|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  16th Meeting: San José, US, 9–17 Jan. 2014 | Document: JCTVC-P1007\_v1 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Scalable HEVC (SHVC) Test Model 5 (SHM 5)** | | |
| *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 | 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) |
| *Source:* | Editors | | |

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

# Abstract

This document serves as a source of general tutorial information on SHVC Test Model 5 (SHM5).

**CONTENTS**

*Page*

[Abstract i](#_Toc373695404)

[1 Introduction 1](#_Toc373695405)

[2 Description of SHVC Test Model 1](#_Toc373695406)

[2.1 General overview of SHVC 1](#_Toc373695407)

[2.2 Coding of the enhancement layer in SHVC 2](#_Toc373695408)

[2.2.1 Inter layer texture prediction in SHVC 2](#_Toc373695409)

[2.2.2 Inter layer motion prediction in SHVC 3](#_Toc373695410)

[2.3 Resampling process of picture samples 3](#_Toc373695411)

[2.3.1 Downsampling process 4](#_Toc373695412)

[2.3.2 Upsampling process 5](#_Toc373695413)

[2.3.3 Cropping windows 7](#_Toc373695414)

[3 Software 7](#_Toc373695415)

[3.1 Software repository 7](#_Toc373695416)

[3.2 Build System 7](#_Toc373695417)

[3.3 Software Structure 7](#_Toc373695418)

[4 Reference 8](#_Toc373695419)

# Introduction

The HEVC scalable extension (SHVC) is currently being 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 draft text [1] and software is also 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 some key features of this standard.

# Description of SHVC Test Model

## General overview of SHVC

SHVC is the scalable extensions of HEVC which currently provide features of temporal, SNR, bit depth, spatial scalability and a combination of these scalabilities. The design of HEVC enables temporal scalability when a hierarchical temporal prediction structure is used. Hence the JCTVC currently concentrates on developing tools to support bit depth, spatial and coarse grain SNR scalabilities in SHVC. Spatial, SNR and bit depth scalabilities are enabled using a layered approach in SHVC. A general block diagram of a three spatial layer SHVC encoder 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 or 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 tools currently supported 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 relatively simple system architecture. Another consideration is to keep the SHVC architecture design 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 resampled reference layer pictures; 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 (resampled if necessary), together with the temporal reference pictures, in the reference picture lists of the enhancement layer picture. 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 = ( ( xRL + 4 ) >> 4 ) << 4 (2‑1)

yRL = ( ( 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.

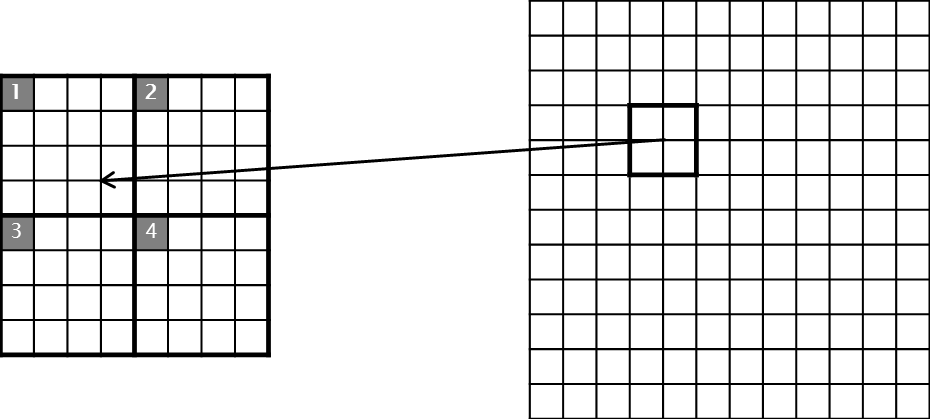


Figure 2‑2 Derivation of collocated 16×16 block in refrence layer

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 coded using a non-HEVC codec.

## Resampling 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 relation 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.

### Downsampling process

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

Figure 2‑3 (a) and (b) show the relative sampling grids 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. Both the enhancement layer and reference layer video sequences are assumed to be in YUV4:2:0 format. As shown in the figure, SHVC support two sample grid alignment schemes between layers: zero position alignment (in Figure 2‑3 (a)) and central position alignment (Figure 2‑3 (b)). When zero position alignment is used, the locations of the luma sample grids of the two layers are aligned at the top-left sample position of the pictures. When center position alignment scheme is used, the locations of the luma sample grids of the two layers are aligned at the center sample position of the pictures. A flag is signalled in VPS extension to indicate which alignment scheme is used.

(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 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 need to be further developed. 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 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 introduces the 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, its collocated position (*x*16*, y*16) in the virtual 16x picture before decimation is first found by using the method specified in sub-clause H.6.2 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.8.4.1.3 and H.8.4.1.4 in [1], for luma and for chroma, respectively. Additionally, the collocated sample derivation for upsampling process is designed to support both of the two sampling grid alignment schemes discussed in section 2.3.1.

As mentioned in section 2.3.1, a flag is 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 up-sampling can be matched with that used in downsampling. In addition to this flag, the vertical phase of the SHVC upsampling process can be further adjusted with additional flags signalled in the bitstream. A key application which can utilize the additional vertical phase adjustment flags 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. Additional phase adjustment can be performed to align the sample positions of the upsampled reference layer picture with the sample positions of the enhancement layer picture. A flag is signalled in the sequence parameter set to indicate whether vertical phase adjustment is enabled. When vertical phase adjustment is enabled, an additional flag is signalled in the slice header (picture level) to indicate the position of the vertical phase, based upon whether the reference layer picture is a top field or bottom field picture.

