture that will be kept after decimation are interpolated. To generate a sample located at (*x, y*) in the enhancement layer picture, first its collocated position (*x*16*, y*16) in the virtual 16x picture before decimation is found by using the method specified in sub-clause H.8.1.3.1.3 in [1]. Then, the sample value at (*x*16*, y*16) is interpolated by applying the appropriate phase filter to the support region in the reference layer reconstructed picture. The process to determine the reference layer support region and the appropriate phase filters are specified in sub-clause H.8.1.3.1.1 and H.8.1.3.1.2 in [1], for luma and for chroma, respectively. Same as in the HEVC motion compensation interpolation process, when the collocated sample position or any of its neighbouring samples used in upsampling are located outside of the reference layer picture, the reference layer picture is padded and the upsampling interpolation process is applied to the padded samples.

As mentioned in section 2.3.1, when non-default downsampling process is used, additional phase offsets are signalled in the bitstream to specify the sample grid alignment between an enhancement layer and its reference layer such that the reference layer sample location (phase) derivation process used in upsampling can be matched with that used in downsampling. In SHVC specification, four syntax elements, *phase\_hor\_luma*, *phase\_ver\_luma*, *phase\_hor\_chroma* and *phase\_ver\_chroma*, are used to specify the luma phase offset in the horizontal direction, luma phase offset in the vertical direction, chroma phase offset in the horizontal direction and chroma phase offset in the vertical direction, respectively.

One of the key operation in upsampling process is finding the collocated position in the reference layer picture. In SHVC, given the location of a sample (luma or choma) in the EL picture, (*xEL*, *yEL*), the collocated RL sample location (*x16RL*, *y16RL*), in units of 1/16 RL samples, is derived as follows:

 (2‑3)

 (2‑4)

where *phaseX* and *phaseY* are *phase\_hor\_luma*, *phase\_verl uma* in case of luma component, and *phase\_hor\_chroma* and *phase\_ver\_chroma* in case of chroma component. *offsetXEL* and *offsetYEL* are the left and top scaled reference layer offset, and *offsetXRL* and *offsetYRL* are the left and top reference region offset, which is to be described in the next sub-section. *horS* and *verS* are the horizontal and vertical inverse scaling factors between the EL and the RL in 16-bit fixed point precision, which are calculated as follows:

 (2‑5)

 (2‑6)

where *WEL* and *HEL* are the width and height of the EL corresponding region and *WEL* and *HBL* are the width and height of the BL corresponding region as shown in Figure 2‑5.

By default, *phase\_hor\_luma*, *phase\_ver\_luma* and *phase\_hor\_chroma* are set to 0. *phase\_ver\_chroma* is set equal to 0 if the encoded video sequences of both EL and RL are in the YCbCr4:4:4 format and is set to rounding (4\**HEL* + *HBL*/2)/*HBL*– 4 if the encoded video sequences of both EL and RL are in the YCbCr4:2:0 format. Those default values are determined according to the relative sampling grids of the EL picture and the RL picture and the chroma sampling grid location related to luma sampling grid as shown Figure 2‑3 (a).

Table 2‑3 provides the recommended phase offset syntax values for cross-layer center position alignment shown in Figure 2‑3 (b).

**Table 2‑3 – Recommended phase offset syntax element values for center position alignment**

|  |  |  |
| --- | --- | --- |
| **Syntax** | YCbCr4:2:0 content | YCbCr4:4:4 content |
| phase\_hor\_luma | (*WEL*\*8+*WBL*/2) / *WBL* − 8 | (*WEL*\*8+*WBL*/2) / *WBL* − 8 |
| phase\_ver\_luma | (*HEL*\*8+*HBL*/2) / *HBL* − 8 | (*HEL*\*8+*HBL*/2) / *HBL* − 8 |
| phase\_hor\_chroma | (*WEL*\*4+*WBL*/2) / *WBL* − 4 | (*WEL*\*8+*WBL*/2) / *WBL* − 8 |
| phase\_ver\_chroma | (*HEL*\*8+*HBL*/2) / *HBL* − 8 | (*HEL*\*8+*HBL*/2) / *HBL* − 8 |