Table 2‑3 indicates the recommended values of the vertical phase adjustment syntax elements for different use cases.



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

**Table 2‑3 –Recommended vertical phase adjustment syntax element values**

|  |  |  |  |
| --- | --- | --- | --- |
| **Reference layer** | **Enhancement layer** | **SPS vert\_phase\_position\_enable\_flag** | **Slice header vert\_phase\_position\_flag** |
| top field | top field | 0 | - |
| bottom Field | bottom field | 0 | - |
| top field | frame | 1 | 0 |
| bottom field | frame | 1 | 1 |
| frame | frame | 0 | - |

### 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. In addition to arbitrary spatial ratio, SHVC also supports the cropping mode between any lower layer and the current enhancement layer. The geometrical parameters defining the cropping window are signaled at the sequence level.

Figure 2‑5 illustrates the sample location relations between the reference layer and enhancement layer. In current SHM, the picture samples upsampling process is first applied to the reference layer picture to form upsampled picture within the scaled windows. Then, the scaled reference layer picture is further padded in horizontal and/or vertical directions to create an inter-layer reference picture having the same resolution as the current enhancement layer picture. In motion mapping process, the motion information outside of scaled base layer window is marked as unavailable by setting the block prediction mode to intra prediction mode. The scaled parameters can also have negative values. In this case, the enhancement layer picture corresponds to a cropped area of the reference layer picture.

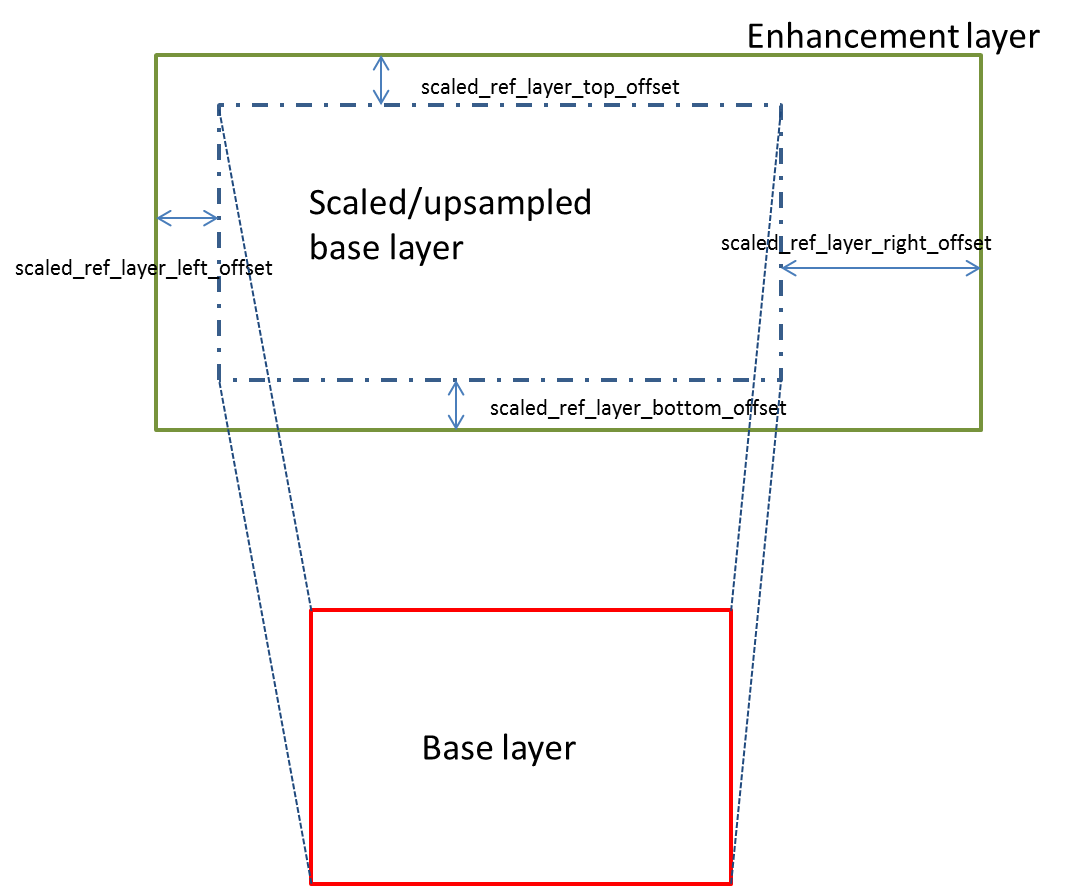


Figure 2‑5 – Relations between enhancement layer and reference layer with ESS

# Software

## Software repository

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

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

For tool integration branch for a company can be obtained by contacting:

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. Yan and M. M. Hannuksela, G. J. Sullivan, Y-K. Wang, “Scalable High Efficiency Video Coding Draft 5”, JCTVC-P1008, 16th JCTVC Meeting, San Jose, US, Jan. 2014
2. G. Tech, K. Wegner, Y. Chen, M. Hannuksela, J. Boyce, “MV-HEVC Draft Text 7”, JCT3V-G1004, 7th JCT3V Meeting, San Jose, US, Jan. 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