Besides the non-default sample grid alignment application motioned in section 2.3.1, another key application which can utilize such flexible phase adjustment method provided by SHVC is interlaced-to-progressive scalability, in which the reference layer picture is an interlaced field picture, and the enhancement layer picture is a frame picture. Figure 2‑4 shows the default luma and chroma sample locations of YUV 4:2:0 interlaced video. Sample locations in the top field correspond to the odd rows in a progressive frame picture and sample locations in the bottom field corresponds to the even rows in a progressive frame picture, for both luma and chroma samples. The relative sample location between two layers for the case of interlaced-to-progressive scalability is different from the progressive-to-progressive spatial scalability case. For interlaced-to-progress scalability, in order to obtain the correct collocated sample position in upsampling, two sets of phase offsets (i.e. two different PPSs) need to be signalled in the bitstream. One is used for the enhancement layer pictures with the reference layer pictures being top field pictures and the other is used for the enhancement layer pictures with the reference layer pictures being bottom field pictures.



**Figure 2‑4** **Nominal vertical and horizontal sampling locations of 4:2:0 samples in top and bottom fields**

Recommended phase offset syntax element values for interlaced-to-progressive scalability shown in Figure 2‑4 is provided in Table 2‑4.

**Table 2‑4 – Recommended phase offset syntax element values for interlaced-to-progressive scalability**

|  |  |  |
| --- | --- | --- |
| **Syntax** | **RL: top field EL: frame** | **RL: bottom field EL: frame** |
| phase\_hor\_luma | 0 | 0 |
| phase\_ver\_luma | 0 | 16 |
| phase\_hor\_chroma | 0 | 0 |
| phase\_ver\_chroma | 0 | 16 |

It’s worth noting that, in SHVC specification, *phase\_hor\_chroma\_plus8* and *phase\_ver\_chroma\_plus8* are signaled instead of *phase\_hor\_chroma and phase\_hor\_chroma*. So the real signaled syntax values are equal to those values mentioned above plus 8. This design is used to avoid the negative syntax value.

### Cropping windows

In the scalable extension of ITU-T H.264 | ISO/IEC 14496-10, Extended Spatial Scalability (ESS) enables a generalized relationship between successive spatial layers. A picture of a reference spatial layer may represent a cropped area of the higher resolution picture and the ratio between successive spatial layers can have any value. Similarly, SHVC also supports the cropping mode between any lower layer and the current enhancement layer. The geometrical parameters defining the cropping windows are signalled at the PPS level.

Figure 2‑5 illustrates the sample location relationship between the reference layer and the enhancement layer. As shown in the figure, two sets of offsets can be used to indicate the sample location between the two layers. The reference region offsets specify the spatial correspondence of the reference region in the reference layer picture relative to the decoded reference layer picture. The scaled reference layer offsets specify the spatial correspondence of the current layer picture relative to the scaled reference region of the scaled reference layer picture. Both the scaled reference layer offsets and the reference region offsets can have negative values. When reference region offsets are negative, the reference layer picture corresponds to a cropped area of the enhancement layer picture. When scaled reference layer offsets are negative, the enhancement layer picture corresponds to a cropped area of the reference layer picture.



Figure 2‑5 – Relations between enhancement layer and reference layer with ESS

### Colour mapping process for colour gamut scalability

SHVC allows the reference and enhancement layers to be represented in different colour spaces. For example, a reference layer may be in Rec. 709 colour space and the enhancement layer may be in Rec. 2020 colour space. In order to improve the coding efficiency of colour gamut scalability, 3D Look-Up Table (3D LUT) based colour mapping is used in SHVC to convert the tri-chromatic samples from the reference layer colour space to the enhancement layer colour space. The 3D LUT divides the input 3D YCbCr colour space into up to 8×2×2 cuboid regions, with up to 8 regions in the Y dimension and up to 2 regions in the Cb and Cr dimensions. Colour mapping coefficients for each cuboid region are signalled in the PPS to define the 3D LUT operation. As shown in equation (2‑3), 12 mapping coefficients are signalled for each cuboid. Uniform partition is applied in the Y dimension, while as shown in Figure 2‑6, non-uniform partition can be used in the Cb and Cr dimensions. When 2 partitions are used in the Cb and Cr dimensions, adaptive Cb and Cr partition thresholds are also signalled in the PPS, along with other colour mapping coefficients.



Figure 2‑6 – Non-uniform partitions of chroma components

The colour mapping process applies the 3D LUT operation to each reference layer sample as part of the inter-layer prediction process. When upsampling is required, colour mapping is applied prior to upsampling. For each input reference layer sample, the cuboid to which it belongs is first determined. This determination is based upon the number of cuboid regions used, the chroma partition thresholds and the YCbCr values of the input reference layer sample. Then, within the cuboid, the mapped sample values in the output enhancement layer colour space are calculated by using the following equation, where , , , and for are the colour mapping coefficients associated with the cuboid.

(2‑3)

In the SHM reference software encoder, the optimal 3D LUT is derived for each enhancement layer picture. Although the current SHVC standard only allows up to 8×2×2 partitions, the SHVC reference encoder can be configured to use more partitions such as 8×8×8. There are two methods supported by the SHM encoder to derive an appropriate number of partitions for each enhancement layer picture. In the first method, the number of cuboid regions for the current picture is determined based on the bit cost of the 3D LUT of all the previously coded pictures. The number of cuboid regions (that is, 3D LUT table size) decreases when the accumulated bits used for the 3D LUT is more than 3% of the total bits used for all the previous coded picture and increases when the proportion is less than 0.5%. In the second method, the optimal 3D LUT size is decided by rate distortion optimization, where the LUT size that provides the best trade-off between bit cost and distortion reduction is selected.

After the number of cuboid regions or 3D LUT size is determined and when the partition number in the chroma dimension is 2, the adaptive thresholds for Cb and Cr components are calculated by averaging the Cb samples and the Cr samples in the reference layer picture, respectively. Then, the colour mapping coefficients are derived for each cuboid by minimizing the distortion between the colour mapped reference layer samples and original enhancement layer samples, namely

(2‑4)

Where , , and, represent the Y, Cb, Cr sample values of enhancement layer picture and , and represent the colour mapped reference layer samples calculated from the reconstructed reference layer sample using equation (2‑3).

Since the colour mapping process is conducted before up-sampling, when the two layers have different spatial resolutions, the matrix coefficients are derived by minimizing the distortion between the colour mapped reference layer samples and the down-sampled enhancement layer samples.

In the SHM common test condition, once per second an IRAP picture is sent. For these IRAP pictures, the SHM reference encoder derives a new 3D LUT and sends a PPS containing the new 3D LUT. For all non-RAP pictures, the SHM reference encoder derives the 3D LUT for the current picture and decides whether to send a new PPS containing a new 3D LUT or not. This is done by comparing the rate distortion cost between using the previously sent 3D LUT and using the new 3D LUT. If a new PPS containing a new 3D LUT is sent, the new PPS replaces the previous PPS containing the previously sent 3D LUT.

# Profile, tier and level

Six profiles are currently defined in SHVC specification:

* Scalable Main,
* Scalable Main10,
* Scalable Monochrome,
* Scalable Monochrome12
* Scalable Monochrome16 and
* Scalable Main 4:4:4.

Profile, tier, and level (PTL) are signalled in both the base VPS and the VPS extension, through the profile\_tier\_level( ) syntax structure. An indexing mechanism is used such that i-th profile\_tier\_level( ) syntax structure signalled is referred to as VpsProfileTierLevel[ i ].

The PTL signalled in the base VPS represents profile, tier and level information of the bitstream that a single layer decoder will receive, e.g. one conforming to a profile defined in Annex A. The tier/level must allow for the higher bitrate of the entire bitstream, including those NAL units with nuh\_layer\_id greater than 0 that the single layer decoder will discard. This PTL signalled in the base VPS is referred to as VpsProfileTierLevel[ 0 ].

When the base layer is internal to the bitstream, at the beginning of the VPS extension, a PTL syntax structure is signalled with profilePresentFlag equal to 0, and is referred to as VpsProfileTierLevel[ 1 ]. It is used to signal the PTL representing the profile, tier and level information of the sub-bitstream, formed by the sub-bitstream extraction process, containing only those NAL units with nuh\_layer\_id equal to 0, e.g. corresponding to the resolution, frame rate and bitrate of just the base layer. The level value of VpsProfileTierLevel[ 1 ] may be lower than or equal to the level value of VpsProfileTierLevel[ 0 ], based upon the bitrate.

Both level values in VpsProfileTierLevel[ 0 ] and VpsProfileTierLevel[1] are signalled because a backwards compatible single layer decoder will be required to handle receiving the bitrate of the entire bitstream, even though it will discard NAL units with nuh\_layer\_id greater than 0. For example, a decoder conforming to main tier level 4 can decode a bitstream with 1920x1080@30fps and 12 Mbps. If an SHVC bitstream contains a base layer with 1920x1080@30fps at 12 Mbps with an enhancement layer of 3840x2160@30fps with a total bitrate of 20 Mbps, the PTL included in the base VPS should indicate main tier level 4.1, because a Main profile main tier level 4 decoder would not be able to handle being given 20 Mbps.

Later in the VPS extension, a PTL syntax structure is signalled for each necessary layer of each output layer set. These are numbered beginning with VpsProfileTierLevel[ 2 ], assuming that the base layer is internal to the bitstream

**Table 3‑1 – VpsProfileTierLevel index representation**

|  |  |
| --- | --- |
| ***i*** | ***VpsProfileTierLevel[ i ] represents*** |
| 0 | Base layer plus extra bitrate of entire bitstream |
| 1 | Base layer |
| 2 .. N | Enhancement layers in output layer sets |

Consider an example bitstream containing two layers, with values nuh\_layer\_id equal to 0 and 1, and two output layer sets. The output layer set with output layer set index (olsIdx) 0 is inferred to include only the base layer, and the output layer set with olsIdx 1 includes both layers. Continuing with the above example, with a base layer at 1920x1080@30fps and the enhancement layer at 3840x2160@30fps, with 12 Mbps in the base layer and 8 Mbps in the enhancement layer, for a total bitrate of 20 Mbps, the values signalled in VpsProfileTierLevel[ i ] for i in 0 ..2 and the association of the VpsProfileTierLevel[ i ] to the existing output layer sets are shown in the tables below.

**Table 3‑2 – An example of PTL signalling**

|  |  |  |  |
| --- | --- | --- | --- |
| ***i*** | ***VpsProfileTierLevel[ i ]*** | | |
| ***Profile*** | ***Tier*** | ***Level*** |
| 0 | Main | Main | 4 |
| 1 | Main | Main | 4.1 |
| 2 | Scalable Main | Main | 5 |

**Table 3‑3 – An example of PTL index setting**

|  |  |  |
| --- | --- | --- |
| ***OLS*** | ***Layer*** | ***PTL Index*** |
| 0 | 0 | 1 |
| 1 | 0 | 1 |
| 1 | 2 |

# Software

## Software repository

The source code for the software is available in the following SVN repository.

<https://hevc.hhi.fraunhofer.de/svn/svn_SHVCSoftware/>

Contact point for software issue report and question:

Seregin, Vadim ([vseregin@qti.qualcomm.com](mailto:vseregin@qti.qualcomm.com))

## Build System

The software can be built under linux using make. For Windows, solutions for different versions of Microsoft Visual Studio are provided.

## Software Structure

The SHVC Test Model Software inherits the same software structure from the HEVC test model HM software, which includes the following applications and libraries for encoding, decoding and downsampling process:

* Applications:
  + TAppEncoder, executable for bit stream generation
  + TAppDecoder, executable for reconstruction.
  + TAppCommon, common functions for configuration file parsing.
  + TAppDownConvert, downsampling functionalities.
* Libraries:
  + TLibEncoder, encoding functionalities
  + TLibDecoder, decoding functionalities
  + TLibCommon, common functionalities
  + TLibVideoIO, video input/output functionalities

# Reference

1. J. Chen, J. Boyce, Y. Ye and M. M. Hannuksela, G. J. Sullivan, Y-K. Wang, “Scalable High Efficiency Video Coding Draft 7”, JCTVC-R1008, 18th JCTVC Meeting, Sapporo, JP, 30 June – 9 July 2014
2. G. Tech, K. Wegner, Y. Chen, M. Hannuksela, J. Boyce, “MV-HEVC Draft Text 9”, JCT3V-I1002, 9th JCT3V Meeting, Sapporo, JP, 3 – 9 July 2014
3. J. Dong, Y. He, and Y. Ye, "Downsampling filter for anchor generation for scalable extensions of HEVC", m24499, 100th MPEG meeting, Geneva, CH, Apr. 2012.
4. E. Alshina, A. Alshin, X. Li, J. Chen, M. Karczewicz, J. Dong, Y. Ye, E. Francois, “AhG13: Performance analysis of scalable systems with different down-samplers”, JCTVC-O0071, 15th JCTVC Meeting, Geneva, CH, Oct. 2013